Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

i-Society 2007

International Conference on Information Society, October 7–11, 2007, Merrillville, Indiana, USA

Proceedings of the International Conference on Information Society (i-Society 2007)

EDITORS
Charles A. Shoniregun, University of East London, UK
Alex Logvynovskiy, e-Centre for Infonomics, UK

Table of contents

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007

Message from the General Chairs

Welcome to the International Conference on Information Society (i-Society 2007). The i-Society 2007 conference is an opportunity for researchers and practitioners to exchange ideas about past, present and future trends in 'Information Society'. The i-Society 2007 received 116 papers from 30 different countries of which 57 were accepted. It is interesting to point out that the authors of the accepted papers are from 24 countries. To evaluate each submission, a double blind paper evaluation method was used: each paper was reviewed by at least three internationally known experts from our Program Committee. Furthermore, a short list of twenty papers was selected to appear in 'International Journal for Infonomics' and the 'International Journal for Internet Technology and Secured Transactions'.

Many people have worked very hard to make this conference possible. we would like to thank all who have helped in making i-Society 2007 a success. The program committee members and referees each deserve credit for the excellent participation. Special thanks go to the General Vice-chair; Technical Program Co-chairs; International Co-chairs; Poster and Demo Co-chairs; Local Arrangements Co-chairs; Publication Chair; Publicity Chair; Industrial Chair.

We would like to thank the authors who have contributed to i-Society 2007, the invited speakers *Antony Satyadas*, *Ian Foster*, *Marcus Rogers*, *Eli Cohen* and *Mark Ciampa* for agreeing to participate in the i-Society 2007. We will also like to acknowledge my appreciation to the following organisations: *Purdue University Calumet*, *ACM*, *BCS*, *e-Centre for Infonomics*, *IEEE*, *IET*, *IT advisory group*, *Kansas State University*, *Microsoft UK*, *St. Francis Xavier University*, *University of East London* (UeL), *Cengage Learning* (formerly *Thomson Publishing Company*) and *Prentice Hall Pearson Professional and Career Publishing*

It has been great pleasure to serve as the General Chairs for the i-Society 2007. The long term go i-Society 2007 is to build a reputation and respectable conference for the international community.

On behalf of the i-Society 2007 Executive members, we would like to encourage you to contribute to future i-Society conferences as authors, speakers, panellists or volunteer conference organisers. We wish you a pleasant stay in Indiana, and please feel free to exchange ideas with other experts.

Prof Reza Kamali
Purdue University, Calumet, USA

Prof Charles R. Winer
Purdue University, Calumet, USA

Conference organisers

STEERING COMMITTEE

Michael Wellman ACM SIGecom
Wolfgang Gentzsch D-Grid, Germany
Kia Makki Florida International University, USA
Liang-Jie Zhang IBM, Watson Research Center, USA
Caterina Scogilio Kansa State University, USA
Charles A. Shoniregun University of East London, UK

EXECUTIVE COMMITTEE

General co-chairs

Reza Kamali Purdue University, Calumet, USA email
Charles R. Winer Purdue University, Calumet, USA

General vice-chairs

Charles A. Shoniregun University of East London, UK
Victor Ralevich Sheridan Institute of technology and Advanced Learning, Canada

Technical program co-chairs

Niki Pissinou Florida International University, USA
Maaruf Ali Oxford Brookes University, UK
Dragana Martinovic University of Windsor, Canada

International co-chairs

Seamus Simpson Manchester Metropolitan University, UK
Rimvydas Skyrius University of Vilnius, Lithuania
Zhixiong Chen Mercy College, Dobbs Ferry, NY, USA

Poster and Demo chairs

Keyuan Jiang Purdue University Calumet, USA
Victoriya Repka Kharkov National University of Radioelectronics, Ukraine
Tony Shan Lead Systems Architect Wachoiva Bank, USA

Local arrangements chair

Keyuan Jiang Purdue University Calumet, USA
Barbara Nicolai Purdue University Calumet, USA

Publicity co-chairs

Sam Liles Purdue University Calumet, USA
Alex Logvynovskiy London South Bank University, UK

Industrial chair

Jen Yao Chung IBM, Watson Research Center, USA

INTERNATIONAL PROGRAM COMMITTEE MEMBERS

Abdullah Abonamah, Zayed University, United Arab Emirates
Maaruf Ali, Oxford Brookes University, UK
Costin Badica, University of Craiova, Romania
Stuart Barnes, University of East Anglia Norwich, UK
Paolo Bellavista, DEIS, University of Bologna, Italy
Shlomo Berkovsky, University of Haifa, Israel
Bharat Bhargava, Purdue University, USA
Peter Bieringer, Deep Space 6, Germany
Roy Boggs, Florida Gulf Coast University, USA
Stephane Bressan, National University of Singapore, Singapore
Rajkumar Buyya, University of Melbourne, Australia
Jiannong Cao, Hong Kong Polytechnic University, Hong Kong
Christer Carlsson, Abo Akademi University, Finland
Malu G. Castellanos, HP Labs, Palo Alto, USA
Patrick Y.K. Chau, University of Hong Kong, Hong Kong
Jeng-Chung (Victor) Chen, National Cheng Kung University, Taiwan
Wen-Chyuan Chiang, University of Tulsa, USA
Young B. Choi, James Madison University, USA
Ta-Tao Chuang, Gonzaga University, USA
Paul D. Clough, University of Sheffield, UK
Fabio Crestani, University of Strathclyde, UK
Sally Jo Cunningham, University of Waikato, New Zealand
Mohammad Dastbaz, University of Greenwich, UK
Reggie Davidrajuh, University of Stavanger, Norway
Jaime Delgado, Pompeu Fabra University, Spain
Mieso Denko, University of Guelph, Canada
Brian Detlor, McMaster University, Canada
Flavia Donno, European Laboratory for Particle Physics (CERN), Switzerland
Olaf Droegehorn, University of Kassel, Germany
Paloma Díaz, Carlos III University of Madrid, Spain
Ephrem Eyob, Virginia State University, USA
Yaniv Eytani, University of Haifa, Israel
Tiziano Fagni, ISTI, Italian National Research Council, Italy
Cristiano di Flora, Nokia Research Center, Finland
Shiwa Fu, IBM T.J. Watson Research Center, USA
Akira Fukuda, Kyushu University, Japan
João Gama, University of Porto, Portugal
Claude Godart, University Henri Poincaré and INRIA, France
Fernando Gomez, University of Central Florida, USA
Mounira Harzallah, University of Nantes, France
Rena Hixon, USA
Sonja Hof, Samtis
Meng-Hsiang Hsu, National Kaohsiung First University of Science and Technology, Taiwan
Lynne Humphries, University of Sunderland, UK
Fidelis Ikem, Virginia State University, USA
Pedro Isaias, Aberta University, Portugal

Hemant Jain, University of Wisconsin - Milwaukee, USA
Makoto Kageto, Nihon Fukushi University, Japan
Hiromitsu Kato, Hitachi Systems Development Laboratory, Japan
Mounir Kehal, International University of Monaco, Monaco
Bach Hung Khang, National Centre for Natural Science and Technology, Vietnam
Fredrik Kilander, IT University in Kista, Sweden
Myoung Ho Kim, Korea Advanced Institute of Science and Technology, Korea
Seong Soo Kim, Texas A&M University, USA
Andy Koronios, University of South Australia, Australia
Herma van Kranenburg, Telematica Institute, The Netherlands
Atul Kumar, IBM India Research Lab, India
Monica Landoni, University of Strathclyde, UK
Guanling Lee, National Dong Hwa University, Taiwan
Matthew K.O. Lee, City University of Hong Kong, Hong Kong
Liz Lee-Kelley, University of Surrey, UK
Dušan Lesjak, University of Primorska, Slovenia
Qianhui Althea Liang, Singapore Management University, Singapore
Jay Liebowitz, Johns Hopkins University, USA
Billy B.L. Lim, Illinois State University, USA
Geng Lin, Cisco Systems Inc., USA
Yaowei Liu, Alcatel Shanghai Bell, China
Claudio Lucchese, ISTI, Italian National Research Council, Italy
Zong-Wei Luo, E-Business Technology Institute, Hong Kong
Nazim H. Madhavji, University of Western Ontario, Canada
Sanjay Madria, University of Missouri-Rolla, USA
Bendick Mahleko, Fraunhofer Gesellschaft, Germany
Mihhail Matskin, Royal Institute of Technology, Sweden
Barry McCollum, Queen's University Belfast, Northern Ireland, UK
Ron McFadyen, University of Winnipeg, Canada
Peter Milligan, Queen's University Belfast, Northern Ireland, UK
Nader F. Mir, San Jose State University, USA
Ali R. Montazemi, McMaster University, Canada
Daisuke Morikawa, KDDI R&D Laboratories, Japan
Bernd Mrohs, Fraunhofer FOKUS , Germany
Daryl G. Nord, Oklahoma State University, USA
Shirley O'Neill, University of Southern Queensland, Australia
Mohammad S. Obaidat, Monmouth University, USA
Salvatore Orlando, Università Ca' Foscari, Italy
Georgios Papadimitriou, Aristotle University of Thessaloniki, Greece
Yang Park, University of Wisconsin - La Crosse, USA
P. Pichappan, Annamalai University, India
Despina Polemi, University of Piraeus, Greece
Wolfgang Prinz, Fraunhofer Institut FIT, Germany
Dimitrios P. Ptochos, National Tecnhical University of Athens, Greece
Matthias Rauterberg, Eindhoven Technical University (TU/e), The Netherland
Oriana Riva, University of Helsinki, Finland
Maytham Hassan Safar, Kuwait University, Kuwait
Karin Sallhammar, Norwegian University of Science and Technology, Norway
Samiaji Sarosa, Atma Jaya Yogyakarta University, Indonesia
Caterina Maria Scoglio, Kansas State University, USA
Ming-Chien Shan, Hewlett Packard Laboratories, USA
Dong Hee Shin, The Pennsylvania State University, Berks, USA
Chi-Ren Shyu, University of Missouri - Columbia, USA
Fabrizio Silvestri, Institute of Information Science and Technologies, Italy
Andrew Simpson, University of Oxford, UK
Kwan-Ho Song, National Internet Development Agency of Korea, Republic of Korea
Michael Sonntag, FIM, Johannes Kepler University, Austria
Heinz Stockinger, University of Vienna, Austria
Moiez A. Tapia, University of Miami, USA
Michael J. Tarn, Western Michigan University, USA
Ewe Hong Tat, Multimedia University, Malaysia
Samir Tata, Institut National des Télécommunications, France
Do van Thanh, Telenor R&D , Norway
Frédéric Thiesse, University of St. Gallen, Switzerland
Abdallah Tubaishat, Zayed University, United Arab Emirates
Bruno Tuffin, IRISA , France
Jari Veijalainen, University of Jyväskylä, Finland
Richard Vidgen, University of Bath, UK
Rosina Weber, Drexel University, USA
David W. Wilson, University of London, UK
Ouri Wolfson, University of Illinois at Chicago, USA
Andreas Wombacher, University of Twente, The Netherlands
David Yang, National Kaohsiung Normal University, Taiwan
Jian Yang, Macquaire University, Australia
David Yang, National Koohsiung Normal University, Taiwan
Chee-Sing Yap, TM Net, Malaysia
George O. Yee, Institute for Information Technology, NRC, Canada
Ping Yi, Fudan University, China
Cui Yu, Monmouth University, USA

Keynote speakers

Antony Satyadas leads worldwide competitive initiatives for IBM. Antony has 23 years (18 in USA, 5 in India) of consulting, marketing, entrepreneur and leadership experience with Government and Fortune 500 companies worldwide. Antony is an expert in intelligent systems modelling, knowledge innovation including portals and collaboration/social computing, BPM, and Service-Oriented Architecture. More recently he has been exploring the 4D web and situational awareness. He has more than 50 publications, member of 10 editorial/advisory boards, 40 program/scientific committees, reviewer for several book publications/journals and has offered more than 20 tutorials in this area. His education is in Marketing, Computer/Cognitive science (MS–92, PhD-abd, University of Alabama, USA), and Electrical Engineering (BS–84, University of Kerala, India).

Ian Foster is the Senior Scientist (Associate Division Director) in the Mathematics and Computer Science Division at Argonne National Laboratory, where he leads the Distributed Systems Laboratory, and he is a Professor in the Department of Computer Science at the University of Chicago. He is also involved with both the Global Grid Forum and with the Globus Alliance as an open source strategist. In 2006, he was appointed director of the Computation Institute, a joint project between the University of Chicago, and Argonne. An earlier project, Strand, received the British Computer Society Award for technical innovation. His research resulted in the development of techniques, tools and algorithms for high-performance distributed computing and parallel computing. As a result he is denoted as 'the father of the Grid'. Foster led research and development of software for the I-WAY wide-area distributed computing experiment, which connected supercomputers, databases and other high-end resources at 17 sites across North America in 1995. His own labs, the Distributed Systems Laboratory is the nexus of the multi-institute Globus Project, a research and development effort that encourages collaborative computing by providing advances necessary for engineering, business and other fields. Furthermore the Computation Institute addresses many of the most challenging computational and communications problems facing Grid implementations today. Foster's honors include the Lovelace Medal of the British Computer Society, the Gordon Bell Prize for high-performance computing, as well as others. He was elected Fellow of the American Association for the Advancement of Science in 2003.

Marcus Rogers is a Professor of Computer Information Technology at Purdue University (West Lafayette) specialising in the area of Cyber Forensics. He is also a faculty member with the Center for Education and Research in Information Assurance and Security (CERIAS). Dr Rogers is a Certified Information Systems Security Professional (CISSP), a former Senior Lead Instructor for (ISC)2, a member of the QA team for the SSCP and CISSP certifications and the co-author of the Law Investigation and Ethics section of the CISSP CBK Review Course. He is also a former police detective with a background in computer crime investigations. His area of interests include Applied Computer Forensics, Cybercrime Scene Analysis, and Cyber-terrorism. He has authored several book chapters, and articles in the area of computer forensics and forensic psychology and sits on the editorial board for several international journals. Dr Rogers is a frequent speaker at international and national information assurance and security conferences, and guest lectures at various universities throughout the world.

Eli Cohen founded the Informing Science Institute (ISI), an international organisation of over 500 members from over 60 countries. The institute publishes 8 journals and, so far, a dozen books, all of which are available online to everyone without charge. The organisation also holds two international conferences each year. ISI is an organisation of colleagues mentoring fellow colleagues. It draws together people who teach, research and use information technologies to inform clients (regardless of academic discipline) to share their knowledge with others. Dr Cohen's background is multi-disciplinary. He holds degrees in and has published research in Management Information Systems, Psychology, Statistics, Mathematics, and Education. He has taught in Poland, Slovenia, South Africa, Australia and the USA. In addition, he has conducted seminars in Fiji, New Zealand, Australia, Hong Kong, Singapore, Malaysia, Thailand and Cyprus. Eli Cohen attended (and taught at) Purdue University Calumet many, many years ago. His talk deals with the megatrend of transdisciplinary work and of work teams.

Mark Ciampa is the Director of Academic Computing and Associate Professor of Computer Information Systems at Volunteer State Community College in Gallatin, Tennessee. He has authored several textbooks for Thomson Course Technology, including Network Administrator: Netware 4.1, Networking Basics, and A Guide to Designing and Implementing Wireless LANs. Mark received his Master's degree in Computer Information Systems from Middle Tennessee State University. He has served as a computer consultant for several state organizations and businesses in computer applications and networking, such as the US Postal Service, the University of Tennessee, and the Tennessee Municipal Technical Advisory Service. He is a frequent speaker at national and regional technology conferences

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Conference program

	Monday 8 October	Tuesday 9 October	Wednesday 10 October	Thursday 11 October
09:30–10:00	Opening ceremony	Keynote address (Ian Foster)	Keynote address (Marcus Rogers)	Keynote address (Eli Cohen)
10:00–10:30	Keynote address (Antony Satyadas)			
10:30–11:00				
11:00–11:30				
11:30–12:00	Networking break			
12:00–13:00	Lunch			
13:00–13:30	'e-Society' Session chair: Roger Wallis	'New enabling technologies' * Session chair: Dragana Martinovic	Plenary talk (Mark Ciampa)	'Intelligent data management' Session chair: Sang-goo Lee
13:30–14:00			'Secure technologies' Session chair: Victor Ralevich	
14:00–14:30				
14:30–15:00				
15:00–15:30	Networking break *			
15:30–16:00	'e-Learning and e-Science' Session chair: Ahmed Sameh	'e-Business' * Session chair: Karsten Boye Rasmussen	'e-Governance, e-Health and e-Art' Session chair: Sahin Albayrak	
16:00–16:30				
16:30–17:00				
18:30–21:00			Conference dinner and Best paper award	

* All Tuesday afternoon sessions will be held at the Purdue University Calumet Campus

SESSION RUNNING ORDER

SESSION 'E-SOCIETY'

Session chair: Roger Wallis

Out of the strong, something to eat(Book of judges 14:14)
Aharon Yadin

Sustainable creativity and the challenge to the IPR regime: threats, opportunities, myths and likely developments
Roger Wallis, Jimmy Halvarsson

Designing and implementing an undergraduate program in information systems security
Dragana Martinovic, Victor Ralevich

How can leaders encourage participation in virtual communities of practice?
Indira Guzman, Nicholas Bowersox

Open source migrations: experiences from European public organizations
Andres Baravalle, Sarah Chambers

Competition between mobile TV and broadcast industries
Imsook Ha, Johannes M. Bauer

SESSION 'E-LEARNING AND E-SCIENCE'

Session chair: Ahmed Sameh

Innovative collaboration
Gillian Rawlings

Integrating VoIP into distance learning
Dannan Lin, Charles Shoniregun

E-Learning Status in Arab Countries
Naseem Matar, Ziad Hunaiti, Zayed Huneiti, Mohammed Al-Naafa

The blend of m-Learning and e-Learning at AUC
Ahmed Sameh

Exploring virtual worlds as an extension to classroom learning
James Braman, Andrew Jinman, Goran Trajkovski

Social exchanges theory applied on a web-based learning community
Maximira Carlota da Silva André, Sérgio Roberto Kieling Franco

Experiential Learning: "Teaching citizenship through database case study application, the hurricane Katrina disaster experience"
Barbara Nicolai

Reducing instructor workload in online classes
Joy Colwell, Carl Jenks, Shoji Nakayama

Formative assessment of the effectiveness of collaboration in GCB
David Villegas Castillo, S. Masoud Sadjadi, Heidi Alvarez, Xing Hang

SESSION 'NEW ENABLING TECHNOLOGIES'

Session chair: Dragana Martinovic

Intellectual scrutinizer for compute, storage, network & system characteristics of Linux system
A. Balamurugan, M. Savithashree, H. Sriram, G. Vidya

Asynchronous network for QAE: community schools of African developing countries
Kenedy Greyson, Mussa Kisaka, Damian Haule

Designing web-based business application with multimedia data
Mohammed Hassouna

A Survey of DRM in digital video
John F. Duncan

CRM data grid services
Yongmin Tang

A switch interaction solution for detecting and isolating ARP spoofing
Dengke He, Jiashang YanJiang Du

Enterprise content management: bridging the academia-industry gap
Sergey V. Zykov

Some design considerations in context aware and ubiquitous computing
Charles Shoniregun, Daniel MacCormac, Fred Mtenzi, Mark Deegan, Brendan O'Shea

An information support service for moderators of SME company networks
Heiko Thimm, Karsten Boye Rasmussen, Kathrin Thimm

SESSION 'E-BUSINESS'

Session chair: Karsten Boye Rasmussen

An analysis framework for the economic potential of process interoperation
Reinhard Riedl, Thomas Keller

A model of evolution and ontological development for trust transferring in e-business
Omer Mahmood, John D Haynes

Clustering e-satisfaction factors in tourism industry
Masoomeh Moharrer, Hooman Tahayori

Motivation in organisations: trends in modern ICT companies
Olatubosun Olubusuyi Ojo

The impact of Internet marketing banks in Tanzania
Happiness Joseph Mbuna, Ali Alao Babatunde

Issues and challenges related to online shopping in Saudi Arabia
Inam Abousaber, Anastasia Papazafeiropoulou and Ziad Hunaiti

Improving outsourcing operations by integrating outsourcing determinant index & outsourcing cycle effectiveness
A. Adnan, S. Arunachalam, A. Cazan

Taxonomy and frameworks for improving outsourcing operations
A. Adnan, S. Arunachalam, A.Cazan

Trust and e-procurement transaction management
Joy Okah, Sonny Nwankwo, Charles Shoniregun

Mitigating effect of number of bidders on perceived uncertainty
Ossama Elhadary

SESSION 'SECURE TECHNOLOGIES'

Session chair: Victor Ralevich

An adaptive routing protocol for censorship-resistant communication
Michael Rogers, Saleem Bhatti

Secure communication with SSL remote access VPN
Olalekan Adeyinka, Charles Shoniregun

A new taxonomy for analyzing smart card-based authentication processes
Ramaswamy Chandramouli

Conceptualising and analysing internet threats using a 4-dimensional hypercube
Jan van den Berg

Towards incentive-based cyber trust
Patrick Amon, Russell Cameron Thomas

Language-based security policy enforcement
Fredrick J. Mtenzi, George S. Oreku, Jianzhong Li

Korea prepares for the upcoming ubiquitous society
Byung Joo Jeong

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

SESSION 'E-GOVERNANCE, E-HEALTH AND E-ART'

Session chair: Sahin Albayrak

Security decadence in electronic voting
Cyril E. Azenabor, Charles A. Shoniregun

Democracy development trends as a framework for ede-mocracy
João Paulo Costa, Rui Pedro Lourenço

Towards advanced e/m-Government platforms
Vassilis Meneklis, Spyros Papastergiou, Chris-tos Douligeris, Despina Polemi

e-Health, New enabling technologies or Intelligent data management
Michelle Marie Goulbourne

Developing a web-based intelligent decision support system for personalized healthcare
Chien-Chih Yu, Wen-Liang Kung, Hsiao-ping Chang

Towards collaborative user-centric healthcare services
Carsten Wirth, Paul Zernicke, Sahin Albayrak

Multimedia evaluation: understanding the user-needs gap
Olatubosun Olubusuyi Ojo

SESSION 'INTELLIGENT DATA MANAGEMENT'

Session chair: Sang-goo Lee

Intelligent data classification techniques
T. Shatovska, V. Repka, A. Kharchenko

Survey of agent oriented software engineering methodolo-gies
Mohd Shkoukani, Rawan Abu lail, Saed Ghoul

A proposed model for agent-oriented software engineering
Mohd Shkoukani, Rawan Abu lail, Saed Ghoul

Controlled semantic tagging – how can topic maps support subject indexing in digital libraries?
Hendrik Thomas, Bernd Markscheffel, Tobias Redmann

The wiki way of knowledge management with topic maps
Tobias Redmann, Hendrik Thomas

Data attribute selection with genetic programming
Gina Hope, Joel Hickman, Taehyung (George) Wang

Knowledge management: problems and prospects
Junainah Mohd Mahdee, Mohammad Poorsartep

Blind detection of statistical watermark using extreme learning machine
Anurag Mishra, Rampal Singh, S Balasundaram

A wiki-system with integrity support for structured data
Jaehui Park, Sang-goo Lee, Jonghoon Chun

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007**

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Out of the strong, something to eat (Book of judges 14:14)

Aharon Yadin

The Max Stern Academic College of Emek Yezreel
Aharony@yvc.ac.il

Abstract Over the past three decades, Information Technology has matured and evolved into an essential part of every organization and every society. Information systems, like any other technology, provide many benefits for the users. However, improper use of information systems can lead to undesired and potentially harmful results. During the research, which prepared background materials for an ethics course in information systems education, data from a reselling academic works website was analyzed. By supporting plagiarism, this site demonstrates unethical and undesired information systems usage. However, after analyzing all data extracted from the website and using some simple statistics on this `bad` information, additional `good` facts were discovered. The research provided some interesting conclusions regarding rating of academic institutes, study areas, and academic work, as was reflected by the variety of papers presented on the website. The results enabled us to gain some insight into the plagiarism phenomenon and its distribution among students from the various academic institutes, as well as among the different disciplines and study fields. Additional revealed information includes the average price requested per paper and its changes over the years, average price per paper in the various disciplines and academic institutes, average number of words per paper (in each institute and discipline) as well as the average number of references quoted in the papers. In this case, even with the negative side effects of information systems usage, additional unique benefits were observed

Keywords Intellectual property rights, e-Learning, Information Systems misuse

1 INTRODUCTION

Technology is sometimes defined as an application for enhancing human capabilities. By using this definition, technology in general and information technology in particular provide additional benefits and value to our lives. Information technology spans a wide range of system types and solutions. This technology, which developed over more than thirty years, provides a large variety of capabilities and benefits for the single user [1, 2], organizations [3, 4], communities [5, 6] and society as a whole [7, 8]. However, in many cases, using technology has some negative side effects. Bugs in information systems and/or improper usage can lead to undesired and sometimes dangerous outcome. In her book [9], Sara Baase draws a comparison between fire and computer systems. Like fire, which was given to humans, enhanced their lives, but also caused some terrible disasters, so computer systems enhanced human lives, but also created undesired and dangerous situations. Raymond Kurzweil sees technology as "a double-edged sword, empowering both our creative and our destructive natures" [10]. In a quote attributed to Albert Einstein, he defines technological progress as "an axe in the hands of a pathological criminal". This quote probably refers to a more destructive technology and not to information technology, but, even so it stresses once again that technology usage has its benefits, but also the potential for negative and harmful side effects.

This research goes one step further. It demonstrates that there are additional benefits even to the potential negative side effects of technology usage.

The information for the research was obtained from a website that resells various academic papers, which were uploaded by students. With its support for plagiarism, this website is an example of information system abuse. Even so, analyzing the information reveals some interesting facts about ratings of academic institutes, learning areas, fields and even subjects. The results provided interesting insight into the plagiarism phenomenon in Israel and the way it developed in recent years. In addition, by associating the academic papers on the website with the various institutes and disciplines, the research reveals the development of plagiarism in each academic institute and each learning discipline.

Furthermore, since the academic institutes in Israel do not share information regarding students' submitted papers, the facts revealed in this research hold the potential of being unique. These new facts represent the additional benefits gained from the negative side effects of technology usage.

2 AN ENHANCED WAY OF THINKING

The main purpose of the research is to suggest an enhancement to the ordinary way of thinking regarding technology usage. In addition to the standard two fold implications of technology usage: benefits, followed by negative side effects; the research suggests and demonstrates additional possible benefits to consider. The research question was "Can one find additional benefits from improper information systems usage?" And it relates to the `bad` results of the technology, but also implies that additional benefits, which are direct consequences of these results, do exist.

3 APPROACH AND METHOD

The research was done on data obtained during August 2006 from an Israeli website (www.smarter.co.il) [11] that resells academic essays, research papers, book reports and bibliographic lists. The website, which provides an infrastructure for cheating, utilizes a large database and clearly demonstrates improper usage of information systems. It should be noted that this is not the only site that is engaged in this type of activity, but this is the newest, so it has less outdated and irrelevant materials. The documents stored and maintained by the site are described using several attributes:

- Essay or research paper name;
- The knowledge discipline or study area (biology, art, communication, etc.);
- The academic institute, where the paper was submitted;
- Publication year;
- Total number of words in the document;
- Number of references (sources) used;
- The required paper price.

The research approach concentrated only on academic essays and research papers, and ignored all other types of documents available on the website. Obtaining the information was straightforward, since the website provides all documents` attributes. The research phases included: downloading all the attributes for all the documents (data that is freely available), filtering out the irrelevant materials, deleting anomalies and documents with illogical attributes and analyzing the remaining documents.

The information was downloaded from the website during August 2006. At that time the total number of documents available on the website was 3,800. After the filtering process only 1,740 documents remained.

4 ANTICIPATED BENEFITS

In spite the fact that this website demonstrates improper usage of information systems, it provides unique information regarding the academic industry in Israel. By building a small database with the information downloaded from the website, one can get insight into the Israeli academic dishonesty phenomenon. In addition, it provides various types of information like: paper price distribution over the years, average paper price in the various knowledge disciplines, average paper price per academic institute, and average number of words in papers submitted in various institutes, knowledge disciplines and over the years. Since, there is no other mechanism in place that provides comparison data between different academic institutes, the results obtained by the research hold the potential of being unique.

5 RESULTS

The results obtained in the research were divided into two categories: qualitative and quantitative.

The qualitative results include:

- Course recycling – There are many incidents in which a course requires a term paper and year after year the requirements for the paper remain identical. It was observed in some of the local branches of foreign universities that operate in Israel.
- Duplicate papers – Many papers appear in the website more than once (papers that look identical in all

Figure 1. Yearly distribution of papers submitted

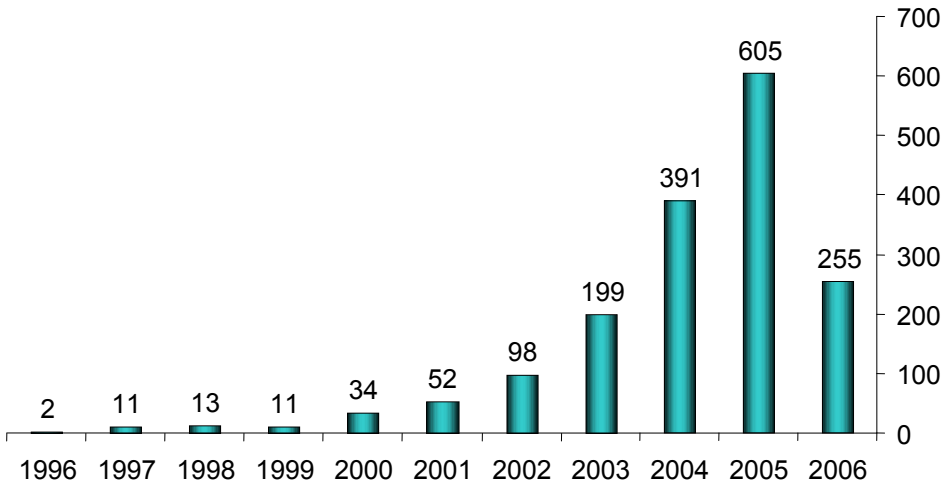


Figure 2. Institute distribution of papers submitted (1)

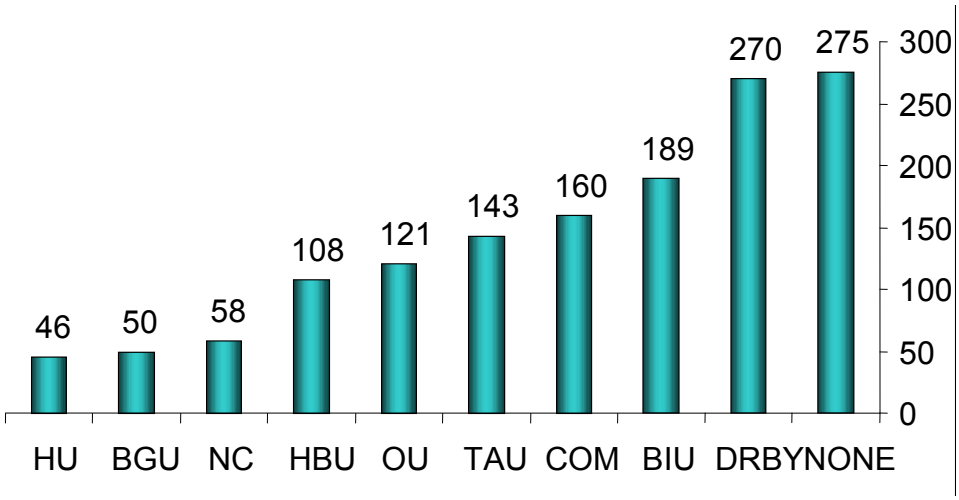
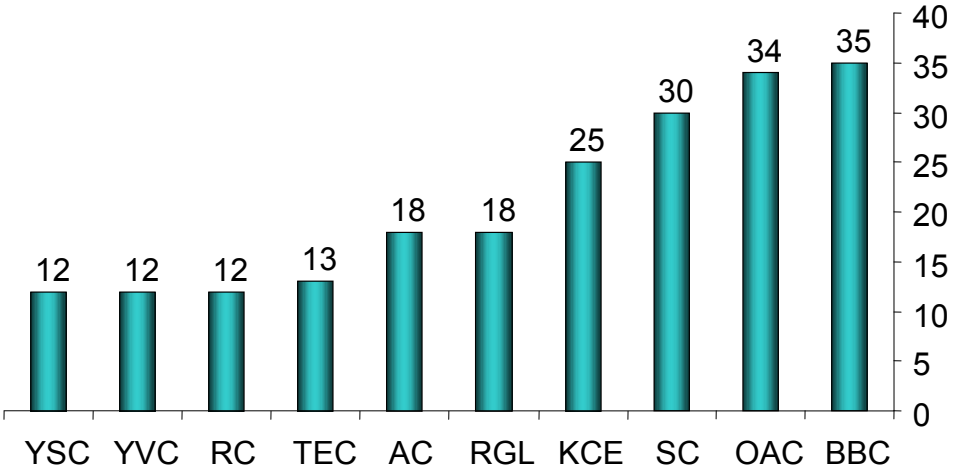


Figure 3. Institute distribution of papers submitted (2)



attributes except one). Explanation for this phenomenon may be (a) a college student uploaded his/her paper and since it did not sell, he/she upload it again, but this time using, what it seems to be, a more prestigious institute (university instead of college for example) and (b) a student who bought the paper, is trying to resell it as if he/she wrote it.

- Paper writers – There is a market for paper writers. These are not students that upload their papers, but ordinary people (probably graduates) that are engaged in writing papers on various subjects for the purpose of selling them later. The important issue here is not the fact that there is a market for writers, but what can be understood for the existence of this market. If people are willing to spend time on researching and writing documents hoping that they will sell, it implies that there are (many) buyers.

The quantitative results include a wide range of analysis reports that provide insight driven from different angles.

5.1 Number of papers

The first quantitative result relates to the number of papers submitted to the website over the years (Figure 1). This research was done during August 2006, so due to the war, the results for 2006 are incomplete, but they were included as well.

The yearly distribution of papers submitted to website raises two conclusions:

Absolute numbers – Considering the fact that there are over 200,000 students in Israel, the absolute number of uploaded papers is (still) quite small. This may imply that the cheating phenomenon is still very limited and well under control.

Yearly increase curve – In spite the small numbers, the curve describing the number of papers increase demonstrates the potential magnitude of the problem. It can be noticed that in the past years, the number of papers almost doubled every year. If this trend continues, it potentially will become a major problem in a couple of years.

A different report (Figures 2 & 3) was produced for comparing the total number of papers submitted by students of the various academic institutes. The full analysis report contains two additional figures that describe the lower end distribution. These two figures were omitted from this paper.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. Papers submitted increase at the Open University

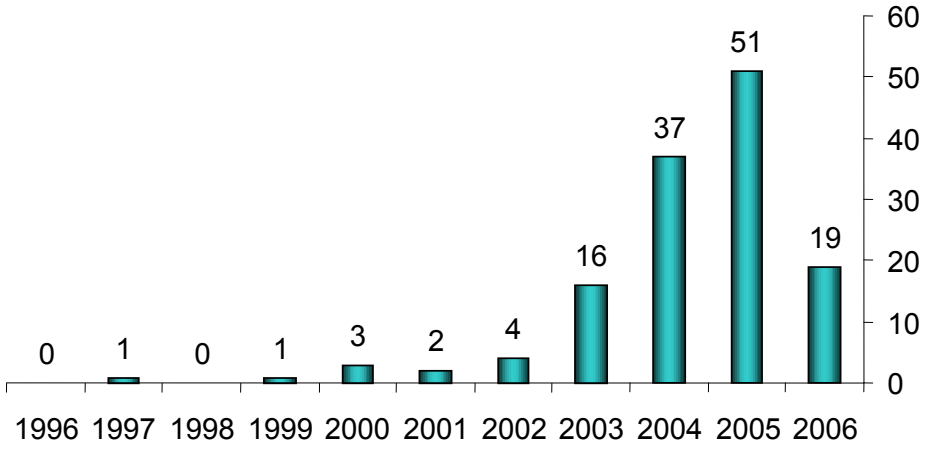
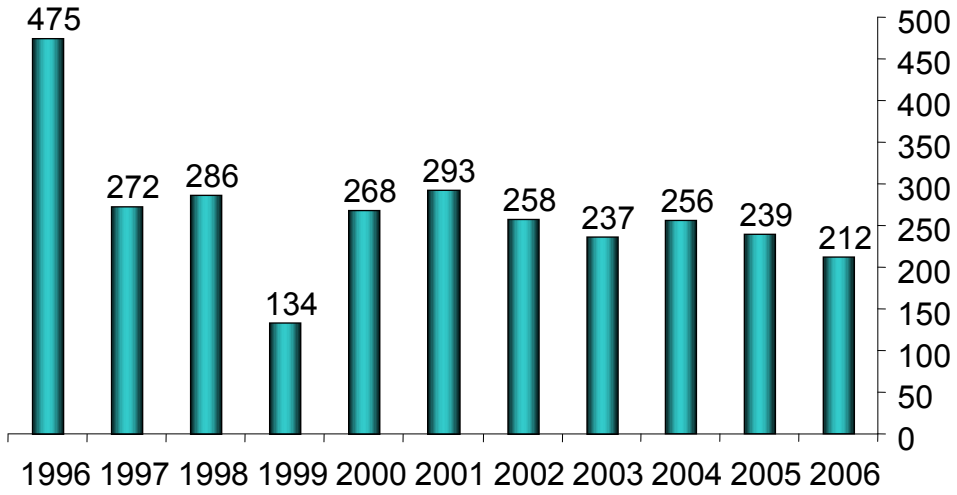


Figure 5. Average paper price over the years



All together there were over 40 academic institutes mentioned in the research, however, only the first 20 are included in the figures. The naming convention used is as follows:

- NONE – No institute was mentioned.
- DRBY – University of Derby extension.
- BIU – Bar Ilan University.
- COM – The College of Management.
- TAU – Tel Aviv University.
- OU – The Open University of Israel.
- HBU – The Hebrew University of Jerusalem.
- NC – Netanya Academic College.
- BGU – Ben Gurion University of the Negev.
- HU – Haifa University.
- IDC – Interdisciplinary Center Herzeliya.
- BBC – Beit Berl Academic College.
- OAC – Ono Academic College.
- SC – Sapir College.
- KCE – Kibbutzim College of Education.
- RGL – Ramat Gan Law College.
- AC – Achva Academic College.
- TEC – Technion - Isreal Institute of Technology.
- RC – Ruppin Academic Center.
- YVC – The Max Stern Academic College of Emek Yezreel.
- JSC – The College of Judea & Samaria.

An additional report included the yearly increase in number of papers per each academic institute. This report includes absolute numbers and it does not take into account the number of students in each institute. The full research paper includes over 40 graphs in this category. For the purpose of demonstrating an example, the report for The Open University is included (Figure 4).

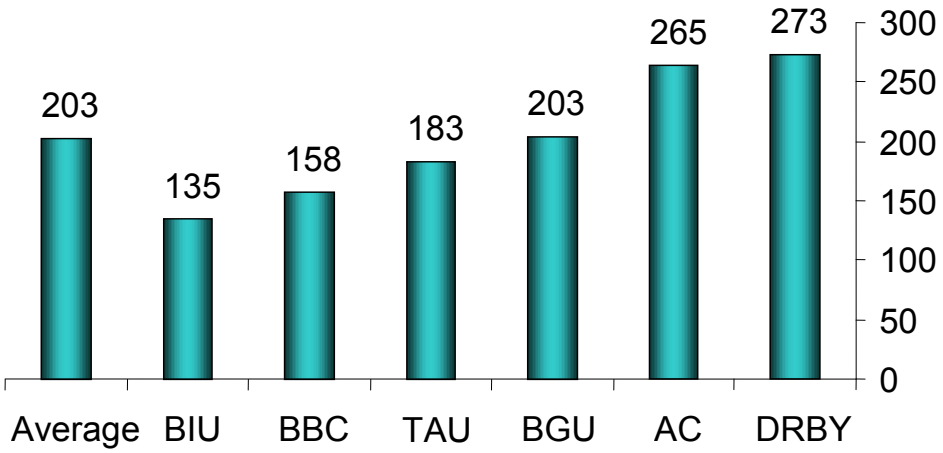
In a similar way, additional reports were produced to demonstrate the distribution of papers over the years per each knowledge discipline and even distribution over the years per discipline and per each academic institute. These reports were not included in this paper.

5.2 Papers price

An interesting report is the distribution of the average paper price (in local currency) over the years (Figure 5). As can be seen, the price is almost constant (with a small decrease over time). The average price for a paper is roughly 50 USD. The relative cheap average price of the documents represents a problem, and it raises true ethical questions. In this case, the price is not a barrier anymore and what prevents students from buying (and cheating), is only their ethics.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 6. Average price of anthropology papers in the various academic institutes



As described previously, additional reports were produced. One report, which is not included here, provides the average price distribution per academic institute. A more elaborated report (Figure 6) provides the price distribution for a specific knowledge discipline per academic institute. Figure 6 is an example of the price of Anthropology papers in all available institutes.

5.3 Average number of words per paper

The number of words is no indication to the quality of the paper, but even so it can provide some insight into the requirements and standards in the various academic institutes. It is possible to produce a report outlining the distribution of the average number of words in the papers submitted over the years and per institute. This report is too general and it does not provide meaningful information. However, a report (Figure 7) that includes the average number of words in papers submitted under a specific knowledge discipline in various academic institutes provides an interesting base for comparison.

Figure 7. Average number of words in Sociology papers

6 CONCLUSIONS AND FUTURE WORK

This research relates to information technology abuse and its consequences. Technology in general and information technology in particular provide a wide range of benefits. Some of these benefits may be used in illegal or unethical ways, which produces the negative side effects of the technology. However, this research suggests and demonstrates that even from the negative side effects, additional benefits can be gained. The data for the research was downloaded from Smarter, a website which resells students` work. By supporting cheating and plagiarism, this website is a demonstration of improper usage of information systems. After analysing all the papers attributes, some interesting facts were revealed regarding the plagiarism phenomenon in Israel, in each academic institute and in each discipline. The overall magnitude of the problem is still manageable (605 papers were uploaded to the website on 2005, by a small fraction of the enrolled 200,000 students). However, the yearly rate at which papers are uploaded is alarming. On average the number of papers is roughly doubling every year, without any significant change in the number of students. Similar facts regarding the phenomenon in each academic institute were observed. The paper outlines one example (for the Open University, in which the yearly rate is identical), but the information gathered provides the base for observations regarding all other academic institutes and academic disciplines. Additional interesting findings include the average price required per paper (~$50) and its changes over the years. Similarly, the research revealed the average paper price in each of the academic institutes and the various disciplines. Similar reports were produces regarding the average number of words and the average number of references used in a paper, per institute and per discipline. All these reports were produced for a specific year and for a range of years.

Future work related to the research described above includes revisiting the site and analysing the documents in order to establish trends and increase in the number of papers submitted. Additional interesting work is related to the parameters used by students for defining the paper price. Is the price based on paper length, number of references, the academic institute prestige or may be the amount of work put into preparing it.

Even with these additional unique benefits, it should be noted that this paper in not advocating or promoting improper use of information systems.

REFERENCES

1. D.L Goodhue, (1995), "Understanding user evaluation of Information Systems", Management Science, Vol. 41, No. 12, pp. 1827-1844.
2. Y. Yoon, T. Guimaraes, and Q. O'Neal, (1995), "Exploring the factors associated with Expert Systems Success", MIS Quarterly, Vol. 19, No. 1, pp. 83-106.
3. S. Shang, and P.B. Seddon, (2002), "Assessing and managing the benefits of enterprise systems: the business manager's perspective", Information Systems Journal, Vol. 12, No. 4, pp. 271-299.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4. G.G. Gable, D. Sedera, and T. Chan, (2003), "Enterprise systems success: a measurement model", Proceedings Twenty-Fourth International Conference on Information Systems, Seattle, USA, pp. 576-591.
5. F. Berkers, (2004), "Rethinking community based conservation", Conservation Biology, Vol. 18, No. 3, pp. 621-630.
6. B. Whitworth, and A. de Moor, (2002), "Legitimate by design: towards trusted virtual community environments", Proceedings of the 35th annul Hawaii International Conference on System Sciences, pp. 2831-2842.
7. J.R. Skees, and B.J. Barnett, (1999), "Conceptual and practical considerations for sharing catastrophic systemic risks", Review of Agricultural Economics, Vol. 21, No. 2, pp. 424-441.
8. B. Wellman, (2002), "Designing the internet for a networked society", Communication of the ACM, Vol. 45, No. 5, pp. 91-96.
9. S. Baase, (2002), "A gift of fire – social, legal and ethical issues for computers and the Internet", Prentice Hall.
10. R. Kurzweil, (2003), "The Promise and peril of technology in the 21st century", CIO.
11. www.smarter.co.il, Aug. 2006

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Sustainable creativity and the challenge to the IPR regime: threats, opportunities, myths and likely developments

Roger Wallis
Jimmy Halvarsson

Dept of Media Technology & Graphic Arts
School of Computer Science and Communication
Royal Institute of Technology (KTH)
SE 10044 Stockholm, Sweden
rogerw@kth.se

Abstract Content industry rhetoric concerning file-sharing has been particularly aggressive ever since the emergence of the Napster P2P service. Even if "fundamentalist" elements in the audio- and audiovisual industries continue to believe that suing randomly chosen file sharers will stop P2P activities. Several developments suggest some type of pending licensing solution which makes these activities legal, rewards creators and satisfies the needs of network operators for whom P2P activities are a significant source of traffic. This paper refers a) to general research in Sweden which indicates that those who download illegally consume more cultural products than those who follow the law, and b) to our own research indicating that heavy down-loaders are willing to pay for such a service to be legal under certain conditions. Prerequisites include non-intrusive DRM systems, a growing awareness amongst politicians and regulators that current strategies cannot stop P2P usage, and are encouraging more anonymous network activities, and a realisation that the whole copyright regime could become a victim if societal supports withers. Our research indicates the music and film industry's legal strategy, and the related rhetoric of "stealing from artists" has backfired. Consumers, particularly younger ones, are becoming less and less convinced that a reasonable share of revenues large media companies receive ever go back to the artists and creators of the works they exploit. This raises even more important issues relating to creativity in society and the IPR regime's ability to survive with societal support. We conclude with a focus on the acute dilemma facing policy-makers, legislators and regulators, via a proposal for a model for sustainable creativity

Keywords Peer-to-peer (P2P), copyright/IPR regime, internet regulation, business models, digital policy, music industry, creativity, cultural diversity.

1 INTRODUCTION AND BACKGROUND

History repeats itself. The catalogue of disruptive technologies that have disturbed the media industries provides a long list of attempts to block developments, rhetoric directed at politicians and the courts, and finally an acceptance of the new technologies with impressive subsequent revenues being generated from new business models.

Such was the case when the phonogram appeared in the late 1800s – pianola manufacturers tried to block the new technology, arguing that it would lead to massive job losses. More strikingly, when radio appeared in the USA in the early 1900s, music publishers took a similar approach. If consumers could hear music over the radio, whenever and how often they wanted, then the value music (and copyright) would collapse, job losses would follow and sales of sheet music would wither (Kusak & Leonard, 2005). The argument is remarkably similar to that we have heard since new copyright legislation for the digital age has been introduced. The radio problem was solved, of course, by forming a copyright collection society, ASCAP, which negotiated blanket licences with radio stations and then forwarded net revenues to the publishers and composers. The case of the VHS cassette (via Betamax) is another milestone in this series of events. So a researcher must ask if we are witnessing yet another repeat or does the current situation involve a paradigm shift? The similarities are striking. The difference is that digital copyright law provides rights holders with greater abilities to control how, when, where, how often etc. consumers can use cultural products, than in the analogue world. And this has created the strategy of suing randomly chosen individual consumers who comprise a tiny percentage of the growing numbers of file sharers.

On the other hand it is becoming increasingly clear that those who engage in file-sharing consume as a rule more cultural

products than those who do not – P2P networks provide a "virtual playground" where consumers can seek out old or new experiences. Data supporting this is available both from the UK (Musically 2005) as well as from Canada (CRIAA 2006) in studies commissioned by the recording industry. The virtual experience seems to trigger off a need for physical products and experiences. Support for this postulate also comes from the observation that attendances at live music concerts has increased between 5 and 10% in Sweden every year since 2002. Record companies have suffered, mainly because their contracts with artists do not normally include a share of extra revenues such as those from concerts and merchandising.

2 CONTENT INDUSTRY RHETORIC – FACTS OR MYTHS?

Two words have dominated the industry rhetoric: piracy and free music. File-sharers seeking their favourite music from other network members have been put in the same category as commercial pirates running illegal CD-factories. KTH research suggests that this has had the effect of decreasing, in particular, younger persons' respect for both copyright and the established music industry. This is enhanced by the fact that many smaller record companies working with lesser-known, local artists, rely on P2P networks for marketing. One such record company enthusiast told us " we release all our recording free as MP3 files over the net. We know that if 100,000 download tracks from an album, we will sell around 7 – 8000 CDs, and many fans will come to the band's concerts" (Lövkvist,M 2007).

That downloading music and films illegally is "free" is also a well-expounded myth. The downloader has to own a computer and, as a rule, pay an ISP or supplier of broadband a monthly fee. As early as 2003/4, KTH researchers concluded the following:

"Swedish consumers' annual spend on finding and downloading Internet music exceeds by over 50% the annual net revenue of the Swedish recording industry. A similar relationship would seem to be valid in the whole of the industrialised world. The music industry's strategy should be to encourage file sharing via collaborating and revenue sharing with telcos/ISPs, rather than aggressively discouraging Peer-to-Peer activities.

The report also noted:

Even a conservative estimate of consumers' annual investment in acquiring music from the Internet results in figures that are far greater than the annual net revenue (revenue minus costs) of the Record industry from selling "legal" CDs.

*The 2003 report estimated that Swedish consumers spent **125 million US dollars** on such activities during 2002, and that the global figure, after a sudden fall-off with the closing of Napster, had risen from **6 billion dollars** to over **11 billion dollars** annually. The latter is the result of the explosive growth of programmes such as Kazaa. These figures can be compared to estimates of the annual profit of the recording industry in Sweden (**82 million US dollars**) and globally (**3.8 billion dollars per annum**). (Landegren.J, Liu.P 2003)*

Another "myth" has concerned the causality between file-sharing and decreased CD sales. In an overview of the many research reports from this area, Edström-Frejman (2007) has observed that most conclusions should be taken "with a grain of salt". Even if some groups who file share regularly have clearly decreased their spending on legal CDs, then other factors can be the cause. One is a shift of spending from CDs to concerts. The other is alternative demands on the wallet, not least costs for mobile phones. (Edström-Frejman (2007)

3 FROM MYTHS TO AS CLOSE AS WE CAN GET TO FACTS

One author of this paper recently concluded a study with in-depth interviews with regular file sharers, divided into 2 age groups, high school students and university undergraduate students. (Halvarsson. J. 2007) The main task of this study was to shed more light on the behaviour of peer-to-peer-users, their similarities and differences, forms of usage, motivation and moral stance. Even if it is illegal to use free services for peer-to-peer sharing of copyrighted material (EC 2001), at least twenty percent of the Swedish population have used this method. Therefore it is highly relevant to examine peer-to-peer users' own thoughts and opinions. Through interviewing a group of university students and a group of high school pupils who currently use peer-to-peer networks or have been using them previously, the study sought to test the overall hypothesis that "if the conditions are right more consumers will be prepared to pay for digital material on Internet". Another goal was to draw conclusions regarding the prerequisites for convincing users that use free systems today to start using alternatives where they pay to download music

The target groups had been chosen because of the observed fact that frequent peer-to-peer sharing occurs in these two groups. The study concluded that several other factors than merely economic issues affect attitudes among persons who regularly download. University students in the study emphasize that factors such as range of choice, user-friendly applications and absence of platform limitations (inter-operability) are important. Another requirement for attractive pay systems was that they should include other forms of media, for example movies and games, and not just music. The most important factor for the high school pupils, however, was more a moral one, the question of who gets what, in other words, how much of the revenues generated actually go to the artists. Regular content industry claims that if you "can get it for free, then you will not be willing to pay for a service" were disproved. Both university students and high school pupils are prepared to pay between one and two hundred Swedish crowns (€11–€20/month) for a pay system that fulfils these demands. The result in general confirms the hypotheses "If the conditions are right more consumers **will** be prepared to pay for digital material on Internet".

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

A key observation here is that consumers demand the same range to choice before being willing to migrate to a pay-on-demand service. Here a problem arises. The audio and film industries have been slow in offering legal downloading services, and then only via client-server solutions. Such solutions are extremely sensitive to sudden surges in demand. The efficiency of P2P networks as systems for distribution suggests that any comprehensive solution to file sharing will entail making P2P legal rather than replacing P2P networks with client-server alternatives where consumers have to form a queue to get what they want.

The reality of the Swedish P2P landscape has become clearer over time, not least thanks to longitudinal studies of the Swedish population's media behaviour from the SOM Institute at Gothenburg University, Sweden. The latest report, in Swedish, entitled "Film, pirates and a sinking ship" (Antoni R. 2007) concludes that a) downloading of film does not scupper the cinema industry, b) that there is a diminishing degree of support in society for a policy of taking down-loaders to court, and c) that the attitude of the authorities towards file sharing can be a major issue in the next Swedish general election. The last point might surprise an observer not familiar with Sweden. The last election in September 2006 saw the emergence of a new political party, with growing support among younger voters, namely the Pirate Party. This was a one-issue party with the demand for the legalisation of file-sharing as its primary agenda.

The most striking SOM results concern the cinema industry. Cinema ticket sales were up in 2006 from 14.6 million (2005) to 15.3 million, almost the same figure as 1991 (long before file-sharing became a popular pastime). In every age group, more regular down-loaders went to the cinema at least once a month, than those who did not. This was most noticeable in older age groups. Amongst 65-85 year olds interviewed, 7 percent of those who never downloaded films illegally visited a cinema once a month or more. The figure for those who engaged in illegal downloading was 33%. A similar trend could be noted regarding CD sales.

One can reasonably conclude from this that the cinema business would collapse without the availability of illegal file-sharing networks where consumers can sample films and decide whether or not they wish to see them on the big screen.

So if this is the case, then what will be the likely developments? The law makes most file-sharing illegal. For the consumers it is hard to know what materials can or cannot be downloaded, i.e. where all rights holders have or have not given their specific permission and cleared the material. Alternative copyright solutions such as Creative Commons thus become more attractive to those who desire their works to reach as many people as possible via as many channels as possible. But even this can lead to long-term problems for the individual creator. There is no way to revoke a Creative Commons license once it has been granted. But we do maintain that some sort of a solution making P2P legal will be inevitable, if the correct prerequisites are in place.

Bearing in mind the intensity of incumbent content industry attacks on Peer-to-Peer technology and its users, one tends to feel sympathy for views such as those expressed by De Cleen (2005):

"Given the power of the copyright lobby in influencing legislation and its prominent position in general debates on culture, a more critical treatment of cultural industry discourse by academics is desirable".

But it is also worth considering some more economic aspects – after all, with P2P distribution, marginal costs for making a copy go down to near-zero. If there is no remuneration system which rewards creators, then creativity can clearly suffer. But do creators get just reward with current legal sites – composers for instance can receive less than the credit card company that enables the transaction every time a song is sold for 99 cents from iTunes!

4 A FURTHER ECONOMIC ANALYSIS; THE CONCEPT OF EXCLUDABILITY

Economists analysing copyright in the digital era have highlighted the dangers of using the law to "heighten excludability", i.e. to use exclusive rights to expand control over what consumers can or cannot do (with or without demands for payment). As excludability increases, so does the initial potential for higher revenues. But so do the costs for policing and implementing such copyright-based demands (Pickard 2004). At some point the costs will exceed revenue increases and the activity will become counter-productive. This has already happened with file-sharing.

There is no publicly available data to indicate that the substantial revenues gleaned by the major record companies and their trade body, the IFPI (RIAA in the USA), from fines and out of court settlements with file-sharers and a variety of new businesses, have led to an extra cent in royalties to artists and composers. The costs have been swallowed up by lawyers' fees and trade bodies' costs. And file sharing is still increasing according to Big Champagne, a monitoring firm regularly used by the major record companies for P2P intelligence. P2P file-sharing grew 14% in the USA in January 2006 compared to the figure for January 2005, with an estimated average number of 7 million simultaneous users (Anonymous, Music and Copyright, 2006). The Swedish SOM report based on data collected during 2006, referred to above, also notes that file sharing in Sweden has increased from 800,000 (regular file sharers over 18) to 1.2 million or 14% of the population.

The heightened excludability argument is significant. If point where costs outweigh revenue increases has already been passed, then this suggests that a speedy solution that legalises and monetises file sharing should be an imperative and attractive business solution.

Economists also observe another danger with a continued heightening of excludability. The purpose of copyright is not only to protect a work, but also to create more innovation and economic value (Towse, 2001). Control that limits the

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

ability to adapt or improve existing ideas can hinder innovation. One can also reach a point where demands placed on consumers clash with general concepts of what is reasonable regarding limitations on ownership of a cultural product. The SONY-BMG "root kit" debacle (Billboard 2006-01-07), involving espionage and risks to a customer's property brings us closer and closer to this point where societal understanding of and respect for copyright could collapse. It has certainly given anti-music industry consumer groups all the free ammunition they could ever dream of.

Copyright is a very vulnerable animal, requiring for its survival a careful balance of interests between users and creators. Current music industry strategies that focus so heavily on what users cannot do, with very little mention of the positive role as an economic incentive to create, could lead to collapse of the very source if income they are supposed to defend. As Frith (2004) points out in his introduction to the 2nd edition of "Music and Copyright" – "the notion of fair use – once essential in the attempt to balance the interests of authors and users of a work – has been systematically marginalised"

5 PREREQUISITES FOR LEGALISING FILE-SHARING ACTIVITIES, WITH A REVENUE SYSTEM FOR RIGHTS HOLDERS

The following prediction is proposed.

File-sharing within a variety of P2P networks will be come legalised within three to five years, based on some form of simple licensing system which allows consumers to share content (and opinions of the same), and to even modify content within acceptable limitations of moral rights. Revenues will be generated and distributed to rights holders according to best possible non-intrusive monitoring of what is actually shared. The use of intrusive Digital Rights Managements (DRM) systems to control and spy on individual's activities will be rejected by consumers.

Let us consider the factors that support this vision via the prerequisites for it becoming a reality:

- The development of refined query and search tools in P2P networks allowing users to find a "needle in a haystack", thus supporting general goals of cultural diversity and creative activity/interactivity, rather than mere short cuts to newly released popular content. Current studies of requests in P2P networks suggest a shift from Top 10 hits to a broader range of above all older and more obscure works.
- A non-intrusive DRM system intended primarily for monitoring usage via collecting aggregated data, rather than for controlling what individual users can or cannot do. Some form of watermarking or tagging will be required to identify works that are swapped. Consumers will reject highly intrusive DRM systems (Blomkvist,U., Fritzell.M., Olofsson. M.2005).
- An awareness amongst legal authorities of a) that file sharing is so widespread in networks that tend to become more and more anonymous that current legal actions cannot stop it, and b) that the possible imposition of some form of compulsory licensing might be necessary if major content owners refuse to voluntarily embrace the notion of legal P2P networks. A related factor is the negative effect on society's overall respect for the law when specific legislation cannot be effectively enforced.
- Growing awareness that actions aimed at chasing users of popular P2P networks such as Kazaa, combined with actions to clog up such networks by filling them with polluted files ("spoofing") will lead avid file sharers to seek more anonymous networks where it is harder to be identified. This development towards "darknets" could cause considerable societal problems as regards fighting other forms of criminal activities and has even caused concern among senior Microsoft researchers (Biddle, England, Peinada & Willen 2002, Edström-Frejman 2002).
- Growing ground-swell pressure from those who are developing new business models for the so-called "free download" environment of the Internet, as well as a growing awareness amongst major content owners that activities within P2P networks are a vital source of marketing intelligence. An increasing need to be able to inform P2P users that specific material is available for consumption, encouraging movements such as Creative Commons.
- Growing concern amongst creators and performers that the emerging "legal download" services give the producers a very large percentage of the proceeds, but leave very little for performers and writers. This was the main reason for the French performers' and artists' organisation, Spedidam, to lobby the French parliament in 2005 to introduce a so-called "global licence" to make P2P legal. Revenues generated would be distributed between producers, performers and authors/publishers (Spedidam 2005).
- Growing appreciation of the need to be able to access existing content of different forms and use it as an inspiration for new ideas, alternatively as base material which can be modified/improved (open content/user generated content).
- A growing number of studies, some even financed by the music industry, that conclude that avid file-sharers include groups of individuals who are the most active consumers of culture and cultural products such as CDs, legal downloads and concert tickets (Beauvillian 2000, Musically 2005, CRIA Canada 2006).
- An acceptance by telecom operators providing Broadband services of the need to be involved in the billing process (revenue sharing, subscription, micro-billing etc.), as opposed to a rejection based on lack of conduit responsibility in accordance with the WIPO 1996 Copyright treaty and the EU Copyright Directive of 2001.
- A speedy and proactive policy by authors' collective management societies to start the process of negotiat-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

ing with those who provide and derive revenues from broadband Internet access services.

6 MORE THAN ONE LEGAL AND REGULATORY REGIME INVOLVED

Can the IPR regime survive the current pressures, with a likely growing degree of mistrust in society, as well as a legal regime which becomes harder and harder to implement?

Here the problem seems to involve the relationship between two different legal regimes, IPR/Patent law and Competition/Anti-trust law.

Copyright law has given particularly large owners or controllers of copyrights, immense power to control how consumers have access. Via DRM systems, one can block a user's ability to manipulate or alter files. To circumvent such control systems is a criminal offence. If one considers that little, if any, creativity appears in a vacuum, but involves borrowing and improving existing ideas, then one can postulate that the essence of the Digital Millennium Copyright Act and its European equivalents pose a latent threat to creativity. This is particularly so within the computer software sector where a growing open-source/open content movement is responsible for most innovation.

At the same time, Competition Law has not been able to limit the growth of vertically-integrated conglomerates. These amass huge repositories of rights and patents and tend to use them as a combined tool to attack others who encroach on their rights. This means a) that a gigantic music publisher with in a media conglomerate no longer actively promotes each work an author has hander over, and b) that actual market power can exceed the market share of different parts of the conglomerate. The latter is particularly evident in the case of music/film companies that control both publishing rights and production rights (to sound recordings or films). Thus we see the types of business deals where record companies threaten to sue a new content aggregator such as YouTube, and then settle with a slice of equity in YouTube a week before it is purchased by Google. The record companies make a windfall on the share, but no revenue goes back to the individual artists and composers whose works were a prerequisite for the deal. Several similar cases are on record during these recent turbulent post-Napster years. That down-loaders have little faith in the desire of the major media companies to financially reward their hero artists is hardly surprising. It does amount to a major threat to sustainable creativity in society and the IPR regime.

7 WILL SOCIETAL SUPPORT FOR COPYRIGHT WITHER?

Our conclusions from available data, qualitative and quantitative, are that the copyright regime itself could become a victim in the wake of the current turbulence, subject to a number of "ifs":

- if the **control function** takes too much precedence over the **economic incentive** function, i.e. if copyrights are used more to control usage than to support and foster new creativity.
- if the creator loses too much control over IPRs to agents and/or intermediaries (this seems to have occurred in the audio/audiovisual industries).
- if **innovation** based on improving existing ideas is hindered by the **degree of control** by rights holders.
- if the users collective does not accept that the "balance" (**control/permitted use**) is reasonable.
- if a reasonable share of revenue generated by agents, using IPRs, is not seen to filter back to the original creators.

If these factors have indeed triggered off a diminishing degree of understanding and respect for the IPR regulatory regime in society, then policy makers and those who implement the policies face a serious dilemma. Can we handle a democratic society where over 14% break the law, and a handful are chosen at random (to create fear /set a few examples) and taken before the courts. They then meet judges whose own children probably also download illegally.

OR

Do we seek ways to achieve a society where file sharing thrives, with reasonable payments to creators based on usage and popularity, comprising a haven for innovation, curiosity etc.

8 AN ATTEMPT TO UNDERSTAND THE POLICY DILEMMA FACING POLITICIANS, LEGISLATORS AND REGULATORS – AND A PROPOSED ANALYTICAL TOOL

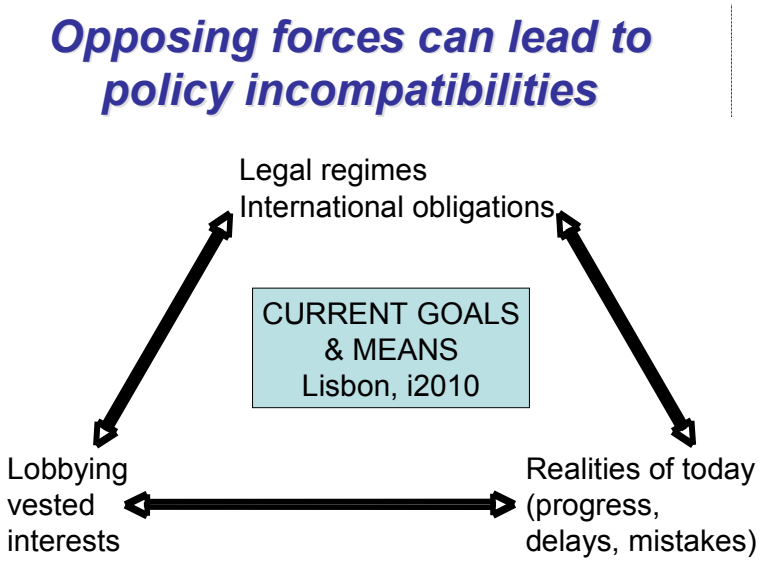
Policy makers, legislators, civil servants are all subject to a number of opposing and sometimes incompatible forces. We have endeavoured to visualise this situation in an attempt to better understand today's confusion as regards what the IPR law says and reality.

One problem involves time frames. Many basic legal instruments such as international conventions and agreements take many years to negotiate and inevitably can be out of phase with technology. This seems to have happened in the IPR arena, with current legislation based mainly on the WIPO copyright treaty from 1996 - years before most people had heard of something called file sharing, even if computer experts had already started using the same technique to share resources (so-called GRID technology). The WIPO treaty negotiations were also subject to considerable lobbying from vested interests, not least the telecommunications industry which was keen to enjoy a lack of "conduit responsibility" in a digital networked world.

At the same time politicians fix overall goals. In Europe the EU has adopted the so-called Lisbon agenda, describing goals for Europe to be the world's competitive Information Society by 2020. Or sub-goals along this route such as the i2010 programme (i2010, European Commission 2006).

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Figure 1. The policy dilemma



On the other hand, there is also an on-going dynamic reality, where metrics point to different rates of change or movements in certain directions. All these have to be taken into account when implementing or revising policies/legislation. Balancing these sometimes opposing forces seems to have been a very tricky process for governments and legislators.

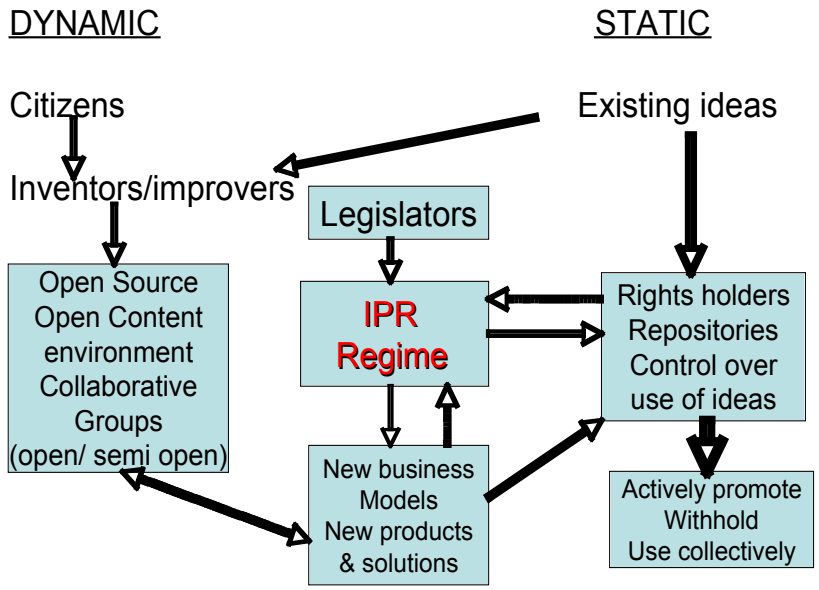
How should decision- makers handle this dilemma)? The answer is to invoke legislation which supports a high degree of innovative dynamics. Maybe one should return to some of the early debates regarding the information society. In an early document from the European Commission on the emerging Information Society (1994) one could read the following:

"IPRs are an important factor in developing a competitive European industry… their protection must continue to be a high priority, on the basis of balanced solutions which do not impede the operation of market forces" (EC 1994). In more than one document from the same period one can read of the need to balance the IPR regime with the need to allow new business models to develop. This does not seem to have happened.

Our attempt to visualise the challenge follows below in figure 2.

For creativity to be sustainable there must be a balanced mix of protection, flexibility to allow improvements to existing protected products, and a need, probably via improved anti-trust measures, to contain large owners of copyrights/ patents whose first priority is not to exploit all they control as actively as possible in the market. We need a flexible regime where groups of innovative creators have considerable freedom to experiment, seek new ideas and improve existing ones without running the risk of legal redress from large owners of copyrights or even patents controlled by so-called patent trolls (Cane.A. 2006). Much can be learnt from the expanding open source /open content movements and the development within these of new business models. Maybe

Figure 2. Sustainable creativity model



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

there is a need to adjust patent and copyright law to put more intellectual property in the public domain if it has not been exploited actively by an owner for a certain period of time. Owners of content should be restrained from withholding materials from the public as well from merely using immaterial rights collectively to restrain others from developing new products and services.

9 FINAL WORD – WHAT IS ACTUALLY HAPPENING?

The music industry is starting to embark on attempts to a) produce a common stand regarding activities in digital networks (presently many different views abound), and b) to consider some form of licensing of "illegal" up and downloading. This became clear after an interesting meeting in Norway in June 2007 where different sectors were present. One journalist wrote after this meeting:

"At the Norwegian summer resort of Kristiansand in Norway last week, representatives of all corners of the British (and global) music business came together to think the unthinkable. That's unusual in itself. What's generally called the "music industry" consists of violently opposed parties: small labels against big labels; publishers against recording rights owners; managers against everyone else. Much time is spent screwing or suing one another. The past sure is tense.

Perhaps for the first time, the entire range of representatives - including the biggest and most powerful recording interests - looked into the abyss, and agreed that what they need to do is very different to what they've been doing for the past 15 years - ever since the sudden growth of public computer networks represented by today's free-for-all internet." (Orlowski.A, 2007).

The meeting also heard that suppliers of broadband services also have problems as they head for new investments in even faster networks – cut throat competition for broadband subscribers means they need others sources of revenue. This could herald a departure from the argument that they have no responsibility for content in their networks (no "conduit responsibility"), to one where they accept an economic responsibility for a premium service which allows consumers to do what they already do, but pay to make it legal. Income could then be shared between operators and rights holders – hopefully via a system which provides an economic incentive to create.

REFERENCES

1. Antoni R. (2007), "Film, pirates and a sinking ship", Annual report on downloading/uploading audio/audio-visual files in Sweden from the SOM Institute, Gothenburg University, Sweden. Available at: http://www.som.gu.se/SOMseminariet2007/presentationerna.htm (accessed 2007-05-28)
2. Beauvillian O. (2000), "Selling Music On Line – an analysis of Napster in Germany", Paper at PopCom 2000, Jupiter Research, Germany
3. Biddle P., England P., Peinado M., Willman B. (2002), "The Darknet and the Future of Content Distribution", Proceedings of the 2002 ACM Workshop on Digital Rights Management, November 18, 2002, Washington, DC. Originally published at http://www.crypto.stanford.edu/DRM2002/prog.html.
4. Billboard (2006), "Sony BMG agrees to DRM settlement" 20060107 pp5-6
5. Blomkvist U., Fritzell M., Olofsson M.(2005), "DRM- Intrusion or Solution?", Proceedings of e-Challenges Conference, Ljubljana in "Innovation and the Knowledge Economy", Ed. Cunningham,P, & Cunningham, M. IOS press Amsterdam pp 925-933
6. Cane A. (2006), "Trolls control the rickety-rackety bridge of intellectual property", Digital Business Supplement, September 20, 2006, Financial Times, London
7. CRIAA (2006), "CRIA Consumer Study of Radio and Music Survey Results" February 2006, POLLARA Inc.Toronto (available at http://support.crtc.gc.ca/applicant/docs.aspx?pn_ph_no=2006-1&call_id=29786&lang=E&defaultName=Canadian%20Recording (accessed 2006-10-22)
8. De Cleen B. (2005), "Piracy on the Internet. A critical analysis of IFPI discourse on music downloading", Proceedings of the 1st European Communications Conference, Amsterdam, November 2005.
9. EC (1994), "Europe and the Global Information Society", The Bangemann Report. Brussels: European Commission.
10. EC (2001) EC (2001) Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society, Brussels.
11. Edström-Frejman A. (2005), "Spoofing, Suing and Sniffing- Major Record
12. Companies'Response to File Sharing", proceedings of e-Challenges conference, Ljubljana 2005 in "Innovation and the Knowledge Economy", Ed. Cunningham,P, & Cunningham, M. IOS press Amsterdam pp 933-941
13. Edström-Frejman 2007, "e-Commerce rhetoric and reality in the music industry. Estimating the real impact of file-sharing activities on CD-sales ", pending conference paper, eChallenges, The Hague, October 2007).
14. Frith S. (2004), "in Music and Copyright, Second Edition", Ed. Frith,S & Marshall, L. Edinburgh University Press, UK
15. Halvarsson J. (2007), "Patterns of behaviour amongst peer-to-peer users of music and other forms of content – Driving Forces, modes of use and morals", KTH masters thesis, Dept of Media technology and Graphic Arts, Stockholm.
16. i2010 European Commission (2006), First progress report available at: http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm (accessed 20070822)
17. Kusak D., Leonhard G. (2005), "The Future of Music – Manifesto for the Digital Music Revolution", Berclee Press, Boston, USA
18. Music and Copyright (2005), "P2P File Sharing" 20060215, Informa Press, London
19. Landegren J., Liu P. (2003), "Usability Factors in the Distribution of Digital Music Services. A study of Online Distribution Services, Identification of important factors for attractive services, and the Interests of consumers, distributors, content owners and music creators", (KTH Media Technology research dept. Stockholm, Sweden, available at
20. www.nada.kth.se/utbildning/grukth/exjobb/rapportlistor/2003/rapporter03/landegren_joakim_03059.pdf (accessed 2007-08-22)
21. Lövkvist, M (2007), Interview Matthias Lövkvist, CEO Hybris records, Sweden April 2007
22. Musically (2005), "Study of file-shares and downloaders", Downloaded summary from http://news.bbc.co.uk/1/hi/technology/4718249.stm (accessed 20060520)
23. Orlowski A. (2007), "Music Biz agrees – stop shooting self in foot", The Register, June 2007, available at: http://www.theregister.com/2007/06/27/kristiansand/ (accessed 20070710)

24. Pickard R. (2004), "A note on Economic Losses due to Theft, Infringement and Piracy of Protected Works" Journal of Media Economics J7(3) pp 207-217

25. Selg H., et.al. (2005), "A study of file-sharers at Swedish Universities", Music Lessons project, KTH, Stockholm (www.musiclessons.se) (accessed 2007-06-01)

26. Spedidam (2005), "14000 Artistes – Interprètes Pour Une Licence Globale Sur Internet", Paris, France (www.spedidam.fr) (accessed 2006-09-10)

27. Towse R. (2001), "Creativity, Incentive, and Reward: An Economic Analysis of Copyright and Culture in the Information Age", Cheltenham, UK, Edward Elgar Publishing

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007**

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Designing and implementing an undergraduate program in information systems security

Victor Ralevich

BAISc (Information Systems Security)
Sheridan Institute of Technology and Advanced Learning
430 Trafalgar Road, Oakville, ON, Canada, L6H 2L1
victor.ralevich@sheridaninstitute.ca

Dragana Martinovic

Faculty of Education
University of Windsor
401 Sunset Ave., Windsor, ON, Canada, N9B 3P4
dragana@uwindsor.ca

Abstract In this paper we present our experiences in developing curriculum for the newly implemented four-year undergraduate program in applied information sciences with the emphasis on all the aspects of information systems security. We discuss importance of having such a program, and scarcity of qualified instructors as one of the obstacles in program delivery. Various levels of student maturity and uneven gender representation are factors that need to be considered in implementation of such undergraduate program. Program curriculum details and useful experiential conclusions are provided.

Keywords IS security, undergraduate program, curriculum

1 INTRODUCTION

Right after the burst of the "dot-com bubble" in 2000 and later, all higher education institutions in North America experienced a sharp decline in the number of new students in university degree programs for computer-related careers. International outsourcing and the recently allowed increase of the skilled "guest workers" (e.g., those participating in the U.S. H-1B visa program) had further negative impact on interest in computer technology studies.

This declining interest in computer-related higher education unfortunately happens simultaneously with an increased need for the specially trained computer professionals from the information systems security field. Having in mind the ever growing dependence of businesses, local governments, and other vital components of societies on global networks, this trend will continue, thus creating for all stakeholders unforeseeable circumstances.

The instability of the IT job market, constant emerging of new computer technologies, and increased security awareness ask for new approaches in related educational offerings, in terms of adjustments in existing programs and development of the new ones.

In this paper we provide a detailed description of the designing process and implementation of one such new program, namely Bachelor in Applied Information Sciences (Information Systems Security) [BAISc (ISS) in the text to follow], offered by School of Applied Computing and Engineering Sciences at Sheridan College Institute of Technology and Advanced Learning, Ontario, Canada.

Up until year 2000, education offerings by colleges in Ontario were restricted to the one-, two-, and three-year diploma programs with possible co-op (work-term) component, as well as, post-diploma programs for certain specialities. In year 2000, Ontario Ministry of Training, Colleges and Universities (further addressed as "the Ministry") announced its support for induction of pilot projects to allow colleges of Applied Arts and Technology to issue the, so called, "applied degrees." However, in order to be able to assess this initiative, the Ministry capped to eight the number of projects that could be approved yearly in the next three years. Furthermore, no college was allowed to develop more than one such program which had to be in the field where the college demonstrated an utmost academic excellence. Other requirements from the Ministry related to "demonstrated demand from students and employers" and programs that will "not duplicate programs normally offered at universities in Ontario" ([4], p.6). These requirements resulted in the working definition of the "college applied degree" as a combination

of "solid grounding in theory and analytical skills with career-oriented, practical education and training" ([4], p.9).

As one of the prestigious colleges in the computer-related field, Sheridan College and in particular, the School of Computing and Information Management (SCIM) was motivated to keep the leading role and be among the first to get the degree program. It was to the Sheridan's advantage that one of the professors in SCIM previously had a unique and successful career as the Chief Scientist and Security consultant in companies like DiversiNet Inc., DocSpace Inc, and XanderTech Corporation. It was only natural that Dr. Ralevich leads the process of accreditation of the Bachelor applied degree program related to the information systems security. One of the first steps in that process was to assemble an advisory committee of information and computer security professionals and industry representatives from the region, in order to open a discussion that would lead to the proposal. From that point on, building on his working experience and IT security- related research, Dr. Ralevich has drafted the initial and subsequent versions of the program map including the courses' contents.

Having specific profile of anticipated graduates as a goal; identifying the components not covered in existing courses as well as differences in expected level and scope (gap analysis); and using skills identified by various information security certification bodies such as SANS Institute and (ISC) as a model, by the end of 2001, Sheridan proposed the first version of the BAISc (ISS) curriculum. (SANS stands for SysAdmin, Audit, Network, Security; while (ISC) stands for International Information Systems Security Certification Consortium)

In this proposal were included letters of support from (a) large and mid-size companies in computer-related industry and in financial sector, (b) security consulting businesses, and (c) IT managers in local municipalities. These support letters also emphasized the prospective demands for IS security skills in the near future.

Based on the proposal and findings of the Quality Assessment Panel members appointed by the Ministry, the program was approved in 2003 and the first generation enrolled in the program had their first term in the fall of 2004.

2 MOTIVATION FOR ISS PROGRAM OFFERING

The demand for people with expertise in information systems security is growing rapidly due to our increased dependence on the Internet and need for heightened security due to the global situation. According to Wendy Cukier, the Associate Dean of Ryerson University's School of Business, "there is a disconnect between the number of jobs and grads, with around 9,000 math, computer science, and information systems grads per year to fill the 89,000 new jobs (roughly 18,000 to 30,000 new jobs per year) that will crop up over the next three to five years." (see [3])

3 ISS PROGRAM GOALS

In the planning of the BAISc (ISS) program curriculum the intent was to provide future graduates with multiple skills, a strong foundation in computer science and engineering disciplines equivalent to level and depth of typical university offerings, significant amount of practical experience in a variety of IT security fields, and soft-skills necessary for such profile of professionals.

In essence, this means that students are offered bachelor degree level of content which, in most instances, matches, and in some instances goes way beyond, recommendations of the final version of the Undergraduate Computer Science Curriculum document published by the Joint IEEE-ACM Task Force on Computing Curricula in 2001, and particularly the domains of:

- Discrete Structures;
- Programming Fundamentals;
- Algorithms and Complexity; Architecture and Organization;
- Operating Systems;
- Net-Centric Computing;
- Programming Languages;
- Human-Computer Interaction;
- Information Management;
- Social and Professional Issues;
- Engineering;
- Science and Numerical Methods.

(see [1]and [2])

The program includes a large portion of security focussed courses, but also deals with security-related issues within each of the core computer science disciplines and topics of interest.

The information systems security part of BAISc (ISS) curriculum exceeds the list of domains mentioned in [5] as possible basis for the Common Body of Knowledge for Information Security. Above all, the program has a strong applied component, common in the college programs, which includes:

- lab work,
- team and individual research,
- participation in course-related projects which gain in complexity with every semester of the program, as well as
- eight months paid internship.

4 ISS PROGRAM DESCRIPTION/ PROGRAM CONTENT AND LOAD DISTRIBUTION

In order to better understand the fundamental principles of the ISS curriculum design and relations and interdependences of the courses, we may, to some extent artificially, divide the entire pool of courses into the seven streams: IS security; core programming courses; mathematics; networks-related

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 1. Representation of streams in the ISS curriculum (1 credit = 1 hour/week)

Stream of courses	Credits	% core	% overall
Security courses	45	37.5	30%
Programming	36	30.0	24%
Mathematics	12	10.0	8%
Networks	9	7.5	6%
Database Management	9	7.5	6%
Topics in Computer Science	9	7.5	6%
Breadth courses	30		20%
Overall	150	100.0	100%

courses; database design and management; computer science; and breadth courses (see Table 1).

More specifically, the program consists of the core courses given in Table 2. Note that between terms 6 and 7 curriculum program includes a co-op component of the program offering an eight-month workplace experience.

In addition to the discipline depth obtained through the core courses, the students also develop broader insight and understanding of social, philosophical and related topics through 'electives' or 'breadth' courses. During their four years of study students are required to take 10 (20% of overall amount of credits) elective courses covering topics such as: Cognitive Psychology; Introduction to Philosophy; Social and Cultural Anthropology; Censorship and Literature; Practical Ethics; etc.

Gradually, but consistently, this program introduces students to the ISS field. In the first two years of the program, students develop critical understanding of the theory and principles of the computer science fundamentals, computer technology, telecommunications technology, and resource protection. Delivery is in class; students study theory and apply what they have learned through labs, team and individual projects, case studies, and problem solving exercises.

Year three builds on this technical foundation to introduce students to increasingly complex and integrated IT security issues. Students learn to evaluate the appropriateness of different approaches to solving problems and develop ability to apply underlying concepts and principles outside the context in which they have been first studied.

Upon completion of year three, students participate in an eight-month paid work placement. At this point they have a strong IT background and entry-level skills in the field of information systems security. Through our relations with employers, we have been able to offer a variety of placements that will allow students to apply what they have learned and expand their technical expertise.

Furthermore, in the workplace, the students will exercise personal responsibility and hone ability to work in a team and make decisions. The co-op positions students presently hold are: Information security intern, Security systems specialist – IT Co-op, Security analyst, consultant, cybersecurity developer, security analyst intern, user access reviewer, internetworking engineer intern, vulnerability research QA intern, and similar. The employers cover a range of Ontario

Table 2. Core courses in the ISS curriculum

Term 1	Term 2
Introduction to UNIX Operating System	Intermediate Object-Oriented Programming
Introduction to Object-Oriented Programming	Computer Mathematics
Finite/Discrete Mathematics	Structured Data Modelling
IS Loss Prevention Methodologies	IS Security Threats and Risk Assessment
Introduction to Communication Networks	
Breadth course: Composition and Rhetoric*	
Term 3	**Term 4**
Structured Computer Organization	Programming Algorithms and Data Structures
Advanced Object-Oriented Programming	Multi-Tier Programming I
Statistical Methods	Internetworking
Database Implementations and Management	IS Forensics and Investigations
Term 5	**Term 6**
Operating Systems Design	Networks and Distributed Systems Security
Systems Programming	Introduction to Cryptology
Multi-tier Programming II	IS Security Auditing
IS Intrusion Detection and Prevention	Secure Software Development
Database Security	
Eight-month paid work placement comes between terms 6 and 7.	
Term 7	**Term 8**
Secure e-commerce Technologies	Computer Security
Malicious Code: Design and Defence	Information Age Ethics
Scientific Computing	Ethical Hacking
Project	Graduation Project
	IS Security Seminar

*Composition and Rhetoric is the only mandatory breadth course

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

municipalities' IT departments, security consulting firms, large retail and telecomm organizations, as well as the well known electronic industry companies in Japan.

In year four of the program, students continue to build their skills in information systems security. The last two terms are mostly dedicated to advanced topics related to IS security practices and the graduation project.

The topics in the fourth year include but are not limited to:

- Study of malware from the design and protection points of view, including code reverse engineering, cryptomalware, and diagnosis of the impact of the malware on the systems and data;
- Capstone computer security course which serves as the synthesis of components previously elaborated in the curriculum, some deeper insight into practical aspects of various facets of the ISS and computing oriented professions, national and international IS security standards, and similar topics;
- Ethical hacking (penetration testing), its practical issues, techniques, methodologies, and limitations, as well as legal and other concerns and practices related to it;
- Ethics in the computer age, which also covers issues of national and international security related legislation, privacy protection, intellectual property and copyright.

In the fourth year of studies, students participate in one major (graduation) project geared towards application of knowledge and skills gained so far, in order to identify, address and resolve information systems security related issues. Teams are presented with cases proposed by some of our industry/ business/ government partners, and have to identify problems, resolve them through implementation of relevant theories, test and present their work. More specifically, work on the project is envisioned as a two-stage process. During the first phase of the project, in term 7, students form teams and come out with the project theme or accept some of the offered project themes. During that time, the students are mentored by the qualified faculty members and helped around project planning and scope of deliverables. Furthermore, each project has external "customers" who partly evaluate the project's relevance and quality of deliverables. External customers are active members of the IS security community, primarily affiliated with the PAC members' institutions, but may also be representatives from academia, or have a proven record of security and computing-related accomplishments.

Through this process all the requirements of the program are fulfilled, namely students will have demonstrated ability to: secure network applications; document, troubleshoot and implement security policies; work collaboratively and individually on IT security tasks of significant complexity; implement IS security-related vulnerability assessments, security audit, IS forensics investigation; develop secure programming solutions in a plethora of object oriented and procedural programming languages and technologies taught and actively used throughout the curriculum, and even participate in security and computer science research if they choose to pursue a Master's degree .

5 FACULTY MEETINGS

The program curriculum was designed as a coherent and logically interconnected body of theoretical and practical knowledge in such a way that successful students would gain awareness of value and importance of all of the components and methodologies used in various courses. In order to accomplish coherence and logical relationship between courses, the courses are tightly-knit and interdependent. For that reason, regular faculty meetings and discussions are necessary for adequate coordination and delivery synchronization of related subjects. The regular (usually monthly) faculty meetings help faculty members to understand better the program overall goals, evaluate compliance of the courses with these goals, adapt to changes in curriculum, and monitor student progress.

6 EXTRACURRICULAR ACTIVITIES IN THE PROGRAM

Since computing field and especially security field rapidly change, curriculum has to be flexible enough to satisfy emerging needs and, often diverse, interests of the students. Furthermore, specialists in this area need to follow recent events on a global scale, be well informed about professional gatherings and associations; and examine results of newest research and development in the security and related fields. To achieve all that, the following informal components were added to the program:

- *IS Security Seminar* – forum where students, teachers or guests discuss issues, experiences and special interests in form of brief lectures, mini-courses or presentations of news in the industry and legislation.
- Feedback from professional gatherings and conferences (i.e., reports, distribution of materials)
- *Guest speakers* (e.g., IT Security professionals, consultants, law enforcement cybercrime specialists, lawyers)
- *Bulletin board* - The students also run their password-protected bulletin board on the Web where they share profession-related news, interests, place questions, clarify concepts, etc.

7 QUALITY CONTROL

All of the program adjustments and modifications have to be approved by the Program Advisory Committee (PAC). The Committee consists of key business and industry representatives who provide ongoing advice to help ensure that program stays current and relevant within the profession. The PAC meets at least semi-annually, with many of its members providing additional assistance on special projects, guest lectures and co-op placements. Novelty of this program is in active students' participation on the Committee where student delegates attend the meetings, and provide their input to the

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

discussion on the course deliverables and other program related issues.

8 APPLICANT SELECTION

Applicants to the program need to satisfy almost the same enrolment requirements as if they are applying for university, i.e. the Ontario Secondary School Diploma (or equivalent), including English, Mathematics, or Science with the minimum average of 65%. The school selects between 30 and 40 new students every school year.

9 MOBILE COMPUTING AND FACILITIES

It is mandatory for all students in this program to have laptop computers leased or bought through the school. These laptops are equipped with a variety of software applications to assist students in developing expertise in information systems security, i.e., multiple partitions with different operating systems, compilers, forensics software, and various simulators. Laptops also enable students to collaborate on-line while working on group projects, to conduct Web-based research and access the subject area on-line lecture notes and chat rooms. Besides the wired and wireless network access available throughout the entire college, students also have 24/7 access to the classroom space dedicated specially to this program. Students also use the Computer Forensics Lab, and the Networking Lab, both of which contain necessary equipment and software.

Sheridan College is part of the SHARCNET, a consortium of colleges and universities in a "cluster of clusters" of high performance computers, linked by advanced fibre optics. Its unique computational infrastructure combines an active academic-industry partnership, enabling world-class computational research. BAISc students may use this network for graduation and other projects.

10 MASTER'S DEGREE OPPORTUNITIES

Our graduates can enrol into distance learning Master's program in Applied Computer Information Systems with a specialty in Information Systems Security at the University of Denver, Colorado, USA. Similar agreements for Master's Degree programs exist with the University of Western Sydney and Griffith University in Australia.

11 DIRECT – ENTRY OPPORTUNITIES (BRIDGING TERMS)

Sheridan College offers a one-term 'bridging' program in order to facilitate the option of the direct entry into the later terms of the degree program to as many qualified applicants as possible.

To be eligible for bridging into Year 2, candidates must have a 2-year diploma in a computer-related program at the college level, with a minimum GPA of 75%, or have completed core courses in the first year of a computer-related program at the university level, with a minimum GPA of 65%.

To be eligible for bridging into Year 3, candidates must have a 3-year diploma in a computer-related program at the college level, with a minimum GPA of 75%, or have completed core courses in the first two years of a computer-related program at the university level, with a minimum GPA of 65%.

Graduates of computer programs must have graduated within the last 4 years to be eligible for the bridging program.

The complete bridging program consists of courses: BAISc Mathematics; IS Security Overview; IS Forensics; Algorithm Design; CPU Architecture and Assembly Language; and, Composition and Rhetoric.

The bridging program for direct admission into the second year of the BAISc (ISS) program consists of the subset of the mentioned courses. For direct entry into the 3rd year, eligible candidates must complete all six offered courses. These courses can be taken during the Sheridan's summer term.

12 PROGRAM CONSTRAINTS AND OBSTACLES

One drawback of the program is that it is offered only in a full-time format and as such does not offer opportunity to the employed potential candidates to study on the part-time basis. Also there are not any on-line course offerings anticipated in the foreseeable future. That restriction is imposed by the Ministry.

One of the major constraints to the growth of the student numbers is in difficulty to attract and hire new faculty with the adequate background and experience in ISS. Most of the experts and practitioners in the field, except for the related IS security certification, do not meet the criteria for teaching in the degree program, such as having a doctoral, or at least master's degree in the related field. Universities have similar problem of recruiting faculty and that is one of the basic reasons for lack of such programs at the undergraduate level in North America or anywhere else. In case of the applied baccalaureate degree such as BAISc (ISS), teachers also need to have a significant practical experience in order to be able to help students to develop practical problem solving skills under 'real life' circumstances. This issue is addressed in the email received by the coordinator of the program from one of the students currently on the internship placement:

[..] I am currently doing vulnerability analysis and alerting. I look for vulnerabilities that are critical, score them accordingly, write a quick description and summary as to what methods and scripts are vulnerable, how to mitigate them by using vendor patches etc., and I send the alerts out to our clients (mostly big big businesses). It is quite interesting and I am learning a lot. Of course, without the knowledge that you've taught us, this would not be possible. I would also like to take this time to thank you all for providing us with the skills that we need to be in this industry. I must say that it has helped me a lot versus a few other new hires here who

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

has (sic) never done security before (Waterloo and University of Toronto) students. (Received on May 29, 2007)

13 EXPERIENCES SO FAR

There have been three generations of students in the program so far, and the forth is on its way in the fall of 2007. The overall experience shows that there are some interesting patterns in student population such as:

1. Those candidates who had some previous experience as students at the college or university level perform better; are more motivated to study and experiment; and have better understanding of the computing concepts.
   When it comes to the level of skills, the differences among students are most obvious in the first two terms. That difference gets levelled as the students get to the more advanced courses covering concepts that are not part of any typical computing-related curricula that some students might have had encountered earlier.
   In the first year, we witness some sort of bipolar grouping of students rather than a typical bell-shaped distribution in terms of experience and skills. At one end we have a group of students who are adequately prepared for programming languages, use of operating systems', or mathematics. At another end are about 25% of students' population who have not had sufficient exposure to such content. However, what contributes to student success is a good cooperation and team work out of class time, in terms of additional problem solving sessions, discussion groups within the school's learning management system (WebCT Vista), and unselfish help from those students who have better insight and stronger background in some of the course topics. Most of those activities are spontaneously organized and handled by the students without much of the teachers' intervention, except when it is necessary. Inexperience and lack of adequate training in mathematics are major causes of attrition in the first two terms. If the initial frustration is successfully attended to, then in the later semesters, students do not have much difficulty to learn and use a rather sophisticated mathematical methodology such as: statistics, number theory, abstract algebra, computational complexity, numerical analysis, etc.
2. Students, on their own initiative, provide a detailed feedback regarding every course at the end of every term and a lot of those comments are taken into consideration in adjustments of the courses' delivery modes and scope for the future.
3. Students who are more involved in some of the IS security or computer science activity and have special topics of interest in which they developed a certain level of expertise, scaffold other students, in spirit of cooperation and sharing of the common goals.
4. Vast majority of the students are male except for just couple of female students each year, most of which drop out after a year. The percentage of female students fluctuates in between 0.3% and 0.8% out of total student population in the program. In later terms there are almost no female students (in the last two years of the program, there is only one female student).

14 CONCLUSION

Information systems security is a rigorous and complex profession requiring high standard of ethical conduct and technical competence. Currently, there are no requirements for certification in information systems security in Canada although there are professional associations that offer educational opportunities, provide certification and promote good practice. Sheridan will encourage its graduates to apply for membership to professional organizations such as the Information Systems Security Association.

An applied degree program in information systems security contributes to both the provincial economic strategy for increased IT strength and to Industry Canada's Canadian Strategy for Electronic Commerce in Canada. The *Ontario Jobs and Investment Board* in its *Economic Plan for Jobs in the 21st Century*, states the need for "investment in the infrastructure and technology 'enablers' to sharpen Ontario's competitive edge and access to global markets" ([6]).

In spite of the seemingly local character of the applied degree initiative we strongly believe that the BAISc (ISS) curriculum has much broader educational value and that it may be used as an innovative model of undergraduate applied computer science and engineering anywhere in North America and worldwide.

REFERENCES

1. "Computing Curricula 2001 – Computer Science, Final Report" (2001), Joint Task Force on Computing Curricula, IEEE Computer Society and Association for Computing Machinery.
2. "Computer Engineering 2004, Curriculum Guidelines for Undergraduate Degree Programs in Computer Engineering" (2004), Joint Task Force on Computing Curricula, IEEE Computer Society and Association for Computing Machinery.
3. Smith, B "Wanted: 89,000 IT employees" (2007), http://www.itbusiness.ca/it/client/en/home/News.asp?id=43124 (6/12/2007)
4. "Increasing Degree Opportunities for Ontarians", A Consultation Paper, Ontario Ministry of Training, Colleges and Universities, 2000
5. Theoharidou, M., Gritzalis, D. (2007). "Common body of knowledge for Information Security", IEEE Security & Privacy, 5, (2), 64-7.
6. "Ontario Jobs and Investment Board, A Roadmap to Prosperity: An Economic Plan for Jobs in the 21st Century", 1999.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

How can leaders encourage participation in virtual communities of practice?

Nicholas Bowersox

Touro University International,
nbowersox@tourou.edu

Indira Guzman

Touro University International,
College of Information Systems
iguzman@tourou.edu,

Abstract This paper attempts to integrate the literature surrounding what this researcher calls the 2nd generation of communities of practice – the virtual community of practice. Specifically, this paper will explore the motivators to participation in these online communities from the perspective of the group leader in the organizational context. This paper poses the question "How can group leaders encourage participation in virtual communities of practice"? Further, it suggests that there are five "virtual community drivers" that encourage participation. They are leader involvement, online interaction, offline interaction, usefulness, and IT infrastructure quality. This paper then suggests that empirical research be conducted analyzing the effects of each virtual community driver on participation. Finally, this paper concludes by suggesting that future empirical research is needed concerning the synergistic relationships between other possible motivators and participation in VCoPs.

Keywords virtual communities of practice, communities of practice, knowledge management, leadership

1 INTRODUCTION

With the growth of information and communication technology (ICT) such as the internet, email, and video conferencing, organizations have become increasingly more efficient and productive. However, not only does computer technologies increase daily productivity rates among employees; they also increase an organization's yearning to digest larger amounts of information with the end desire of being able to share that information across the entire organization. As such, terms such as knowledge management, knowledge society, and the information age have become dominant metaphors in today's organizational settings. Because of this, it may come as no surprise that there is an increasing desire to emphasize knowledge sharing techniques and strategies that will foster improved performance and effectiveness. The emphasis on knowledge management through collaborative means is an excellent manner in which to achieve this [E. C. Wenger & Snyder, 2000]. Perhaps one of the most popular methods by which to share such large amounts of organizational information is through informal learning environments such as communities of practice.

1.1 Communities of Practice (CoPs) Defined

Communities of practice (abbreviated as CoPs hereafter) are defined in the early works of Lave and Wenger [1991, p.98] as "a set of relations among persons, activity and world, over time and in relation with other tangential and overlapping communities of practice". This definition centered on the idea of apprenticeship in which CoPs were viewed as a form of socialization into a community [Kimble & Hildreth, 2005]. This assumes a unidirectional process by which newer community members integrate themselves into the community's practices. Lave and Wenger [1991] state that newcomers move from a state of "legitimate peripheral participation" into that of "full membership". During legitimate peripheral participation, newcomers engage in several roles at the same time to invoke varied degrees of experience and interaction. Eventually members of the community become recognized as they learn the rules and boundaries which guide that community.

Although this definition of CoPs is accurate, perhaps a more modernized and simplified definition is provided by Kimble & Hildreth [2005]. They define CoPs as "groups of people bound together by a common purpose and an internal motivation", often with long-term objectives in mind. Consider this definition in the organizational context. Applying the

keywords of the definition provided by Kimble & Hildreth [2005], it can be assumed that the various departments of any organization comprise a CoP (i.e. human resources, finance, and marketing). For example, let's consider an example such as the finance department of Microsoft. Each employee working in finance for the Microsoft Corporation has a common purpose: to successfully control, monitor, and manage the financial assets of the company. Some employees may serve as financial analysts looking at corporate financial statements while others may be in charge of long-term budget forecasting, but in essence their purpose is one in the same. In addition, they are internally motivated to do the best they can to ensure that Microsoft maintains it competitive advantages for many years into the future. As a result of this example, it can be assumed that the practice and purpose of CoPs may be construed as having always existed, even before being formally identified as such.

1.2 The History of CoPs and the Social Learning Theory

The early work of the social learning theory was attributed to Bandura [1977]. In general, social learning theory emphasizes the importance of observing and modeling the behaviors, attitudes, and emotional reactions of others. According to Bandura [1977], learning would be exceedingly difficult and hazardous if people had to rely solely on the effects of their own actions to inform them what to do. Through socialized learning, employees of a company are able to share information and knowledge in an effective manner. Learning that takes place in a CoP is viewed as a social process by which members become active participants in the community they are part of.

Etienne Wenger is perhaps one of the most prominent theorists in linking social learning theory to CoPs. Although his theory does not seek to replace existing theories such as Bandura [1977], it does come with its own set of assumptions and its focus. In his book titled Communities of Practice: Learning, Meaning, and Identity, Wenger [1998] precludes four reasons as to why learning should be social, rather than individual in nature:

1. We are social beings.
2. With respect to valued enterprises, knowledge is a matter of competence.
3. Knowing is a matter of participating in the pursuit of such enterprises. We should actively engage in the world.
4. Our ability to experience the world and engage with it as meaningful is ultimately what learning is about.

The culmination of the four premises above conclude that learning is not so much individual as it is an individual acting as a participant in a social community.

Wenger's primary focus of his social theory is that learning should be viewed as social participation where participation "refers not just to the local events of engagement in certain activities with certain people, but to a more encompassing process of active participants in the practices of social communities" [1999, p. 31]. Wenger [1998] also discusses four components necessary to surmise that social participation is a process of learning. They are:

1. Meaning: a way of talking about how individuals experience the world around them through their individual and collective abilities
2. Practice: a way of talking about shared historical and social frameworks, resources, and perspectives that can sustain mutual engagement in action
3. Community: a way of talking about the social configurations that our enterprises are designed in
4. Identity: a way of talking about how learning changes who we are in communities

Figure 1 below represents a visual model of Wenger's [1998] components of the social theory of learning. The four elements – meaning, practice, community, and identity – are interchangeable in regards to their relationship to learning. For example, switching any of the elements with learning still allows the figure to make sense.

Much of the scholarly research work conducted on CoPs is based on Wenger's social theory of learning. For example, vKimble & Hildreth [2005] explored the relationship be-

Figure 1. Components of the social theory of learning: an initial inventory [adapted from Wenger, 1998, p.5]

tween knowledge management and CoP's using data collected from a case study on a large international corporation. Specifically, the article discusses how the social relationships and shared artifacts inherent to the company's virtual communities of practice (VCoPs) can be linked to Wenger's concepts of a participation-reification duality. Their case study found that shared artifacts were important in the process of creating, sharing, and transferring knowledge through the VCoP as well as facilitating social participation, which is important in building and maintaining personal relationships between VCoP group members.

Ardichvili, Page & Wentling [2003] conducted a qualitative study of the motivators and barriers to participation in VCoPs. They argue that active participation is a critical ingredient to the successful functioning of any type of CoP. Further, they describe CoP participation as an economic model of supply and demand. In other words, the supply of knowledge provided by the knowledge givers must be sufficient to meet the demand for the knowledge seekers. Therefore, social participation is critical. Overall, the results of their study found that knowledge flows easily when employees view knowledge as a public good that benefits the entire organization. Finally, employees will participate more when they are geographically dispersed and are trying to integrate themselves more quickly into their work environment.

1.3 Determining Characteristics of CoPs

According to Wenger [1998], CoPs can be characterized using two broad categories: structural characteristics and dimensions of practice. Structural characteristics attempt to define how CoPs are established, whereas dimensions of practice explain how members join CoPs. Both types are equally important in defining, managing, and cultivating CoPs, and, as such, are described in greater detail below.

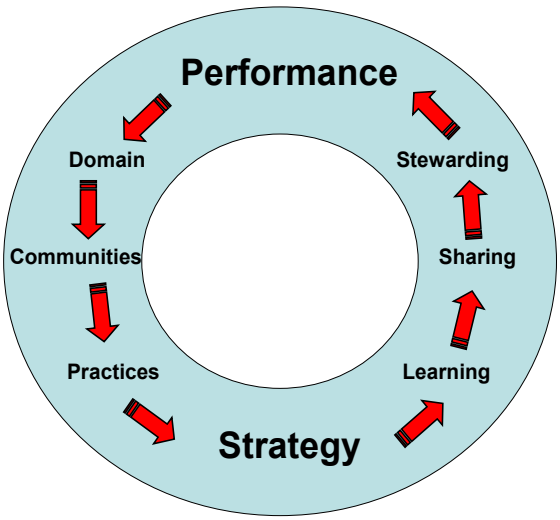
There has been much research work focusing on identifying CoP structural characteristics. Wenger [2004] defines the three elements of a CoP as the domain, community, and practice. He defines domain as "the area of knowledge that brings the community together, gives it its identity, and defines the key issues that members need to address" [p.4]. Further, the domain of a CoP helps to recognize the "area" of knowledge to be studied, rather than identifying tasks to be accomplished. The goal in developing the domain is to take the strategy of the organization and develop it into a set of domains of knowledge which should then be able to connect the strategy to the daily work.

Wenger [2004] describes the community as "the group of people for whom the domain is relevant, the quality of the relationships among members, and the definition of the boundary between the inside and the outside" [p. 4]. This community is more than a group of people sharing similar interests. Rather, it is a group of people fostering high levels of interaction in an attempt to discover new knowledge, transfer existing knowledge, and solve problems. This structural characteristic occurs after the knowledge domains are present. It is here that community members are recruited and those with more extreme experience may take the lead in further developing and growing the community.

The third element as defined by Wenger [2004] is practice which is "the body of knowledge, methods, tools, stories, cases, and documents which members share and develop together" [p.4]. Practice takes place after the domains of knowledge and community members are established. It is here that community members are engaged in the development of their practice through various means which may include community speakers and community meetings. This structural characteristic involves finding ways to maximize the amount the knowledge available through efficient use of the resources at hand.

The combination of these three elements enables CoPs to effectively manage their knowledge. According to Wenger [1998], domain provides a common focus, community builds relationships that enable collective learning, and practice anchors the learning in what people do. Because CoPs are organized into domains of knowledge catered to specific members that practice within them, they are well-positioned to add sustainable strategic value to the organization. Figure 2 below depicts how knowledge management is a strategic activity that starts with a strategy and ends with a strategy. Strategy is connected to performance through knowledge.

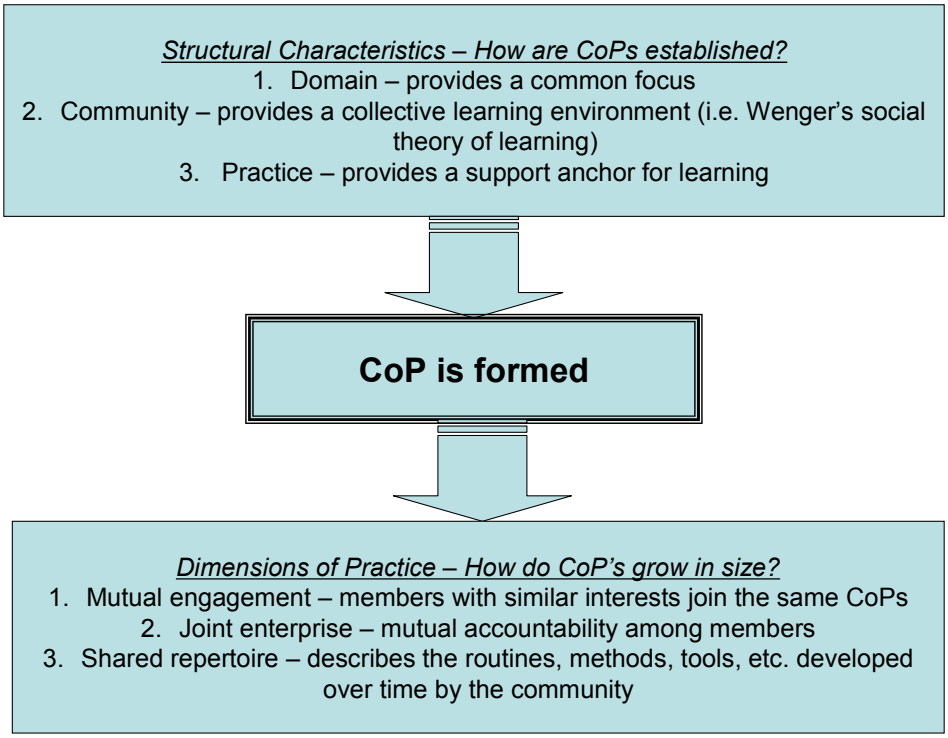
Figure 2. The doughnut model of knowledge management [adapted from Wenger, 2004, p.3]



Aside from the structural characteristics of CoPs, it is also important to mention the dimensions of practice. Wenger [1998] states that there are three components of practice for a CoP: mutual engagement, joint enterprise, and shared repertoire. Mutual engagement refers to the notion that practitioners with the same interests and ideas will typically be members of the same CoPs. Joint enterprise reflects the notion that beyond stated goals there is mutual accountability among community members. Finally, shared repertoire includes routines, methods, tools, stories, gestures, symbols, and other such actions and objects that the community has developed over time. The figure below addresses the evolution of the CoP, beginning with its structural characteristics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. The evolution of the CoP: from conception to the growth of community members [adapted from Wenger, 1998]



that lead to it conception and then its dimensions of practice leading to its growth in community members.

By examining the above figure, reconsider the Microsoft Corporation finance example described earlier. Each employee working in finance for the Microsoft Corporation has a common purpose or domain: to successfully control, monitor, and manage the financial assets of the company. As such, they form a community of employees that collectively work together each day to ensure the financial success of the organization. Thus, social interaction among one another is common, promoting a team-oriented learning environment. Practice, or the knowledge, methods, tools, stories, cases, and documents, within the community provides an anchor for collective learning to occur. Since domain, community, and practice are present, the right environment exists for a CoP to form. As the Microsoft Corporation grows over time, new finance employees enter the department. Because they share the same interests towards corporate finance and are committed to the financial success of the corporation as the employees who have been around for some time, they become part of the CoP. Over time, they integrate themselves into the collective network of community members, while at the same time learning the routines, policies, and practices that comprise the community they are a part of.

2 THE RISE OF VIRTUAL COMMUNITIES OF PRACTICE

Technological advancements are undoubtedly allowing employees that are geographically separated from one another the opportunity to become part of a community within the organization. These communities are known as virtual communities of practice (abbreviated as VCoPs hereafter) and have become quite popular with the onset of the global information age. As defined by Allen, Ure, & Evans [2003], VCoPs are

"physically distributed groups of individuals who participate in activities, share knowledge and expertise, and function as an interdependent network over an extended period of time, using various technological means to communicate with one another, with the shared goal of furthering their practice or doing their work better" [p.7].

VCoPs are essentially the same as CoPs; however, members use technologies such as the internet, email, and videoconferencing to maintain "virtual contact" with one another, whereas traditional CoPs employ face-to-face methods for member communication. Most activity in a VCoP is done via a website through some form of a message board where members view and post messages to one another. It is not uncommon for most communication to occur via computer-mediated means. However, other methods such as telephone and face-to-face contact can occur.

2.1 VCoPs as Knowledge Sharing Mechanisms

It is no surprise to both scholars and practitioners that knowledge sharing is an innovative force that can lead to long-term competitive advantages for organizations wishing to embrace it. Connelly & Kelloway [2003] define knowledge sharing as a set of behaviors that aids in the exchange of information to others. In today's globalizing business economy, interpersonal means of communication are becoming rare. Instead, virtual environments are becoming the cost-effective and time-sensitive norm to knowledge sharing for several reasons.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Among the central reasons as to why VCoPs are important is because they have the potential ability to transfer an organization's tacit knowledge – the source of its competitive advantages [Dougherty, 1995]. Tacit knowledge is that knowledge that is often based on years of experience and is not easily codifiable into a useable form. Horvath [1999] states that tacit knowledge is often buried within the stories people tell and that VCoPs are an excellent means by which to share this tacit knowledge. VCoPs allow employees to communicate anytime and anywhere through virtual telecommunications.

2.2 Participation in VCoPs

According to Ardichvili, Page & Wentling [2003], "one of the critical factors in determining a virtual community's success is its members' motivation to actively participate in community knowledge generation and sharing activities" [p. 64]. Many studies, such as those of Connelly & Kelloway [2003], suggest the importance of a work environment that stresses positive social interaction and knowledge sharing. Organizations like this give rise to employees who are knowledgeable about company rules, regulations, and procedures. Further, these types of employees better understand and trust their co-workers, and are more willing to work with them on team projects. Still other studies [Ciborra & Patriota, 1998] show that employees are unwilling to share knowledge and participate in positive social interaction cultures. Holthouse [1998] instead argues that the successful (or unsuccessful) transfer of knowledge is a by-product of the organization's knowledge management system. Amidst all of this confusion, though, there is little substantiating evidence to support why employees of an organization choose to participate in VCoPs. The paragraphs that follow will attempt to shed some light on this subject.

For a VCoP to have activity, it is critical that all members take an initiative in participation. These two words – activity and participation – are important and deserve further explanation. Koh, Kim, Butler, & Bock [2007] delineate activity in a VCoP as posting activity and viewing activity which may be done through various means to include live audio/video streaming, message boards, and online chats. Activity such as this is a necessary and critical component to any VCoP. Participation is a two-fold definition and entails that members be willing to both share and use existing knowledge. Therefore, when members share their knowledge, they are participating in posting activity; when they use knowledge that is available on the VCoP, they are participating in viewing activity. But what exactly is meant by sharing and using knowledge? Simply put, sharing knowledge implies that the "owner" of the knowledge is willing to allow others to use it. Knowledge sharing can be defined as the activities that involve gathering, absorbing, and/or transferring product and/or service information between organizations and customers, alliance partners, and/or employees [Chen & Barnes, 2006]. Those using the knowledge will gain increased levels of understanding and efficiency into the business policies, practices, processes and procedures, thereby allowing them to more greatly contribute to the firm achieving it competitive advantages in the marketplace.

Active participation helps to maintain the socio-technical nature of this online environment [Koh et al., 2007]. Ardichvili, Page, & Wentling [2003] link employee participation in VCoPs around three central themes which are discussed below.

1. Employees must willingly participate to share knowledge – The first reason why employees participate in VCoPs is to share knowledge. Many employees often feel a desire and passion to educate others and give back to the company. Further, these types of employees disregard information hoarding as an obsolete technique for corporate success. Essentially, these types of employees are adding to the supply of knowledge in VCoPs.
2. Employees must willingly participate to use knowledge – If employees are willing to share knowledge, then it only makes sense that other employees are willing to use that knowledge. One of the primary reasons for the existence of VCoPs is to help disseminate knowledge across the organization. Today's competitive marketplace has forced a strong demand on the use of both new and existing knowledge.
3. Employees must willingly participate to use technology – In order to effectively use the full functions of the VCoP, employees must be willing to use the technology that comprises it. For a virtual community, members should feel comfortable in using a computer, the internet, and various other web-based technologies. Technology acts as a necessary facilitator to the flow of knowledge.

2.3 Motivators to Successful VCoPs

The success of VCoPs in the organizational context is undoubtedly based on several factors. Just as in traditional CoPs, participation by employees is a necessary factor. Participation is necessary to create a social learning environment where the sharing of knowledge can occur. Thus, social participation is a process of learning [Wenger, 1998]. Studies have yielded many similar success factors to participation which we will term "motivators". Koh et al. [2007] proposes 4 motivators for successful VCoPs: leader involvement, offline interaction, usefulness, and the IT infrastructure quality. I propose to add a fifth motivator: online interaction. These are discussed in greater detail below.

The first motivator is leader involvement. Leader involvement is perhaps the most important factor that encourages employees' use in VCoPs. When leaders stay involved, employees are more willing to take an active role in posting and viewing comments. In other words, they are more willing to share and use the knowledge provided by the VCoP. This is supported by Allen et al. [2003] who states that "active participation in communities by upper-management clearly indicates that the organization has made a commitment to VCoPs and serves to motivate others to participate" [p. 37]. Further, leaders must show involvement by providing the overall guidance and support that will build, maintain, and grow the community [Fontaine, 2001]. Finally, Koh et al.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

[2007] states that leadership involvement is necessary to promote trust among community members.

Online interaction deals with the level of interaction that community members face with each other while being in touch through the computer. Online interaction does not necessarily imply that members are online at the same time. With advances in computer-mediated communications, members of a VCoP may be able to stay in contact through asynchronous forms of interaction such as the use of websites, electronic bulletin boards, and email. Synchronous forms of interaction may include live chat and videoconferencing. Online interaction is important in that it is the defining characteristic of a VCoP.

Offline interaction, such as face-to-face interviews, is another equally important element of VCoPs. Although collaboration in a VCoP is often done via computer-mediated technology in an online environment, offline interaction among community members helps to establish working bonds, trust, and communication skills that may otherwise be difficult to obtain. Due to physical separation, offline communication may not always be possible. It should, however, be maximized whenever possible.

The perceived usefulness of VCoPs is also a critical motivator to employees' participation. Employees must be willing to see a benefit in their use, and the perceived benefit must be greater than the cost of maintaining them. For example, members must feel that if they post questions for help on a particular topic, they will receive helpful feedback from other members. In addition, members should be given ample time to contribute to VCoPs. If the employees are provided time to access VCoPs, then the supply and demand for new and existing knowledge should increase over time.

The final motivator for successful VCoPs is the IT infrastructure. The mention of this as a motivator comes as no surprise. Without the technology, VCoPs would not be able to properly function, and would cease to exist. Employees would not be able to willingly share and use knowledge that should otherwise be available. The infrastructure is equally important to the VCoP as is the physical space to a traditional CoP. Because the IT serves as the basis for a virtual community, it must first be able to satisfy the users' needs [Koh et al., 2007]. According to Koh et al. [2007], the response time of the system should be satisfactory to sufficiently allow for member interaction. In addition, the system should be user-friendly and reliable. As such, the IT infrastructure helps motivators of VCoP participation increase both the level of posting and viewing activity. Therefore, the quality of the IT infrastructure acts as a moderator in the relationship between the VCoP motivators mentioned above and the participants willingness to share and use the knowledge available on the VCoP.

3 IMPLICATIONS FOR RESEARCH

Because VCoPs are much less antiquated than traditional CoPs, additional research needs conducted to determine motivating factors. Future practitioners should consider further exploring the motivators described by Koh et al [2007]. Once again, they are the impact of leader involvement, online interaction, offline interaction, usefulness, and IT infrastructure quality as predictors of employees' willingness to share knowledge, use knowledge, and use technologies assisting in the daily functions of VCoPs. This leads to a potential research question and a corresponding set of hypotheses aimed at allowing us to better understand how leaders motivate employees to participate in VCoPs.

Research question: How do leaders influence employees' participation in VCoPs?

As a result, a valid set of testable hypotheses may include:

Proposition 1: Employees' willingness to share knowledge (DV) is positively related to leaders' involvement in VCoPs (IV).

Proposition 2: Employees' willingness to use knowledge (DV) is positively related to leaders' involvement in VCoPs (IV).

Proposition 3: Employees' willingness to share knowledge (DV) is positively related to the level of online interaction between members in VCoPs (IV).

Proposition 4: Employees' willingness to use knowledge (DV) is positively related to the level of online interaction between members in VCoPs (IV).

Proposition 5: Employees' willingness to share knowledge (DV) is positively related to the level of offline interaction between members in VCoPs (IV).

Proposition 6: Employees' willingness to use knowledge (DV) is positively related to the level of offline interaction between members in VCoPs (IV).

Proposition 7: Employees' willingness to share knowledge (DV) is positively related to the usefulness of VCoPs (IV).

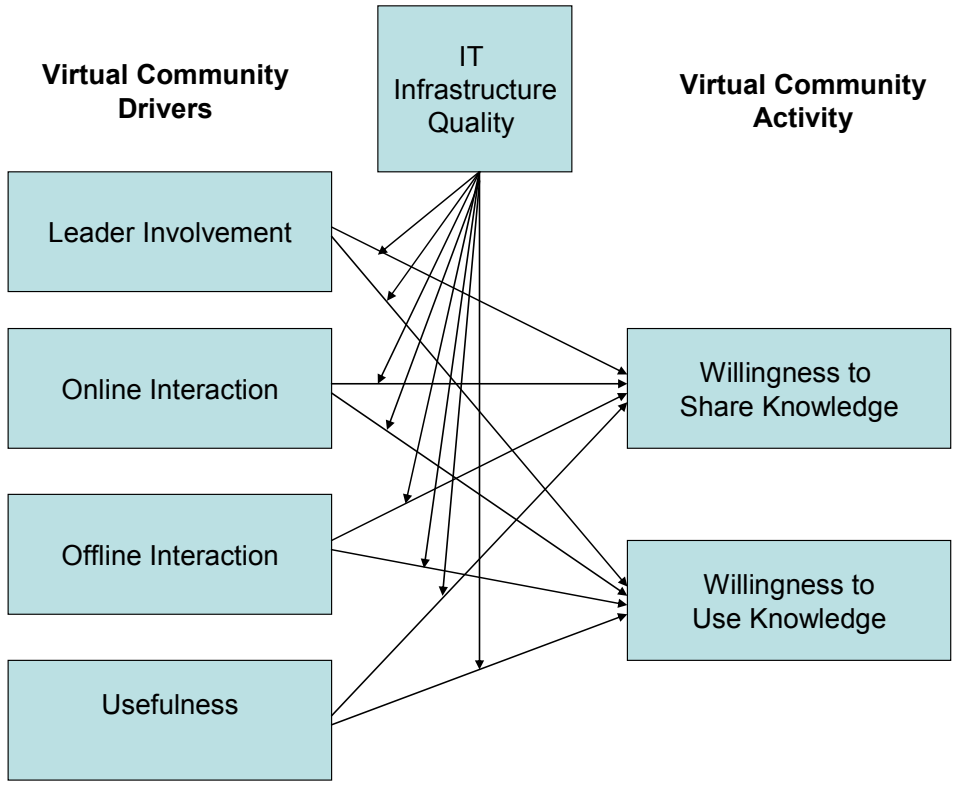
Proposition 8: Employees' willingness to use knowledge (DV) is positively related to the usefulness of VCoPs (IV).

Proposition 9: The quality of the IT infrastructure in VCoPs moderates the relationship between leader involvement (IV), the level of offline interaction (IV), and the usefulness (IV) with the employees' willingness to share (DV) and use (DV) knowledge.

Figure 3 below represents a graphic representation of the proposed relationships that exist between the independent, moderating, and dependent variables in the above hypotheses.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. Virtual community stimulation structure [adapted from Koh et al, 2007, p.70]



4 FUTURE AREAS FOR RESEARCH

Because studies considering leadership and VCoPs as relational variables are relatively scarce, there are several potential avenues for future research. Researchers may want to consider revisiting the work of Koh et al. [2007] to determine if additional motivators play a key role in determining employees' willingness to participate in VCoPs. Researchers may also choose to examine the barriers to VCoP participation. Finally, researchers might consider investigating how the various leadership styles of the firm's management (i.e. transformational and transactional leaders) can affect participation in VCoPs.

The first avenue to consider for additional research is be to determine if additional factors act as motivators to employees' participation in VCoPs. For this paper, there was a focus on the research work conducted by [Koh et al., 2007]. Other researchers, however, have also attempted to examine motivators to participation in virtual knowledge-sharing communities. Research by Ardichvili et al. [2003] found that employees may be willing to share knowledge because there is a moral obligation and community interest to do so. In addition, contributing knowledge allowed some employees to feel as if they were "experts" in their field, while others felt as if they were "giving back" to the organization. Wasko & Faraj [2005] also found that employees are willing to share their knowledge when it will benefit their reputation. In regards to the use of such knowledge, Ardichvili et al. [2003] found that new employees were willing to use the knowledge to get acquainted much faster. They also found that the knowledge was always available, and it keeps them apprised of developments in their profession. It is important to note that some if these reasons for knowledge use may fall under the "usefulness" motivator developed by Koh et al. [2007]. In any case, this area is worth re-examining.

Equally important is understanding the barriers to participation in VCoPs. Just as there is a limited amount of research on the relationship surrounding motivators to participation in VCoPs, there is a lack of research on barriers to participation in VCoPs. The work of Ardichvili [2003] found that most employees would not share knowledge because they were afraid of posting something incorrectly, that no one would view it, or because they believed in hoarding all available knowledge. Others stated that the process to post information to VCoPs was time-consuming. Finally, many feared both posting and viewing information because of security reasons.

Finally, researchers may want to investigate how the leadership styles of firms' management affects employees' participation in VCoPs. Two primary leadership styles have emerged in the mainstream literature: transformational leadership and transactional leadership. Transformational leadership was first introduced by Burns [1978] and studied extensively by Bass [1985]. Bass [1985] defines transformational leadership as the leadership style that inspires followers to exceed their own self-interest for the good of the organization. In contrast to the transformational leader, the transactional leader clarifies followers' roles and what must be done in order to obtain desired outcomes and goals [Bass, 1985]. Future studies could be done to determine if transformational leaders – or motivational leaders – are better equipped to encourage employee participation in VCoPs versus transactional leaders – authoritative leaders.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

5 CONCLUSION

KM scholars and practitioners have urged companies to find more effective ways at sharing knowledge to better create and maintain competitive advantages in today's hostile marketplace. With the aggressive onslaught of modern technologies, VCoPs provide an efficient means by which to achieve this. However, the leaders of the organization are crucial elements in assuring that employees actively participate in VCoPs . Leaders are the driving force in establishing the cultures, systems, and boundaries that promote such knowledge sharing throughout the organization.

This paper chronicles recent exploratory research designed to examine the role of community drivers as enablers of knowledge sharing in VCoPs. Because little research has centered on this topic, future empirical research is needed concerning the synergistic relationships between motivators and participation in VCoPs. Finally, it is important to note that research should be conducted in various types of organizations that utilize VCoPs to enhance the generalizability of the findings.

REFERENCES

1. Wenger, E. C., and Snyder, W. M. (2000), 'Communities of practice: the organizational frontier', Harvard Business Review, Harvard Business School Press, Boston, 1/78/7.
2. Lave, J., and Wenger, E. (1991), Situated Learning: Legitimate Peripheral Participation, University Press, Cambridge.
3. Kimble, C., and Hildreth, P. (2005), 'Dualities, distributed communities of practice and knowledge management', Journal of Knowledge Management, Emerald Group Publishing, Bradford, 4/9/12.
4. Bandura, A. (1977), Social Learning Theory, Prentice Hall, Englewood Cliffs.
5. Wenger, E. (1998), Communities of Practice: Learning, Meaning, and Identity, Cambridge University Press, Cambridge.
6. Wenger, E. (1999), 'Learning as social participation', Knowledge Management Review, Melcrum Publishing, London, 6/1/4.
7. Ardichvili, A., Page, V., and Wentling, T. (2003), 'Motivation and barriers to participation in virtual knowledge-sharing communities of practice', Journal of Knowledge Management, Emerald Group Publishing, Bradford, 1/7/14.
8. Wenger, E. (2004), 'Knowledge management as a doughnut: shaping your knowledge strategy through communities of practice', Ivey Business Journal Online, Ivey Publishing, London, 3/68/9.
9. Allen, S., Ure, D., and Evans, S. (2003), 'Virtual Communities of Practice as Learning Networks: Executive Summary (pp. 50)', Brigham Young University Instructional Psychology and Technology Department, The MASIE Center, Brigham Young University.
10. Connelly, C. E., and Kelloway, E. K. (2003), 'Predictors of employees' perceptions of knowledge sharing cultures', Leadership & Organization Development Journal, Emerald Group Publishing, Bradford, 5/24/8.
11. Dougherty, D. (1995), 'Managing your core competencies for corporate venturing', Entrepreneurship Theory and Practice, Blackwell Publishing, Oxford, 3/19/23.
12. Horvath, J. A. (1999), 'Tacit knowledge in the profession', In Tacit knowledge in professional practice, Laurence Erlbaum, London.
13. Ciborra, C. U., and Patriota, G. (1998), 'Groupware and teamwork in R&D; limits to learning and innovation', R & D Management, Blackwell Publishing, Oxford, 1/28/11.
14. Holthouse, D. (1998), 'Knowledge management research issues', California Management Review, University of California at Berkley, Haas School of Business, Berkley, 3/40/4.
15. Koh, J., Kim, Y.-G., Butler, B., and Bock, G.-W. (2007), 'Encouraging participation in virtual communities', Communications of the ACM, Association of Computing Machinery, New York, 2/50/7.
16. Chen, L. Y., and Barnes, F. B. (2006), 'Leadership behaviors and knowledge sharing in professional service firms engaged in strategic alliances', Journal of Applied Management and Entrepreneurship, H. Wayne Huizenga School of Business and Entrepreneurship; http://www.huizenga.nova.edu/jame/ (June 26, 2007).
17. Fontaine, M. (2001), 'Keeping communities of practice afloat', Knowledge Management Review, Melcrum Publishing, London, 4/4/6.
18. Wasko, M., and Faraj, S. (2005), 'Why should I share? Examining social capital and knowledge contribution in electronic networks of practice', MIS Quarterly, Management Information Systems Research Center, Minneapolis, 1/29/23.
19. Bass, B. (1985), Leadership and Performance Beyond Expectations, Free Press, New York.
20. Burns, J. M. (1978), Leadership, Harper and Row, New York.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Open source migrations: experiences from European public organizations

Andres Baravalle
Department of ICT
The Open University
andres@baravalle.it

Sarah Chambers
Department of Computer Science
The University of Sheffield
s.chambers@dcs.shef.ac.uk

Abstract The landscape of public organizations in Europe is diverse and complex. Public administrations differ in the services that they provide and on their characteristics, but they all rely on computing to deliver their services, even if it is to varying degrees. This paper analyzes the experience of a group of European public organizations investigating the possibility of supporting their services through the use of Open Source. Open Source is software developed inside a community committed to producing software that is free to use, modify and redistribute. The group under examination is composed of a number of public administrations varying in size, from four different countries. While the motivation for starting the migration varies across the members, the results from the different experiences are consistent and show that Open Source is a realistic opportunity to consider. Technical, strategic, and environmental aspects that arose during the migration have been investigated and analyzed

Keywords open source, public organizations, case studies, European Union

1 INTRODUCTION

The landscape of public organizations in Europe is diverse and complex. Public organizations cover a vast area of activity, varying greatly in size, in the type of services provided and in wealth.

The increasing use of and reliance on computing and communications technology in order to deliver services to the public has focused attention on how this phenomenon can be supported and accelerated. Interest in *e-governance* [1, 2] and the technological innovations coming from, for example, mobile telephony, the Internet and, more generally, ubiquitous computing [3, 4] offer both opportunities and challenges. Technological innovation comes at a price, and in many countries the resources available to are insufficient to support what they need to do.

To overcome this problem, one solution that is being actively investigated is the use of Open Source software. Six main projects, for a total cost of €9.6 millions, have been funded by European Union through CORDIS (Community Research & Development Information Service, www.cordis.lu) on the study of different aspects of Open Source between 2001 and 2006. Furthermore, more than 80 projects that involve the development of Open Source have been funded by CORDIS in the same period, the largest to date being RODIN (€4.4 million).

Currently there is very limited amount of literature focussing on actual Open source software migrations [5, 6] and in this paper we present the experience of a group of European public organizations investigating the possibility of implementing supporting their services through the use of Open Source.

1.1 The Open Source movement

The Open Source movement [7] is an offshoot of the Free Software movement [8] and advocates the freedom to use, modify and redistribute software, on both pragmatic and philosophical grounds.

Freedom to use, to modify and to redistribute are the main attributes of Open Source software: in some cases, for example, the user will have the freedom to use and redistribute the software, but not to change it and thus the software does not qualify as Open Source. This is the case, for example, of freeware and shareware [9], that are merely royalty-free. Proprietary software is, regardless of price, software which does not provide all the liberties of Open Source.

Research on Open Source has already examined in depth the business models that drive the development of Open Source [10, 11, 12, 13, 14, 15]. We will focus on the practical implications of the migrations that we studied.

2 PROJECT OUTLINE

The research described in this papers was undertaken as part of the COSPA project, a 2 years and half project funded by the European Commission to investigate the viability of using Open Source on desktops within European public organizations [16]. The consortium consisted of 15 partners throughout Europe from public organizations, academia and industry.

Real-life experiences of public organizations adopting Open Source have been collected [17] by our group as part of its involvement in the project and are described and discussed within this paper.

The collection of the information about the experience and the motivations of the public organizations was divided in two parts. The first part ran between August 2004 and January 2005, and focused on collecting administrative business processes used within the public organizations, which resulted in over 300 processes being collected. The second part ran between February and March 2005, and involved collecting reports about their open source experiences written by the public organizations, using questionnaires and interviews where necessary.

2.1 Public organizations involved in the study

This paper is based on data collected by organizations located within Denmark (Hanstholm), Hungary (Torokbalint), Ireland (South West Regional Authority) and Italy (Consorzio dei Comuni and Province of Pisa).

Beaumont Hospital is a hospital located in Dublin, Ireland. In 2000, Beaumont Hospital had an obsolete software environment based on proprietary software, and was at the very least unclear its observance on software licenses. Ensuring compliance to the copyright legislation by acquiring obsolete software did not make sense from neither financial nor technical perspective. On the other hand, the budget required to update the current software and to acquire the correct number of licenses was estimated to be 1 million euros, which was deemed to be unaffordable by the hospital. Use of software with a low or null cost of acquisition, such as royalty-free software and Open Source, became the focus of the new strategy. A full migration plan was prepared, including the migration of the backoffice and middleware applications and a migration of the desktop software in many departments.

Consorzio dei Comuni dell'Alto Adige is a consortium of more than 150 public organizations in the province of Bolzano, Italy. Almost all these are small and the consortium manages their IT infrastructure, including approximately 2,500 desktops. The first migration of server software to Open Source was in 1997, when Linux and the Samba file sharing technology were adopted as the new bases of their network infrastructure, in a incremental migration process that required six years. The migration of the desktop software began with the COSPA project and was focused on changing the office suite to OpenOffice.org.

Hanstholm is a Danish municipality of approximately 70 employees. During the COSPA project, a number of desktops have been migrated to StarOffice, a proprietary derivative of OpenOffice.org and a migration of the middleware was ongoing at the end of the COSPA project.

The Province of Pisa (Italy) decided in 2003 to proceed towards the adoption of Open Source and towards the promotion of the development of Open Source solutions that could be reused by others. The province has thus developed Open Source applications for internal use and a transition to OpenOffice.org was done during the COSPA project.

The South West Regional Authority comprises of three public organizations based in the South West Region of Ireland. The main objectives of its Open Source experimentation is to reduce the costs of IT services and ensure its citizens could gain benefits from open access. The initial experiments with Open Source began in May 2004, with a trial of OpenOffice.org, involving 50 desktops.

Torokbalint is a Hungarian municipality of nearly 40 employees. Employees are using rather old computers. A network is present but its use is mainly for incoming and outgoing communication, not for internal communication. Economic motivations are driving the adoption of Open Source. The main aspect of the migration is the replacement of Microsoft Office with OpenOffice.org; Torokbalint is already using Linux on all their servers, and a partial migration of the desktops is ongoing.

3 DISCUSSION

The public organizations within the study submitted experience reports describing their use and experiences of Open Source to date. These were considered in conjunction with the requirements obtained by studying the more than 300 business processes collected from these organizations and the software requirements that they explicitly stated. The subsections that follow focus on a qualitative analysis of the issues that arose, while more information on the process of data collection and on the raw data itself can be found in the deliverables 2.4/2.5 [17] and 4.3 [18] of the COSPA project. More information on the Open Source software that has been analysed for suitability for use in the public administrations has been included in the deliverable 2.1 [19] of the project..

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

3.1 Strategic aspects

Efficient use of resources is one of the main topics that the public organizations need to consider and different strategies may be required according to the temporal horizon. The public organizations within the project could focus on medium to long term objectives, thanks, at least in part, to the funding that they received towards the experimentation costs.

Savings have been clearly identified by some partners, mainly in licensing costs. Beaumont Hospital identified savings of 8 million euros (over a five year period). Consorzio dei Comuni identified 1.5 million euros of savings (the cost that would be needed to buy a proprietary alternative of the Open Source software that they are using). At Hanstholm they estimate a saving of 41,000 euros per two year cycle, that correspond to the amount currently required to upgrade their office suite according to the licensing policy of their IT supplier. Torokbalint estimates 17,000 euros of savings for the next year. For the South West Regional Authority and for the Province of Pisa and Genova, economic savings are, again, in licensing costs.

Amongst the advantages, the public administrations reported a simplification in the acquisition and management of software. Using Open Source software *decreased the internal cost of license management* (as no licenses are required) and tracking the number of licenses that have to be purchased can be costly and time consuming. Moreover, using Open Source allowed the deployment of the same configuration regardless of the individuals use of the tools without wasting money whilst *simplifying the installation and maintenance* procedures which can be labour intensive and costly.

Cost of labour and technology can strongly influence the cost of a migration. Open Source tends to be more labour-intensive than proprietary software, and it might be less beneficial in countries with a high labour costs. Instead, amongst the test locations, it was perceived that the possibility to use legacy hardware systems, and was limiting the technological obsolescence of their systems.

The *types of software used* need to be considered as well, as they can have major implications for the viability of a migration. Public organizations that are using varied and highly specialized software find it more difficult to migrate to Open Source. Contrastingly, public organizations with a low level of software use (as Torokbalint) find migration much easier.

All the partner public organizations deployed incremental migrations, with different departments migrating at different times. This approach appears to have the advantage of phasing the effort and of improving the experience by trial and error. At Consorzio dei Comuni, the long term plan is to remove all the proprietary software still in use (wherever possible), but in the short term, some departments still use proprietary software, such as the accounting system. Replacing the software before the end of their life cycle is not considered as delivering sufficient benefits.

Interoperability is a further issue that public organizations need to consider. Public organizations are interconnected in different ways [18] and need to exchange inputs and outputs, both internally (internal communication) and with any other organizations or individual (external communication). Deferred communication also needs to be considered: a public organization needs to be able to have a memory and to query it, not to lose the information that has been acquired over many years.

The number, the importance and the type of interconnections of the public organizations has to be considered too, to ensure that it will still be possible to receive inputs and provide outputs from and to other organizations.

Three possible scenarios are discussed in the next paragraphs, based on the type of relation between the organizations. From an operative point of view, we consider that both external and internal communication obey the same rules.

In the first scenario, both *organizations have their own autonomy and similar contractual power*, or in any case a possibility exists that agreement can be reached on how to exchange data. Public organizations using different formats for internal use can decide a data standard to be used for external data interchange. When a consensus can be easily reached, a transition to Open Source will not cause major problems for communication. For example, public organizations that are using OpenOffice.org can easily agree to export the documents in several other data standards, including different Microsoft data standards.

In the second scenario, a *public organization which has limitations on the viable options* and is, at least for the communication with other organizations, bound to the use of specific data standards or software. This is the case, for example, when a public organization is dependent on another organization: making a transition to Open Source implies finding a way to adapt to the situation.

If the use of a specific data standard is mandatory for their communication, it might be possible to use software that exports data to that standard. If the use of a specific software package is required, it needs to be considered thoroughly. In some cases it is possible to make proprietary software coexist alongside Open Source. For example, it might be possible to make full transitions to Open Source operating systems, running the legacy applications through emulators such as Wine [20]. In other cases, public organizations may decide to make a migration of the office suites, but not of the operating systems, or a migration of the operating system, but not for all the computers.

For example, Torokbalint is required to use specific software to communicate with the central government and its banks, which is provided to them without cost. These software are used by nearly 50% of the users but runs only on a proprietary operating system, and that needs to be considered.

In the last scenario, the *public organization has a high contractual power* or even full power on deciding the specific data

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

standards or even the software to be used for exchanging inputs and outputs. For example, this may be the case for large public organizations. Suppliers will be more willing to conform to the requirements of the public organization, and other public organizations may decide to follow the example. This is the case for the Province of Pisa which is involving local companies and other public organizations in a common migration scenario.

During our analysis of the administrative business processes, we found that only 2.5% of the processes reported by the public organizations were using *open data standards* (freely available and written by *super partes* organizations). If a similar percentage could be confirmed across Europe, the attention to interoperability would be essential. Public organizations face the real risk of being enclosed in multiple proprietary environments where interoperability is difficult and laborious, or a single software environment over which they have little or no control.

Open Source adoption can be effectively slowed by *interoperability* problems. This has been experienced, to differing degrees, in all the locations but at a higher level when the public organizations do not have high contractual power or are not wholly committed to an Open Source migration. On the other hand, proprietary software does not ensure interoperability between different versions of the same software. For example, Consorzio dei Comuni experienced difficulties in accessing documents created 10 years ago with a popular proprietary text editor that is no longer supporting its own legacy data formats. Using Open Source software allows to overcome this type of problems [21], ensuring that it will be possible to support any legacy format whenever necessary.

The reports from the public organizations also showed that they feel that *flexibility* is one of their main requirements and that being capable of adapting to different contexts is the key to their success [22]. Public organizations are interested in not being linked to just one architecture, Open Source or proprietary. Only in 34% of the software requirements analyzed, public organizations gave a preference to just one operating system.

To conclude, public organizations need to integrate applications in their middleware. Consorzio dei Comuni has been involved in Open Source for a number of years and stated that from a technical point of view they can now easily develop customized solutions that meet their internal needs.

3.2 Environmental aspects

Public organizations take their decisions in the context of a specific environment. Political, legal and self-preservation aspects are considered at the management level and can strongly influence the evaluation.

Different political considerations play a role in the various perceptions of Open Source. From this perspective, Open Source can increase the *inclusiveness of the electronic experience*. Much Open Source software is distributed free of charge; thus a public organization that is using Open Source does not force its citizens to purchase software in order to communicate with it.

Promotion and support of Open Source has an impact not just on the public organization itself, but can affect the *local economy* and this is also considered at the political level. The vitality of local businesses can be impacted because the public organizations will require, in many cases, support for the implementation of Open Source. In the Open Source model there is a shift from acquiring a product in the global market, to acquiring a service, very often in the local market. The side effects in the local market appear to have influenced in a positive way the acceptance of Open Source, as reported by some partners of the COSPA project so far [18].

Open Source by definition provides the seeds for cooperation, which can be a key factor for transforming services. Nevertheless, the approach towards cooperation can vary greatly in different contexts.

Public organizations which acquire or develop Open Source are free to pass it on to other public organizations. Thus Open Source can be strategic for lowering the cost of software acquisition, as solutions developed by or for other public organizations can be reused. Reusing and improving software can also have a high *impact on both the quality and variety* of new digitally-based services provided, especially in the context of e-government.

Alternatively public organizations may prefer not to establish links with other public organizations. From a Darwinian point of view, cooperation and competition are both possible strategies and the choice of one or the other strictly depends on the environment in which the organizations reside.

A concept that has occurred repeatedly in our study is that the decision of the migration is linked to the perceived *personal consequence of the failure of the project*.

Interviews held at the beginning of the research with a number of managers in UK public organizations indicated that their positions are not viewed as being secure in case of failure. A strong personal responsibility links the managers with their decisions, and that makes them easily subjects of Fear, Uncertainty, and Doubt (FUD) [23]. Decisions may then be affected by considerations of not only of what is best for the public organization, but possibly what is the safer option for the individual decision maker. The immediate consequence is a lower propensity to migration, and innovation in general. At the same time, managers of public organizations feel that choosing a reputable company is a way of covering themselves if something goes wrong. The point of view is based on the fact that they would be better able to justify their decision if something went wrong with a reputable company or product than if they had chosen something different or new.

It is important to note that these views were localized and in most of the other locations the management was not afraid of the consequences of a failure in the migration with respect to their job. It was not clear if this can be linked to the pos-

sibility to hide the failure or to the unlikeliness that it would affect their job.

To some, pirate software is a real and practical alternative both to proprietary software and to Open Source. The advantage of pirate software, notwithstanding ethical considerations, is clear: a very *low cost of acquisition*, that tends towards to the cost of the storage medium plus the cost of transfer, as can be verified checking for "second hand" software in web sites such as e-bay (www.ebay.com).

In some contexts *it is realistic to estimate that proprietary software is, in fact, not an option under consideration*, but that organizations are willing to consider only software with a very low cost of acquisition. The GNI per capita in 2004, as published by the World Bank [24] is $ 8,270 in Hungary, and $ 2,350 in the FYR of Macedonia, compared to $ 26 120 in Italy and $ 34 280 in Ireland, but the price of software is only loosely linked to the local economy and is quite constant across Europe. Therefore, proprietary software in less wealthy countries can be far less appealing. Moreover, the level of enforcement of copyright legislation may vary across countries, making in some cases pirate software a low cost-opportunity option as it is not likely to be actively persecuted.

Statistics on software piracy published in 2004 by Business Software Alliance (BSA) [25, 26] show that the estimated piracy level is at 29% in the UK, compared to 37% across the European Union and an average of 70% in Eastern Europe. The other countries represented in the COSPA, apart from Denmark (26% of estimated piracy) all have an higher estimated piracy rate: 49% for Italy, 42% for Hungary and 41% for Ireland.

A possible consideration is that public organizations in countries where piracy is more widespread are more likely to be affected by the problem. It is possible that the decision in favour of experimentation with Open Source might be linked, in some cases, to the need for action on software piracy.

3.3 Technical aspects

When considering any change of software, technical considerations are of major importance: software unable to meet the technical requirements is unlikely to provide a suitable solution.

It is important to know what the *requirements* are and to analyse whether the software can meet them. However, identifying the requirements may be far from trivial: the effort required may depend on numerous factors, including whether the software will be used to replace a manual system, existing software with similar functionality or whether increased functionality is required.

From our study it appeared that many users simply do not need lots of the advanced features that the general software such as office applications provide. It could be the situation for example that an office suite that is currently in use in the organization is far more advanced than an Open Source alternative, but if the requirements were simple and the funds limited, this could be a good alternative.

Different members of staff may have different requirements of the same software packages and not all staff opinions are necessarily equally influent. For example at Beaumont Hospital the senior management was not happy to use the newly installed e-mail system and was dissatisfied with the office suite. While these users were few in number, they have a large influence on the software that is used and consequently a migration back to the software packages they were previously using is a possibility.

Also, there might not be a suitable Open Source option. This concurs with the results of the Open Source trials by the UK government that were published in 2002 and 2004 [27, 28].

Further studies might investigate if the lack of advanced software requirements is linked to a Taylorization of the work process, and if so what are the implications connected.

The public organizations report that one of their main preoccupation is to have a software environment that is *secure, reliable and stable*. It needs to be secure, meaning that it must be possible to operate only actions that have been allowed. It needs to be reliable, that is to work in a predictable way, so that to the same set of procedures always corresponds the same, predictable reaction (e.g. the software should not crash). Finally, it needs be stable, meaning that it should not be subject to sudden or extreme evolutionary change.

Thus public organizations have to deal with the possibilities of contrasting requisites. At the present time, having a secure and reliable software environment often means constant software and/or hardware updates. Software updates depend on the support cycle of the product, often forcing an upgrade or a migration to newer products. Similarly, hardware updates are a recurring preoccupation for public organizations, both for the cost of hardware itself and for the cost of labour required for the set-up. This has been mentioned in their reports by Beaumont Hospital, Consorzio dei Comuni and Torokbalint, and it is likely to be relevant to most public organizations.

Open Source can help the public organizations in this context. Open Source operating systems offer systems to manage and automate software updates using custom repositories. Public organizations can set up repositories of updates and force automatic update of the computers, which can be tailored to their specific requirements.

Similarly there are companies that specialize in support for legacy software, which are no longer supported by the original distributor. Finally, graphical environments such as XFCE or IceWM can be used on quite old computer hardware, whilst still taking advantage of the security updates.

Training in the Open Source technologies is another important issue and different approaches have been used. In Torokbalint, the training was delivered one-to-one, during

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

the migration, due to the very small size of the public organization. In the other locations, training has been carried out using a two-tier approach, with courses for internal staff who later spread the acquired knowledge to the entire organization through internal courses.

To conclude, changes in the IT infrastructure need to address the conflicting interests of users and project managers [29, 30], in order to minimize the possibility of failing.

4 CONCLUSIONS

The future scenario of Open Source is not yet clear; what is instead taking shape is a *digital divide*, between organizations that decide to use proprietary software and organizations that have a widespread use of Open Source.

Many successful Open Source projects are or can be used in the public organizations. Public organizations can use on their desktops a Open Source *operating system* (e.g. Linux), a Open Source *desktop environment* (e.g. Gnome and KDE), an Open Source *office suite* (e.g. OpenOffice.org), and Open Source *software for e-mail and web navigation* (e.g. Mozilla). Similarly, the IT infrastructure can use Open Source software such as Apache, Bind, Samba. However, no sufficiently advanced groupware solutions are currently available, and the same applies to middleware applications such as payroll system, business analysis software, financial analysis software, to name a few.

Several different licenses are used in the Open Source community, to release software and documentation. In July 2005, we performed an analysis of the software map at SourceForge (sourceforge.net), the widest existing repository of Open Source (including more than 65,000 Open Source projects) to study the diffusion of the different licenses. At that time, we found that nearly 69% of the software included was released under the GPL license, 11% under the LGPL and 7% under the BSD license. A previous analysis by Wheeler [31] showed similar results: in 2003, 71% of SourceForge Open Source projects (45,000) were using the GPL, 10% the LGPL and 7% the BSD license. It needs to be noted that the software can be released (and often is) under more than one license and these data refers to software that included the GPL as possible license. Furthermore, in Wheeler's earlier analysis [32] of Red Hat Linux 7.1 (Red Hat was and still is the most popular Linux vendor) he found that nearly 50% of its code was released under the GPL only.

The GPL requires that the rights that a user received with a GPL software must be granted by any program derived or linked to it, which means from a practical point of view, that developers including GPL software in their projects must release their work *under the same conditions*. The LGPL is used mainly by libraries (shared components), and allows to use it in any type of software, but requires that any modifications to the software are released under the same conditions. The Open Source movement is developing a new environment, that is taking advantage from the licensing conditions to spread.

Innovation can spread both inside the Open Source community and amongst developers of proprietary software, but the two groups cannot easily communicate. Open Source and proprietary technology currently cannot be easily mixed. Users can use Open Source and/or proprietary applications, but in most cases Open Source applications cannot have proprietary components and vice versa. Public organizations that decide not to use proprietary software are enrolling in a community that is open to the members and very closed to the non-members.

Further directions for our research will include an in-depth analysis of the diffusion of different types of Open Source licenses. Understanding Open Source and its implications cannot be limited neither to the study of the economical aspects, nor of the technical ones, but should include a more in-depth overview of strategic aspects.

Open Source is a phenomenon that public organizations could be looking towards in the future as a viable way forward to deliver high quality administrative services to citizens in a cost-effective manner.

However, there are many issues that need to be explored and resolved before it will be possible to exploit these opportunities to the full. One of these is the relationships between the users and developers of Open Source. Whilst standard Open Source packages such as office applications are of a generic nature and can, in many cases, be adopted into an administrative environment there will be many areas where new software, compatible with other Open Source applications and open data standards needs to be developed. These may be replacing propriety packages, bespoke solutions or may be new applications to support a new administrative function or a manual activity. Thus, it is essential that an effective communication framework be established that will bridge the gap between the public organizations and Open Source developers.

The data collected does not cover all types of public organizations in all countries of the European Union. However, the consistent nature of the data collected leads the authors to believe that it is likely to be representative of a much wider community of public organizations. What may differ is the amount of progress different public organizations have made towards supporting their administrative processes with IT. Thus. there is a possibility that the more experienced public organizations might be able to assist others in adopting solutions for processes that are essentially the same.

REFERENCES

1. West, D. M. (2004). 'E-government and the transformation of service delivery and citizen attitudes', Public Administration Review, 64 (1), pp. 15-27.
2. Edmiston, K. D. (2003). 'State and local e-government - Prospects and challenges', American Review of Public Administration, 33 (1): 20-45.
3. Jessup, L. M., & Robey, D. (2002). 'The relevance of social issues in ubiquitous computing environments', Communications of the ACM, 45 (12), pp. 88-91.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4. Banavar, G, & Bernstein, A. (2002). 'Software infrastructure and design challenges for ubiquitous computing applications', Communications of the ACM, 45 (12): 92-96.
5. Waring, T., & Maddocks, P. (2005). 'Open Source Software implementation in the UK public sector: Evidence from the field and implications for the future', International Journal of Information Management, 25 (5), pp. 411-428. \
6. Fitzgerald, B., & Kenny, T. (2003). 'Open Source Software the Trenches: Lessons from a Large-Scale OSS Implementation', ICIS 2003, pp. 316-326.
7. Perence, B. (1999). 'The Open Source Definition', DiBona, C., Ockman, S., Stone, M. (Eds.), Open Sources: Voices from the Open Source Revolution, pp. 171-184.
8. Stallman, R. (2004). 'The Free Software Definition', Free Software Foundation, http://www.gnu.org/philosophy/free-sw.html (July 6th, 2007)
9. Werbach, J. L., & Dreben R. N. (2003). 'The Accidental Licensor: Advanced Issues in Software Licensing', ACCA Docket, 21 (2), pp. 54–71.
10. Bonaccorsi, A., & Rossi, C. (2003). 'Why Open Source Software can succeed', Research Policy, 32, pp. 1243–1258.
11. Haruvy, E., Prasad, & A., Sethi, S.P. (2003). 'Harvesting altruism in Open Source Software development', Journal of Optimising Theory and Applications, 118 (2), pp. 381–416.
12. Hertel, G., Niedner, & S., Herrmann, S. (2003). 'Motivation of software developers in Open Source projects: An Internet-based survey of contributors to the Linux kernel', Research Policy, 32, pp. 1159–1177.
13. Lakhani, K., & Wolf, R. (2003). 'Why hackers do what they do: Understanding motivation and effort in Free/Open Source Software Projects', http://ssrn.com/abstract=443040 (July 6th, 2007)
14. Lerner, J., & Tirole, J. (2001). 'The Open Source Movement: Key research questions', European Economic Review, 45, pp. 819–826.
15. Lerner, J., & Tirole, J. (2002). 'Some simple economics of Open Source', The Journal of Industrial Economics, 50 (2), pp. 197–234.
16. Kovacs, G.L., Drozdik, S., Zuliani, P., & Succi, G. (2004). 'Open Source Software and Open Data Standards in Public Administration', W. Elmenreich et al. (Eds.), Proceedings of the 2nd IEEE International Conference on Computational Cybernetics, pp. 421-428.
17. Chambers, S., Baravalle, & A., Holcombe, M. (2005). 'Analysis of Requirements for OS/ODS Applications in the Public Administration', Deliverable 2.4/2.5 of the Cospa project, http://www.cospa-project.org/download_access.php?file=D2.45-AnalysisOfRequirementsForOSandODS.pdf
18. Baravalle, A., Chambers, S., & North, S. (2005). 'Experience report on the implementation of OS applications in the partner PAs', The Cospa Project, http://www.cospa-project.org/download_access.php?file=D4.3-ExperienceReportOnTheImplementationOfOS.pdf (July 6th, 2007)
19. Free University of Bozen/Bolzano (2006). 'Catalogue of available Open Source tools for the PA', The Cospa Project, http://www.cospa-project.org/download_access.php?file=D2.1-CatalogueOfOSSUsed-InThePA.pdf (July 6th, 2007)
20. Hnizdur, S., & Briscoe-Smith, C. P. (2003). 'The IDA Open Source Migration Guidelines', http://www.netproject.com/docs/migoss/v1.0/ (July 6th, 2007)
21. Holtgrewe, U. (2004). 'Articulating the speed(s) of the Internet - The case of open source/free software', Time & Society, 13 (1), pp. 129-146.
22. Carroll, G. R. (1989), 'Ecological Models of Organizations', Administrative Science Quarterly, Vol. 34, No. 3, pp. 503-507.
23. Wikipedia (2005). 'Fear, Uncertainty and Doubt', Wikipedia, http://en.wikipedia.org/wiki/FUD (July 6th, 2007)
24. The World Bank Group (2005). 'Data Query', World Bank, http://devdata.worldbank.org/data-query/ (July 6th, 2007)
25. BSA (2004). 'Major IDC Study Finds 37 Percent of Software in Use in the European Union Is Pirated', http://www.bsa.org/eupolicy/press/newsreleases/Major-IDC-Study-Finds-37-Percent-of-Software-in-Use-in-the-European-Union-Is-Pirated.cfm (July 6th, 2007)
26. BSA (2004). 'Global Piracy Study Press Releases', BSA, http://www.bsa.org/globalstudy/pressreleases/ (2nd January, 2005)
27. Office of Government Commerce (2002). 'Open Source Software Guidance on implementing UK Government Policy', http://www.ogc.gov.uk (July 6th, 2007)
28. Office of Government Commerce (2004). 'Open Source Software Use within UK Government Version 2, 28th October 2004', http://www.ogc.gov.uk (July 6th, 2007)
29. Sauer, C. (1993). Why Information Systems Fail: a case study approach, Alfred Waller, Henley-on-Thames.
30. Drummond, H, Hodgson, J. (2003). 'The chimpanzees› tea party: a new metaphor for project managers', Journal of Information Technology, 18 (3), pp. 151-158.
31. Wheeler, D. A. (2005). 'Make Your Open Source Software GPL-Compatible. Or Else', http://www.dwheeler.com/essays/gpl-compatible.html (July 6th, 2007)
32. Wheeler, D. A. (2005). 'More Than a Gigabuck: Estimating GNU/Linux›s Size', http://www.dwheeler.com/sloc/redhat71-v1/redhat71sloc.html (July 6th, 2007)

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Competition between mobile TV and broadcast industries

Imsook Ha

Information and Communications University
Daejeon, South Korea
Quello Center for Telecommunication Management and Law
Michigan State University
East Lansing, MI 48824, USA
haimsook@msu.edu

Johannes M. Bauer

Department of Telecommunication, Information Studies, and Media
Quello Center for Telecommunication Management and Law
Michigan State University
East Lansing, MI 48824, USA

Abstract The purpose of this study is to examine how mobile TV competes with existing media and how regulatory measures affect the intensity of this rivalry. To address this question, this study uses niche theory in which media are modeled as competing and coexist on scare resources. Most previous niche theory related studies did not reflect the effects of market changes and regulation issues for competition. In this study, we suggest a media competition map that attempts to visually display the position of media in terms of competition and superiority, and we measure the impact of regulatory policy on competition in the mobile TV market. Overall, the results reveal that cable TV shows a high level of generality and superiority on cognitive and affective dimensions. But terrestrial mobile TV (T-DMB) shows the widest diversity and highest superiority on gratification opportunities. T-DMB is superior to satellite-based mobile TV (S-DMB) with intense competition. But if T-DMB becomes a fee based service, it will lose its competitive superiority to S-DMB. The retransmission of TV program as deregulation on S-DMB seemed to have an insignificant effect.

Keywords Mobile TV, Niche theory, The use and gratifications, T-DMB, S-DMB

1 INTRODUCTION

As a result of the appearance of new media, existing media's dominant status would be marred and new media would fill up that niche, or they would replace the functions existing media have made. In this context, this study is proposed to analyze competition between existing and new media. By looking over the tendencies of the study on complementary relations between new media and existing media, the media with functional similarities are likely to replace the other but try to specialize in certain features or seek out new market and resources to avoid competition (Dimmick & Rothenbuhler, 1984). Also media have co-evolved and integrated into a third complex media with existing media based on the new media environment. Although new media have more or less taken over the existing media, the existing media survive and some are expected to continue to coexist with the new. (Fidler 1997)

On the other hand, the research results on media substitution can also be divided. First, the broad options provided to the viewers due to multi-media and multi-channels might lead to the replacement of some media. Many TV viewers gradually get out of habitual and ritualistic exposure, but choose the channels which would maximize their utility and satisfaction while meeting the instrumental purpose. (Jeffres 1978, Jeffres & Atkin 1996). The other part is about media substitution when the existing media has similar functions with the new one. The new media can provide the same function more effectively, nullifying and replacing the conventional media with new ones. Some studies showed that the internet would reduce the influence of TV (Coffey& Stipp 1997, Viswanath, Fergurson, and Perse 2000) Also the growing number of cable subscribers would drag down the ratings of terrestrial TV. (Baldwin et. al. 1992, Krugman & Rust 1987, Albarran & Dimmick 1993, Dimmick 1993, Dimmick et al. 1992, Weimann 1996). Henke (1989) said the VCR would replace TV viewing, especially entertainment programs, or patronization of theaters.

The broadcasting environment, called the era of digital technology, is changing more rapidly. In particular, given that TV is not simply limited within the territory of broadcasting, but rather crossing over the field of communication, broadcasting and communications convergence is ongoing. This kind of diverse change, crossing over terrestrial TV, cable TV, and VOD (WebTV), brings about the expansion of mobile TV. Mobile TV utilizes personal portable receivers or automobile receivers to allow reception of TV, radio, and data broadcasts of multi-channel broadcasts. This service appeared with the need to apply mobile, personal, multimedia broadcasting due to changes in environment, the coming of the digital broadcasting era, and the expansion of the broadcasting market. In 2005, South Korea started a mobile TV service, called S-DMB and T-DMB, on May 1 and December 1, respectively. Accordingly, when the service territory of mobile TV is reviewed, competition with existing media is expected.

Therefore, this study is intended to analyze the competitive relation between mobile TV and existing media in the environment of broadcasting and communications convergence. To analyze the competitive relation, this study conducted niche theory in which media are modeled as competing and coexist on scare resources. It has been applied to many areas of research in the media market. However, research on media applying the niche theory contains a few limits. First, previous studies can find only which medium has relative superiority, or whether the breadth of resources is wide through niche value. They have the disadvantage of an inability to provide integrated perspectives to consider whole objects in competition simultaneously. Second, market competition in the media market is getting complex and fierce due to emergence of multi-media and multi-channels. Therefore the importance of policy to regulate market competition is emphasized. The competition structure of industry could be changed according to how policy for competition is set up. But previous studies explained only the current status of competition and didn't reflect a change of competitive situation. In this regard, this study will suggest a media competition map based on niche theory. This will enable easily integrated understanding of the relation between new media and existing media, and a change of competition structure according to the change of the market environment. Under this background, this study has the following purposes. First, this study conducts a competitive relation between mobile TV and existing media through niche theory. Second, this study found the major regulatory policy on competition in the mobile TV market and analyzed the change of competition structure caused by the change of regulation.

2 LITERATURE REVIEW

The theory of Niche was evolved by ecologists to answer questions concerning how populations compete and coexist on limited resources in an ecological community. (Dimmick, J., Rothenbuhler E., 1984) When two organizations use the same resources, they are in competition. Three key concepts in niche theory are niche breadth, overlap, and competitive superiority. Niche theory is not only limited to biology, but it is a general theory about coexistence and competition. Despite the application whether it is animal or organizations, when two populations compete with limited resources, the theory is applicable. Therefore the theory of niche could often be used to explain competition and coexistence among media industry. Dimmick et al. (1984) first applied the niche theory to analyze advertising as resources for the competition between media. They were conducted on the competition among newspaper, TV, radio, and outdoor advertising industry over the same advertising resource from 1935 to 1980. Dimmick and Patterson(1992) investigated TV, cableTV and radio industry by the emergence of the cable. Due to those strengths, the study combining gratification and niche theory were conducted more actively. Dimmick (1993) analyzed the competition between media by estimating gratification. He has done a research on the competition between TV, cable TV, and VCR. Preliminary result shows that TV came out strong with wider niche breadth in affective and cognitive, but VCR in gratification opportunities. Dimmick et al. (2000) has analyzed competition between e-mail and telephone at the level of gratifications derived by consumer. The results indicate that since they used e-mail, forty-eight percent of respondents reported using the less. A wider spectrum of needs is being served by the telephone, whereas e-mail provides greater gratification opportunities.

Another important concept is the **uses and gratifications**. It considered one of the most widely accepted theoretical frameworks to study media adoption and use (Lin, 1996, Kang, 1999). These psychological motives motivate the audience to purposefully select certain media, or media contents, in other to satisfy a set of psychological needs behind those motives (Blumler 1979, Katz et al. 1974) Much previous research specifically investigated in the areas of new media (Atkin et al., 1998, Morris & Ogan,1996, Rafaeli, 1986, Newhagan & Rafaeli, 1996) But, due to theological, methodological uncertainty of concept in these studies, relationship between media adoption and gratification was not found clearly. Accordingly, many researches to refine the concept of gratification have processed. Greenberg (1974) divided the concepts of gratification into gratification sought (GS) and gratification obtained (GO). According to Rayburn (1984), there is strong correlation between these two gratifications, but they do not coincide always. Furthermore, many studies attempted to clarify what kind of relationship two gratifications build with other variables such as contact, adoption, and dependency with media. Palmgreen (1981) provide the result that while gratification sought(GS) does not divide viewers of American major television- network news program, gratification obtained(GO) is successful in predicting choice of news program. Based on these previous studies (Rayburn 1984, Wenner 1982) they concluded that rather than gratification sought (GS), gratification obtained (GO) is far more useful in explaining choice of media or program. Therefore, for finding competitive relations between media by applying the concept of gratification to niche theory, it is more suitable to use gratification obtained (GO) rather than gratification sought (GS). Gratification obtained (GO) as niche dimension is divided by cognitive and affective factors. Affective is related to feeling or emotional states and

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

cognitive factor related to the mental process is involved in knowing, learning, and understanding thing.

Another key concept in gratification is **gratification opportunities**. This concept was found first in questionnaire survey on gratification obtained from media. Respondents, after using media, pointed to the various selections, flexibility of time to use media as factors for satisfaction in addition to cognitive, affective dimensions of gratification. As a result of factor analysis, these factors were divided into different ones separate from cognitive, affective dimension (Dimmick 1993) When using media, users evaluate whether those media exist in the time and space they can use, and whether they can use media conveniently and flexibly, which is just gratification opportunity. User considers even the possibility to choose media and space situation together with whether used media can meet what he or she pursues. Therefore, media compete for gratification opportunity together with many sub-levels of gratification they can provide to users. When using existing media like television, movie, radio, and newspaper, users should be, in allocating their leisure time, in conformity with comparatively strict and limited timetable. Compared to this, new media provides more various choices and flexibility to users. So far, we could found that in niche theory, 'gratification' of uses and gratification approach becomes niche resources. On the macro level, users evaluate usable media on three levels of gratification sought (GS), gratification obtained (GO), and gratification opportunity, serving as the basis for choice of media

3 REGULATORY POLICY ON COMPETITION IN KOREA DMB MARKET

The DMB in Korea has faced a regulatory setback in its initial stage. Because of the absence of a clear concept of convergence in relevant policy and regulation and the convergence service, the DMB in Korea has faced overlapping regulation in one case and non-regulation in other cases (Shin 2006). In a process of introduction to S-DMB, it has brought about a sharp conflict between the broadcasting carrier and telecommunications service provider. The Korean Broadcasting Commission (KBC) plays an important role for the regulation of the DMB. They have postponed the schedule of approval for S-DMB businesses with reasons of insufficient demand, and the shrinking of the T-DMB market, and prohibited from the entrance to the broadcasting market. Coping with this situation, the S-DMB providers insisted deregulation for the preoccupation effect on the world mobile broadcast market and fair competition with existing broadcast media. After an introduction to S-DMB, a sharply confrontational situation continued between broadcast market participants about the matter of time to be introduced S-DMB, the proper number of service providers, the method to select service providers, and retransmission of terrestrial TV programs. First of all, allowing the retransmission of terrestrial TV programs to S-DMB has constituted the big conflict for existing TV broadcasters. KBC left the matter of retransmission as a contract between providers. The existing TV broadcasters, including KBS, MBC, and SBS, which were all given T-DMB licenses, opposed the retransmission of their programs on S-DMB. The S-DMB provider, TU Media, argued strongly for retransmission of terrestrial TV programs to its S-DMB subscribers in order to attract more subscribers. Their demands can be summarized as followed. In terms of viewer's rights to choose the channel, most viewers want to watch terrestrial TV programs through S-DMB because terrestrial TV programs are common property of the viewer and broadcaster. Secondly, all charged broadcastings except S-DMB are conducting real-time retransmission of terrestrial TV programs. In order to facilitate requirements for fair competition, retransmission should be allowed to S-DMB. Thirdly, decline of profitability of satellite DMB creates depression with associated industry fields including program provider and manufacturers. (TU Media). However, still facing strong opposition from S-DMB, terrestrial TV would not sign retransmission contracts until their T-DMB services became stable in the mobile television market in competition with S-DMB.

Another policy issue for competition is whether or not to allow T-DMB service providers to have pay-service. This is related to the cost of T-DMB gap-fillers (Lee 20005). In the case of S-DMB, its subscription fee includes gap-filler expenses but T-DMB providers insist that they should have a partial fee-based service because in the early development stage, profits from advertising could be fragile and not cover the cost of installing gap-fillers and other expenses. In this point, T-DMB provider intends to adopt various pay services in addition to terrestrial TV programs, which are provided free. They make request for profits using various business models such as monthly flat-rate service, paid service for selected channels, etc. Viewers will be able to directly vote for quiz shows or popular music channels as well as download the contents they like. Also, they can purchase products like clothes or bags that are featured in TV programs or advertisements. In this case, the size of communication traffic, using back channels, will increase and become a new source of revenue for mobile service providers (DMB portal) But, as T- DMB was born on free service because it adopts VHF frequencies, it is all potential users of T-DMB service that the principle of 'free service of T-DMB' should be kept without fail. Therefore, until now, KBC has experienced difficulties in finding solutions to the T-DMB cost problem

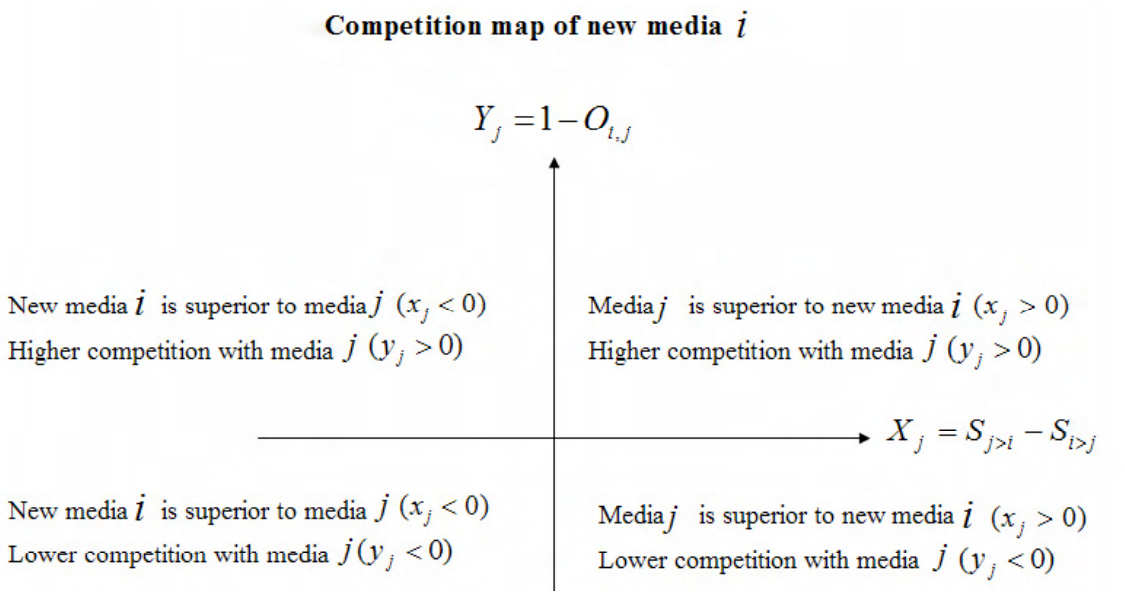
4 METHODOLOGY

The DMB in Korea has faced a regulatory setback in its initial stage. Because of the absence of a clear concept of convergence in relevant policy and regulation and the convergence service, the DMB in Korea has faced overlapping regulation in one case and non-regulation in other cases (Shin 2006) In this study, the measures of niche breadth, overlap, and competitive superiority developed by Dimmick (1993) are used to conduct competitive relationship among competitors.

First, the niche breadth indicates the degree to which a medium is capable of gratifying a relatively broad or relatively narrow spectrum of statements on a gratification dimension.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure1. Competition map of new media



$$B = \sum_{n=1}^{N} \frac{\left(\frac{\left(\sum_{k=1}^{K} GO_n \right) - KI}{K(u-I)} \right)}{N} \quad (1)$$

The perceived similarity in gratifications derived from two mediums is measured by the following equation 2. Overlap is an inverse measure. If the value is low, it means two mediums have similarity and higher overlap in gratification. To that extent, competition gets fierce. It can be considered as the substitution. The lower limit of the overlap is zero and indicates that the gratification niches of two mediums completely overlap.

$$O_{i,j} = \frac{\sum_{n=1}^{N} \sqrt{\sum_{k=1}^{K} \frac{(GO_i - GOj)^2}{K}}}{N} \quad (2)$$

Though the overlap indicates competitive relation of two mediums, it cannot be find which medium is superior in gratification. Therefore, Schoner's alpha(1974) is used to measure the competitive superiority between two mediums. A medium that obtains a significantly higher superiority score than another medium is superior in providing gratifications to the audience members. The difference in superiority between two means on a gratification utility dimension may be tested for significance using a t test for correlated groups. Formula measuring competitive superiority is as follows.

$$S_{i>j} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K}(M_{i>j})}{N}, S_{j>i} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K}(M_{j>i})}{N} \quad (3)$$

We designed a two-dimensional competition map with two axes of competition and superiority based on niche equations. The x-axis indicates the gap between superiority scores of two media and the y-axis is computed by inverse value of the overlap index. Centered by criteria medium i, it enables integrated analysis on comparative relation with medium j. Position of comparative medium j for criteria medium i can be defined as follows.

$$\begin{cases} x_j = S_{j>i} - S_{i>j} \\ y_j = 1 - O_{i,j} \end{cases}$$

Reviewing the characteristics of the map,

(1) When medium j is located in the area of $y_j > 0$ (first and second quadrant), competition between medium i and medium j is high.
(2) As y_j increases, degree of competition between medium i and medium j increases gradually.
(3) When medium j is located in the area of $x_j > 0$ (first and fourth quadrant), medium j is superior to medium i. ($S_{j>i}$ is greater than $S_{i>j}$)
(4) When medium j is located in the area of $x_j < 0$ (second and third quadrant), medium j is inferior to medium i.
(5) A s the value of x_j increases gradually, competitive superiority of medium j to medium i increases gradually. Figure1 shows the characteristics for four quadrants in media competition map.

In this study, we denote mobile TV (T-DMB and S-DMB) as a criteria medium. TV, Cable TV, and VOD are comparative media in the media competition map. In order to analyze competition relation among media, this paper suggests the following research questions:

Research question 1: *What is the competition relation with existing media caused by the appearance of mobile TV?*

- Does mobile TV have superiority (or inferiority) to existing medium j? ($x_j < 0$ or $x_j > 0$)
- Does mobile TV have higher competition (or lower competition) with existing medium j? ($y_j > 0$ or $y_j < 0$)

Research question 2: *How competition relation between mobile TV and existing media changes according to regulation policy for competition? (How will position* (x_j, y_j) *change?)*

- If it becomes possible for retransmission of terrestrial TV programs on S-DMB, how will competition in the broadcasting market change?
- If T-DMB becomes fee-based services, how will competition in the broadcasting market change?

5 RESULTS

In this study, we denote mobile TV (T-DMB and S-DMB) as a criteria medium. TV, Cable TV, and VOD are comparative media in the media competition map. In order to analyze competition relation among media, this paper suggests the following research questions This study is made up of only experienced users in TV, cable TV, VOD, T-DMB, and S-DMB as the concept of gratification obtained (GO). Subsequently, work for filtering inexperienced users through the questionnaire on whether or not to use was conducted in advance. This study assumes that the resource for media is gratification. Therefore, questionnaire items focused on factors for gratifications found in previous studies. And additionally, questionnaire items were added through studies on mobile TV. Finally six items of cognitive dimension, six items of affective dimension, and four items of opportunity were deducted. After confirming that three factors are classified into sub-level using 30 data collected for a preliminary test, this main survey was executed. To do this work, we conducted an online survey to evaluate the research model for 10 days. A total of 1,235 samples were gathered by a professional market research firm, Pollever (www.pollever.com). All the questionnaires used in this survey have been validated in previous studies. The majority of the sample substantially matches current mobile TV users.

For classifying the degree of gratification into sub-levels, factor analysis was conducted Research of gratification commonly employed a confirmatory factor analysis. However, as this study added items on gratification with mobile TV, a new broadcasting media which were not dealt with by existing studies, exploratory factor analysis was used. As for the methods of factor analysis, we conducted principal axis factor analysis with VARIMAX rotation. These methods are generally used in factor analysis for the gratification niche theory. (Albarran and Dimmick 1993, Dimmick 1993, Dimmick et al. 2000) After reviewing the result of factor analysis obtained for each medium and inspecting the reliability with Cronbach's alpha, we removed items deteriorating reliability of each satisfaction level or items belonging to each different dimension.

Accordingly, satisfaction of media user was classified into three factors. These factors are similar the three factor found in repeated studies of Dimmick(1993). The first factor is 'Cognitive', which includes five items on degree of satisfaction with news and utility of communication as information. The second factor is 'Affective', which consists of five items on amusement and relaxation acquired through the use of media. The third factor is 'gratification opportunity', which indicates degree of satisfaction with time, space accessibility and choice of contents in the use of media.

Next, three factors, niche breadth, niche overlap, and competitive superiority are calculated for each medium. The results of niche breadth are shown in Table 1.

Table 1. The results of niche breadth

	TV	Cable TV	VOD	T-DMB	S-DMB
Cognitive	0.67	0.69	0.59	0.60	0.58
Affective	0.69	0.73	0.62	0.63	0.60
Gratification opportunities	0.55	0.58	0.61	0.63	0.61

As shown in Table 1, in cognitive and affective dimensions, cable TV exhibited the broadest niche breadth, followed by terrestrial TV, but S-DMB shows the narrowest breadth. T-DMB shows a high degree of diversity on the opportunities.

Table 2. The results of niche overlap

Media	Cognitive	Affective	Gratification Opportunities
TV / CTV	0.96	0.72	0.9
TV / VOD	1.29	1.25	1.55
TV / T-DMB	1.11	1.14	1/58
TV / S-DMB	1.38	1.42	1.69
CTV / VOD	1.26	1.19	1.61
CTV / T-DMB	1.23	1.17	1.59
CTV / S-DMB	1.42	1.47	1.62
VOD / T-DMB	1.01	0.95	1.54
VOD / S-DMB	1.09	1.05	1.47
T-DMB / S-DMB	0.94	0.91	1.01

Table 2 shows the result of niche overlap. High competition across dimensions was found between cable TV and terrestrial TV, followed by T-DMB and S-DMB. Therefore, two

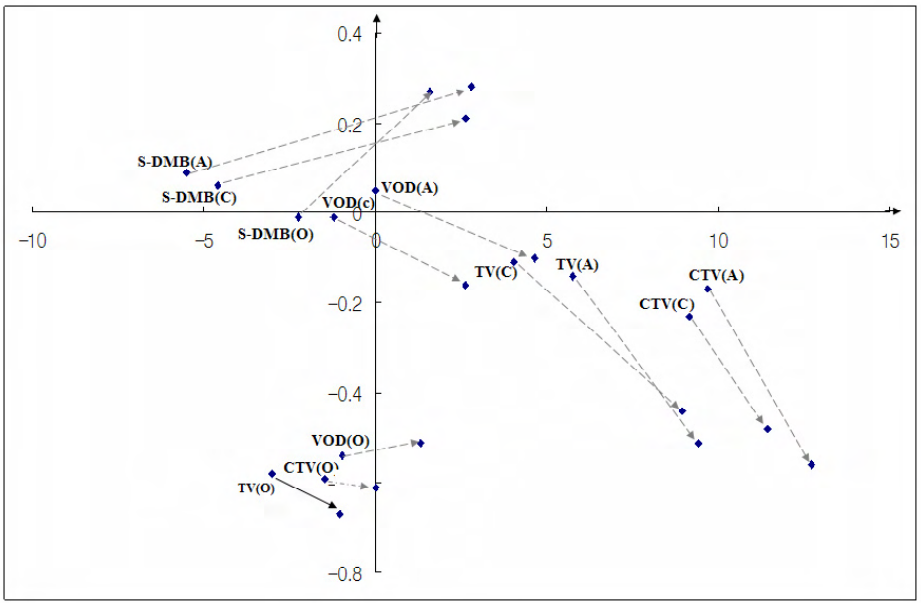
Table 3. The results of superiority

Media	Cognitive	Affective	Gratification opportunities
TV / CTV	CTV	CTV	CTV
TV / VOD	TV	TV	VOD
TV / T-DMB	TV	TV	T-DMB
TV / S-DMB	TV	TV	S-DMB
CTV / VOD	CTV	CTV	CTV
CTV / T-DMB	CTV	CTV	T-DMB
CTV / S-DMB	CTV	CTV	NS
VOD / T-DMB	T-DMB	NS	T-DMB
VOD / S-DMB	VOD	VOD	NS
T-DMB / S-DMB	T-DMB	T-DMB	T-DMB

NS=non-significant difference between mediums

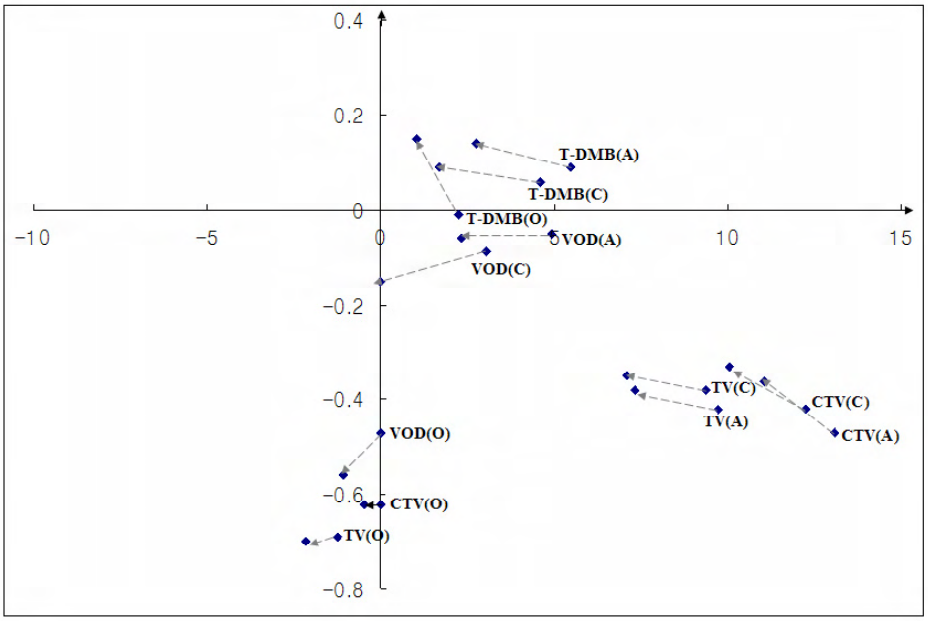
Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Competition map of T-DMB



* TV(C)= Position of TV in cognitive; TV(A)=Position of TV in affective ; TV(O)=Position of TV on opportunities
CTV(C)= Position of CTV in cognitive; CTV(A)=Position of CTV in affective ; CTV(O)=Position of CTV on opportunities
VOD(C)= Position of VOD in cognitive; VOD(A)=Position of VOD in affective ; VOD(O)=Position of VOD on opportunities
S-DMV(C)= Position of S-DMB in cognitive; S-DMB(A)=Position of S-DMB in affective ; S-DMB(O)=Position of S-DMB on opportunities
* - - - - > = position movement according to change in regulations

Figure 3. Competition map of S-DMB



* TV(C)= Position of TV in cognitive; TV(A)=Position of TV in affective ; TV(O)=Position of TV on opportunities
CTV(C)= Position of CTV in cognitive; CTV(A)=Position of CTV in affective ; CTV(O)=Position of CTV on opportunities
VOD(C)= Position of VOD in cognitive; VOD(A)=Position of VOD in affective ; VOD(O)=Position of VOD on opportunities
T-DMV(C)= Position of T-DMB in cognitive; T-DMB(A)=Position of T-DMB in affective ; T-DMB(O)=Position of S-DMB on opportunities
* - - - - > = position movement according to change in regulations

pairs are substitutes because they serve the similar gratifications. Results of lower competition were found when comparing new media: S-DMB, and existing media: TV and cable TV. S-DMB was considered as complementary of cable TV and terrestrial TV.

Table 3 shows the result of competitive superiority. As a characteristic result, cable TV shows superiority to all media in cognitive and affective dimensions. This is a result explaining that in the current Korean media market, cable TV functions as the most superior medium in cognitive and affective dimensions. Contrarily, in gratification opportunity, T-DMB shows superiority to other media. Meanwhile, T-DMB is superior to S-DMB across all dimensions.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

We suggested a media competition map to view the change of competition relation among media according to market change. T-DMB and S-DMB are defined as criteria medium. TV, cable TV, VOD, and S-DMB are considered as comparative media

First, we assume that T-DMB becomes fee-based services. The result is as follows in Figure2

As shown in Figure 2, S-DMB is located in the second quadrant before changing position. This means T-DMB and S-DMB compete fiercely and T-DMB has competitive superiority. But it moves to first quadrant after T-DMB becomes a fee based service $\left(\left(x_j < 0, y_j > 0\right) \Rightarrow \left(x_j > 0, y_j > 0\right)\right)$.

It shows that the degree of competition with S-DMB increases but T-DMB falls into competitive inferiority. Even if T-DMB provides fee-based services with better content than now, the pay service will cause a big loss of gratification. Also after the position change, lower competition in cognitive and affective dimensions was found between T-DMB and existing media (TV, cable TV, and VOD). But there is not a relatively big difference in position movement in gratification opportunities. Second, we assume that S-DMB can provide real time retransmission of terrestrial TV programs

As shown in figure 3, T-DMB is located in first quadrant before the position change. After S-DMB is able to provide retransmission of terrestrial TV programs, T-DMB is still in the first quadrant. It is shown that T-DMB has continuing competitive superiority. Also, S-DMB is still inferior to existing media $x_j > 0$. It is to say that even though retransmission as deregulation is allowed to S-DMB for fair competition, it cannot still escape from an inferior competition situation.

6 CONCLUSION

This study considered how new media compete with existing media, and how new media establishes competitive relations according to regulation policy for competition. We conducted the niche gratification theory, in which media compete and coexist on limited gratification resources and a media competition map to display the change of media competition in regulations. Our results were thoroughly reviewed and the following strategies are proposed to compete among media based on the regulatory policy on competition

First, Cable TV has the widest niche breadth and highest superiority on the cognitive and affective dimensions compared to all other media. It shows that cable TV functions most broadly and in a superior position in the Korean broadcasting market. Recently Korean cable TV has made considerable growth. So far, the program provider in cable TV mainly organized a retransmission of terrestrial or foreign programs, but program providers like On media, CJ media, etc. have, lately, produced their own programs, like drama and music show. Intense competition across all dimensions was found between terrestrial TV and cable TV. Therefore user satisfaction of cable TV has increased dramatically due to multi-channels and good-quality content, compared to terrestrial TV.

Second, S-DMB shows the narrowest niche breadth and lowest superiority among all media in cognitive and affective dimension. Also, high competition was found when comparing S-DMB and T-DMB. In actuality, the S-DMB provider, which is in the inferior position on the side of consumer gratification, consistently has asked for retransmission of terrestrial TV programs for fair competition with T-DMB. From this point, this study assumes the possibility for retransmission of S-DMB to forecast future market competition. The results showed that S-DMB is still in an inferior position on cognitive and affective dimension. After the retransmission of S-DMB, though conditions of higher competition with T-DMB were found, it still couldn't have competitive superiority across all dimensions Therefore, even if it is possible for S-DMB to rebroadcast terrestrial TV programs, it is found not to be able to provide more satisfaction compared to T-DMB.

Third, though T-DMB is inferior to cableTV and TV in cognitive and affective dimension, it scored the widest breadth and most superiority in gratification opportunities. The gratification opportunity has features of time-space accessibility and diversity of choice. In a time space situation in which CableTV and TV cannot be used, the easily portable T-DMB can be superior to existing media. This result is similar to Dimmick's study (2000). He insisted that new media brought diverse choices and control over consuming time to consumers by providing higher gratification opportunities. But if T-DMB becomes a fee-based service, the competitive superiority across dimensions will be lost. Even if contents of T-DMB provided as a charged service are far better than now, the pay service will cause a big loss of gratification felt by consumers.

Finally, in the case of regulatory aspects, the current policy structure in the Korean broadcast market seems to confine the possibility of broadcasting and telecommunication convergences. With regard to DMB regulations, the Korean Broadcasting Commission (KBC) has not been clear in its position concerning whether or not to allow S-DMB service retransmission of terrestrial TV program and T-DMB service providers to have a pay-service. Therefore, this study is significant because it examines how important regulatory issues affect competition among broadcasting media..

ACKNOWLEDGEMENTS

"This work was supported by Korea Research Foundation Grant funded by the Korean Government (MOEHRD)" (KRF -2006- 612-H00003)

REFERENCES

Albarran, A. B., Dimmick, J. (1993). "An assessment of utility and competitive superiority in the Video Entertainment Industries", Journal of Media Economics, 6, 45-51

Atkin,D., Jeffres, L., Leuendorf, K. (1998), "Understanding internet adoption as telecommunications behavior", Journal of Broadcasting & Electronic Media, 42, 317-336

Baldwin, T. F., Barrett, M., and Bates, B. (1992), "Uses and values for news on cable Television", Journal of Broadcasting and Electronic Media, 36, 225-234

Blumer, J.G. , Katz, E. (1979), "The role of theory in uses and gratifications studies", Communication Research, 9, 9-36

Coffey, S., Stipp, H. (1997), "The Interactions between computer and television usage", Journal of Advertising Research, 37, 61-67

Demers, D. (1994), "Relative constancy hypothesis, structural pluralism, and national advertising expenditures", Journal of Media Economics, 7, 31-48

Dimmick, J. (1993), "Ecology, economics and gratification utilities", In Alexander, A and R. Carveth(Eds), 1993 Media Economics: Theory and Practice. Hillsdale, NJ: Jawrence Erlbaum Associates, Inc.

Dimmick, J., Rothenbuhler E. (2000), "The gratification niches of personal E-mail and the telephone: competition, displacement and complementarity", Communication Research, 27, 227-248

Dimmick, J., Rothenbuhler E. (1984), "Quantifying competition among media industries", Journal of Communication", 34, 103-119.

Dimmick, J., Patterson, S. J., Albarran, B. (1992), "Competition between Cable and broadcast industries: A niche analysis", Journal of Media Economics, 5, 13-30

Dimmick, J., Wallschaeger, M. (1986), "Measuring corporate diversification: A case study of new media ventures by television network parent companies", Journal of Broadcasting & Electronic Media, 30, 103-119

Ferguson, D. A., Perse, E. M. (2000), "The world wide web as a functional alternative to television", Journal of Broadcasting & Electronic Media , 44, 155-174.

Fidler, R. (1997), "Media morphosis: Understanding new media", Thousand Oaks, CA: Pine Forge

Greenberg, B. S. (1974), Gratifications of Television viewing and their correlates for Britich Children. In Blumler, J.G., and E, Katz(eds) 1974. The use of mass communications: Current perspectives on gratifications research bevely Hills: Sage

Henke, L. L. , Donohue, T.R. (1989), "Functional displacement of traditional TV viewing by VCR owners", Journal of Advertising Research, 29, 17-21

Jeffres, L.W. (1978), "Cable TV and viewer selectivity", Journal of Broadcasting 22, 155-170

Jeffres, L.W., Atkin, D. (1996), "Predicting use of technologies for consumer and communication needs", Journal of Broadcasting & Electronic Media, 40, 318-330

Kang, M., and Atkin D. J. (1999), "Exploring the role of media uses and gratifications in multimedia cable adoption", Telematics and Informatics, 16, 59-74

Krugman, D.M., Rust, R.T. (1987), "The impact of cable penetration on network viewing", Journal of Advertising Research, 27, 9-13

Lee, S. W. (2005), "TV in your cell phone: The introduction of digital multimedia broadcasting (DMB) in Korea.", Annual Telecommunications Policy Research Conference

Lin, C.A. (1996), "Looking back: the contribution of Blumler and Katz's uses of mass communication to communication research", Journal of Broadcasting & Electronic Media, 40, 574-581

Morris, M., Ogan, C. (1996), "The internet as mass medium. Journal of communication ", 46(1), 39-50

Newhagen, J.E., Rafaeli, S. (1996), "Why communication researcher should study the internet: a dialogue", Journal of communication ,46(2), 4-13

Palmgreen, P., Wenner, L., Rayburn, J. (1981), "Gratification discrepancies and news program choice", Communication Research, 8(4), 451-478

Rafaeli, S. (1986), "The electronic bulletin board: a computer-driven mass medium", Computers and Social Science ,2, 123-136

Rayburn , J. D. , Acker, T. (1984), "Media gratification and choosing a morning news program", Journalism quarterly, 61. 149-156

Rayburn, J.D., Palmgreen, P. (1984), "Merging uses and gratifications and expectancy value theory", Communication Research, 11, 537-562

Schoener, T. W. (1974), "Some methods for calculating competition coefficients from resource utilization spectra", The American Naturalist, 108, 332-340

Shin, D. H. (2006), "Prospectus of mobile TV: Another bubble or killer application? ", Telematics and Informatics ,23, 253–270

Weimann, G. (1996), "Cable comes to the holy land: The impact of Cable TV on Israeli viewers", Journal of Broadcasting & Electronic Media, 40, 243-257

Wenner, L. A. (1982), "Gratifications sought and obtained in program dependency: a study of networking evening news and 60minutes",Communication Research ,9,539-560

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Innovative collaboration

Gillian Rawlings

Edge Hill University

Abstract The focus of this paper is Innovative Collaboration and how this is achieved through using the successful assimilation of moodle, into the blended learning environment to help educators create effective online learning communities. Moodle is a course management system (CMS) - a free, Open Source software package designed using sound pedagogical principles, to help educators create effective online learning communities. [6]

Teaching and learning in any subject can sometimes become mundane but in order that our students are motivated it is sometimes necessary to use an approach which makes the process more innovative. The method adopted to overcome this has been the use of **virtual learning groups** using internet based communication tools (moodle) to enable learners who would otherwise physically be unable to meet, to come together in cyberspace and discuss moral issues relating to computer systems.

Over the past two years a collaboration has taken place with our third year students at Edge Hill University with international students in the teaching of Professional, Legal and Ethical Aspects in Software Engineering (PLEASE). PLEASE focuses on the legal, ethical and social aspects of computing. The ethical strand of this module, which aims to develop moral reasoning in the learners, has often proved to be the most difficult for students to grasp and consequently has had a de-motivating effect on some learners.

Successful collaboration depends on both the technology and the ways in which the technology is used. The technology alone will not deliver the desired benefit. Ill-considered use of the technology may have results which are the opposite of what you set out to achieve.

This paper describes the first cycle of the study, the results obtained, lessons learned and how that has informed the approach used in the next study.

Keywords Innovative collaboration, virtual learning groups, blended learning, internet based communication.

1 PHILOSOPHY OF MOODLE [7]

The design and development of Moodle was guided by a particular philosophy of learning, a way of thinking that is referred to as "social constructionist pedagogy". [7]

The constructivism point of view maintains that people actively construct new knowledge as they interact with their environment. Everything that is read, seen, heard, felt, and touch is tested against prior knowledge and if it is a viable concept, may form new knowledge that one may carry with them. Knowledge is strengthened if this is used successfully in a wider environment.

Constructionism asserts that learning is particularly effective when constructing something for others to experience. This can be anything from a spoken sentence or an internet posting, to more complex artifacts like a painting, a house or a software package.

Social Constructivism extends the above ideas into a social group constructing things for one another by collaboratively with one another. Members of this sort of group are learning all the time about how to be a part of the collaboration. So when in discussion the motivations of individuals are examined in more depth.

2 SEPARATIST, CONNECTED AND CONSTRUCTED BEHAVIOUR

Initially the individual members of a group try to remain 'objective' and 'factual', and tend to defend their own ideas using logic to find holes in their opponent's ideas – a separatist approach. Connected behaviour is a more empathic approach that accepts subjectivity, trying to listen and ask questions in an effort to understand the other point of view. Constructed behaviour is when a person is sensitive to both of these approaches and is able to choose either of them as appropriate to the current situation.

In general, a healthy amount of connected behaviour within a learning community is a very powerful stimulant for learning, not only bringing people closer together but promoting deeper reflection and re-examination of their existing beliefs.

All of these issues help focus on the experiences that would be best for learning from the learner's point of view, rather than just providing them with information that they need to know. It can also help realise how each participant in a course can be a teacher as well as a learner.

Moodle itself does not make people behave in a certain way, but it is the use of moodle as a virtual learning environment (VLE) that supports the collaboration between them. Moodle was used with a group of third year students specifically on a module that was concerned with the legal, social and ethical implications of computing as a profession in today's workplace.

3 PROFESSIONAL, LEGAL AND ETHICAL ASPECT OF SOFTWARE ENGINEERING (PLEASE)

Collaboration has taken place with international students in the teaching of Professional, Legal and Ethical Aspects in Software Engineering (PLEASE). PLEASE focuses on the legal, ethical and social aspects of computing. The ethical strand of this module, which aims to develop moral reasoning in the learners, has often proved to be the most difficult for students to grasp and consequently has had a de-motivating effect on some learners.

Computing and Information Systems is an area of practical activity, which in different ways, employs and affects a large number of people in society. It is vital that students are aware of the most pressing professional, legal and ethical issues in the workplace of today.

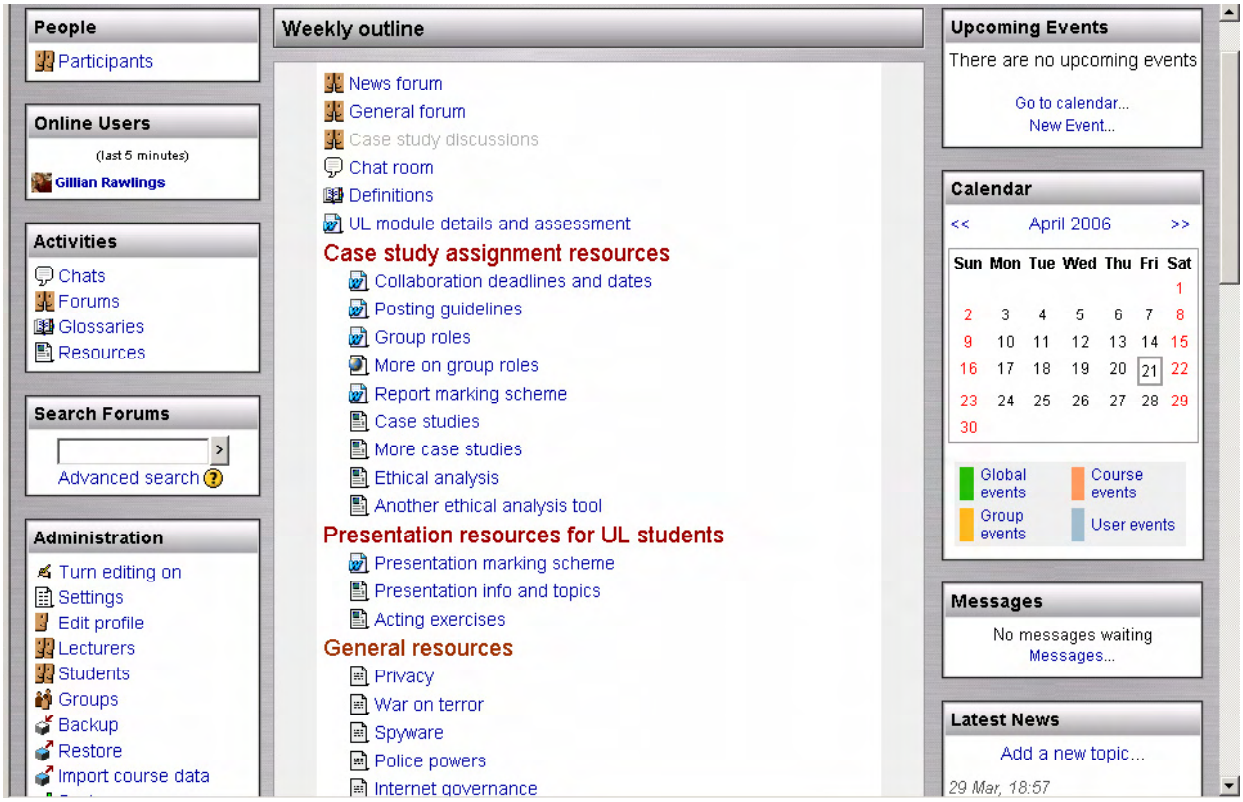
PLEASE develops an understanding of the professional and legal constraints within which computing specialists operate, using a 'discussive' environment as the vehicle where the students will be confronted with social and ethical issues of using technology in place of, or supporting, human abilities. The module develops a mature attitude to working as an ethical, environmentally aware computing professional, through critically reviewing germane issues.

For the students at Edge Hill University the collaboration needed to assess specific Learning Outcomes, which required the students to be able to:

- critically discuss future development and deployment of computing and information technologies and assess the possible ethical, legal and professional issues invoked;
- understand the application of relevant laws governing the IT and Computing Industries; justify their actions and decisions as computer specialists via rational appeal to ethics, law and professional codes of conduct.

It also required them to be able to verbally express personal ethical principles. Each group of students were required to produce a report, and give a presentation to their module tutor. The report needed to demonstrate a wide understanding of the issues under discussion, critical reasoning, and

Figure 1. Snapshot of moodle site used for collaboration



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

analysis and synthesis of material. Each group consisted of 3 students from Edge Hill and 3 from the collaborating university. For the first cohort all the groups were chosen at random, and as there was a strong recommendation from this cohort to allow any subsequent cohorts to form their own 'home' groups, this was implemented. The final pairings were allocated randomly.

There were a total of 78 students in the first study, which made 26 groups of 3 students from each institution. The ratio of male/ female students was 2:1. There were no identified specific learning difficulties within any group. There was however a difference in the average age of the students from the partner institution, in which the students' average age was 24 years, to those at Edge Hill whose average age was 33 years old.

The site was formatted to allow easy access to all relevant information, and included access to course material. There were areas which could only be accessed by all the tutors involved.

Contact between the institutions involved was initially via a conference on Professional Ethics. The first cohort ran from January to March 2005. This paper looks at the lessons learnt from this initial experience and how that informed the second cohort that ran from February to March 2006, and the third cohort which ran from February to March 2007

There were both formative and summative assessments built into the collaboration. The formative assessment required the students to choose a case study from those listed on the moodle site, and by means of postings, analyse their chosen scenario. There were a series of lectures at both institutions to give a clear understanding of the implications of ethical theories.

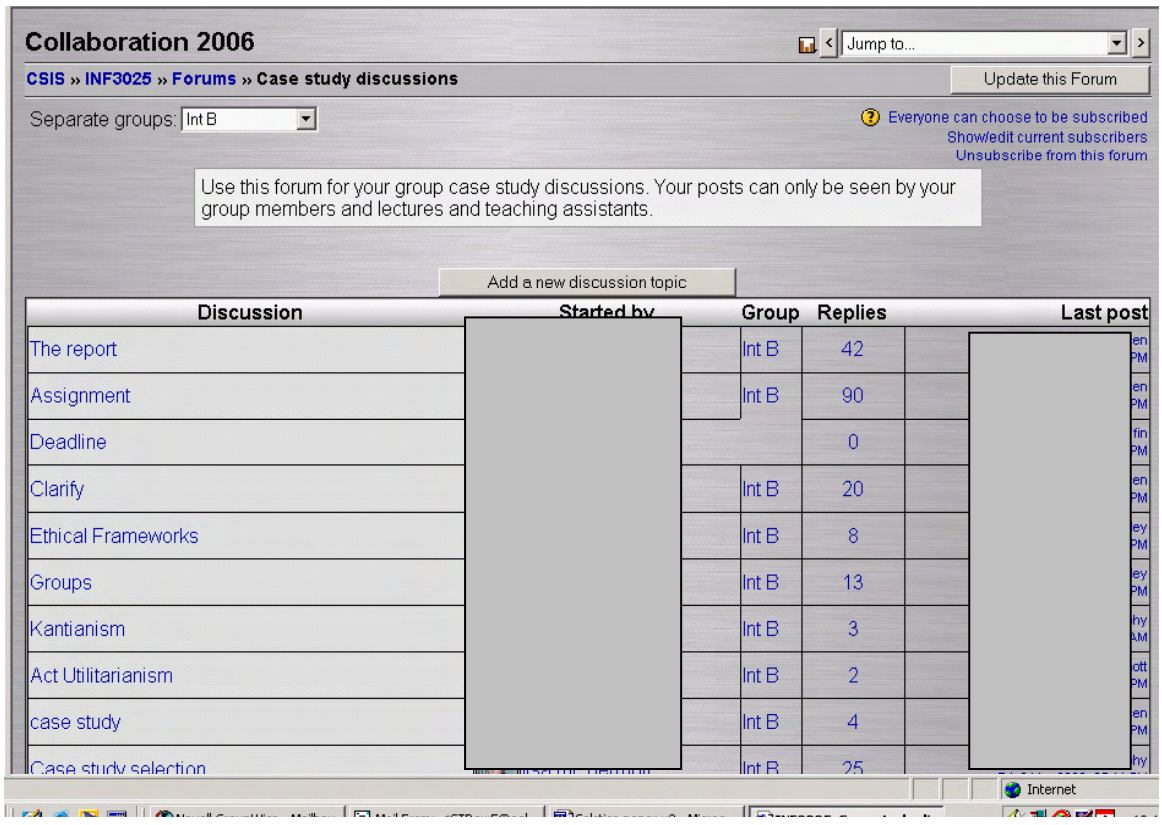
Students were given very clear guidelines on posting to the site. The focus was to be on the topics posted, but they were advised to bring in related thoughts and material, other readings, or questions that occurred from the ongoing discussion. Two substantive messages or three or four smaller posts (*about 450 words per person per week for three weeks*) were required for assessment. Posts were assessed according to the criteria below, and needed to reflect an understanding of the ethical theories as they were applied to the case studies. Each posting was graded according to the analysis of cognitive presence described in table 1.

Table 1. Posting descriptors

Descriptor	Indicator	Category
Triggering events	Suggestive	Recognising a problem
Exploration	Uncertain	Information exchange Suggestion for consideration Brainstorming Leaps to conclusions
Integration	Interim	Group begins to plan Connecting ideas – synthesis Creating solutions
Resolution	Committed	Firming solutions Defending solutions

The tutors received a copy (as an e-mail) of every posting to the site. Posts were organised into discussion topics, and were added as threads to that topic.

Figure 2. Postings for group B



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 2. Collaboration timetable for 2006

	Tasks for students	To be completed by:
1	Register with moodle Add/ update personal profile	24th February
2	Explore moodle · start a discussion · reply to a discussion	3rd March
3	Form into groups of 3 · Note: this is on a voluntary basis, but students will be allocated to a group if there is no input	3rd March
4	Groups to socialise on-line · build up a group identity · build up group trust	3rd March
5	Select case study Inform tutor of selection	3rd March
6	Team building exercise	10th March
7	Start assignment	13th March
7	Hand-in completed assignment	Wednesday 29th March

Because this type of communication reduces the range of cues we have for building a mental picture of the people with whom we are interacting and for making judgements about what their words actually are meant to say, all participants are required to add a personal profile and a photograph, and some time is initially given for everyone to socialise on-line.

3.1 Feedback from first cohort

Comments were very positive. Many students commented on the improvement of their own communication and team working skills. All feedback on improving the module was constructive, and focused on the use of the VLE (Moodle). The fact that Easter came at a crucial time in the collaboration meant that the students had to keep in contact (and motivated) when some had planned holidays. This did indeed develop their communication skills!

Students commented that they felt the tutor support was good. This is an important element when collaborating using a VLE. It is very much the case that each and every day all tutors need to be monitoring the postings (and there were in the region of 2000 postings for the first cohort and this number increased each year) and giving constructive feedback when required. The overwhelming feeling was that the module was very enjoyable as it was 'different' from any other they had studied.

3.2 What changed and why

The next time this module ran (semester 2, 2006) the scheduling of the collaboration was very straightforward, with no holiday period to consider. One of the more difficult elements when collaborating internationally is the difference in holidays and religious festivals. The students started 'talking' to each other earlier, and on a more social basis. Although this was available the first year not many students engaged early enough – in fact some did not log on to moodle until the collaborative element of the coursework kicked in.

There was an increase in the number of students in the second cohort, from 68 to 87, which made 29 groups of 3 students from each institution. The ratio of male/ female students was still 2:1. Again there were no identified specific learning difficulties within any group. There was still however a difference in the average age of the students from the partner institution, in which the students average age was 25 years, to those at Edge Hill whose average age was 32 years old.

Because students from the first cohort used MSN at various stages to chat, instant messaging and a chat room were provided on moodle for all future collaborations. These 'conversations' could therefore be included as postings (and therefore assessed) if required.

A timetable was provided, which gave clear instructions on how the collaboration was to proceed. The collaboration was further defined by pairing the students within the groups, with one student from each institution working on a particular ethical theory within the group.

The team building exercise in task 6 was used for formative assessment of postings on moodle which helped the students adjust the level/contents of postings before the assessment started.

Some of the comments from the student when asked what they had gained from taking this module were:

- Using ICT in this way increases my control of when and where I work
- The importance of teamwork and a good insight into ethical priorities
- Better communication and teamwork skills
- To work through difficult situations with people I had never met before
- The confidence to work with people through collaboration

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

- An understanding of the difficulties of 'remote working' and what works and what does not work
- Understanding of other peoples views from other culture/backgrounds.
- Confidence to put my views across
- That professional, legal and social aspects of computing are an important factor in IT and IS, and are largely underestimated by the majority of computer users
- Sound understanding of legal aspects in context

4 WHY COLLABORATE USING A VLE?

The type of collaboration described above can be classed as networked learning.

"Network learning is learning in which information and communications technology (ICT) is used to promote connections: between one learner and other learners, between learners and tutors; between a learning community and its learning resources. Some of the richest examples of networked learning involve interaction with on-line materials and with other people. But we do not see the use of on-line materials (such as World Wide Web resources) as a sufficient characteristic to define networked learning. There is a danger that such a definition would soon embrace all forms of learning that use ICT". [4]

The obvious strengths are that it supports relatively high degrees of interaction between the learner and other learners, between the learner and tutor, and with on-line learning resources. In conventional forms of higher education, interaction with peers and tutors usually requires co-presence. Networked learning supports interactivity *and* flexibility over the time and place of learning. Interaction needs to be through well-designed tasks. The learner has time to consider what others have been posting, to reflect, to use other sources, and to prepare their own contributions. All postings are stored as a permanent record. This contrasts with 'real-time' interactions - such as in face-to-face seminars - where there is much less opportunity to consider and prepare one's argument, and much of what was said - good and bad – would not be available to the learner for any subsequent reflection.

One important advantage shown in our collaboration is that those students who have previously shown evidence of under-participation in face-to-face events do not show the same behaviour when using the on-line environment of a VLE. People who are not quick or confident in face-to-face debate can sometimes find themselves 'liberated' by the less intensive demands of communicating in this way.

There are however some limitations. Text-based communications can be seen as a drawback. The use of 'emoticons' (such as a J to represent good or I am happy with that) are commonplace. Text-based messages may not have the expressive richness of a quick and lively verbal exchange. On the other hand, well-crafted text can be richer than off-the-cuff discussions.

One important feature of Moodle is that postings are 'held' for 20 minutes before they appear on the site. This gives the author time to change any aspect of the posting, or to delete it all together! The decision was taken to allow a considered response to all postings, and to allow others to assimilate their replies. When discussion or group work extends over days rather than minutes, it can be hard and slow to build a consensus around a decision that needs to be taken, but working to a strict deadline does somewhat concentrate the mind.

It is common to find a mix of kinds of language and contribution in an on-line discussion. Some contributions may be long, deep, analytic and thoughtful. Others may be much more spontaneous and flippant. Clear guidelines on postings are an absolute necessity, as described earlier.

5 SUMMARY

Most of the claimed strengths of networked learning using the internet as a platform have their roots in both the technology and the ways in which the technology is used. The technology alone will not deliver the desired benefit. Ill-considered use of the technology may have results which are the opposite of what you set out to achieve.

The balance of evidence from our cohorts of students who have been involved in this type of collaboration shows that the majority enjoy and approve of the experience. A minority take a more negative view. An explanation for this seems to be that they are not prepared to work outside the normal structure of lectures and seminars, which means they do not participate fully in the experience.

Motivation is a key concept. Katzeff [5] stresses motivation is a critical factor for instructional design and for learning to occur the learner must be motivated to learn and motivation is created by four individual factors:

- challenge,
- fantasy,
- curiosity and
- control

and three interpersonal factors:

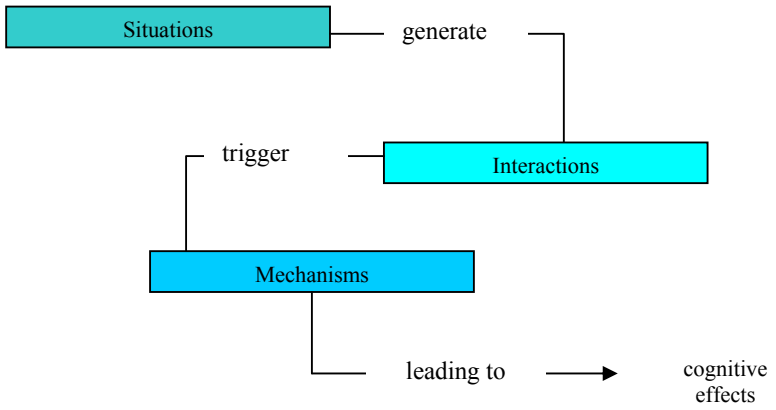
- co-operation,
- competition and
- recognition.

This type of innovative collaboration meets all these aspects.

Collaboration can offer benefits for learning in at least two ways. Firstly collaboration can lead to learning. For group members to collaborate necessitates them in articulating and explaining their ideas to each other. Articulation 'externalises' ideas for scrutiny by the group member him/herself, as well as by the other members of the group. Explaining one's ideas and sharing perspectives and viewpoints encourages each group member to examine their own ideas in the light of others' views.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure3. Situations, interactions, mechanisms and effects



Secondly the experiences of collaboration may help develop important personal transferable skills, including learning how to collaborate. This can also include communication, co-ordination, and self-management skills, which are seen as important for the workplace, where professional work is increasingly project-based, team-based and distributed.

Dillenbourg [2] offers an excellent account of collaboration in learning processes from a cognitive psychology perspective. He is especially interested in problem-based tasks and looks at both paired and group-based collaborations. Figure 3 shows how he links learning situations, interactions, cognitive mechanisms and cognitive effects when groups engage in solving problems. This model is useful in that it emphasises the *indirect* connection between a collaborative learning situation and its learning outcomes.

Some of the cognitive mechanisms, he suggests are:

Conflict or disagreement referring to when diverging viewpoints lead to verbal interactions in order to resolve a conflict. A slightly 'softer' take is where group members pose alternative propositions.

(Self-)explanation as a process of giving an explanation where there can be learning gains for both the person explaining and for the person hearing the explanation.

Internalisation which happens when a conversation leads to a progressive integrating of ideas under discussion into one's own reasoning.

Appropriation is an interesting notion in that Dillenbourg suggests that interpretations or playing-back of our ideas, by others to ourselves, can actually help us to gain a richer understanding.

Shared cognitive load is a principle of economy, in that group-based work can allow members to spontaneously share out a task, to avoid redundancies and to optimise effort matched to the skills or knowledge within the group

6 CONCLUSION

The kinds of collaborative learning situations we can design for on-line activity need to be examined to see how likely they are to generate interactions among the group from which cognitive mechanisms may be triggered or enabled.

Learners should not be expected to generate their own effective ways of collaborating. They need clear guidance about how to participate in a group learning situation. Successful collaboration depends on both the technology and the ways in which the technology is used. The technology alone will not deliver the desired benefit. Ill-considered use of the technology may have results which are the opposite of what you set out to achieve.

This study is part of ongoing research. Using virtual learning groups does seem, so far, to indicate that it has had a positive effect on the experience of the users. Learners seem to be more motivated by the experience of collaborating with other institutions as there seems to be an increased element of competition involved, over and above that normally associated with working in groups. Research currently underway concentrates more on the analysis of gender and its relationship to performance when using a VLE.

REFERENCES

1. Anderson T.. (2004), 'Theory and practice of online learning', Athabasca University, Canada.
2. Dillenbourg P. (1999), 'Collaborative learning: cognitive and computational approaches (advances in learning and instruction)', Pergamon Press, Amsterdam & New York.
3. Garrison D. R., Anderson T., and Archer W. (2001), 'Critical thinking and computer conferencing: A model and tool to assess cognitive presence', American Journal of Distance Education 15 (1).
4. Goodyear P. (2001), 'Networked learning in higher education project', JCALT, UK.
5. Katzeff C. (2000), 'The design of interactive media for learners in an organizational setting – the state of the art', Proc. NordiCHI 2000, Stockholm, Sweden.
6. Dougiamas M. (1999), 'Wecome to moodle!', Moodle – a free, open source course management system for on-line learning, Martin Dougiamas, www.moodle.org (2007)
7. Dougiamas M. (1999), 'Philosophy', Moodle – a free, open source course management system for on-line learning, Martin Dougiamas, www.moodle.org (2007)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Integrating VoIP into distance learning

Dannan Lin, Charles Shoniregun

University of East London
Docklands Campus
4-6 University Way
London E16 2RD

Abstract The knowledge based economy is the driving force behind the fast growing e-distance learning. The new economy and globalisation require people to learn the knowledge not only in timely fashion but also in the way of cost effective. Distribute knowledge require mass interconnectivity to achieve ideal result. Nowadays the most common ways of communicate between tutors and learners are telephone and email. Telephone has been serving us well in the past, however the cost of running such service is too high when consider its distance relative- VoIP. The VoIP (Voice over Internet Protocol) has been progressing since 1990s. Due to high usability and cost advantages, VoIP is certainly becoming the future telecommunication backbone. The real technical problems of using VoIP in distance learning are security and quality of service. In this paper we introduce the general security issues that are likely to be involved in the VoIP implementation. We also proposed a SIP (Session Initial Protocol) based VoIP model that could utilise the bandwidth on both peers and thus achieve ideal Quality. And at the end, we conclude that VoIP can make positive contribution to the distance learning.

1 INTRODUCTION

The distance learning is not the same as e-learning. E-learning is a methods, it is used to describe the way of distribute education where all the interactivities are taken place in virtual environment such as network based interactive classroom. Distance learning is similar but it is more rely on the offline support or we could say it is 'virtualless '. (Tsai et al.,2002) Nonetheless as the network technology progresses, distance learning are now merging with e-learning. E-distance learning is becoming increasingly popular. E-distance learning allows learners to stay off-site and study at their own paces; therefore it requires asynchronous communication between learners and tutors. The existing communication methods are based on the emails and telephones. Learners once sign up for the distance learning they would be given study materials, a set of phone numbers and email addresses of tutors. These are the free numbers. The cost of free phone services is passed to the learners and included in their tuition fees. However the free 0800 numbers is not cheap, if we could use VoIP to replace the free phone numbers, there would be significant cost saving.

Our research aims to identify the feasibilities of VoIP integration in distance learning. In this paper we also identified the true technical obstructs of using VoIP in distance learning. We proposed a VoIP communication model to illustrate the VoIP implementation in distance learning. At the last we concluded that VoIP is highly usable in distance learning.

2 METHODOLOGIES

Distance learning is so different from conventional on-site education that requires intensive study to identify what factors are needed to achieve satisfied result. Based on the findings from other researchers we were able to identify the key obstacle in distance learning was the inefficient communications. (Grill 2005 & Mclean 1999)

Our research must focus on the improvements of efficiencies of e-distance learning communications. By taking these factors into consideration we could build up a model that has positive impact on the e-distance learning.

VoIP technology has several advantages over traditional telephones. Cost saving is the main reason people use VoIP. Our research must ensure that the cost of running VoIP does not exceed traditional phone services. Therefore the model we were going to propose must be not only practical but also cost effective.

3 OVERVIEW OF VOIP PROTOCOLS

Many researchers and industries reports have concluded that e-learning can be at least as effective as conventional class training. (Zhang et al., 2004; Kruse, 2004 and Nicholson et al., 2007)

VoIP based applications are successful in both civic and business use. (Varshney et al, 2002) The success of VoIP as viable alternative for service requires that educational organisations

to adopt relevant to technology and management. The development of VoIP has been proliferated significantly. The traditional distance learning is telephone based, with limited geographic dispersion between tutors and learners. There are many successful stories about VoIP, A prominent focus of VoIP customer service centre has been the achievement of integration of customer care and management services. (VoIP News, 2005) we believe that E-distance learning service could be better with VoIP.

Although adding VoIP service to E-distance should not be difficult, there is a problem of choosing right protocol. Our research shows that there are two popular protocols in voice services. They are H.323 and SIP (Session Initial Protocol). They have their own advantages over one another.

Before we could suggest any VoIP protocols to be used in E-distance learning, we have to make sure that the protocol we choose is simple, easy to manage and yet has high upgradeability in the future.

The H.323 and SIP were both designed to provide multimedia services over IP network. They use TCP (Transmission Control Protocol) and UDP (User datagram protocol) as call signalling. Voice or video streaming is based on the RTP (Real Time Protocol). They both support wide ranges of codec such as G.711 or G.729.

Both H.323 and SIP require a server for call set up. H.323 server is called gatekeeper and SIP uses proxy server. One thing worth to mention here is that SIP's proxy server can bridge two user agents directly. Having said that, SIP can be used in the way similar to P2P service as user agents do not require intermediary servers while H.323 relies on gatekeepers in the middle to keep the media stream alive.

Despite their similarities on protocol concept, the H.323 was based on telephony protocols while SIP is text based protocols similar to HTTP. SIP is much simpler than H.323 because H.323 is written in binary code and such format is not user friendly to system developers. Although SIP can be used in VoIP, it was not entirely designed for transmitting voice only; in fact the transmission session is relied on another session description protocol. That being said SIP is more flexible it could do text based instance message service or live video streaming. (Chen et al., 2006; Leonard et al., 2003; Ho et al., 2001)

E-distance learning will have more features in the future. If we choose H.323 based VoIP then we would loss the flexibility. Part from that over complicated telephone protocols is difficult for most of the programmers. Therefore we would use SIP in our proposed VoIP enabled E-distance learning model.

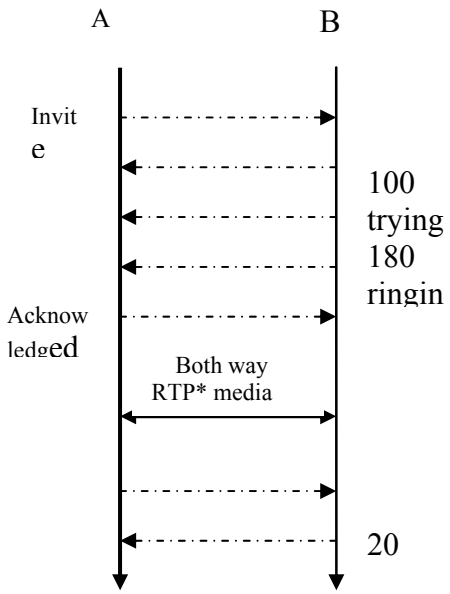
SIP technology is user-centric. We often relay on different applications or physical devices to keep communicating. However it is hard to manage email, instance messenger, video conference and mobile phones all together because we do not know which one is most efficient to achieve best result. Different applications or devices also increase the burdens to the users; the results are missing calls or messages, thus decrease in productivity. SIP only require one application or one device.

In E-distance learning there are scenarios that would only be benefited from SIP. For example, Learners could contact their learners through web portal simultaneously initiate emails, instance messages and VoIP calls to appropriate instructors to maximise responsiveness. Learners and tutors can also upgrade to video conference when audio along is not enough. (Only if there is enough network resources) When talking about learning, tutor's comment is always important, learners could invite instructor into a file sharing session that instructors could make comments about learner's work.

We identified the External factors that would have impacts on VoIP enabled the E-distance learning communication. Prior experience in VoIP use and online relationships is an important factor. A lack of VoIP and internet use obviates the issues involved in developed ease of use system.

The figure 2 shows typical SIP call flow

Figure 1. standard SIP call flow

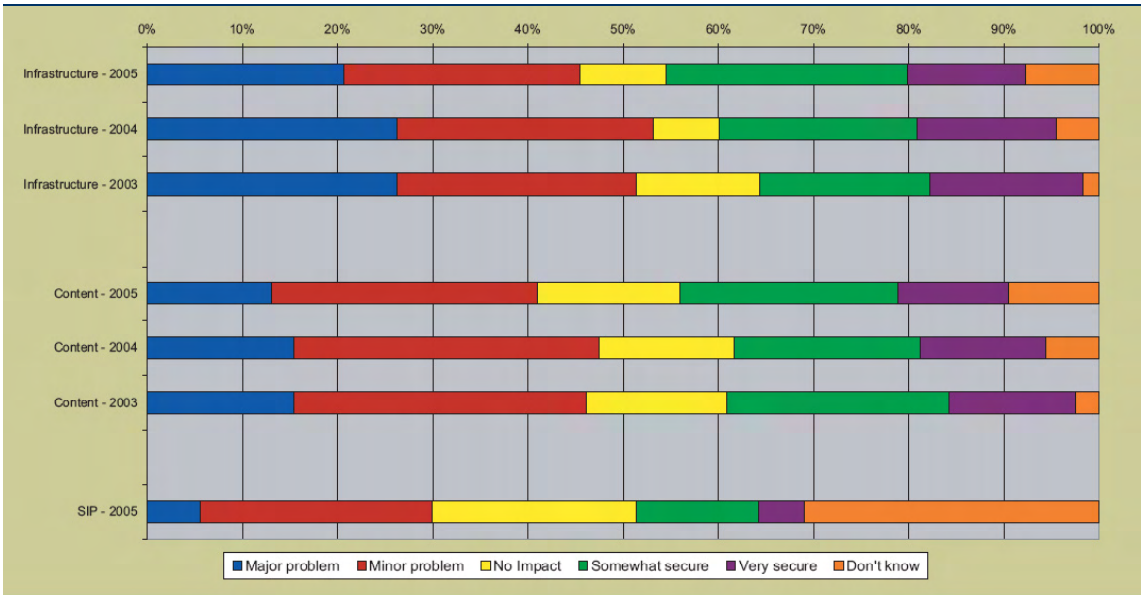


4 SIP SECURITY

SIP service involves both PBX and endpoints. Reliable encryption and authentication method is always recommended and they are common technical approaches to alleviating security problems. Secure communication between two endpoints using authentication and encryption have been using for a long time, but it is still suffering from high security risks. The typical attacks on SIP network such as automated enumeration of SIP extension and user name is not a difficult task. SIP service is required to be exposed to certain extend and this is its nature. Enable user authentication and usage by using INVITE and REGISTER is absolutely necessary. Nonetheless, such method is not always help, the best security practice we recommend here is to use VPN (Virtual Private Network) to separate from normal data network.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Extent to which security is a concern for the infrastructure, for conversation content, and by use of VoIP. (Taylor 2006)



Since the SIP is endpoint based such security feature must be built into the endpoint. For example the encryption of Transport Layer Security (TLS) has to be supported by two endpoints otherwise it is not going to work. However the according to industrial survey SIP is considered to have less security concerns.

Security is always important, there is yet to have standardized solutions. Most of the security solutions are defined by the vendors. Therefore protect the application as well as the network by using encryptions; firewalls and session control are important elements in SIP security.

5 SIP QOS (QUALITY OF SERVICE)

When consider VoIP, QoS and securities are always the topics attract most of the attention.

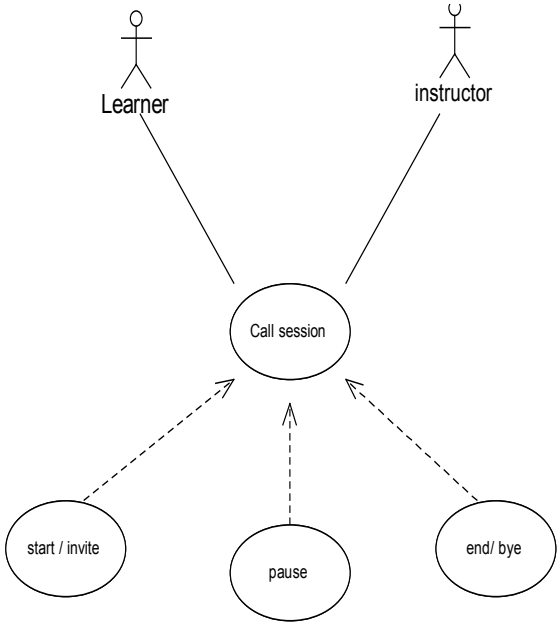
Quality of services often associates with the amount of bandwidth available to it. Although majority of the people are having broadband connections, there are still large numbers of people rely on 56k dial up connections; besides even the broadband cannot guarantee the QoS. (Quality of Service) (Ahuja et al., 2001 & Peden et al., 2001) VoIP is two ways communication that is to say assumes that two end points are talking at the same time and each side consume 10k bandwidth then total bandwidth consumption would be 20k. VoIP does not consume significant amount of bandwidth, 56k dial up internet connection will be sufficient but due to the nature of the network our voice quality will be affected by the network traffic. (Lin, 2007) To study the bandwidth control of VoIP in E-distance learning is beyond the scope of this research. We would have to find way to work around this.

There are more than just technical issues involved. For example, if the distanced learner is temporarily located in China. And he wants to establish a call session with tutors. He would occasionally experience poor QoS. The reason behind this is not purely technical, it is because that China service providers are currently battling with foreign VoIP service providers such as Skype therefore they made their own VoIP more reliable on their side of network and degrade the quality of Skype if any Skype calls initiated from within their network. On contrary if tutor is calling from networks out side of china then the voice quality would not be affected. However this is not practical due to the nature of distance learning.

Traditional telephone system has constant QoS but QoS for VoIP needs to be managed. We believe web based portal for E-distance learning is the best solution for SIP enabled VoIP service. The QoS are largely affected by the bandwidth and network structure. Since alter network structure involves large capital investment and that is against our cost saving objective, we would focus on the bandwidth saving. And the bandwidth availability is determined by situation of network congestion and overload.

The normal way to utilise the network is to use CAC (Call Administration Control) to limit loading as well as to design

Figure 3. Simplified SIP user case diagram



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

the network capacity to meet the forecasted traffic and QoS requirement.

Our research has shown that by adding extra function to the voice application we would be able to achieve bandwidth saving and thus reach ideal QoS. The method we used is shown below.

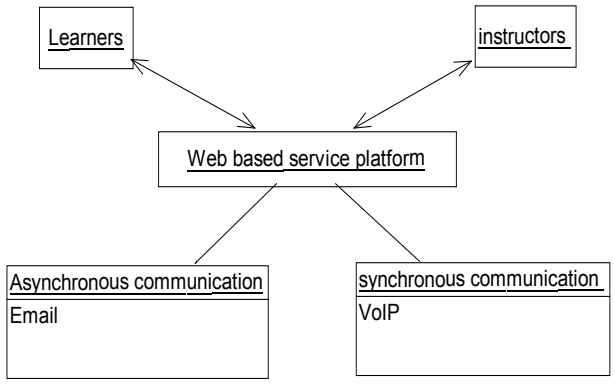
Normal SIP call flow would involved messages such as invite, trying, ringing, acknowledge and bye etc. We could manually create a pause function on the each side of speech, that is to say, learners should keep silence when instructor is talking. This kind of "slim" VoIP could save the bandwidth up to half. Put it simple, it works in the same way as walkie-talkie that user push pause after he/.she finishes his/her side of talking. Similar but more sophisticated technique is called silence suppression and voice activity detection. (Boger et al., 2007) However such techniques only work when there is absolute silence on the phone call. Consider the situation that a call from noisy place, the absolute silence is unattainable.

Bandwidth consumption by VoIP is not significant therefore this method is only temporary, the future study would focus on more sophisticated bandwidth control.

6 PROPOSED MODEL OF VOIP E-DISTANCE

With our SIP pause function in place we could build the VoIP enabled E-distance learning model. (Figure 4)

Figure 4. Web based E-distance learning



As we mentioned previously SIP is a text based protocol; it can be easily handled by the network programmers. The SIP messages are easily read therefore makes network QoS monitoring much easier. Accessing SIP enabled voice system is similar to the way we use instance messenger. According to our past experience build up a MSN messenger active platform based chatting application is not a difficult task and it can be handling easily by most of the senior programmers.

7 MEASURE THE VOIP ENABLED E-DISTANCE PERFORMANCE

VoIP for E-distance learning assessment needs to start from end user's perspective. End users will not accept any QoS that lower than traditional telephone system. The first part of measurement should be on the technical part. That include the comparison between PSTN calls VoIP calls. This measures factors such as network echo, background noise and network delay etc. By using our slim VoIP technique introduced previously we could save the bandwidth potentially.

Despite the technique elements in QoS, network management should focus on the user's aspect. For example, users would expect to handle the SIP calls same as they do with traditional telephones. Users would also expect to locate the IP address of the tutors easily and further more the entire system must have a backup plan to against unexpected network down time.

Security management is also required. VoIP is different from PSTN (Public Switched Telephone Network), PSTN obtain its security by physically seal off its networks. VoIP is built over the IP network therefore it is exposed to the all kinds of threats. Although firewalls are recommended to the PCs of both tutors and learners, the website portal security is also need to be considered as first priority. SSL (Secure Socket Layer) and other common encryption methods need to be used to protect user sign in/off and transmission contents.

The SIP enables VoIP system is flexible for the future upgrade. It requires network management to adapt to the changes as well. The future E-distance learning communication method would be unified messaging service that integrated voicemail, fax and email. When design a web portal it is important to keep future upgradeability in mind.

8 CONCLUSION

The E-distance learning's shift to internet telecommunication is inevitable. SIP enabled VoIP service has the advantages of simplicity and upgradeability. By introduce such pioneer service to E-distance learning will certainly leverage the data network. There are several advantages of using VoIP instead of traditional telephone system. Although there are issues concern about the user's trust and network management, consider the potential of cost saving and increase in productivity all these problems can be forgiven.

The model we proposed here highlights the importance of benefiting users in the context of E-distance learning. Design and managing such service web portal can be highly complex and providing data network support for voice must be carefully planned and managed to ensure the traditional service level are sustained. QoS and Security are critical and they require constant monitoring. The best practise is to establish solid foundation for VoIP today and ready for the future expansion.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

9 FUTURE STUDIES

VoIP enabled E-distance learning is still in early stage of its development. Unlike other business VoIP applications that are exist in the market for a long time. The future study should focus on more sophisticated bandwidth control as well as security. The use of SIP will certainly encounter numerous problems such as creating online directories, learners need to be able to located their tutors online through directories and such directories are not just lists of tutors' IP addresses because SIP service needs to be portable, therefore, the portability of SIP will certainly be the next problem to solve.

REFERENCES

Ahuja, S.R., Ensor,R (2004) VoIP: What is it Good for? , ACM press, volume 2,issue 6,P48-55.

Boger,y (2007) Fine-tuning Voice over Packet services, RADCOM Ltd.,

accessed on: 07/05/2007,

available at: http://www.protocols.com/pbook/pdf/voip.pdf

Chen,J J., Lee ling., Tseng Y C., (2006) Integrating SIP and IEEE 802.11e to support handoff and multi-grade QoS for VoIP applications, proceeding of the 2nd ACM international workshop on Quality of Service & Security for wireless and mobile networks , ACM press, P.67-74.

Grill, Grandon T., (2005) Distance-learning strategies that make sense, part1: a micro analysis, Volume 2005, issue 3, P1.

Ho, J M., Hu, J C., Steenkiste, P.,(2001) A conference gateway supporting interoperability between SIP and H.323, Proceedings of the ninth ACM international conference on Multimedia. ACM press, P.421-430.

Kruse,K.,(2004) Calculating E-learning Cost. E-learning GURU, retrieved on 15/05/2007, available at : http://www.e-learningguru.com/articles/art5_4.htm .

Leonard,J.,Riley, E., Staman, E M.,(2003) Classroom and support innovation using IPvideo and data collaboration techniques, proceeding of the 4th conference on information technology curriculum, ACM press, P.142-150.

Lin dannan(2007) VoIP development in SMES, proceed to ICITST London conference.

McLean, Robert S.,(1999) Meta-communication widgets for knowledge building in distance education,Computer Support for Collaborative Learning, International Society of the Learning Sciences.

Nicholson,D., Hamilton,D., McFarland, D., (2007) Learning: The Interactive learning model and learning combination inventory. Consortium for Computing Sciences in Colleges, volume 22, issue 6, P8-17.

Peden, M., and Young, G.(2001) From voice-band modems to DSL technologies, John Wiley & SON,inc. Intrnational Journal of Network Management, volume 11,issue 5.

Taylor,S (02/2006) 2005/2006 VoIP State of the Market Report, Network Associates, Inc,

Accessed on:04/05/07,

available at: http://www.webtorials.com/main/resource/papers/sotm/paper2/2005-2006-VoIP.pdf

Tsai, S., Machado, P., (2002) E-learning, Online Learning, Web-based Learning, or Distance Learning: Unveiling the Ambiguity in Current Terminology, ACM press, volume 2002, issue 7. p.3.

Varshney,u.,Snow,a.,Mcgivern M.,Howard,C.,(2002) Voice over IP , ACM, volume 45,issue 1, P89-96.

VoIP News(25/07/2005) Purdue University Researchers Conduct Study on Web-based Customer Service that seamlessly connects customers to the call center over the internet While on a company's web site, VoIP News, accessed on: 17/05/07,

available at: http://www.voip-news.com/art/8y.html

Zhang, D., Zhao, J.l., Zhou, L.,Nunamaker, J.F., (2004) Can E-learning Replace Classroom Learning. Communication of the ACM, volume 47, issue 5 , P.75-79.

i·Society 2007

E-Learning Status in Arab Countries

Naseem Matar
Ziad Hunaiti

Anglia Ruskin University, Chelmsford CM1 1SQ, UK

Zayed Huneiti
Mohammed Al-Naafa

University of Ha'il, Ha'il, Saudi Arabia

Abstract This paper discusses the status of e-learning in Arab countries located in the Middle East. The study was based on navigating the official web sites of universities in the region. The list of the universities included was obtained from the Ministry of Higher Education in each country surveyed. The purpose of the study was to provide an analytical overview of where the use of e-learning and the management of learning in Arab countries currently stand. The primary aim of the study was to give an evaluation of the status of e-learning by gathering facts about its use, its adoption and its penetration in Arab universities. To achieve this, the services provided by universities through their web sites were analyzed.

Keywords E-learning, Middle East

1 INTRODUCTION

Electronic Learning or what is referred to as "e-Learning" is defined as the delivery and acquisition of education or training in electronic format using electronic media. E-Learning now plays an important role in shaping the national educational curricula of many countries. Most educational institutions have implemented this technology not only to achieve delivery-mode diversity but also to enhance the learning processes of their students, while reaping significant financial benefits for themselves. Many Arab universities throughout the Middle East are taking gigantic steps in their use of e-learning to enhance higher education. The countries surveyed in this investigation were found to be heading in the same direction as far as implementing this technology was concerned. It was also found, however, that the process was rather slow and still at initial stages in some cases due a number of factors such as, political peculiarities; rigidity of government agendas; levels of economic development and technological challenges. The political situation in the region has been facing many challenges. These challenges have had a significant bearing on the different sectors of human activity with education being a case in point. So influential has their impact been that it has affected other areas especially those that have a close relationship with education such the country's economy, technology and financial status. (LAS, 2003).

Governments play a major role in the administration of educational affairs in their countries, with many universities and educational establishments being directly supervised and supported by their individual ministries of higher education. As a result, national strategies and action plans designed to lay the groundwork for educational standards have been established. To say that the current status of education and economic development in many countries is the direct result of government-agenda influence is not an overstatement. In some cases government strategies have helped fast track the progress of some universities, while in others they have served to hamper progress.

Table 1. Number of universities in each country

Country	Population	Total Universities
Egypt	71,236,631	29
Kuwait	2,630,775	3
Jordan	5,282,558	26
Bahrain	723,039	2
Lebanon	4,509,678	22
Oman	2,424,422	3
Palestine	3,259,363	16
Qatar	795,585	3
Saudi Arabia	23,595,634	24
Syria	19,046,520	6
UAE	3,870,936	25
Yemen	20,764,630	12

The economic and financial situation of the region has also gone through many vicissitudes thanks to changes in the po-

litical arena. Most countries experienced economic growth in mid fifties and sixties due to the discovery of oil which led to accelerated economic growth and development. World Economic Forum on the Middle East (2006).

Educational funding suffered a slump when government priorities changed and more emphasis was placed on building infrastructure in such areas as industry, agriculture and healthcare. Soon governments in the region realised that in order to accelerate and sustain development, more attention to education must be paid. Consequently, they gave education its rightful place among their priorities. Many universities were built in order to train the local population and produce much needed manpower in the fields of science and technology. Today there are more than 171 universities serving a population of nearly 158,139,771. The first university in the region, the Al-Azhar University in Egypt established in 988, is considered to be the world's second oldest university after the University of Al Karaouine in Morocco, which was established in 859. Table 1 shows the number of universities in each country included in the study, as well as the country's population estimate. The figures are based on a March 2007.

To keep pace with the rapid demand for education in their countries, Arab governments in the region began to urge individuals and private businesses to set up their own colleges and universities and even passed legislation to make it easier for them to do so. Such universities were placed under the ministry of education of the countries where they were built. This was done in order to make sure that high standards were maintained and that quality was not compromised. Thanks to government regulations and the sharing of expertise and experience among key players in the educational field, the quality of education in the region as a whole has been improving (LAS ,2003). The chart depicted in figure 1 shows the proportions of public and private universities in the region.

2 FROM TRADITIONAL APPROACHES TO E-LEARNING

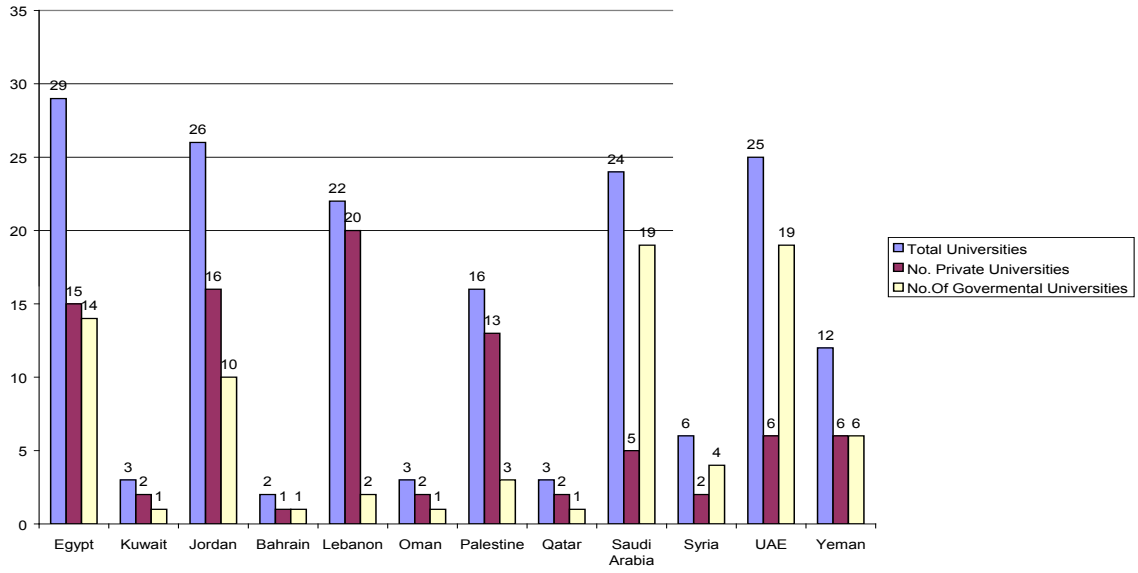
E-learning as a method of delivering instruction started during the Second World War. At that time, motion pictures were introduced as instructional aids. They gave some assurance that trainees using this mode of instruction would receive the same training regardless of their geographical locations in the world. With time and with technological advances, other types of media and equipment began to come on the scene. Computers, CD's and PowerPoint presentations are good examples of current technology. The latest technology to lend itself to e-learning has been the internet which is now the largest single resource repository for educational establishments (Bascsich ,1997)

At first many Arab universities decision makers in considered e-learning an additional burden on their institutions' financial resources. They were convinced that applying this technology would require them to incur costs on such things as changing network infrastructures at universities, adapting university labs to suit this new technology, buying relevant software and conducting staff training. (Bates, 2000)

Following a substantial amount of research carried out in countries within the region, important facts to the contrary began to appear and the educational community began to view e-learning in a different way. Facts first presented at a conference held in Beirut in 1998 under the title "The Beirut Declaration of the Arab Regional Conference on Higher Education" (UNESCO, 1998, p.44) served as an eye opener.

It was clear from the stated facts that demographic changes in the region would bring about new challenges. With a projected population growth of around 2.5% for the period 2000 – 2010, a much higher growth rate than that of the entire world estimated at 1.2% and that of developed countries estimated at 1.5%, there was no doubt that Arab countries needed to do more to cater for the unprecedented surge in demand for educational resources that would ensue. Larger

Figure 1. The proportions of public and private universities in the region



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 2. Middle East Internet Usage and Population Statistics (Internet world statistics)

Middle East Internet Usage and Population Statistics						
MIDDLE EAST	Population (2007 Est.)	Usage, in Dec-00	Internet Usage, Latest Data	% Population (Penetration)	(%) of M.E.	Use Growth (2000-2007)
Bahrain	738,874	40,000	152,700	20.70%	0.80%	281.80%
Jordan	5,375,307	127,300	629,500	11.70%	3.20%	394.50%
Kuwait	2,730,603	150,000	700,000	25.60%	3.60%	366.70%
Lebanon	4,556,561	300,000	700,000	15.40%	3.60%	133.30%
Oman	2,452,234	90,000	245,000	10.00%	1.30%	172.20%
Palestine (West Bk.)	3,070,228	35,000	243,000	7.90%	1.30%	594.30%
Egypt	71,236,631	450,000	5,000,000	7.00%	15.30%	1011.10%
Qatar	824,355	30,000	219,000	26.60%	1.10%	630.00%
Saudi Arabia	24,069,943	200,000	2,540,000	10.60%	13.10%	1170.00%
Syria	19,514,386	30,000	1,100,000	5.60%	5.70%	3566.70%
United Arab Emirates	3,981,978	735,000	1,397,200	35.10%	7.20%	90.10%
Yemen	21,306,342	15,000	220,000	1.00%	1.10%	1366.70%

populations would require the building of more educational establishments, the expansion of current colleges and universities, and so on. It was stated in the Beirut declaration that as populations grew and demands for higher education increased, many universities had been forced to increase their enrolment figures, despite not having appropriate financial and human resources. This obviously led to a significant drop in educational standards and pedagogical outcomes. Despite the fact that more and more private universities had been established as urged by governments in order to solve the problem, the educational systems still failed to cope. It became imperative, therefore, that other solution avenues needed to be explored.

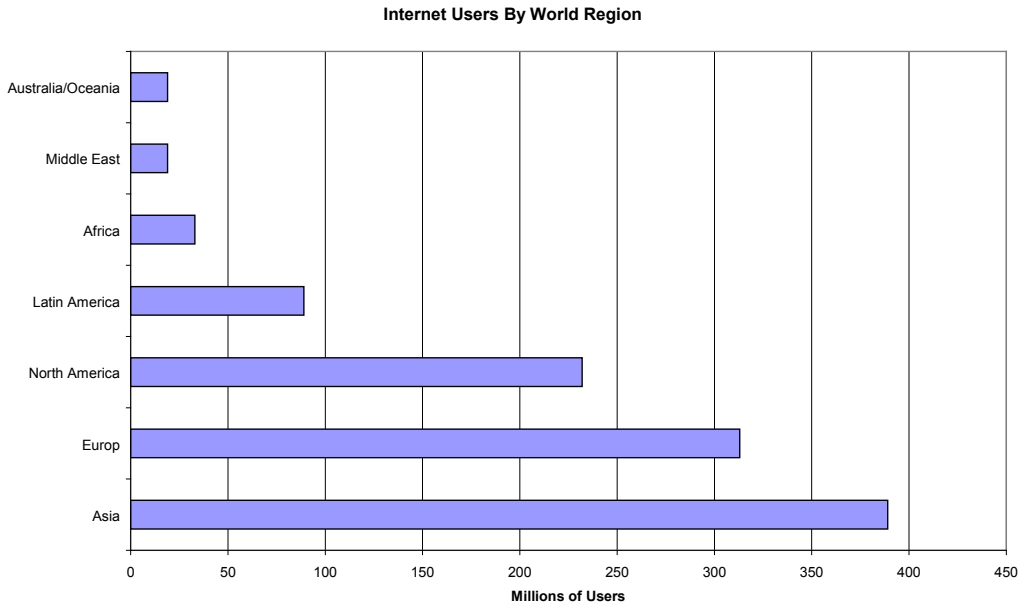
Based on the outcomes of extensive research and as proven by the experience of some national universities, it emerged that not only would e-learning guarantee the maintenance of high educational standards, but that it would also produce the desired educational outcomes (Peter, 2000).

An added advantage of e-learning over other solutions is that it is far less expensive. Building new schools or establishing evening classes within the universities' programs is obviously great deal costlier. Thanks to all this knowledge, implementing e-learning technology in the Arab world has now become the natural solution of choice for institutions seeking to leverage educational standards and pedagogical outcomes (Hedberg, ,2001).

3 LAYING THE GROUNDWORK FOR E-LEARNING IN ARAB COUNTRIES

Today the most popular media for providing e-learning is the internet. By using the different facilities that this technology provides such as web pages, e-mailing, discussion groups, blogs, chatting, Learning Management systems, Multimedia and VoIP (Clark, 2000), the way of delivering education in the Arab world has changed. The internet was first used in Arab countries sometime between 1993 and 1995 (Dewa-

Figure 2. Internet usages in the Middle East against its use in other regions of the world (internet world statistics)



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 3. Internet growth level in the Arab countries between year 2000 and 2007

Middle East	Population (2007 Est.)	Usage, in Dec-00	Internet Usage, Latest Data	% Population (Penetration)	(%) of M.E.	Use Growth (2000-2007)
Bahrain	738,874	40,000	152,700	20.70%	0.80%	281.80%
Iran	70,431,905	250,000	7,500,000	10.60%	38.70%	2900.00%
Iraq	27,162,627	12,500	36,000	0.10%	0.20%	188.00%
Israel	7,237,384	1,270,000	3,700,000	51.10%	19.10%	191.30%
Jordan	5,375,307	127,300	629,500	11.70%	3.20%	394.50%
Kuwait	2,730,603	150,000	700,000	25.60%	3.60%	366.70%
Lebanon	4,556,561	300,000	700,000	15.40%	3.60%	133.30%
Oman	2,452,234	90,000	245,000	10.00%	1.30%	172.20%
Palestine(West Bk.)	3,070,228	35,000	243,000	7.90%	1.30%	594.30%
Qatar	824,355	30,000	219,000	26.60%	1.10%	630.00%
Saudi Arabia	24,069,943	200,000	2,540,000	10.60%	13.10%	1170.00%
Syria	19,514,386	30,000	1,100,000	5.60%	5.70%	3566.70%
United Arab Emirates	3,981,978	735,000	1,397,200	35.10%	7.20%	90.10%
Egypt	71,236,631	450,000	5,000,000	7.00%	15.30%	1011.10%
Yemen	21,306,342	15,000	220,000	1.00%	1.10%	1366.70%
TOTAL Middle East	193,452,727	3,284,800	19,382,400	10.00%	100.00%	490.1

Table 4. Middle East against growth against the rest of the world

Middle East region	Population (2007 Est.)	Pop. % of World	Internet Users, Latest Data	% Population (Penetration)	Usage % of World	Use Growth (2000-2007)
Total in Middle East	193,452,727	2.90%	19,382,400	10.00%	1.80%	490.10%
Rest of the World	6,381,213,690	97.10%	1,074,147,292	16.80%	98.20%	200.30%
WORLD TOTAL	6,574,666,417	100.00%	1,093,529,692	16.60%	100.00%	202.90%

chi, 2001). Since then, its usage has grown exponentially as can be seen in table 2.

After the internet's introduction in Arab countries, many universities adopted the technology. They made changes to their network infrastructure and computer equipment in order to adapt them for use with this new technology. The purpose was to make the best use of it in providing services to their faculty members and their student communities. Most, if not all of them, set up official web sites through which they could disseminate essential information to current and prospective students.

One of the reasons for this low internet penetration in Arab countries is the weak regional Information and Communication Technology (ICT) infrastructure. As revealed at the Beirut summit, most Arab countries lack adequate ICT infrastructure. Services are concentrated in cities, where, in some countries, only 20 to 30% of the people live while the immense majority (around 70% to 80%) are scattered in smaller rural communities where basic essentials such as telephone lines and electricity are either irregular or non-existent (UNESCO, 1998, p.44). In some countries up to 75 % or more of the country's phone lines are found in the capital city alone. Other reasons cited included the high cost of using internet services, which depends on the availability of phone lines, and the cost of using the phone lines themselves. In Arab countries, the cost of using a phone line is considered expensive and when you add the cost of using the internet itself, the whole thing becomes rather astronomical (Dewachi, 2001). Figure 2 is a comparative illustration of internet usage in the Middle East against its use in other regions of the world.

The percentage of internet users in the Middle East is still very small when you compare it with some of the other regions in the world. This probably explains why most universities and other educational establishments have been slow to adopt e-learning and thus have deprived the region's population of reaping the full benefits that emanate from using this mode of delivering education (Jesshope, et al,2000).

Internet technology use in Arab countries is minimal, but nonetheless, the number of users is growing exponentially. This can be pinned down to advances in other areas of human endeavour such as improved income levels, increased awareness and a more educated population. Table 3, taken from (www.internetworldstats.com), shows growth levels per country in the region between 2000 and 2007. Comparing these figures to the rest of the world and summarizing, the position of the Middle East is as shown in table 4.

It is clear from these figures that the rate at which the number of users is growing in the Middle East is far higher than that

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. Apportions LMS adoption by universities per Middle-Eastern country

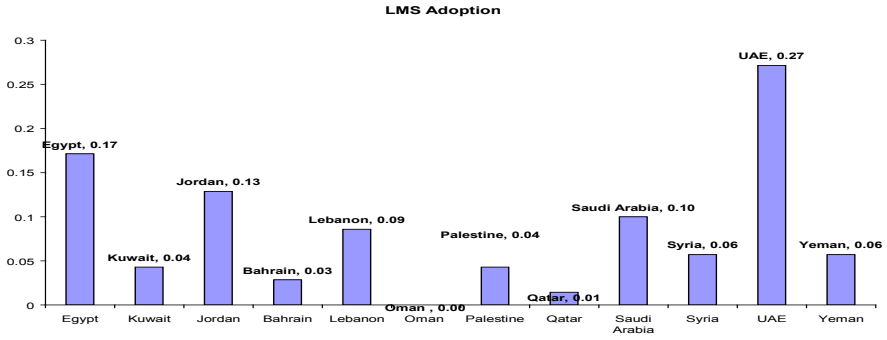
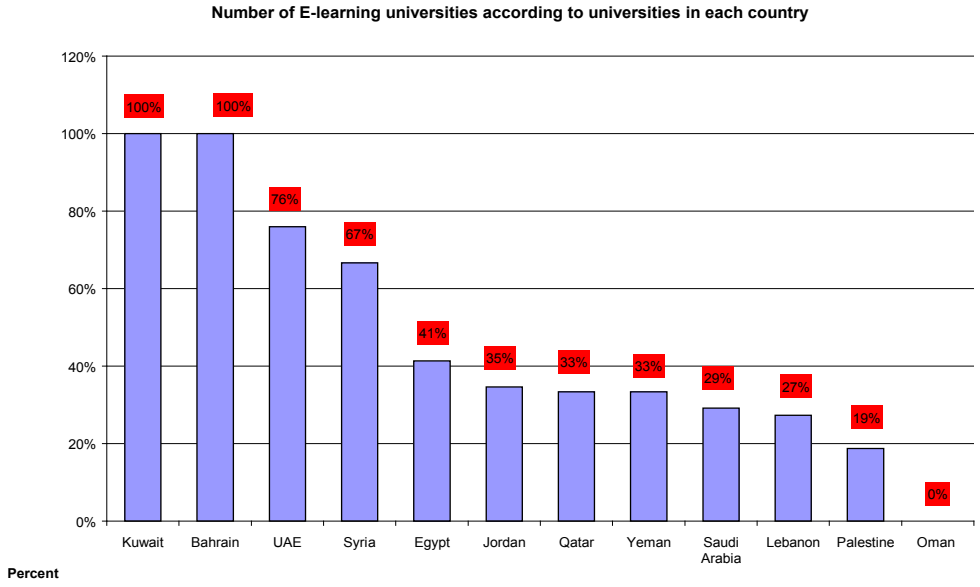


Figure 4. Percentage of universities using E-learning in each country



of the rest of the world. This is because of the fast-paced development of internet infrastructures in the region that usage is growing so fast. The availability of improved tools for producing web pages has encouraged many educational establishments and instructors to set up their own web sites and use them to disseminate information to their students either as additional resources, or as original course materials. (Wegerif, 2004)

As instructors in different universities designed more and more courses, the need arose for those courses to be grouped together and managed under a single system where student access and learning supervision could be facilitated. The result was the creation of what have come to be known as Learning Management Systems (LMS) or Learning Content Management System (LCMS).

4 IMPLEMENTATION OF LEARNING MANAGEMENT SYSTEMS

To implement e-learning, different factors should be taken into account. Some of these are (1) having a Learning Management System in place and (2) having a proper e-content that should reflect the pedagogical needs and pursued goals of the courses to be taught. With so many Learning Management Systems on the market, many universities found it easy to adopt e-learning for their instructors and students. The presence of open-source Learning Management Systems also allowed some low-income universities to adopt the technology without being unduly bogged down by financial or platform concerns (Calverley, 2004).

This survey showed that 41% of the region's universities have adopted the use of Learning Management Systems. Of these, 58.5% are public universities with the remainder being private universities. The United Arab Emirates (UAE) has the highest adoption of LMS among Arab countries (Figure 3). With 19 universities having installed LMS, it represents 27% percent of all LMS usage in the region's universities. The overall high adoption rate in the Arab world is largely due to the tremendous economic growth and development of the public and private sectors in the UAE under the auspices of that country's government, World Economic Forum on the Middle East (2006).

From this chart, we can rank (from highest to lowest) each country's universities' LMS adoption as follows.

UAE → Egypt → Jordan → Saudi Arabia → Lebanon → Yemen → Syria → Palestine → Kuwait → Bahrain → Qatar → Oman.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 5. Number of universities running online programs in each country

Country	Egypt	Kuwait	Jordan	Lebanon	Palestine	Saudi Arabia	Syria	Yemen
Online University	3	1	1	2	1	1	4	1

To give an overview of adoption levels per country according to the number of universities using e-learning, the following chart is presented:

Kuwait → Bahrain → UAE → Syria → Egypt → Jordan → Qatar → Yemen → Saudi Arabia → Lebanon → Palestine → Oman

In Europe, 66% of the universities have LMS. Compared to this, the Middle East's penetration of 41% is relatively small. This can be ascribed to the following reasons:

- High costs of using phone lines and internet services
- Lack of personal computers for university students at home
- Lack of funds to initiate such technological development
- Lack of proper network infrastructure
- Lack of technological equipment such as equipped labs and proper hardware
- Lack of technical staff capable of initiating and consolidating electronic courses
- Lack of interest in e-learning technology among educational decision makers

Despite the small penetration of 41%, the fact that 58.5% of public universities are using LMS to manage their e-learning is a good indication of the support that government ministries of higher education are giving their countries' educational institutions. It is indicative of how seriously these ministries are taking e-learning as they strive to ensure that educational standards and quality are not compromised. This support ensures that the door remains wide open for those institutions that have not adopted e-learning yet to come on board.

5 E-LEARNING METHODS OF DELIVERY

Based on surveys carried out, we found that most universities using LMS are blending e-learning with traditional course delivery methods. The word Blend means "mix", so in other words, colleges and universities are using a mix of traditional learning with e-learning. Another term for blended course delivery is hybrid course delivery. The extent to which each component (e-learning or traditional) is used in the blending depends on the course being taught and course materials required for doing so. Courses that have a very heavy theoretical content will have a larger e-learning component than those that have a heavy practical content. Blended e-learning consists of a number of stages:

At stage one; students are introduced to the training goals of the course and its contents. Stage two marks the beginning of e-learning per se and allows students to acquire basic knowledge in the given subject. Stage three draws on traditional training to allow students to practice the skills and target behaviour they will have learned. It also gives them the chance to share their experiences with each other and engage in discussions that help clarify doubts. In addition, they have the option to resort to their lecturers for more authoritative and thus reliable explanations. The forth and last stage is a return to electronic format comprising reviews, exercises, tests, etc. to allow participants to consolidate the knowledge and skills acquired during the first three stages.

These four stages are largely applied in most Arab universities where a number of universities are also using what is known as the "Online" e-learning Method. In this method (where programs are usually identified as Online Degrees), the first 3 stages are done online with no traditional classroom attendance (Thorpe, 2002).

At stage four, some universities require student presence on campus to sit mid-term and final examinations. Our survey revealed that out of the total of 70 universities providing e-learning in the region, 15 also run 20% of their programs online. Table 5 shows the number of universities running online programs in each country.

The biggest challenge faced by the online method of learning in this region is educational-policy hostility. Most countries still refuse to recognized let alone accredit online degrees. In fact, there appears to be on-going effort by authorities to discourage students from enrolling in online degree programs offered by many European, Australian and north American universities. The first Arab Online University was established in Kuwait in the year 2000. Being the pioneer, it was greeted with scorn in many countries. It was only after it had proven itself that it started gaining acceptance. Today some countries in the region such as Kuwait, Jordan, Bahrain, Saudi Arabia and Lebanon have even accredited its degrees (AOU,2002).

Accreditation did not come easy. The Arab Open University (AOU) had to work hard to prove the quality of its degrees by producing highly knowledgeable and capable graduates who were as good as their counterparts who got their degrees the traditional way. The university spared no effort in exposing how baseless the false impressions harbored by many people regarding online learning were. As a result, online learning was de-stigmatized and acceptance followed. Since its inception, the aims of AOU have been:

- To ensure quality education
- To upgrade teachers
- To democratize education
- To prepare students for the workplace (AOU,2002)

The gigantic step taken by Arab Open University paved the way for other universities to follow. Nowadays the number

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

of universities accepting a blend of traditional and online learning is on the rise. The Al-Baath University, the University of Aleppo and the University of Damascus in Syria; the University of Science and Technology in Yemen; and the Al Quds Open University in Palestine have all taken to online education. In the next few years, we are likely to see more countries and universities accrediting e-learning degrees.

6 OPEN SOURCE OR LICENSED APPLICATIONS

During the past ten years, we have seen relentless efforts by companies to build and adapt their LMS to the ever-increasing challenge for improving the quality of e-learning processes. New applications and activities have been developed in tandem with the fast development of e-learning technology. Three of the most used LMS applications in the region are, WebCT, BlackBoard, and Moodle. Our survey shows that universities differ broadly in their choice of LMS and the choice is predominately tied to a university's financial standing as well as the support it receives from government though such entities as the Ministry of Higher Education. The study shows that countries that are hard up financially and are bogged down by financial and economic burdens usually opt for open-source systems such as Moodle (Calvo et al, 2001). More affluent countries such as the United Arab Emirates, Saudi Arabia, Qatar and Bahrain, on the other had, prefer licensed programs like WebCT and BlackBoard. Table 6 shows how many universities in each country have chosen a particular type of LMS.

Table 6. Universities and LMS

Country	Moodle	WebCT	BlackBoard	Not Defined
Egypt	12	×	×	×
Kuwait	1	×	1	×
Jordan	3	×	4	2
Bahrain	1	1	×	×
Lebanon	2	2	1	3
Oman	×	×	×	×
Palestine	3	×	×	×
Qatar	×	×	1	×
Saudi Arabia	3	4	×	×
Syria	×	×	×	4
UAE	1	14	4	×
Yemen	4	×	×	×

According to this table, 44% of the region's universities use licensed programs while 41% use open-source programs and 14% remain undefined. By further breakdown the data we obtained, we were able to determine what percentage of universities using LMS were government run and what percentage were in private hands. This is depicted in the chart below.

7 CONCLUSIONS AND RECOMMENDATIONS

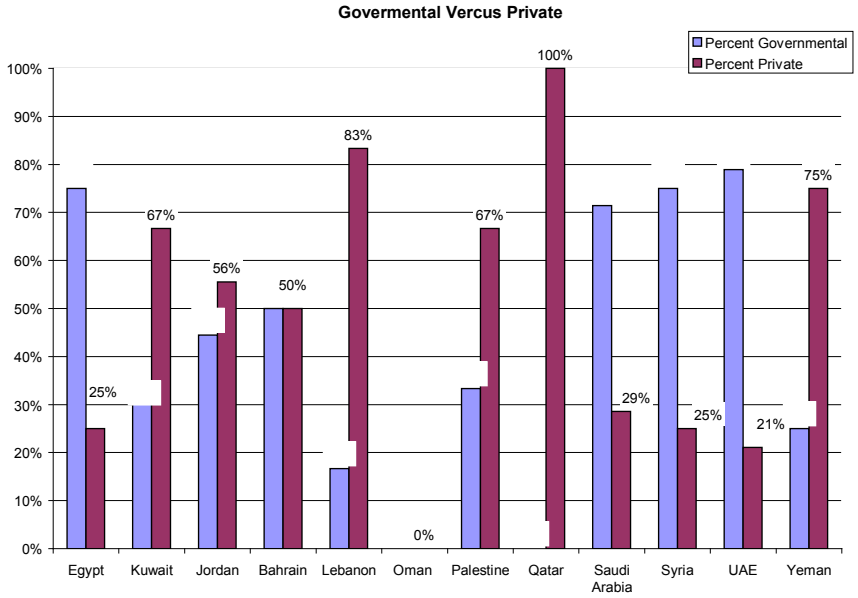
A statistical analysis of the data compiled during this study revealed many important facts about e-learning in the Middle East. Despite the fact that the region as a whole is adopting e-learning at a very accelerated pace, more studies need to be carried out to learn more about the full potential of its use in the Middle East. It is sad to see that most researchers ignore this part of the world but, understandably, this is due to the political and economic instability that currently beset most countries in the region. The following points summarize of some of the important facts revealed by this study which, as already mentioned above, was based on examining the web pages of each university in the region:

- One of the major obstacles hampering e-learning adoption in the region is the low penetration of internet usage which represents a mere 2% of the world's total.
- Interest in e-learning technology is gaining momentum in the Middle East due to the ever-growing number of internet users.
- Interest in e-learning is as equally strong among public universities as it is among private institutions.
- Many are the countries in the gulf region that use licensed Learning Management Systems; some countries, however, are still heavily dependant on free open-source programs such as Moodle.
- Pursuing online degrees is still despised in many Arab countries as evidenced by the scorn with which the (AOU), the first university in the region to offer online degrees, was greeted.
- Some universities are adopting online degree programs within their curriculum; as such, the prospects of educational flexibility for both full-time students and working adults look very promising.
- There is need to take Iraq into account as it is also an Arab country within the Middle East. Due to the chaotic situation the country is now going through, it was not possible to include it in this study.
- In conclusion, we recommend that more research be carried out to fully explore the situation in the Middle East. The focus should be on the following aspects:
- The exact penetration of e-learning in Middle-East universities as indicated by the amount of electronic courses provided by these universities
- The pedagogical value of and concerns regarding the electronic courses being provided.
- The types of infrastructure needed by universities and the costs for adopting e-learning.
- The underlying reasons why some universities in the region have been reluctant to adopt e-learning.
- The inclusion of Iraq in any future studies, as e-learning might be a good answer to that country's educational needs too. This would require a through evaluation of Iraq's current infrastructure.

Many other aspects can be studied with the aim of obtaining information that can be vital to the setting up of the right infrastructure necessary for the provision of services that best

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 5. Percentage of Governmental versus Private Adoption for E-learning in each Country



meet the high-standard-education needs of learners in the region.

REFERENCES

1. Peter, J. (2000), 'How E-Learning Will Transform Education'; http://www.edweek.org/ew/ewstory.cfm? Slug=02stokes.h20 (10 Nov 2005).
2. Abouchedid, K., and Eid, G. M. (2004). 'E-learning challenges in the Arab world: Revelations from a case study profile', Quality Assurance in Education, Volume 12 – Number 1 – 2004 – 15- 27.
3. Abdelraheem A.., (2006), 'The implementation of e-learning in the Arab Universities: Challenges and opportunities', proceeding of the 7th APRU Distance Learning and the Internet Conference 2006, Tokyo, Japan.
4. Arab Open University (AOU) (2002), 'Partnership Fact sheet' ,http://www.ouworldwide.com/pdfs/AOUpartnership.pdf (16 Jun 2007).
5. Dewachi, A., (2001), 'Overview of Internet in Arab States', Arab Region Internet & Telecom Summit, Muscat, Oman, 28 – 30 May, http://www.itu.int/arabinternet2001 (16 Jun 2007).
6. The league of Arab states (LAS) (2003),'Document DT/2 and DT/2', Economic Department, and Division of Basic Services Sector, pp. 1-6.
7. Charnistski, C. W., Molinaro, J. A., Corabi, J., & Nolan, K. (2003). 'Comparing Student Achievement in a Graduate Level Research Methods Course Using Face-to-Face and Web-based Instruction: Result of a Pilot Study'. In Rosset, A. (Eds.) World Conference on E-Learning in Corp., Government., Health., & Higher Education (ELEARN), Association for the Advancement of Computing in Education (AACE), 174-177, VA, USA
8. Wegerif, R. (2004). 'The role of educational software as a support for teaching and learning conversations', Computers & Education, 43 (1-2), 179-191.
9. Calverley, G. (2004), 'Book review: Online Education, Learning Management Systems, Global E-Learning in a Scandinavian Perspective (Morten Flate Paulsen)'. Educational Technology & Society, 7 (2), 141-144.
10. Hedberg, J. G. (2001), 'The online and digital experience: reassuring higher-order learning outcomes'. In L. R.Vandervert, L. V. Shavinina & R. A. Cornell (Eds.), Cybereducation, the future of long distance learning, New York: Mary Ann Liebert, 219-236.
11. Thorpe, M. (2002). 'Rethinking Learner Support: the challenge of collaborative online learning'. Open Learning, 17 (2), 105-119.
12. Clark, J. (2000), 'Collaboration tools in online learning environments', ALN Magazine, 4(1), http://www.aln.org/alnweb/magazine/Vol4_issue1/Clark.htm (16 Jun 2007).
13. Jesshope, C., Heinrich, E. & Kinshuk (2000), 'Technology Integrated Learning Environments for Education at a Distance', Proceeding of the DEANZ 2000 Conference, 26-29 April 2000, Dunedin, New Zealand.
14. Calvo R.A., Sabino J., Ellis R. (2001), 'OpenACES: the open source solution to e- learning', Moving Online II, Gold Coast. Australia.
15. Bascsich (1997), 'Re-engineering the Campus with Web and related technology for the Virtual University'. Paper presented at the Annual Conference on Flexible Learning on the Information SuperHighway, May 19-21, Sheffield Hallam University, UK.
16. Bates, A.W. (2000), 'Managing Technological Change: Strategies for College and University', San Francisco: Jossey Bass.
17. World Economic Forum on the Middle East (2006), 'The Promise of a New Generation', Sharm El Sheikh, 20-22 May 2006, http://www.weforum.org/pdf/SummitReports/middleeast2006.pdf (16 Jun 2007).
18. WebCT (2007),www.webct.com (16 Jun 2007).
19. Blackboard (2007), www.blackboard.com (16 Jun 2007).
20. Moodle (2007), available at: http://Moodle.org (16 Jun 2007).
21. World Internet Statistics, Enrique de Argaez website; http://www.internetworldstats.com (March 2007).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

The blend of m-Learning and e-Learning at AUC

Ahmed Sameh

The American University in Cairo,
P.O.Box 2511, Cairo, Egypt
sameh@aucegypt.edu

Abstract By the year 2008, The American University in Cairo (AUC) is planning to move to its new campus site which will feature blanketed 100% wireless one-of-a-kind educational M-learning (Mobile Learning) pervasive Dome. The wireless Dome is expected to provide new opportunities for improving both the teaching and learning processes at AUC. Sustained connections over time, within the Dome, will intensify students' Learning Process. Mobile Content can-and must be- planed before the move to the new campus. This mobile educational content will be accessed from within the Dome through iPods, MP3 players, i-Mate PDAs, Cell phones, and Smart Wireless Devices. WiFi and WiMax technologies currently exist for providing an effective mobile access for M-Leaning with wireless broadband. Enhanced course management systems such as WebCT will still be around in the new campus with more interactive E-Learning features added. But AUC will experience a slight Shift from E-learning to M-learning as a result of the new wireless Dome. Consequently, AUC will experience a novel blend of M-Learning and E-Learning. To answer the question: What is the right blend? Several pilots and trials have to be investigated to measure the effectiveness of the various blends.

Keywords WiFi, MiMax, Ontology, Content Management, Mobile Devices

1 RESEARCH CONTEXT

The American university in Cairo is a private institution currently located at the heart of Cairo. In 1998, its Board of Trustees approved the relocation of the university campus from the crowded and limited down town campus to 250 acre space in the outskirts of Cairo in what is now known as New Cairo. This new Campus will be in the heart of this newly developed urban suburb. Construction of the university's new campus started by 2002 and expected to be completed by the fall of the year 2008. The planned new campus site is designed to be most technologically advanced educational and learning space. The Information and Communication technology (ICT) infrastructure was designed to adopt the most advanced learning teaching technologies for years to come. A high powered converged IP network infrastructure was designed and the necessary cabling and wireless infrastructure has been contracted out to international vendors. Although most of the technological features that will be installed in the new campus has already been adopted and operational in the down town campus, the magnitude and sophistication that will be available in the new campus will require major thrust and cultural changes that necessitate the establishment of many awareness and adoption training for the university constituency. Many pilot high tech class rooms and systems have been planned for installation prior the transfer to the new campus. Within this context this research paper has been developed as part of the on-going efforts to orient and institutionalize the coming cultural changes in the new campus.

2 M-LEARNING: AN EVOLUTION OF E-LEARNING

Evidences attest that the mobile revolution is finally here. Wherever one looks, the mobile penetration and adoption is irrefutable: cell phones, PDAs (personal digital assistants), MP3 players, iPods, portable game devices, handhelds, tablets, and laptops abound. No demographic is immune from this phenomenon. In fact, it is estimated that by early year 2007, two and a half billion people, all over the world, will be walking around with powerful computers in their pockets and purses. They often do not realize that, because they call them something else. But the fact is that today's high-end cell phones, iMate-PDAs, MP3 players, iPods, portable game devices, handhelds, tablets, and laptops have the computing power of a mid-1990s personal computer (PC)—while consuming only one one-hundredth of their energy. Currently there appear to be a nascent educational wireless market. Mobile learning (m-learning) is positioned to address the growing public concern about traditional teaching and learning processes. The trend toward increased mobility and nomadic in traditional learning and training environments such as campuses, offices, and workplaces is driving demand for a new breed of mobile learning services. There is a sense that higher educational institutes like AUC located in a Country like Egypt-where mobile penetration is vastly greater than fixed-line and PC penetration combined-should place mobile learning at the top of their technological educational agenda.

Much debate has focused on the definition of Mobile Learning (m-learning) though. Is it about mobile learners? Is it about small personal devices? Is it about communication and collaboration? Is it about context sensitivity? Is it a natural extension to e-learning? Insofar as students have traditionally used their time on public transport to catch up on required reading or last-minute revision, Mobile Learning has been with us for quite a while. However, today's technology has significantly extended the scope for learning on the move, and the term "m-learning"" has gained serious currency in describing wireless-enabled learning strategies and processes across the entire gamut of instructional delivery. Current emphases appear to be in remote just-in-time applications, but there are also many instances of m-learning blended into more traditional instructional scenarios. Having material on your phone or palmtop means that it is always accessible to you. Whenever you have a spare five minutes, you can use it to practice some learning, just as you might choose to play solitaire whilst waiting for a train or bus. Learning materials that are colorful, engaging and stimulating make the learner want to go back and practice many times. In fact, m-learning has moved from being a theory, explored by academic and technology enthusiasts, into a real and valuable contribution to learning.

For the purposes of this paper, we will take the lead from the learner, and define "m-learning" as making use of whichever devices and technologies surround our learners, in an attempt to empower and enrich their learning, wherever and whoever they are. It is known that m-learning can empower and engage. It is also known that the engagement and motivation can continue beyond the initial 'gadget honeymoon'. We know that learners are more comfortable engaging in personal or private subject areas via a mobile device than via traditional methods. When it first became widespread, one of the biggest failings of e-learning was the assumption that it could become everything. Teachers were no longer required. Anything could be taught using it. Success was only about 'broadcasting' good quality learning materials. We now know that this is not true, and that good teachers, communication, collaboration and discovery-activities are essential. The good news about m-learning is that it is very difficult to make the same mistakes because the devices being used are that much less powerful than PCs. There is clearly no single solution. Screens are smaller. Many have no keyboards. Connection speeds are slower. Processing power is weaker. There is no single 'platform' or set of features that dominates. The learning you can do on an iPod or MP3 player is very different from what you can do with SMS. In the light of this, we have found it very helpful to describe mobile learning not as a single thing, but rather as a collection of new tools that can be added to a tutor's teaching toolbox, to be assembled as required to achieve specific aims. Some of these tools are:

- SMS (text messaging) as a skills check, or for collecting feedback, or call for wireless conference calls
- audio-based learning (iPod, MP3 players, pod-casting)
- Java quizzes to download to color screen phones
- focused learning modules on a iMate-PDA
- media collection using a camera phone
- online publishing or blogging using SMS, MMS (picture and audio messages), cameras, e-mail and the web.

Blending such M-learning and digital-based game learning tools into E-learning have the potential to achieve what e-learning has tantalizingly offered but never truly achieved: always on learning, accessible to the masses, but tailored to the individual. In this paper, we are suggesting wide possibilities of blends of both E-learning and M-Learning to be used at for AUC's new campus.

The structure of this paper goes as follows: section 3 looks at the current state of m-learning at various campuses. Section 4 describes a vision of both m-learning and E-learning at the new campus. The need to plan for mobile contents is explained in section 5. Section 6 elaborates on available wireless technology options. The proposed testbed along with a number of experimental pilots/trials are described in section 7. The implementation of two pilots/trials is described in section 8, and the conclusion in section 9.

3 M-LEARNING AT VARIOUS CAMPUSES

In order for a technology to improve learning, it must fit into students' lives, not the other way around. Despite what some may consider mobile devices' limitations, University Level students are already inventing ways to use them to learn what they want to know. If educators are smart, they can figure out how to deliver their product in a way that fits into these students' digital lives—and their mobile devices. The combination of wireless technology and mobile computing is resulting in escalating transformations of the educational world. The question is, how are the wireless, mobile technologies affecting the learning environment, pedagogy, and campus life? To answer this question, we must assess the current state of affairs by surveying the cyber-culture on various campuses. Most of the US 3,913 accredited colleges and universities haven't launched initiatives that recommend or require students to use handheld computers. Yet hundreds are experimenting with how to enhance learning with the mobile devices—hoping to leverage the coming convergence of wireless networks, Web services, and enterprise applications. Some pilot projects, like those at Western Carolina University in North Carolina [1] and Loyola College in Maryland [2], are sputtering for lack of funds, or because they aren't central to the college's technology strategy yet.

Bleeding-edge universities such as Harvard are actively exploring how wireless can enhance learning and teaching [3]. Innovative audio-learning products available for download to devices including Apple iPods, MP3 players, iMate-PDAs, BlackBerry, and smart wireless devices are used by many students. Moreover, at Harvard, each student is equipped, at its enrollment in University, with a Tablet PCs of his/her own, which he uses while attending lessons, doing the assigned homework, and enhance his/her learning.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Duke University [4] has been giving an Apple iPod to each incoming student as part of an initiative to encourage use of technology on campus. The devices - more regularly used as music players - come preloaded with university-related content, while a special website modelled on Apple's own iTunes site allow students to download faculty-provided course content, including language lessons, music, recorded lectures and audio books. The move is a pilot programme between Apple and Duke University that will be evaluated at a latter stage, with Duke covering the $500,000 project out of funds set aside for technology innovation [4]. The university hopes to motivate students to think creatively about using digital audio content and mobile computing environments to advance educational goals, with a growing number of faculty members showing interest in adding audio and video components to their courses.

With a 100% wireless campus, a diverse and growing list of iPod projects that enhance teaching and learning, and a new library and Information Technology Center, Georgia College & State University (GC&SU), Georgia's Public Liberal Arts University, is embracing wireless educational technology [5]. They claim to be leaders in speaking to their students in a language they can understand and enjoy. They claim to be one of the largest and most diverse users of iPod technology in higher education in the world! The iPod project at Georgia College & State University has been running since Fall 2002. In keeping with the Liberal Arts Mission of GC&SU, the iPod allows learning to take place outside and inside regular classroom meetings, to enhance the face to face classroom experience. Each participating professor received an iBook with iTunes, and an iPod. Using the iBooks and iTunes, the professors gathered their audio files together. Staff assist in digitizing the audio portions of a video of a lecture and to digitize an audio files, converting both to MP3 format for use with the iPod. iPod quickly took over the classroom as a portable learning tool, allowing anywhere, anytime access to speeches, audiobooks, and lectures. Soon photos and podcasts expanded teaching possibilities, and video is evolving the experience even further [5]. The opportunities are endless for teachers to seamlessly create, organize, distribute, and access all kinds of learning materials. Loading files, photos, notes, and songs onto one's iPod is easy. The iPod dock connector on the bottom of the iPod lets one connect, sync, and recharge quickly using the included USB cable. Or for iPod, one can leave the optional iPod Universal Dock connected to his/her Mac or Windows PC, and sync and charge every time he/she dock the iPod. And since the dock works with the new Apple Remote, one can now connect the Universal Dock to a stereo system, powered speakers, or a TV, and control his/her audio books, slideshows, and video from across the room.

In the summer of 2006, the Stanford Learning Lab (SLL) developed a few rough prototypes for mobile learning [6]. The SLL staff chose foreign language study as the content area, hypothesizing that mobile devices could help provide sorely needed opportunities for review, listening and speaking practice in a safe, authentic, personalized and on-demand environment. In fact, the field of mobile learning is at present characterized, at most of other universities, by a proliferation of pilots and trials that allow mobile technologies to be tested out in a variety of learning contexts. The sustained deployment of mobile learning will depend on the quality of these pilots and trials, which includes evaluation methodology and reporting. At this stage, we examine current evaluation practice, based on evidence drawn from conference publications, published case studies, and other accounts from the literature. We also draw on our work in collecting case studies of mobile learning from a range of recent projects. Issues deserving examination include the apparent objectives of the pilots or trials, the nature of the evaluations, instruments and techniques used, and the analysis and presentation of findings. This section reflects on the quality of evaluation in mobile learning pilots and trials, in the broader context of evolving practices in the evaluation of educational technologies.

An analysis of 12 international case studies in [6] reveals that reasons given for using mobile technologies in teaching and learning on campuses relate principally to:

- Improving access to assessment, learning materials and learning resources,
- Increasing flexibility of learning for students,
- Compliance with special educational needs and disability legislation,
- Exploring the potential for collaborative learning, for increasing students' appreciation of their own learning process, and for consolidation of learning,
- Guiding students to see a subject differently than they would have done without the use of mobile devices,
- Identifying learners' needs for just-in-time knowledge,
- Exploring whether the time and task management facilities of mobile devices can help students to manage their studies,
- Reducing cultural and communication barriers between staff and students by using channels that students like,
- Wanting to know how wireless/mobile technology alters attitudes, patterns of study, and communication activity among students,
- Making wireless, mobile, interactive learning available to all students without incurring the expense of costly hardware,
- Delivering communications, information and training to large numbers of people regardless of their location,
- Blending mobile technologies into e-learning infrastructures to improve interactivity and connectivity for the learner,
- Harnessing the existing proliferation of mobile devices services and their many users.

A review of another 27 projects documented in the proceedings of M-LEARNING 2003 to 2006 [7] shows a similar spread of objectives, with a predominance of objectives identifying or targeting changes in teaching and learning:

- to enable students to look at course information any time and anywhere,

- to ensure that every student can access content independently of the channel he or she chooses to use,
- to ensure that classroom-based pupils benefit from the experience of a field trip being undertaken by their peers,
- to explore the potential for individualized mobile learning - revision material tailored to the needs of the individual,
- to provide learners with a flexible context-awareness system that can react to their needs,
- to provide immediate feedback through interactive tests: the user knows in real time if their choice is correct,
- to allow interactive screens encouraging art gallery visitors to respond to the art on view,
- to set of innovative games, materials and activities which will motivate reluctant young learners,
- to provide user-friendly m-portal that is powerful and empowering, and encourages active participation by its users,
- to enhance interactivity and cooperation while preserving the traditional advantages of face-to-face encounters,
- to provide informal learning with multiple media,
- to investigate how self-produced videos, made with a digital video camera and later viewed on handheld mobile computers, can support informal learning,
- to provide video and still images giving additional context for art gallery works on display, opportunities to listen to an expert talk about details of a work, with the details simultaneously highlighted on the screen,
- to enhance the audio presentation of a multimedia museum guide by using the mobile device screen to travel throughout a fresco and identify the various details in it,
- to explore how context-dependent learners' knowledge concepts are,
- to evaluate fragmentation in mobile learning based on students' deep and surface approaches to learning,
- to capture learners' thoughts, views and behaviors in a mobile learning setting,
- to remain at the cutting edge of educational technology by helping to shape a new generation of multimedia tours in art galleries,
- to investigate whether an integrated set of learning tools would be useful, which tools would be adopted and the contexts in which the tools would be used,
- to find out in which arenas handhelds are used, how and why they are used, and what role they can play,
- to find out what the future take-up of new services and facilities on mobile phones and other technology devices might be,
- to find out whether young adults would be willing to use their phones for literacy and training courses learning,
- to understand the range of actions and opportunities open to mobile learners, and seek ways of extending this range to support what learners want to do – even if they themselves do not yet know what that is.

4 M-LEARNING AND E-LEARNING AT AUC

By the year 2008, AUC is expected to move to its new blanketed 100% wireless campus that will provide a novel educational M-learning pervasive Dome. The wireless Dome is expected to provide new opportunities for improving both the teaching and learning processes at AUC. It is expected that sustained connections over time, within the Dome, will intensify students' Learning Process. Mobile Content can-and must be- planed before the move to the new campus. This mobile educational content will be accessed from within the Dome through iPods, MP3 players, i-Mate PDAs, Cell phones, and Smart Wireless Devices. WiFi and WiMax technologies currently exist for providing an effective mobile access for M-Leaning with wireless broadband. It is expected that Enhanced WebCT will still be around in the new campus with more interactive E-Learning features added. But AUC will experience a slight Shift from E-learning to M-learning as a result of the new wireless Dome. Consequently, AUC will experience a sort of a blend of M-Learning and E-Learning. To answer the question: What is the right blend? Several pilots and trials have to be investigated to measure the effectiveness of the various blends.

As a private university, AUC with its high-social class of students, the student population is extremely technology savvy and everywhere you go, you find students with iPods, MP3 players, PDAs, smart wireless notebooks and devices. If the university starts streaming some of its educational material through streaming servers, to be available either for on-line interaction or for download and play later by students, new channels of teaching and learning will be established at the new AUC Campus. A new era of Pervasive Education at AUC can begin by providing such Educational spaces. Most of the students with their culture background will see this as a novel way of learning and will wetness a technology that fits into their lives, not the other way around. New opportunities for improving both the teaching and learning processes at the new campus will be discovered and revealed by setting up pilots, trials and case studies to investigate objectives similar to ones presented in the previous section at other leading universities. Pilots and trials are also set to investigate the right blend of M-learning and E-learning. Issues deserving examination include the apparent objectives of the pilots or trials, the nature of the evaluations, instruments and techniques used, and the analysis and presentation of findings. Also reflects on the quality of evaluation in these pilots and trials, is the broader context of evolving practices. The pilots and trials should identify the right blend of M & E learning with even traditional learning and develop a strategic plan around them. At the initial stages, AUC will initially concentrate the initial pilots & trials on specific areas such as the English and Arabic Language institutes (ELI, and ALI), PVA, History, Arabic literature, film, Psychology, Political science, Philosophy, Music, Linguistics, Islamic studies, comparative literature, and sociology.

In the light of this, we have found it very helpful to utilize mobile learning not as a single thing, but rather as a collection of new tools that can be added to a AUC tutor's teach-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

ing toolbox, to be assembled as required to achieve specific aims in the target subject areas. Some of these tools are: SMS (text messaging) as a skills check, or for collecting feedback, or call for wireless conference calls; audio-based learning (iPod, MP3 players, pod-casting); Java quizzes to download to color screen phones; focused learning modules on a iMate-PDA; media collection using a camera phone; and online publishing or blogging using SMS, MMS (picture and audio messages), cameras, e-mail and the web.

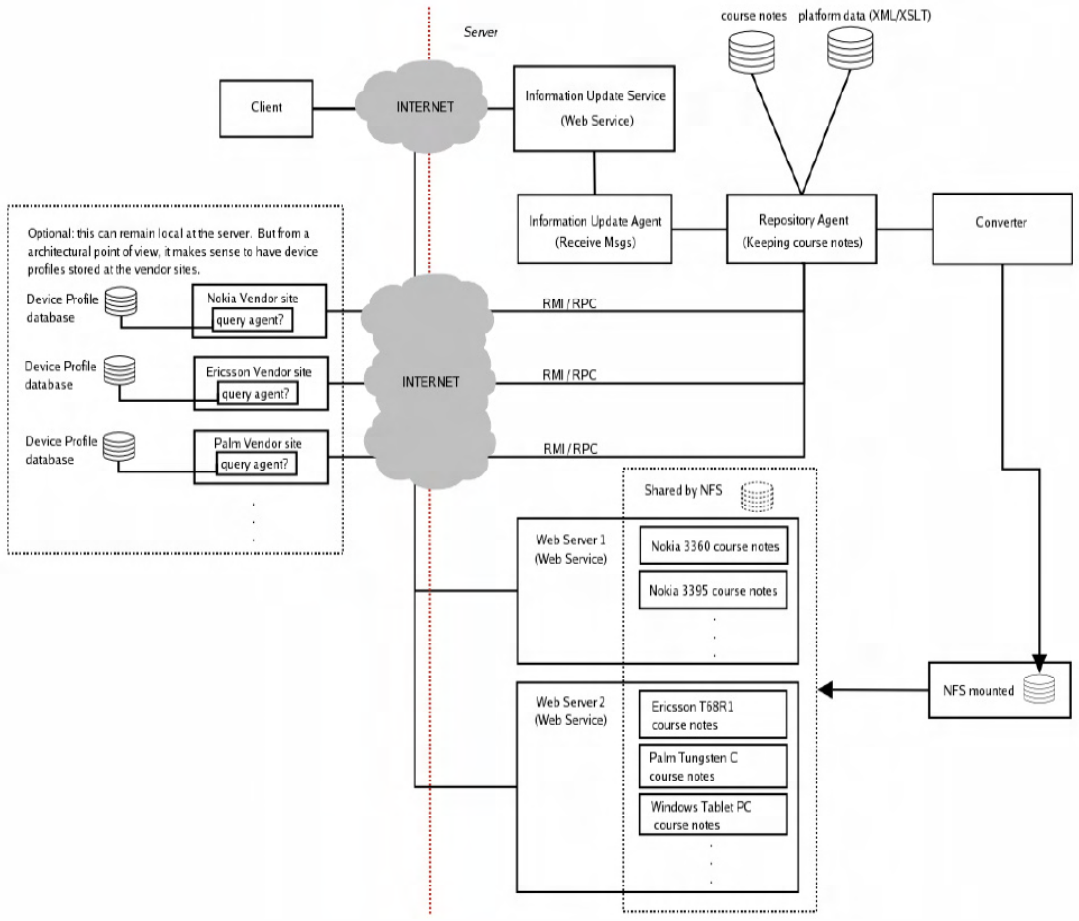
These enhanced capabilities mobile technologies offer for rich social interaction could totally transform how AUC students learn. The most exciting and lucrative opportunity may be in the delivery of educational content to learners allowing to provide new form of learning that does not exit any campus else in Egypt. The combination of wireless connectivity and educational content delivered according to the learner's location, requirements, and skills level may be an application that kills for the new campus of AUC. Mobile devices allow everyone-and not just mainstream consumers, but also those way out on the periphery-to learn and exchange ideas. Active involvement from content owners and providers can help connect a new, more universal social mind, not to mention help build the demand for educational content.

In fact, Egypt is moving towards introducing infrastructures for m-learning technologies. For example, earlier this year the Egyptian government has launched a very promising wireless project in Alexandria[8]. Egypt's government has announced its deployment of a trial 3G network that is used to provide fixed wireless internet access to Egypt's schools. Lucent company will work with its partner TeleTech to deliver a CDMA2000 1xEV-DO and WiFi trial network. The trial network complements Telecom Egypt's CDMA2000 1X network in the El-Makes area and provides a solution that meets the challenges associated with deploying a wireline-based broadband network in Egypt. Students and teachers at El-Makes Alexandria school are now able to access the network via wireless desktop or laptop computers [8]. While WiFi provides connections within the confines of a building, it is the CDMA2000 1xEV-DO network that transports the data to and from the Internet. AUC plans for its new campus comes in-line with that (see section 5).

5 MOBILE CONTENT

For the purpose of this paper we have chosen to start with pilot trials for two areas: ALI and History. We have developed pilots and trials specifically for the two courses ALI234 and HIST222 using the tools described above (SMS, audio, etc.). We investigated the two courses from the AUC catalog and their WebCT archives, identified how the above mobile learning tools can be applied to these two courses: how some of these courses' contents can be put in mobile content format; Sound files on Tracks.; Cameras for Historical images; Images for Arabic alphabet and words; Reading audio Arabic literatures, translations, Arts and architecture images, video and images of field trips to historical sites, Poetry and Novel audio books, Painting, and studies of Qur'an, Linguistics of

Figure 1. The overall architectural diagram



Arabic, Arabic Morphology, Arabic Syntax, Sira-Hadith, and Tafsir.; and Islamic law. As for the HIST222 course: mobile material for Arab History; Islam, Civilizations, Pharaonic Egypt, European history, and American History were identified.

Our approach is to set up several pilots and trials to measure the effectiveness of various blends for the above mobile contents with existing WebCT materials. After each pilot/trial experiment, there is a need to conduct surveys and analyze feedback to draw conclusions. In performing our experiments, we have used cell phones, PDAs (personal digital assistants), MP3 players, iPods, portable game devices, handhelds, tablets, and laptops as mobile devices. We have also identified other non-mobile material that is to be delivered through the traditional WebCT. Learning materials that are colorful, engaging and stimulating make the learner want to go back and practice many times. Innovative audio-learning products available for download to devices including Apple iPods, MP3 players, iMate-PDAs, BlackBerry, and smart wireless devices are used. ALI234 and HIST222 faculty members showed interest in adding audio and video components to their courses. Our approach realize that each course is different and each faculty is different, and as such there is great flexibility in choosing which lessons of the class will use which tools and how it will be blended into existing lesson material.

We have gathered some lectures' audio files together after digitizing the audio portions of a video of a set of ALI234 and HIST222 lectures and digitized audio files, converting both to MP3 format for use with the iPod. iPod as a portable learning tool, allowed anywhere, anytime access to speeches, audio books, and lectures. Soon photos and podcasts expanded teaching possibilities, and, video is evolving the experience even further. The opportunities are endless for teachers to seamlessly create, organize, distribute, and access all kinds of learning materials. Loading files, photos, notes, and songs onto one's iPod is easy. For example, in HIST222, one might be interested in using larger handheld device like a iMate-PDA for his/her m-learning, it helps to have existing, relevant content to include on as templates. This is a real challenge because the small, compact nature of all good m-content means that a motivated student can work through lots of it! We have created over two modules (Arab History and European History)of trialled and tested, curriculum mapped content to run on any PocketPC device. Along the way we have developed several templates and frameworks which embed the best practice and learning design that seems most effective (see sections 6 & 7). These templates have been added to a PC-based authoring tool that lets tutors create their own versions, putting in their own content but using our interactivity. The simplest example of this would be a quiz or a PowerPoint-type presentation, though there are several more 'game-like' activities, all of which can now be created by any tutor using our authoring system. But the content is only the first part. There are three other aspects of creating PDA-ready learning materials that we can help with: installing the materials to the device, helping the student navigate around your materials, and tracking and reporting on use.

6 WIRELESS TECHNOLOGY OPTIONS

AUC new campus will have the following Internet access methods known to date: Dial-up, ISDN, DSL, Cable, Wi-Fi, WiMAX, Satellite, Fiber Optic (T-1/E-1), Power-line Internet. Figures 2 &3 show the main network types and the technologies associated with each using these access methods. Personal Area Networks (PAN), Local Area Networks (LAN), Metropolitan Area Networks (MAN) and Wide Area Networks (WAN) are mapped against ranges and throughputs. In the new campus we are especially interested

Figure 2. Wireless technologies: network type, range and throughput

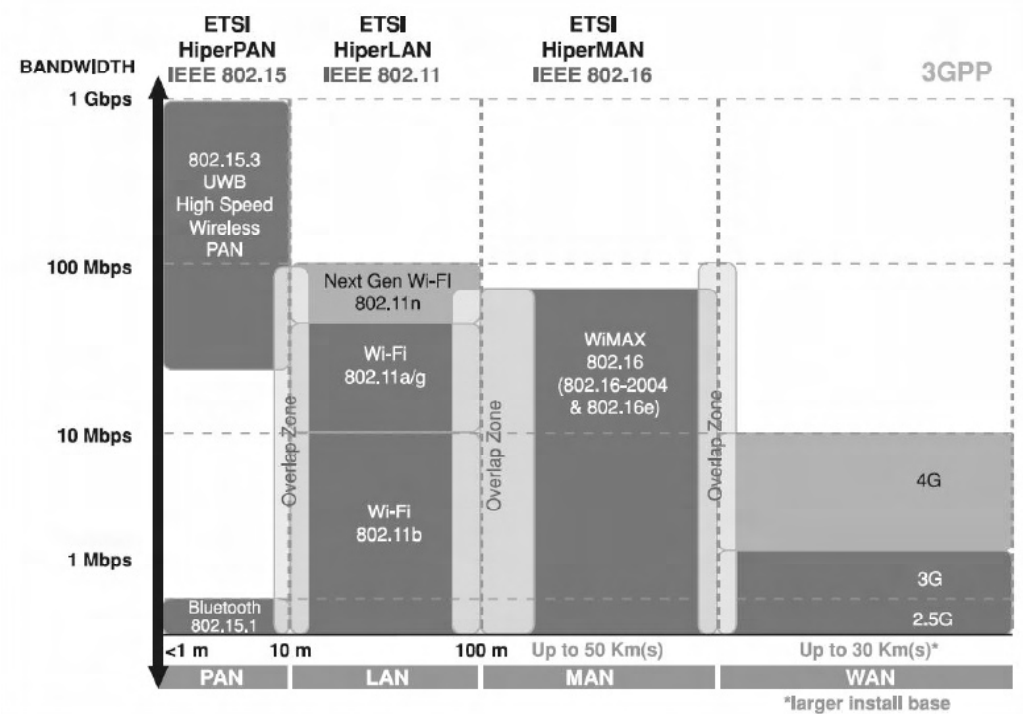
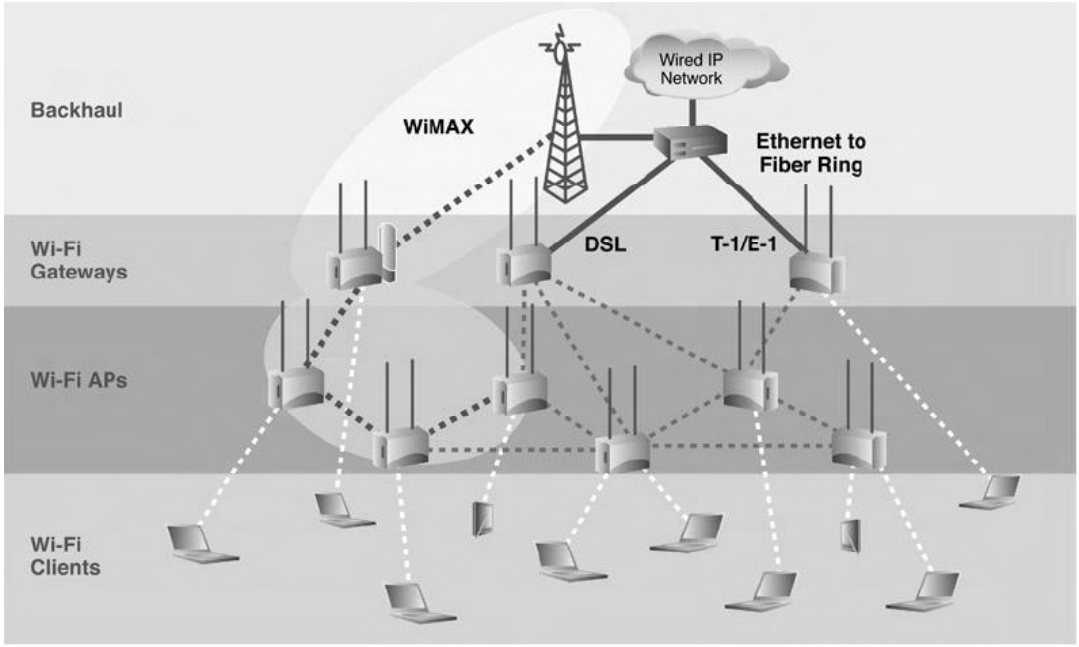


Figure 3. WiMAX and Wi-Fi integration



in those technologies associated with the LAN, MAN and WAN network architectures.

6.1 Wi-Fi

WiFi is an ultra high-speed wireless Internet connection usually available within a radius of a few hundred feet. By setting up multiple access points or ``hot spots," schools can make wireless Internet access available throughout their buildings. Wi-Fi is a high-speed data networking technology that provides an "over the air" interface between a wireless client and a base station or access point. This technology forms the Institute of Electrical and Electronics Engineers (IEEE) 802.11 standard for wireless LANs. The essential subsets of this standard include 802.11a, 802.11b and 802.11g. 802.11a broadcasts in the 5GHz spectrum (the GoE has not yet approved the use of this standard for public), is designed for coverage of up to a 50 meter radius and delivers up to 54Mbps. It operates on twelve channels thus minimizing interference. 802.11b and g both broadcast in the 2.4GHz spectrum and are designed for coverage of up to 125 meters. 802.11b delivers up to 11Mbps while 802.11g delivers up to 54Mbps. Both 802.11b and g operate on three channels thus potential sources of interference must be considered at the design stage. Mesh networks (proposed 802.11n standard) can also considered as a possible architectural implementation of Wi-Fi. Currently, due to proprietary systems and lack of standardized methods of implementation and security, pre-standard Wi-Fi networks suffer from interoperability issues and connecting can be confusing for users.

6.2 Wi-MAX

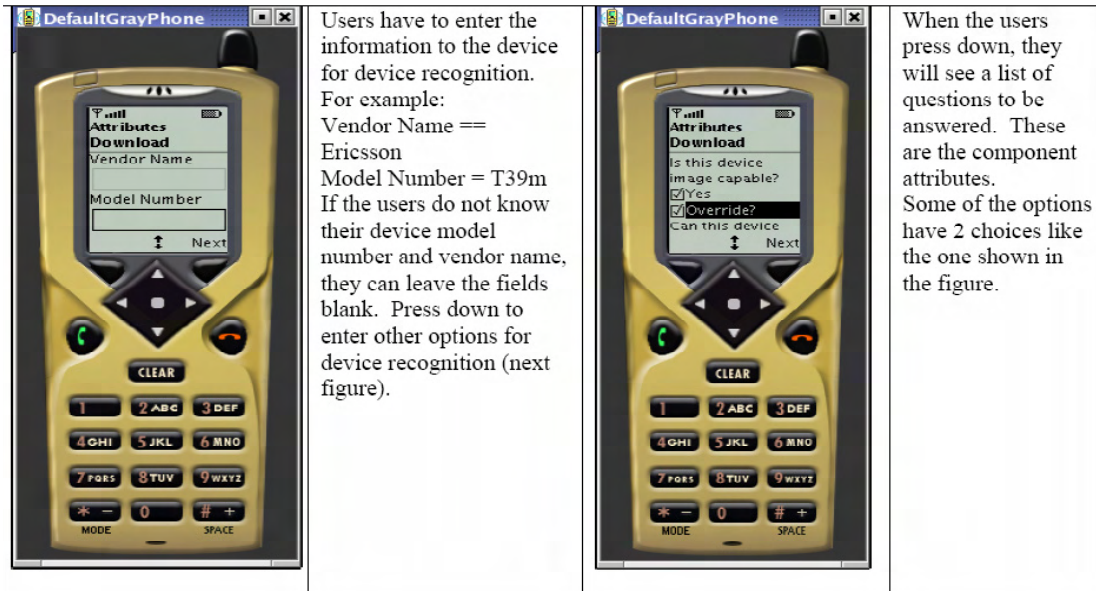
As a rapidly maturing broadband delivery technology, technology currently supports the ratified 802.16-2004 WiMAX specification for fixed broadband internet access. This specification also provides for an interoperable, carrier-class solution for the "last mile". A wireless alternative to cable modem, xDSL, Tx/Ex, OC-x and similar wireline technologies, WiMAX provides:

- A wireless backhaul alternative
- Point-to-multipoint technology
- Non Line of Sight Connectivity (NLoS)
- Service Area up to 30 miles (typically 4-6 mile radius)
- Up to 280Mbps per base station

IEEE802.16-2004 is the current standard for WiMAX to provide a carrier-class fixed wireless solution. The newly announced IEEE802.16e WiMAX standard will enable support for both fixed and mobile broadband wireless access. Intel recently released its codenamed "Rosedale" WiMAX chip-set, the first of its kind. Currently, communication product manufacturers are incorporating this chipset in their WiMAX products. WiMAX supports guaranteed service levels and Quality of Service (QoS) and provides built-in security using PKI certificates to authenticate base stations and link level encryption of data (3DES). While WiMAX pre-standard products have been on the market for some time the first 802.16 standard-compliant products are emerging at time of writing. The 802.16e amendment to the basic specification will enable a base station to support both fixed and mobile broadband wireless access. This aims to fill the gap between high data-rate WLANs and high mobility cellular WANs. Robust bandwidth, QoS support and low cost make WiMAX an excellent choice for long distance backhaul applications such as linking Wi-Fi enabled mesh networks and hotpots to the Internet. Figure 3 shows that by leveraging the Wi-Fi and WiMAX open broadband wireless standards and implementing Mobile Computing architectures, Broadband can quickly and cost-effectively be deployed to areas not currently served. This can be done with little or no disruption to existing infrastructures at AUC. Standards-compliant WLANs and proprietary Wi-Fi Mesh infrastructures are currently being proliferated widely throughout the world. Currently, standards-compliant WiMAX products are available to provide NLoS backhaul solutions for these

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. PDA devices with System Running



local networks. WiMAX subscriber stations are also available to provide access to customer premises such as schools and other educational institutions. Next generation WiMAX and Wi-Fi enabled clients (due in 2007), will be able to directly access legacy proprietary Wi-Fi Mesh networks, newly deployed standards-compliant Wi-Fi Mesh networks and WiMAX networks. New forms of wireless protocols are overcoming challenges of terrain, infrastructure and finance. It is proposed that, by leveraging open broadband wireless standards, like Wireless Fidelity (Wi-Fi*) and Worldwide Interoperability for Microwave Access (WiMAX*), and implementing Mobile Computing architecture, it is now possible to make dramatic strides in this direction.

Wireless m-learning is being implemented throughout the developed and developing world. While much good work has been done, there is a sense that much remains to be achieved. While this may be disappointing, it may also be seen as an unprecedented opportunity. Due to technological advances, particularly in the areas of Wi-Fi, WiMAX, Mobile Computing and Voice over IP (VoIP), there are opportunities to reach audiences that heretofore were too remote or for other reasons beyond the digital divide. While technology is now allowing us to access this ever-increasing audience for m-learning, we must ensure that we can reach them in a truly holistic sense. This depends on us understanding the geographic, economic and cultural complexities of those we want to embrace. WLAN will provide users with wireless high-speed access while it gives service providers great opportunities to stimulate growth in the wireless data market. Broad coverage and easy access will be critical factors for acceptance and growth of public WLANs by: -Adding online/offline functionality which allows users to work anywhere, any time, and to work without disruption, even when network connections are interrupted, and -Providing intelligent roaming capabilities when moving from hotspot to hotspot which means users won't waste time reconnecting or lose critical data because of dropped connections, -Enabling the flexibility to access data and applications on various computing devices, whether they are laptops, desktops, handhelds or servers, , and -Tuning applications to conserve power and maximize performance delivers fast application execution and lets users worry less about running out of battery power.

7 PROPOSED TESTBED FOR PILOTS/TRIAL

The overall structure of the testbed is shown in figure 1. The diagram shows that through appropriate converters we can serve other target devices such as Nokia, Ericsson, Palm, and Tablet PCs. The diagram shows that the WiFi wireless access co-exit with other traditional accesses methods such as RMI, RPC, Web Services, and agent-based access methods. Mobile contents such as sound files, historical images, images of Arabic alphabet and words, audio Arabic books, arts and architecture images and videos, paint images are all stored in appropriate XML format on a separate disk from the specific platform data. The server exposes all its services as web service format. Mobile devices use Jini (Java-based) or UPnP (.net-based) service discovery protocols to discover and connect to such services. Ever client stores its specific device profile locally on the device and activate its client side Jini or UPnP to discover available services. All mobile contents are NFS mounted and exposed to Jini and/or UPnP. The diagram shows the actions of the converter in generating specific devices course notes (e.g. Nokia 3360 course notes, Palm Tungsten C course notes, etc.). The ALI234 & HIST222 pilots/trials are implemented on the testbed. Audio files and Illustration images were collected for the two courses as described in section 4 above. Five specific experimental pilots/trials are initially identified for testing on the testbed:

- An experiment for testing collaborate learning in ALI234: During a "translation" lesson, each student had to download an English sentence and translate it to Arabic and broadcast his answer to everybody else through the server. Students collaborate in reaching the most accurate right answer. This activity takes place as an exercise among the students participating

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 5. The Experimental System in Action with a Number of Snapshots

in the class, and could be supervised by the faculty through the server.

- Three blending experiments. One uses SMS and WebCT. WebCT required an attending mentor to observe the navigation through the "Nahwoe" lesson in the ALI234 class, where a student might run into a stembling block during his review of the WebCt material. Faculties and Mentors can help if they are contacted through SMS. A stempled student can send a quick SMS question that will ease his way through the rest of the WebCT material.

  Another uses Java quizzes and WebCt. WebCT provides quizzes to students attending to E-learning sessions on WebCT. If a student is away from a WebCT session and would like to check his comprehension of the material, he can download a java quiz to his color screen phone and run it. This pilot is prepared for "Pharaonic Egypt" lesson in HIST222 course.
- The third involves using focused iMate module with WebCT. Focused modules on an i-Mate-PDA are to complement the WebCT material of a course list HIST222. The simplest example of this would be a PowerPoint-type presentation, though there are several more 'game-like' activities, all of which can now be created by any tutor using our authoring system. But the content is only the first part. There are three other aspects of creating PDA-ready learning materials that we can help with: installing the materials to the device, helping the student navigate around your materials, and tracking and reporting on use.
- The fifth experiment involves a field trip to a historical Egyptian "The Citadel" in HIST222. Each student produces his own documentary about the trip using his own self-produced photos, videos. Each captures his thoughts, views, and behavior in the documentary.

The overall purpose of such experimental pilots and trials is to investigate how to integrate set of learning tools, and how the tools will be adapted for the context of its use.

8 TESTBED AND PILOT/TRAIL IMPLEMENTATION

We started with two pilots/trials (see snapshots on figure 4 and 5) , one for an ALI Arabic course ALI234 using Nokia 9500 mobile communicator, and the other for a History course HIST222 using i-Mate PDA. We have implemented our own Pilot system for two target i-Mate PDA and Nokia 9500 Cellular phone devices. For two target courses: ALI234 and HIST222. Both Audio steaming and downloads are provided. One on a Nokia 9500 Cellular phone and the other on an i-Mate PDA. Screenshots of the two Pilots/Trails are shown in figures 4 &5 of some of the experiments described in the previous section. With the RealOne Player for Mobile Devices you will be able to both class and exercise material .Access your favorite RealAudio and RealVideo files made available for your mobile device. Drag-and-drop your music files from your RealOne Player on your PC to your mobile device (PocketPC only). Access news, sports and entertainment updates. Download content to your mobile device including music videos, travel guidance, auto reviews and much more. The aim of the m-learning project is to develop a prototype system to provide modules of learning via portable technologies which are already owned by, or readily accessible for, the majority of AUC students. The prototype will seek to attract students to learning and assist in the development and achievement of life long learning objectives.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

9 CONCLUSION

What is the right blend? In this paper we have designed and implemented a testbed for trying several pilots and trials to measure the effectiveness of various blends in ALI234 and HIST222. These two courses are out of the School of Humanities. Our intension is to spend the summer 2007 preparing and testing more courses from the schools of Business and Engineering & Sciences. The pilots/trials have answered a number of important questions such as how m-learning has improved access to assessment, learning materials and learning resources, explored the potential for collaborative learning, for increasing students' appreciation of their own learning process, and for consolidation of learning, guided students to see a subject differently than they would have done without the use of mobile devices, identified learners' needs for just-in-time knowledge, explored whether the time and task management facilities of mobile devices can help students to manage their studies, investigated how wireless/mobile technology alters attitudes, patterns of study, and communication activity among students, explored the potential for individualized mobile learning, revision material tailored to the needs of the individual, allowed interactive screens encouraging art gallery visitors to respond to the art on view, sat a set of innovative games, materials and activities which will motivate reluctant young learners, provided user-friendly m-portal that is powerful and empowering, and encourages active participation by its users, enhanced interactivity and cooperation while preserving the traditional advantages of face-to-face encounters, investigated how self-produced videos, made with a digital video camera and later viewed on handheld mobile computers, can support informal learning, provided video and still images giving additional context for art gallery works on display, and opportunities to listen to an expert talk about details of a work, with the details simultaneously highlighted on the screen.

REFERENCES

1. www.fctel.uncc.edu/
2. mlis.state.md.us/2001rs/budget_docs/All/Operating/R00B27_-_Coppin_State_College.pdf
3. www.hbsp.harvard.edu/b01/en/elearning/elearning
4. www.oclc.org/news/publications/newsletters/oclc/2004/265/downloads/duke.pdf
5. www.accesselearning.net/
6. www.usnews.com/usnews/edu/elearning/directory/elearn1a_1305.htm
7. www.cs.ucd.ie/staff/mbertolotto/home/SelectedPublications.htm
8. www.washingtonpost.com/wp-dyn/content/article/2005/06/15/AR2005061502144.html

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e•Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Exploring virtual worlds as an extension to classroom learning

James Braman
Department of Computer and Information Sciences
Towson University, Towson MD 21252
jbraman@towson.edu

Andrew Jinman
Communications Learning Division
Twofour Studios,
Estover, Plymouth, PL6 7RG
andrew.jinman@twofour.co.uk

Goran Trajkovski
Information Technology Programs
South University
709 Mall Blvd, Savannah, GA 31406
gtrajkovski@southuniversity.edu

Abstract Through the exploration of virtual worlds such as Second Life we present our preliminary investigation into its use as an extension to a college classroom environment. Through our project we intend to transcend traditional web-based and classroom learning platforms. We have begun to experiment within these spaces in a collaborative effort to investigate new and innovative ways for various educational and teacher-student interactions.

Keywords Virtual classroom, online learning, Second Life, virtual worlds

1 INTRODUCTION

Transcending the traditional distance learning methods used by many educational institutions, we present the use of virtual worlds as a medium for learning and information dissemination. New evolving technological advancements have made it possible for the creation of such computer mediated spaces that not only allow communication in a traditional sense of human-computer interaction, but also through the use of virtual spaces where one can visualize their interactions over space and time and in three dimensions. For our research we have chosen the virtual world of Second Life® as a platform for educational use. Other educational paradigms currently employed by many online or hybrid courses can also be extended into virtual spaces. Virtual worlds are becoming a novel new reality for the establishment of communities, social interactions, and for information diffusion. From these elements we are beginning to utilize Second Life as a platform for education and are exploring its usage as an experimental space in a college level setting.

Web-based distance learning has been in use for many years incorporating many aspects of the classroom such as lectures, videos, slide shows, online collaboration, and voice capabilities to name a few. Some have even used virtual spaces as an enhancement for online learning to create virtual laboratory sessions, recreating a "safe environment" for simulation and enhancing student learning and to provide a "Satisfying laboratory experience" [1]. Others see these spaces as an extension for communication of both verbal and nonverbal channels of expression over tradition web based methods [2]. Second Life has also been discussed as a new form of narrative through various interactions [3]. As distance and web based learning paradigms evolve and become even more popular in the college curriculum, new innovative ways of online learning should be explored. Computer games and other multi-user virtual environments (MUVEs) have been used in the classroom and studied in the past; with our project we hope to create a persistent virtual learning space that serves as an extension to classroom learning [4]. In recent years educators have been looking to the "Metaverse" for advancing class room activities beyond the capacity of the traditional classroom environments. Neal Stephenson's 1992 novel Snow Crash introduced the term 'Metaverse' a term given to the work of completely immersive 3D environments where people interact, socialize, work and are entertained [5]. We contend to establish a presence in Second

Life where research can be conducted in real time where we can also extend classroom learning through non-traditional, graphical distance learning strategies.

We see virtual environments as a platform where educators and students can interact on a more personal and natural level compared to tradition web interface environments. Using an immersive virtual space such as Second Life we can enhance the student-teacher relationship by allowing students to interact with visual representations of the instructors. This in turn relieves many anxieties created within the traditional classroom setting, providing a comfortable environment for student expression, creation and interaction that can enhance learning. Student-teacher relationships are often enhanced by certain nonverbal visual cues observed by students in the classroom setting which can be virtually simulated through user's avatars [6].

2 WHAT IS SECOND LIFE?

Second Life, created by San Francisco-based Linden Lab® is one of the leading 3D virtual worlds. Originally released in 2003, it has evolved into a popular interactive space with now over nine million users worldwide connected via the internet [7] [8]. The Second Life platform is a freely downloadable application, and with an account and internet connection, users can log into the environment to explore the many rich and interactive spaces. Each user is represented by an avatar in which one may interact with the environment, objects or other users. Each avatar is completely customizable allowing users the opportunity to give his or her self a unique physical appearance. The "residents" of Second Life can visit many areas by simply "teleporting" or even by walking around. There are also several modes of communication either through standard chat, private messaging, through note cards (small in-world text files) and by general notifications. One controls their avatar with the mouse and arrow keys to travel and to interact with other avatars or objects. Creating objects is also easy as Second Life has a built-in modeling tool allowing for highly customizable objects and shapes to be created. The immersive capabilities of Second Life make it a useful tool as sound is spatially projected, while also using environmental conditions such as clouds, natural elements and land masses help to give it a realistic look and feel. The Linden dollar is the standard currency of Second Life which can be traded for real US currency. Many users own successful virtual businesses in many categories ranging from virtual real estate to clothing stores, scripting services to photography. The exchange rate for the Linden dollar changes daily. At the time of this writing, L$268 is equivalent to about one US dollar [9]. This idea that virtual money can equate to real income coupled with the amount of users has inspired many real world companies to advertise and market virtual goods to in-world residents. Many users learn how to use the Second Life modeling tools to create elaborate objects constructed from many basic shapes such as furniture, clothing, art work, homes and much more. Almost any object that one can think of in real life can often be modeled in some form and incorporated into the Second life environment. On the Second Life web page it is even advertised that "Second Life is a 3D online digital world imagined and created by its residents" [10]. The content of the world has emerged from the interaction and collaboration of the in-world residents.

3 CURRENT WORK

After locating to a new area within the main grid of Second Life, we have begun to build several experimental spaces as part of our endeavours in studying the educational uses of virtual and immersive realties. We have named our joint project "The Extension Project". As its name implies, we are using this project to see how a virtual world can be used as an extension to the tradition brick-and-mortar classroom. We currently have four buildings as part of this project: The Extension building, Virtual Display Center, Conference Center and Scripting Center.

The Extension Building (see figure 1) is constructed in a unique style that utilizes video stream capabilities for its meeting center where live video feeds can be viewed from within the room. As one of the original buildings, it serves as the focal point for the virtual campus providing information for current and future in-world projects. The Virtual Display Center (see figure 2) is reserved for a display gallery for both student and teacher research posters that are displayed in a poster session format along the walls of the building. We also have a three level conference center and experimental lecture hall (see Figure 3). Within this space we currently house several art projects from various students along with virtual artwork displays from multiple members of the Second Life Art Community. Our largest Lecture hall is within this building where we can seat up to thirty users. The newest building is currently being setup for students to learn the Linden Scripting Language (LSL) and other building tools for experimentation purposes. We also hope to use this space in a more collaborative effort to engage other professionals to work together in scripting efforts within Second Life.

We are also currently integrating objects in-world with the 2D web for further enhancement. Objects can be touched by avatars directly which will launch the user's web browser to specific websites. Other research has shown that Second Life can be successfully intergraded into other external platforms such as blogging tools, databases and web browsers as well as being used to produce other media like machinima [3]. Through our work we plan on integrating various technologies to enhance education and collaboration in these virtual spaces.

4 INITIAL EXPERIMENT

As part of our initial experimental description within Second Life, various activities were planned in order to provide a basis for further exploration in a virtual world [11]. For this initial examination in a college classroom setting, a group of students from Towson University's computers and creativity course used the platform as part of a lab activity. The course content focuses on the creative use of technology and the expression of ideas through various digital mediums.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 1. Extension Building [http://slurl.com/secondlife/Dreyfus/224/192]

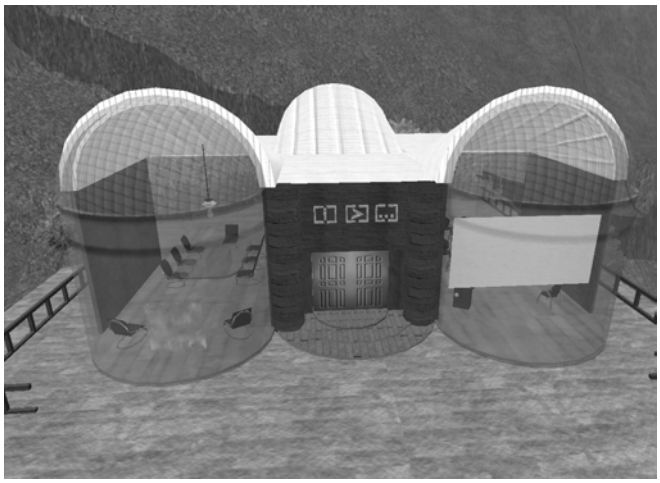
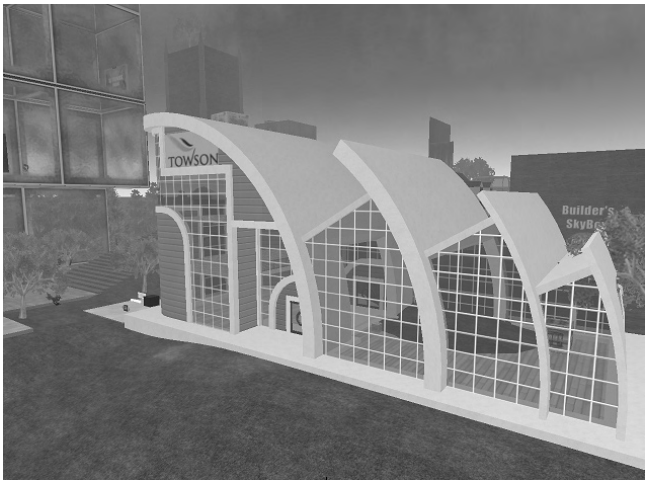


Figure 2. Virtual Display Center [http://slurl.com/secondlife/Dreyfus/224/192]

Figure 3. Conference Center [http://slurl.com/secondlife/Dreyfus/224/192]

Figure 4. Scripting Center [http://slurl.com/secondlife/Dreyfus/224/192]



Second Life is a beneficial and naturally creative computer mediated space whose usage falls nicely into the scope of the course objective. From this initial experiment with both the instructor and class situated in Second Life it was our goal to collect some preliminary feed back to further refine our future goals for the Extension Project.

5 A VIRTUAL FIELD TRIP

As part of a class lesson we chose to spend the lab period immersed in Second Life. This initial testing in a classroom setting would serves as our basis for improvement for future labs. We also wanted to see how students would understand the environment and how quickly they could learn to use the controls. We also wanted to explore different areas of interest from the student's perspective as well as bring our class objective to the virtual space as part of a hybrid class experience (part online and face-to face interaction). Even though we all were located in the same room physically we were also all situated in a virtual space locally. First students we directed to the Second Life web page where they needed to setup a basic account and to create their avatar. Students generally had no difficulties with this process and were able to successfully pick a name for their avatars, and also choose one of the several default avatar appearances currently available. Next students were able to log in with their new account at which time students were initially asked to proceed to the area designated for the Extension Project. Students were initially disoriented and took several minutes to understand how to manoeuvre the controls; many students because frustrated until they could move about the area. The camera controls that allowed the student to "see" what the avatar was observing proved most difficult to master as it took time to learn the mouse controls and to understand how they related. The students were then asked to first explore the open space within the Virtual Display Center to look at several of the posters that were produced by previous student projects from the Cognitive Agency and Robotics Lab at Towson University. These posters were photographed from the real physical lab and uploaded and converted into posters for the Display Center in Second Life. Next students entered the lecture hall were they were surprised to see that several of their semester projects had been uploaded into the building and were being displayed in a poster session format. They were then asked to have their avatars sit down in the newly placed Lecture hall seating where several slides of various pictures were shown to demonstrate to the student the ability for slide show presentations from within a virtual setting. The next part of the tour within the Extension Project area was for the students to view and explore the various 3D art works that were purchased or donated from in-world

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

artists to demonstrate the applicability of creativity and art from within Second Life. Students were also asked to explore the Extension Building and to experiment with the video stream capabilities within the building to view a previously recorded video of a tour of the Extension Project. The next exercise was a field trip to several areas within the Second Life world. First we visited a recreation of Dublin, Ireland where students were first introduced to virtual stores and accessories that could be purchased for their avatars. At this point students became more inclined to edit and experiment with aspects of their physical appearance. Next we visited a dance club where students conceptualized how scripts could animate their avatars to dance and interact in different ways while listening to live streaming audio and chatting with other residents. We also visited a large public sandbox area where the creation and experimentation of objects are encouraged in a communal space. In sandbox areas users are free to experiment with objects and scripts without having to own large amounts of virtual land themselves. In the sandbox the students were shown how to create and manipulate simple object and how simple textures could be applied to create more sophisticated items. Some students were even able to insert pre-made scripts into their newly created object to make them rotate. This exercise was interesting to students and allowed them to understand how content was actually added to the Second Life world. At one point another user who was in the area as the class created an object with a script that was spawning objects automatically along with particle and smoke effects which soon caused a generally slowing of the simulation. This slowing down and delay in an area is often referred to as "Lag". Due to the increasing lag of the sandbox area, interaction and communication with the class became almost impossible and one students computer actually had to be restarted. Some students were able to transport themselves back to the Extension Project Land before they became stuck and their computers became unresponsive. It was at this point that students became frustrated but also seemed intrigued when they made the connection that Second Life is truly an interactive space where the actions of individuals can truly affect the other users. The students were warned before logging into Second Life that their actions could be viewed by the class and by other users logged into the environment. They were asked to respect other users while engaged in the class activity.

6 DIFFICULTIES

It is important to note that there were some technical difficulties encountered in this initial experiment. Initially the program was installed in the entire lab where it was tested and passed initial testing. Approximately three weeks had passed before it was actually used, when it was discovered that Linden Lab released an important update to the software at which point the program needed to be re-installed. Due to security settings, only lab technicians could re-install the program which did take extra time. Once implemented in the classroom we noted that some computers (including the main instructor workstation) would not run the program and adjustments had to be made before the start of the class. Also students noted that due to the hardware limitations, the quality of the graphics were generally poorer then average. Bandwidth did not seem to be an issue but has yet to be tested in a much larger classroom or with multiple classrooms where all the students would be logged on concurrently.

7 PRELIMINARY REACTIONS

At the end of the lab period students were asked several questions as part of a survey to better understand their conceptualization of the space and how to improve Second Life usage in the future. Students noted that navigation and learning to use the controls was the hardest aspect of using Second life. It has been noted in other studies that many virtual based realities are often hard to learn requiring varying degrees of time for individuals to adjust [12]. Many students did note that they found the social aspects of Second Life to be the most interesting and enjoyed interacting with different people and visiting different places. Several students said that they were surprised that so many people use Second Life and that one can actually make real money from selling virtual products. From the class, only two students had previously heard of or used Second life. Everyone in the class did agree that Second Life could be successfully used for educational purposes and several agreed that it seemed a logical next step in education. Since this particular class revolves around the creative use of technology, students also agreed that the viewing of 3D artwork was beneficial from within virtual spaces since it can be globally shared and discussed.

8 FUTURE WORK

We currently are working towards integrating more classroom projects, research and in class activities through Second Life. One immediate goal is to involve more students from a larger class to gather more concrete statistical data. We are also currently looking for ways to expand our land area to expand our capabilities and to attract more users. Larger "sand box" areas are being constructed so students can further practice modeling and scripting without being constricted within our current space. We have been looking into various new sources for funding the project and for the possibility of hiring in-world landscapers and builders to enhance the appearance of our space to give it a more natural look. Currently our work has formed several ties with other Second Life academic projects and professionals. With collaboration between The Extension Project, Twofour Communications Learning Division and Leicester University's Media Zoo, run by Professor Gilly Salmon, we aim to coordinate between our research models in Second Life. Our second phase of the project also involves studying the immersive affects of virtual spaces and its potential to extend the quality of traditional web-based learning by creating more social and collaborative spaces. As part of continuing work in virtual spaces we are asking the following questions and are looking to apply the results to improve distance education:

1. How can we develop a rich interactive virtual environment for distance learning and collaboration?

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

2. What are efficient and affective methods to get both instructors and students to embrace virtual environments for learning?
3. How can virtual objects create environments through emergent interactions and be used as extensions to the real classroom?
4. How has the culture of such virtual worlds developed? How do they impact the real world?
5. How can virtual worlds be used to study cyber-anthropology and other social phenomena?
6. How do we perceive ourselves and conceptualize our interactions in a computer mediated space?

We are particularly interested to see if all aspects of 2D web based distance learning can be restructured to fit in a 3D environment. Also as an extension of human-computer interaction and human perception we will be looking into the possible break down of anxieties in normal human-human interactions in real life interactions as it relates to avatar-avatar interactions. Will certain social phobias also be present in 3D virtual spaces? Through our continued work within Second Life we hope to establish the viability of such virtual worlds as an extension to tradition classroom structures.

9 CONCLUSIONS

In this paper we have presented various possibilities of using Second Life as an extension to classroom based learning. We have also described our current and future work within Second Life along with various goals related to aspects of human-computer interaction, immersion, social interactions and perception in these virtual spaces. We have also discussed our preliminary results from our first investigation in a real classroom environment. We see the future of education and online interactions revolving around new mediums and metaphorical constructs. Second Life as a platform for experimental learning is a step towards higher education in virtual mediums. We will be further pursuing virtual mediums such as Second Life to extend classroom learning and class collaboration.

REFERENCES

1. Leitner, L. Cane, J. (2005) A Virtual Laboratory Environment for Online IT Education. SIGITE '05 Oct 20-22. Newark, New Jersey.
2. Robbins, S. (2007) A Futurist's View of Second Life Education: A Developing Taxonomy of Digital Spaces. 2007 Second Life Education Workshop. Chicago
3. Bakioglu, B. (2007) Collaborative Story-telling: Performing the Narrative of the Griefer. 2007 Second Life Education Workshop. Chicago
4. Delwiche, A. (2006). Massively multiplayer online games (MMOs) in the new media classroom. Educational Technology & Society, 9 (3), 160¬172.
5. Stephenson, Neal (1992) Snow Crash, Bantam Books, USA.
6. Richmond. V, McCroskey, J. (2004) Nonverbal Behavior in Interpersonal Relations 5th edition. Pearson Education.
7. Rymaszewski, M. Au, W. Wallace, M. Winters, C. Ondrejka, C. Batstone-Cunningham. (2007) Second Life the Official guide. Wiley Publishing, Inc. Indianapolis, Indiana.
8. Second Life (2007) User Statistics. Accessed August 31, 2007. http://www.secondlife.com
9. LindeX: Market Data. (2007) Second Life Econimic Statistics. Accessed September 6, 2007.
10. SecondLife.com. (2007) Home page. http://www.secondlife.com
11. Braman, J. Jinman, A.Trajkovski, G. (2007) Towards a Virtual Classroom: Investigating Education in Synthetic Worlds. The AAAI Fall Symposium. Emergent Agents and Socialities: Social and Organizational Aspects of Intelligence. Arlington, VA.
12. Boyd, C. (1995) Human and Machine Dimensions of 3D interfaces for Virtual Environments. CHI '95 Mosaic of Creativity. Doctoral Consortium. ACM.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Social exchanges theory applied on a web-based learning community

Maximira Carlota da Silva André
Sérgio Roberto Kieling Franco

Federal University of Rio Grande do Sul, South of Brazil

Abstract This paper addresses to the knowledge building process of English as a foreign language materialized in a web-based learning community and based on the theory of social and cognitive exchanges proposed by Piaget [1973]. The community named *English for Presentations* is carried out in Paltalk. As for the Social Exchanges theory, it has provided us the necessary background for the understanding and mapping of the exchanges that i) take place in the three levels: 1) rhythms, 2) regulations and 3) cooperative constructions between students and the knowledge-object they are focused in, and that ii) enhance and promote learning. We shall present and discuss in the following pages some of the exchanges (in the three categories) carried out in EFP community. We shall also discuss the impact the exchanges methodology has had in the construction and promotion of individual and collective knowledge. The participants are from five continents, varying in age, sex, culture, religion and professional background, but with common interests or needs, which vary from learning the language itself (English), and learning how to make presentations in this language. Both goals were fully achieved by the orientation of the social exchanges theory and methodology used.

Keywords Web-based learning community; social exchanges theory; knowledge building; foreign language acquisition

1 INTRODUCTION

Languages constitute a crucial part in the human nature, and education plays an important role in the construction of this knowledge. Considering this, Education should be understood in its dimension of plurality, singularity, permanent transformation and integration, through the numerous and successive processes of interaction in any environment it takes place: live or virtual environments under the mediation of internet. An educational experience that does not consider our condition of uneasy beings, owners of a big plurality and individuality, inserted in circumstances that are altered and organised chaotically and accidentally every single moment needs to be reviewed in order to liven up, facilitate and promote the integration of the different participants, of the unusual, and unexpected.

Having this in mind, and considering that knowledge building does not occur from the unique experience of the objects, nor from an innate pre-formed construction in the human being, but from successive constructions along with constant elaborations of new structures, as proposes the theory of Jean Piaget [1], we move to promote and analyse the exchanges between participants on a web-based learning community named English for Presentations. The Community, created at Paltalk, aims to provide students a richer environment of exchanges and in doing so, enhance their learning. We shall map the social and cognitive exchanges participants offer and accept or refuse towards the object (English and English for Presentations) and towards the other participants, as well, so we can analyse exchanges formation, occurrence, and development, as well as, their implication in the knowledge building process in web-based learning communities.

1.1 The English for Presentations (EFP) web-based learning community

Aiming to offer a broader range of interactions to the learners who 1) seek to practise and improve their skills in English, and who 2) find it difficult to achieve it in their home countries due to the status of a foreign language, that is, not used as a means of communication among people in their daily lives, the present web-based learning community is created. English for Presentations is to be understood as a community for both: help students to improve and master the language, and acquiring the skills of a presentation in English, and mostly, a place in which students can freely produce, exchange, and cooperate.

Piaget´s theory claims that the richer the exposure and contact to the object one aims to learn the faster it may be achieved. In this sense, a web-based learning community, open to the 5 continents, seems to fit the condition for the acquisition of a foreign language and is set in order to provide these people the exercise conditions. The interactions range in levels and forms, but as we'll see all of them constitute a succession, being part of the same thing: exchanges.

1.2 EFP planning & management

The community was planned to take place on a web-based environment which is able to offer students all kinds of interactions, that is, written and oral comprehension and production. Besides writing, reading, speaking and listening, students could choose to interact in public or privately with their peers and with the teacher. A further interaction the tool offers is related to the possibility to express feelings through emoticons and other popular visual signs we are used to find and use in communication tools. Students could choose to: raise their hands and get into the queue to speak, send whispers or private messages to other participants, and get engaged in a conversation to exchange something, accept or refuse an offer of exchange, as well as observe only.

As this community was fully open and never had its door *closed* or *locked*, it would be visited by newcomers all the time and in all meetings, whenever the community was in action, from beginning to the end. If, by one hand, this helped making it a live exchanges experience, on the other hand it brought us some difficulties we could overcome by a team work. Some regular students who were already used to offer some help in managing the community, were officially invited to take part of it as co-administrators. This showed to enhance their commitment towards EFP community and gave the community a stronger status of collaborative work. Some would welcome the newcomers whilst others would inform them about the rules of the community, still others would keep an eye on the troublemakers, and some would promote knowledge exchanges.

The facilities offered by Paltalk along with the team work described above provided participants a good exchange experience for 4 months as it was concerned to last. In addition to the 18 regular members, it would count with distinct groups of 15-20 people on a daily basis. Some of these people used to join EFP once a week, others occasionally, from times to times as their agendas allowed them to, but keeping the records of all classes as we moved on.

1.3 EFP rules & norms

Every social exchange as well as every cognitive exchange or mental operation must be based on some rules and norms in order to guarantee this intellectual exchange or cooperative operation. As for the rules we have the following equation:

$$(rx = sy) + (sy = ty) + (ty = vx) = (rx = vx) \quad (1)$$

Where: 1) a participant x acts upon a participant y. This action constitutes rx (or y proposes a ry upon x); 2) y (or x) demonstrates a satisfaction (positive, negative or null) which we will call sy; 3) this satisfaction compels y towards x (or the inverse), constituting a debt named ty; 4) this debt or obligation constitutes a virtual value for x = vx (or vy for y).

The equilibrium conditions (always in relation to any qualitative exchange) are the following, then: 1) it is necessary that x and y have a values scale in common, making the evaluations of rx and vx for x comparable to the evaluations sy and ty por y, and 2) the achievement of the equation before and the second equation that follows:

$$(vx = ty) + (ty = ry) + (ry = sx) = (vx = sx) \quad (2)$$

The final equilibrium suggests therefore that w are able to alter the order of the two continuations (equations) as in:

$$rx = sx = tx = vy \text{ [For Equation I]} \quad (3)$$

And as in:

$$vy = tx = rx = sy \text{ [For Equation II]} \quad (4)$$

In the case of the intellectual exchanges, the terms and relations acquire the following meanings: 1) the participant x proposes an utterance rx (true or false in several degrees); 2) the participant y agrees (or not) constituting sy; 3) the agreement of y makes him/her continue the exchange between y and x, constituting ty; and, 4) the engagement of y attributes the utterance rx a value or validity vx (positive or negative), that is, makes it valuable (or not) concerning the future exchanges between the same participants.

Concerning the Equation I, we will find the following specifications: 1) the equality [rx = sy] means that x and y agree on the utterance given; 2) the equality [sy = ty] implies that y feels obliged to follow the utterance he/she had recognized as true, that is, x is able to maintain the utterance rx as a permanent value.

Concerning the norms, they are constituted by the exchange rules above and by the values of exchange first offered to students at EFP: a group of 14 lessons, which will be underlining their exchange values, that is, students will be exchanging doubts, ideas, utterances, examples, questionings, and satisfaction or not based on these values. The norms of the present learning community also attempts to propose students an environment of cooperation, against coercion or authoritative systems which imply a submission of the students to produce what and only what is expected them

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

to produce or repeat. Instead these systems, mostly found in educational systems, we make use of cooperation methodology demanding mutual respect and collaboration. These rules constitute the norms that regulate this community, and which promote students a rich experience of exchanges as we shall see later.

1.4 The social exchanges theory

According to Piaget´s theory [2] the balance or *equilibrium* concept, that is, the last stage of the construction of a new knowledge, only makes sense on an auto-regulation perspective, which evokes dialectical processes in turn. This is mostly due to the sequences of disequilibrium and auto-regulation in progress imply the intervention of the conflicts meant to be antagonists at the beginning, being overcome in the end by a reorganization process that constitutes the balanced synthesis. In all events that we may find an interaction between a citizen and his/her object, even if this citizen is several citizens, and this the object is the same for all of them, knowledge does not grow or is built on an innate perspective, nor from the objects only, but *in the interaction between them*. It is therefore this interaction that allows the objective exteriorization and the reflexive internalization – in the sense of the continuous and dynamic regulations, that is, knowledge building.

The thing is that all these processes that underline knowledge building depend at the same time on the maturation and on the external or educative transmission, obeying a constant development. That is why language is not learned or acquired in blocks, but through a regular succession of operations and levels of organization. The social relations in cooperation constitute groups of operations, as every logical operation which is made by the individual upon his/her exterior world, being the rules of such groups of operations that define the final equilibrium form.

The theory Piaget proposes is helpful for us to understand how knowledge is built in terms of the constant and regular exchanges the participants do with the object under their interest, and by proposing a full methodology of analysis of these processes by the equations and explanations presented earlier we are able to identify these exchanges implications not only in the final results, but also during the processes of exchanges themselves.

Exchanges formation and development: Rhythm, Regulation and Cooperative Exchanges

These three different levels of exchanges are understood to take place as a succession of one another rather than 3 distinct operations with no relation among them. According to this theory all exchanges play an important role for the knowledge building process and should not be discarded as less or more relevant than the others.

As for the first movement we will easily find the formation of rhythms of exchanges, in which participants establish how the interactions will develop among them in the community. The formation and development of exchanges at this level is to be found during the whole experience as it is what maintains and constitutes a learning community free of coercion or authoritativeness.

The next movement, named regulations, are concerned to the exchanges that show some reciprocity of thought between the exchangers or a common operation they make together, but which have not showed to achieve a final equilibrium in the data or that we are able to observe. Actually, intellectual operations do seem to be very rare to be observed in spontaneous exchanges environments unless an authority demands participants to evidence it somehow.

The latter, understood as cooperative exchanges, would be the exchanges we can observe an intellectual operation being processed in reciprocity or by a common operation; the final movement of regulations we can say. As we have already mentioned they are rare to be observed in spontaneous situations of exchange but not impossible. We shall see next some examples of the three kinds that took place in EFP learning community.

2 EXCHANGES AT A RHYTHMS FORMATION LEVEL

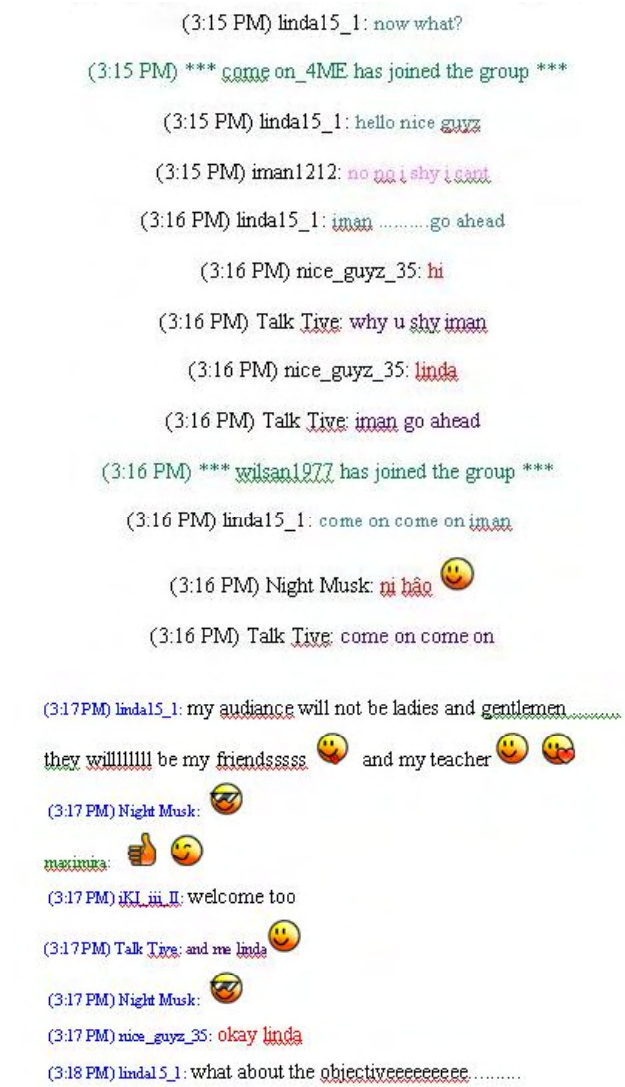
Based on the following interactions or exchanges, for the formation of the rhythms of the community we are able to check participants appealing for the interaction. W can also observe the basic conditions for the succession of these exchanges to deeper levels as it is clear the evidence of:

2.1 Spontaneous interest/action from the participants

We are able to see this in their speeches:

a. The agreement in following the values of exchange offered at the beginning by the community EFP and of the norms of the community, as well. Students seem to have mutual respect and collaboration.
b. The acceptance of participating in a web-based community despite the differences they may find concerning religion, culture, language, age, professional background, and son on.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics



But if the formation of these rhythms of exchange is the first and a constant movement, it is not the only one. We shall see now some other exchanges, at the *regulations level* now.

3 EXCHANGES AT REGULATIONS LEVEL

As for the interrelations exist therefore 2 extreme kinds of relation: the coercion, implying and authoritative and submission system, and the cooperation system which implies the equality of rights or autonomy, as well as, the reciprocity among different people.

The exchanges at this level may occur in both systems, but in the former, these regulations do not achieve what we understand by equilibrium of the intellectual operation, for they can only reach an approximate conservation of the operations that take place. When they take place in the second system, of cooperation, mutual respect and reciprocity, they may reach the final form of equilibrium, constituting truly intellectual cooperative exchanges. The difference therefore between one and another is only how explicit they are presented to us or how well we are able to observe them.

It is interesting to observe that these regulations exchanges base most of the exchanges interactions. Maybe it happens due to the difficulty in observing the final level of exchange, and also due to the use of coercion in most live or web-based educational experiences. It is important to attempt that the norms that are necessary to achieve any exchange based on cooperation must be far from being misunderstood with coercion. One provides the basis of the exchanges, the values that are considered to be relevant and of interest of the participants, the latter does not allow operations at the level of cooperation at all.

The following in an exchange at a regulation level about what is considered to be *signalling* in a presentation.

We see the formation of a new rhythm (in yellow) with the proposition of Saraswat, and how it becomes to be a regulation exchange (in green), by the proposal of other values some participants (Linda, Davvis) offer. Then (in blue), we observe the occurrence of an intellectual exchange as it shows all steps of the equation I provided earlier. Let´s see.

François and Saraswat have a common scale of intellectual values, they are able to understand each other and the meaning of the words they use. Based on the Equation 1 we find:

- the equalily (rx = sy), for François and Saraswat seem to agree on the offer proposed,
- the equalily (sy = ty), for François shows he recognizes as a valid value the offer of Saraswat;
- the equality (ty = vx), in the moment the utterance *rx* receives a validity of being conserved , that is, x may keep rx identically;
- the equality (vx = ty) which means the value of rx is to be always recognized by y;
- the equality (ty = ry) in which y obligation is seen to be applied by him on a new utterance: ry.

4 EXCHANGES AT INTELLECTUAL COOPERATION LEVEL

Below we will find some intellectual cooperative exchanges when Imran and Abu semm to work together for solving proposed by Khadija. Alher, Haann and Darkraw offer several values, which seem to be accepted by Khadija, what comes to constitute all the levels and conditions of a cooperative exchange. Let us see.

maximira: What about SIMPLIFYING?
(2:46 PM) Khadija48: can you explain it to me?
(2:46 PM) darkraw_2005: amplifying
(2:46 PM) imrankhan920: by leaving out useless words
(2:46 PM) Abu Sami: by Using graphics
(2:46 PM) Khadija48: ohhh
(2:46 PM) imrankhan920: or redundant words
(2:47 PM) alher_2: *paraphrase*
(2:47 PM) hanan_10_1: in other words
(2:47 PM) hanan_10_1: make simpler or easier or reduce in complxity or extent
(2:47 PM) darkraw_2005: illustrating
(2:47 PM) Khadija48: I see,, thanksss
(2:47 PM) hanan_10_1: nyou´re wlcome
maximira: ANY MORE EXAMPLES?
(2:48 PM) hanan_10_1: personaly speaking
(2:48 PM) mohammad_kei: honestly ...
(2:48 PM) hanan_10_1: to speak the truth
(2:48 PM) mohammad_kei: to tell u the truth
(2:48 PM) Abu Sami: How does this simplfy?
maximira: "To put it simply" [abu]
(2:48 PM) Abu Sami: thanks teacher,,
(2:48 PM) mohammad_kei: clap for mohammad and Abu!
(2:49 PM) darkraw_2005: clap
maximira:

Next we shall see an exchange of all levels, but mostly of cooperation, all because of one value that was offered by one of the participants. Let us look at it closely.

maximira: Jane Eyre has to make a presentation this coming Monday for her class.. at school
maximira: and what´s the topic Jane?

(3:18 PM) Valentino_nl: oh i see, nice
(3:19 PM) aanhalima: solar system
(3:19 PM) aanhalima: ?
(3:19 PM) Valentino_nl: aha ok, the solar system

(3:19 PM) Teachers_pet1: jane, are you going to use PC and a projector?
maximira: good question teachers_pet!
(3:20 PM) Teachers_pet1: i see - the old fashion way
(3:20 PM) Valentino_nl: with PowerPoint
(3:20 PM) Teachers_pet1: ty max
(3:21 PM) Teachers_pet1: i would use a Powerpoint slids and a projector, too

(3:21 PM) jane eyre_1: solar system
(3:21 PM) jane eyre_1: topic is solar system

(3:22 PM) Teachers_pet1: suzi, i think you are enjoying this topic, aint you?
(3:22 PM) Teachers_pet1:
(3:22 PM) suzaann92002: yes i am ,Teachers

(3:22 PM) jane eyre_1: yes
(3:24 PM) jane eyre_1: i dont understand
(3:24 PM) jane eyre_1: pls repeat again
(3:24 PM) david_young: give us your objective jane
(3:24 PM) jane eyre_1: sorry
(3:25 PM) Valentino_nl: what is the presentation about Jane
(3:25 PM) david_young: solar system
(3:25 PM) david_young: the objective
(3:26 PM) david_young: your goal

maximira: Who has suggested this topic to you Jane?
maximira: your father. ok!
(3:27 PM) david_young: what do u want us to learn about your presentation
maximira: ok jane!!
maximira: now it is clear to me!
(3:28 PM) david_young: ok
(3:28 PM) david_young: i got it
(3:28 PM) david_young: i think so
(3:28 PM) jane eyre_1: yes

As we see the above exchanges can have the following configuration:

1) Utterance offered by Jane Eyre to the teacher
2) Socialization of the utterance to the group
3) Acceptance of the value by Valentino
4) Second presentation of the topic as an answer to Aanhalima
5) Confirmation by Valentino of the information given by Aanhalima
6) New value offered by Teachers_pet1 as an answer to Jane Eyre
7) Acceptance of Teachers_pet contribution
8) Inference of a (possible) oral return from Jane Eyre to Teacher´s_pet
9) Contribution of Valentino to Teacher´s_pet and explaantion concerning Jane Eyre's speech
10) Feeling of satisfaction of Teachers_pet for receiving a positive feedback to his previous offer
11) Agreement of Teacher´s_pet to the methodology chosen by Jane Eyre
12) Proposition of a new value between Teachers_pet and Suzaann
13) Jane Eyre expresses lack of understanding
14) David and Valentino repeat to Jane Eyre the proposed value
15) possible (oral) explaantion by Jane Eyre concerning her objectives
16) Acceptance of David for the explaantion offered by Jane Eyre

Such configuration makes it clearer for us to visualize the principles proposed by the theory of social Exchanges with the difference that we are not talking about exchanges between 2 individuals, but a group of. Exchanges like the previous ones in which we have different groups interacting orally and by written messages origiante other exchanges and some may lack in the formalization of the equations that we use to characterize the exchanges. The difficulty in paying attention to all oral utterances under development at the same time we go through written exchanges, as well, is something we need to learn how to deal with. The following chart presents the interactions in accordance to the principles of what constitute a cooperative exchange.

1º) the individual(s) x propose an utterance rx (true, false or null) [1,2, 9,15]

2º) the individual(s) y agree (or not) = sy; [3,4,5,6, 14, 17]

3º) this agreement (or the lack of it) from y makes them (x and y) to continue the exchanges, where we achieve ty; [7, 8 , 11, 16]

4º) the engagement of y attributes the utterance rx to be a value vx (positive ou negative), that is, that makes it a valid one (or not) in relation to the future exchanges among the same individuals [10;12]

The second condition for achieving the equilibrium, shown by the Equation 1:

$$(rx = sy) + (sy = ty) + (ty = vx) = (rx = vx)$$

is also achieved, in the exchanges between Jane Eyre and Teachers_pet: in (6) Teacher´s_pet signs the equality (rx + sy), recognizing the value offered by Jane Eyre. In (8) Jane Eyre answers Teacher´s_pet, leading him to achieve the second principle (sy = ty) in (11). Then, we must infer from the data the engagement and commitment born from the previous principles, what makes us complete the cooperation exchanges equation.

Let´s continue with one last cooperative exchange achieved based on the reciprocity of thoughts about how to make a good presentation and what length of time it should last.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

1)	(3:31 PM) david_young: maximir usually how long the presentation should be taken
2)	(3:32 PM) cmt_6: *Maximira, you want us try to start presentation*
3)	(3:32 PM) aanhalima: depends on the subject david
4)	(3:32 PM) Valentino_nl: and learning goals of the presentation maybe Maximira?
5)	(3:32 PM) cmt_6: *aboy soliar system* (3:32 PM) cmt_6: *about**
6)	(3:32 PM) david_young: I mean like the topic solar system
7)	(3:34 PM) gadget19: what is the subject of presentation you guys are talking about please
8)	(3:34 PM) ojossoandores: maximira do you open the room everyday?
9)	(3:34 PM) tieuxi: please tell me ,we present free topic or we have assigned topic
10)	(3:34 PM) Valentino_nl: the solar system – is the topic
11)	(3:34 PM) suzaann92002: almost everyday ojos exept suday and Wednesday
12)	(3:34 PM) gadget19: I see. Thanx Vale..
13)	(3:34 PM) aanhalima: I agree Vale. it also depends on the goals
14)	(3:35 PM) gadget19: except
15)	(3:35 PM) ojossoandores: oh thank suzaann
16)	(3:35 PM) david_young: so it can vary.
17)	(3:35 PM) Valentino_nl: sometimes you are not given much time
18)	(3:35 PM) david_young: I see guys.. thanks
19)	(3:35 PM) suzaann92002: ty gadget
20)	(3:35 PM) suzaann92002: maxi you´ll give us only 5 min, isn´t it?
21)	(3:34 PM) aanhalima:
22)	(3:35 PM) Valentino_nl:

The value offered by David is fully accepted by Anahalima and Valentino, to whom David answers back. The exchange continues then between Anahalima and Valentino, where the first accepts the offer given by her partner, Valentino. David is in the movement and expresses his acceptance towards the offers given. Valentino adds a new value for which Anahalima totally agrees. The second condition is also fully reached. Let's see:

- The equality (rx = sy) exists and is expressed in the proposition of Valentino, accepted as true by Anahalima and David.
- The equality (sy = ty) is seen in Anahalima utterance that accepts Valentino´s value and complements it
- The equality (ty = vx) makes the offer rx be conserved. David and Anahalima seem to keep the rx (offered by Valentino) as a true and permanent one.

5 DISCUSSION

Contrarily to what one may think role-to-role community networks do not consist only of like-minded people – such as BMW 2002 fanciers – or of people with complementary roles – such as violinists and cellists. People from different cultures, with different social and professional background, different nationalities and age may find it interesting to interact all together and exchange their views, experiences and knowledge about something or a common need like how to make presentations in English, for instance. To see how others around the world come to understand and go along (or not) with what you think and say about something is a pretty fascinating experience in life, and web-based learning communities can offer people that. While such communities are abundant now, they are flourishing on the Internet and will become even more abundant as the Internet's capabilities develop.

At English for Presentations Community students subscribed without previously knowing who they would interact with, they could not know even if there would be someone from their country or people sharing similar experiences and expectations. They subscribed because of the subject they would deal with: English for presentations. As our proposal was based on a constructivist approach of learning, students could participate in many different ways. They could choose to produce oral language or just listen to others, some would only produce written statements, and still others would just sit and observe during the first weeks. This could vary of course. Ones who were most used to listen to others only, would eventually speak to the group or increase their participation asking others questions about the subject. But most of them would attend our meetings at least 3-4 times a week, for an hour or an hour and a half each, which is far more time than it usually takes place in non-virtual educational classrooms. We believe this is due to the theory and methodology we have chosen: constructivism. According to it students must be given an active role in the play. In other words, they should not do only what the teacher had planned them to do, at the moment and on the way the teacher had planned too, but what they feel like doing at a given moment or what they are able to do. Listening exercis-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

es, grammar drills, written dialogues work and are currently planned in constructivist classes as well. The difference relies on "when" and "how" it happens. Teachers and tasks should be fluid and flexible in order to fit this approach that relies mostly on students needs. Well, we have attempted to this and let things go to check how effectively learning could take place. We provided students some guidance and all assistance they needed to. Having achieved positive and meaningful results in a real short length of time (about three months) we headed to write this paper about what constituted our meetings and how exchanges (interactions) took place despite all differences they had, and difficulties all of us faced when experiencing something for the first time.

Participants then, from different parts of the world, as the community was opened to Europe, Asia, America, Africa and New Zeeland continents would find their way to exchange and build knowledge individual and collectively through the theory of social exchanges proposed by Piaget. During the experience we could clearly see the rhythms, regulations and cooperation exchanges that took place. Students' active participation was crucial to this learning we achieved composed by all the three levels of exchange we are aware of according to this theory.

We have seen some examples of their efforts on the previous pages to build knowledge by sharing/exchanging their personal and professional experiences no matter what it took them. We have also seen in this experience native speakers and English teachers cooperating to beginners, to people from poor countries, and with a poor English. All of this because all are seen and understood in EFP as equal human beings before being British, German, French, African, Brazilian Arabic, whatever. Everyone was fully respected in every sense and was given the right to choose and decide upon when and how he/she wanted to participate more effectively or on a different way. This kind of rule helps to make people feel confident and respected. And when people feel like that our chances to have active participants who will enrich the experience and make it full of different rhythms, regulations and cooperation that is, different kinds of exchanges, increases nearly 100%.

As for the *web-based learning community efficiency* we are able to say it can work and provide excellent results when it is taken into consideration a constructivist theoretical background as the one we have adopted here and when it is respected the nature of the human being, I mean, when students are given the chance to decide and act upon their learning, choose and test things/structures/etc. they have been presented to on their rhythm (not on the rhythm of the teacher) and making use of the strategies they have chosen to best fit their needs (not the teacher).

Related to *foreign languages acquisition* we are convinced it can be fully increased on web-based learning communities due to the richness of interactions and exchanges we find in this kind of environment, and no other. Though there is a lot of room to be explored yet, we can say foreign languages acquisition has much to win when opened to a more interdisciplinary discussion. In other words, when it considers working along with other subject matters complementarily, in opposition to the instructional paradigm of the instruction by itself.

As for knowledge building, we have seen its focus relies upon learning instead of teaching. In EFP it was our main goal to promote this student-active learning process and we did. Each student has received proper guidance whilst among an infinitum amount of possibilities of things to go through, study and master during their personalized constructions and learning.

6 SOME PRELIMINARY CONCLUSIONS

The data collected on the web-based learning community named 'English for Presentations' has shown us it is possible to establish a cooperation educational system based on mutual respect and cooperation instead of coercion, still found in most educational experiences. We can achieve that by giving a student the chance to be an active participant. Even though there must be rules to follow so we can reach this cooperative level, a crucial feature is about the understanding we must have that every participant is able to: think and reason, choose and test their utterances (when and how they feel like doing it). They must feel respected in every sense, and not treated like robots with no feelings or empty boxes which do not have much inside.

The results have shown that they interact and exchange at least 5 times more than in traditional teaching environments increasing their learning speed, as well. The theory proposed by Piaget of Social Exchanges is able to map and cover the exchanges that take place among participants on a web-based learning community and that has shown to enhance the development of languages acquisition. In other words, the more one acts the more he/she will learn. In this experience students do act and participate actively. Their participations were mapped and understood in terms of exchanges, occurring in three basic levels: rhythms, regulations and cooperative exchanges. While in traditional teaching environments it is hard to establish and find cooperation, mostly due to the excessive authoritarianism of the system as a whole, we have proved on this web-based learning community it is possible to reach. We have reached cooperative exchanges based on mutual respect. And we have reached this by giving people a chance to try it out. It seems to us no matter what nationality or culture one has, he or she will always be a human being in first place – always in need of receiving, giving or exchanging something.

The last but not the least, online relationships and online communities have developed their own strength and dynamics. We have verified that participants in this online group have strong interpersonal feelings of belonging, being wanted, obtaining important resources, and having a shared identity. This is a time for individuals and their networks. Autonomy and opportunity seem to rule today's community game.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

To sum up, more studies are certainly welcome in this field. Though Social Exchanges Theory applied on Web-based learning communities to foreign languages acquisition has much to be explored yet, as mentioned previously we are convinced that it certainly can help accelerating learning processes related to the acquisition of foreign languages.

REFERENCES

1. Jean Piaget.(1973) Sociological Studies.
2. Castells, Manuel. (1998) The Information Age: Economy, Society, and Culture. (Three volumes), Oxford: Blackwell.
3. Cressey, Donald R. (1960) The Theory of Differential Association: An Introduction. Social Problems 8, Nr 1. English-Lueck, Jan (1998): "Technology and Social Change: The Effects on Family and Community." Paper presented at the COSSA Congressional Seminar. Available at: http://www.sjsu.edu/depts/anthropology/svcp/CossaP.htm
4. Foucault, Michel. (1984) Space, Knowledge and Power. In: Rabinow, Paul (ed.): The Foucault Reader, New York: Pantheon Books, pp. 239-56.
5. Geser, Hans. (2001) On the Functions and Consequences of the Internet for Social Movements and Voluntary Associations. Zurich, (2nd release). Available at: http://socio.ch/movpar/t_hgeser3a.htm
6. Ling, R. (2000c) Direct and Mediated Interaction in the Maintenance of Social Relationships. In Sloane, A. and van Rijn, F. (eds.): Home Informatics and Telematics: Information, Technology and Society. Kluwer, Boston, pp. 61 - 86.
7. Wellman, Barry. (2001) Physical Place and Cyber Place: The Rise of Personalized Networking. International Journal of Urban and Regional Research, 25. Available at: http://www.chass.utoronto.ca/~wellman/publications/individualism/ijurr3a1.htm
8. Garcia, L. Costa. A.C.R., Franco, S. (2004) Virtual Learning Communities based on Piaget´s Social Interaction Theory and supported by peer-to-peer networks. Available at: http://rocha.ucpel.tche.br/valores/values-for-vlc.pdf

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Experiential Learning: "Teaching citizenship through database case study application, the hurricane Katrina disaster experience"

Barbara Nicolai

Purdue University Calumet
Hammond, IN 46323
bnicolai@calumet.purdue.edu

Abstract The Summer Institute for Teaching Excellence (SITE) met for the first time in June 2006. The goal was to designate faculty as Fellows representing Purdue University Calumet in teaching excellence. The main objective of the summer institute was to provide new teaching theories and an environment to share instructional methods among the Fellows. The concept of experiential learning was presented as a method to expose students as an active participant to citizenship.

One of the learning outcomes of the retreat was to develop an experiential learning module which would be integrated into a selected course. As a professor of database modeling and implementation, a "real-life" case study was chosen to test the new experiential learning concepts. Following the aftermath of the Hurricane Katrina Disaster, a case study was developed by Professors Nicolai and Winer (Nicolai, Winer 2005). This case study was chosen to be used as the research base for a database application system. A student team of 13 members developed a fully functional prototype of over 50 screens, supporting all the necessary services of the affected population after a national disaster. The ideal of citizenship as an experiential learning experience involved students in a real world crisis situation and applied their skills to fill a technology need that "makes a difference."

1 SUMMER INSTITUTE FOR TEACHING EXCELLENCE

The Summer Institute for Teaching Excellence (SITE) met for the first time in June 2006. The goal was to designate faculty as Fellows representing Purdue University Calumet in teaching excellence. The main objective of the summer institute was to provide new teaching theories and an environment to share instructional methods among the Fellows. The concept of experiential learning was presented as a method to expose students as an active participant to citizenship.

2 EXPERIENTIAL LEARNING

What is experiential learning?

"Experiential Learning refers to learning activities that involve the learner in the process of active engagement with and critical reflection about phenomena being studied." (NSEE, 2006)

Experiential learning has come to mean two different types of learning:

1. learning by yourself and
2. experiential education [experiential learning through programs structured by others] (Smith, 2003).

1. Experiential learning by yourself

Learning from experience by yourself might be called "nature's way of learning". It is "education that occurs as a direct participation in the events of life" (Houle, 1980, p. 221, quoted in Smith, 2003). It includes learning that comes about through reflection on everyday experiences. Experiential learning by yourself is also known as "informal education" and includes learning that is organized by learners themselves.

2. Experiential education

(Experiential learning through programs & activities structured by others)

Principles of experiential learning are used to design of experiential education programs. Emphasis is placed on the nature of participants' subjective experiences.

An experiential educator's role is to organize and facilitate direct experiences of phenomenon under the assumption that this will lead to genuine (meaningful and long-lasting)

learning. This often also requires preparatory and reflective exercises.

Experiential education is often contrasted with didactic education, in which the teacher's role is to "give" information/knowledge to student and to prescribe study/learning exercises which have "information/knowledge transmission" as the main goal.

3 DESIGNING EXPERIENTIAL LEARNING COURSES & COURSE COMPONENTS

The seminar presented by Dr. Lee Artz, Center for Instructional Excellence, Purdue University Calumet, addressed Designing Experiential Learning Courses & Course Components. Following the National Society for Experiential Education Standards of Practice, the primary course components for experiential learning were: Intent, Planning, Authenticity, Reflection, Training, Monitoring, Assessment and Acknowledgment. (Artz, 2007)

- Intent identifies experience, site and experiential process that will be demonstrated, used or attained in the course overview. The experiential activities must also meet course objectives.
- Planning demonstrates a clear connection between the experience on site, course content, objectives, schedule, and student achievement.
- Authenticity involves reciprocity of goals between course and site that includes experience and site participation in course design as exemplified by a Site Proposal/Agreement.
- Reflection involves scheduled and structured student reflection activities/assignments.
- Training includes orientation and training for students to gain necessary skills/knowledge for site experience/work as exemplified by Forms and Checklists.
- Monitoring identifies visits/consults or other regular faculty/site mentoring and continuous improvement or collaboration with an appropriate agency representative.
- Acknowledgement includes end-of-course evaluations of students ratings of course objectives, site personnel surveys of student contribution and faculty report/review of course.
- Assessment is evaluated through grades earned with the achievement of course objectives. (Artz, 2007)

4 THE CASE STUDY: THE GOVERNMENT AGENCY

Organizational Background

The Township Trustee's Office (Township) is a part of State and local government's programs to supply indigent population with needed services. They are responsible for providing life's basic support needs to the community's poor and indigent people from all walks of life who find themselves in need of temporary assistance.

The Township support to indigent individuals and families is called poor relief or general assistance. This type of general assistance requires the largest segment of resources, both human and financial. State statutes clearly define what public aid can and must be provided and the process to assess those same services. General assistance services are provided through purchase orders (vouchers) for education, food, shelter (rent or mortgage), utilities (gas, electric, water), medical to include prescription drugs, burial, household, clothing, furniture and transportation.

A Township office has four major goals:

- Service to the indigent in as an efficient manner as possible
- Work within the guidelines of state statutes
- Promote positive human relations
- Work as a team to supply services to the indigent population with respect and concern. (Nicolai & Winer, 2005)

Supporting a Natural Disaster

This case provides a view into existing government programs that supply indigent population with all the necessary services to provide basic life support to people who find themselves in the need of temporary assistance. These governmental agencies historically have minimal if non-existent information technology solutions to support their services. The purpose of this case is to research a typical local government program and discover the best technological solution that can be sustained in adverse conditions as experienced in a national disaster, either natural or man-made. Some of the primary issues discussed in this case are: immediate availability of information and the disbursement of that information, organization of a command center outside the normal office boundaries, connection to State and Federal agencies, triage of medical assistance, plan for disposal of life support items, water, food, bedding, rescue, and an electronic network system that can survive in the emergency environment.

We are sitting in the aftermath of one of the most deadly national disasters facing our nation since September 11, 2001. Hurricane Katrina has shown that, as a nation, we are not yet prepared for the monumental challenge of providing the basic life support to the victims of a national disaster. The government has established the Federal Emergency Management Agency (FEMA) to ensure that the effected indigent people get the basic necessities such as food, shelter, clothing and services to keep in contact with their friends and family members. With Hurricane Katrina as an example of what could happen at the local agency level, the new system must not only be able to serve the indigent client base with everyday assistance and but also be able to provide an emergence response to tens of thousands of natural disaster victims. After a natural disaster or similar devastating situation, the number of people requiring assistance grows exponentially. The current system would never be able to help facilitate the rapid growth of a large client base. The current system requires a client to interface with eight different departments before general assistance may be given. This is the first part of the new system that will need to be corrected. The system

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

currently lacks an emergency plan that expedites meeting immediate needs as basic shelter, food, clothing, and medical services for a large number of people in a short amount of time.

The improved Information System model should correct at least two major problems with the outdated system currently in use: 1) a lack of an efficient way to process an application and be able to manage a natural disaster client base, 2) An absence of an evacuation and response plan readily available. The first goal would be to stream line the eight different processes an applicant must complete before be given any form of assistance. This would include having all functions computerized instead of hard copies being sent from department to department until the eight steps are completed. This change would alleviate some of the redundancies and their associated costs. There would also have to be controls added to the system that help manage the budget and create an emergency coffer to help in times of mass crisis. The second goal would be needed to help expedite the response time for large groups of people in a larger support area who are in need of assistance. The system would have to manage their own client base as well as link to the client base systems in their perimeter. The system application would have to address issues of food, water, shelter, transportation, medical and equipment inventories. Having this functionality would provide the information that would assist the rescue and disaster coordinators to disperse the available resources based on the local area with the greatest need. This system would also have to interface with the federally funded system controlled by FEMA to help enhance and restore inventories for the local regions in need. This application could help the transition from local government agency support to federal government agency support by providing an informational and tracking system of the victims of the disaster. (Nicolai et al., 2005)

5 COURSE OBJECTIVES

The learning objectives for CIS 354 Relational and Object-Oriented Database Modeling are:

This course discusses the functions and components of database management systems and the role of databases in the Systems Development Life Cycle. Both relational and object oriented database techniques are discussed. Data modeling tools presented include enterprise models, entity-relationship diagrams, the data dictionary, object diagrams, and normalization techniques. Also, the role and function of the Database Administrator are addressed. This course provides the student with a variety of methodologies for database analysis and design, with documentation heavily emphasized. The assigned case study will be a team approach with much of the analysis and design accomplished using the Relational and Object-Oriented Database Methodology.

Students will be able to:

- Comprehend the basic concepts and definitions of the database environment.
- Interpret and evaluate the database development.
- Synthesize information regarding the Entity-Relationship Model of the database design.
- Comprehend the enhanced E-R Model and Business rules.
- Diagram object-oriented modeling structured approaches
- Design a logical database model following a relational model.
- Comprehend a physical database design following a relational model.
- Apply the principles of use-case analysis to develop use-case realizations that model the collaborations between instances of the identified classes to an unstructured, real-life problem.

6 EXPERIENTIAL APPLIED RESEARCH

In introducing this case study, "Indigent Population Database Project. Can our existing government agencies provide life support to our indigent population and survive a national disaster?", the students were exposed to a complex and yet meaningful business problem in order to formulate the learning outcomes for this course. Instead of beginning the course with reluctance to learning, the students enthusiastically researched the current crisis of Hurricane Katrina. The teaching challenge was to integrate the foundational business platform of the local governmental agency that supports the indigent population into the natural disaster module. It was here where basic concepts of the Enterprise Model, Business Functions to Entity Matrix and Preliminary Entity Relationship Model acted as the foundation for the natural disaster analysis. Slowly the students were able to understand how the relational database model created the structure for the final application product. The learning experience was energized by the students seeing the connection of what they were able to produce and how they could affect a large population in need with their technology expertise. In order to complete the design of this system, the students had to do vast research into the Hurricane Katrina Disaster. Research came from news media, internet resources, interviews with the American Red Cross, and presentations by the IT consultant of the township government office. The students had to not only analyze the information gathered but categorize the information as pertinent to the final design. This experience met fully with the standards and practices as presented in the seminar attended by the professor. (Artz, 2007)

7 STUDENT LEARNING OUTCOMES

With each of the learning objectives, the case study is providing a real-world application of their skills:

- Comprehend the basic concepts and definitions of the database environment.
- The students are learning how to apply the concepts of entities and attributes to the natural disaster issue. They are discovering how to walk through a data flow within the business of the government agency and decipher the needed functions that would drive the business.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

- Interpret and evaluate the database development.
- After producing the Data Flow Diagram, the students take the next step in identifying what kind of data needs to be organized in categories of information. This carries the student to the next step in the logical model where entities are identified and relationships are analyzed.
- Synthesize information regarding the Enterprise Data Model and Entity-Relationship Model of the database design.
- Some of the entities discovered for the agency support indigent populations are CLIENT, SERVICE, STAFF, SUPPLIES, VENDOR, LOCATION.
- Comprehend the enhanced E-R Model and Business rules.
- The ERD is enhanced to include super-type/sub-type relationships. The entity SERVICE can be divided into Medical Service, Housing Service, Transportation Service, Education Service, Job Assistance Service and Supplies Service. This case study allow for detailed development using the Enhance E-R model, enabling the student to practice the EERD concepts to the modeled design of the application.
- The remaining learning objectives were taught in the remaining weeks of the semester and were not address in this paper.
- Diagram object-oriented modeling structured approaches
- Design a logical database model following a relational model.
- Comprehend a physical database design following a relational model.
- Apply the principles of use-case analysis to develop use-case realizations that model the collaborations between instances of the identified classes to an unstructured, real-life problem.

As a teaching professor, it is always my goal to transfer my knowledge base to the students of my class. Presenting this case study whose subject matter touches such a vast majority of lives, provides me with an invaluable insight into seeing citizenship in the classroom at work. Learning, though important and necessary, must be built with a topic that will promote excitement and acknowledge the students as a stakeholder of the education experience. In providing a real-world experience for the students to use as a tool for developing their skill set, engages the student in an atmosphere of "meaningful work."

8 SUMMARY

"Experiential Learning refers to learning activities that involve the learner in the process of active engagement with and critical reflection about phenomena being studied." (NSEE, 2006)

The student experiential learning experience was so successful that the case study was repeated for the Fall 2006 semester. The results of teaching this same case study to the modeling class produced even a more professional and higher skill set from the students. My colleague, Professor Sam Liles, who teaches senior network design, applied the same concept and topic of a Hurricane Katrina Disaster case study in the Fall 2006 semester. The product of the networking class was a 260 page wireless networking solution that could be used with the database model. As a result of this combined effort of network and database solution, we plan to apply for a grant with a working prototype that could be applied to a national scale.

As a result of the student satisfaction and successful learning outcomes, the Computer Information Technology Department is adapting experiential learning into the various tracks of the programs to provide applied learning and research in the curriculum.

ACKNOWLEDGEMENTS

Acknowledgements are given to the CIS 354 Relational and Object-Oriented Database Modeling Class who are using this case study in the Fall 2005 semester. It is through their excellent efforts that the research regarding Hurricane Katrina and its effects on indigent populations was provided in this case study. They are currently designing the new application system for a typical township servicing an indigent population.

The members of the CIS 354 Fall 2005 course are as follows: Brandon Adamson, Daniel Aguinaga, Sofia Angelova, Nathaniel Bennett, Sylvia Clayton, Irvin Cross, Parama Danoesoebroto, Artay Dates, Jonah Fields, Adam Harsha, Zbigniew Jasek, Kuldoon Khraisat, Brian Luyster, David Miloshoff, Erik Novotny, Michelle Orr, Nenad Petrovic, Eric Piscione, Adam Radloff, Elton Roberts Jr., David Wright.

The members of the ITS 360 Database Distributed Application Architecture and Design Fall 2006 course are as follows: Brian Bilow, Matthew Corban. Shaun Cruz, Omorodion Steve Eguasa, Jennifer Gibson, Sumaiya Jabeeen, Robert Mathias, Evelyn Miramontes, Christie Olson, Katheryn Puntillo, Edwin Trinidad, Shannon Zaronia

REFERENCES

Artz, Lee Dr. (2007). Designing Experiential Learning Courses & Course Components, Center for Instructional Excellence, Purdue University Calumet, April 9, 2007.

Associated Press (2005, September 10) FEMA nixes debit card plan. Concerns about staffing, claims legitimacy are reasons. Post Tribune, p. A7

CIDRAP (Center for Infectious Disease Research & Policy). (2005, September 2). Hurricane Katrina sparks fears of disease outbreaks. Retrieved September 9, 2005, from http://www.cidrap.umn.edu/cidrap/content/fs/food-disease/news/sep0205hurricane.html

DeNoon, Daniel (2005, September 01) Katrina's Aftermath: Health Situation Worsens

Specter of Disease—Physical Health—Enfolds Hurricane Disaster Area

Retreived September 9, 2005 from http://mywebmd.com/content/Article/111/109883.htm

Lussier, Charles (2005, September 06). Schools expect unprecedented numbers

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Retrieved September 8, 2005 from http://www.2theadvocate.com/stories/090605/new_numbers001.shtml

National Center for Missing & Exploited Childern (2005, September 08)

Help Identify People Missing as a Result of Hurricane Katrina. Retrieved September 08, 2005 from http://www.missingkids.com/missingkids/servlet/PageServlet?LanguageCountry=en_US&PageId=2077

Nicolai, B., Winer, C. (2005). A Case Study: Indigent Population Database Project. Can our existing government agencies provide life support to our indigent population and survive a national disaster?, Purdue University Calumet, Hammond, IN.

NSEE, National Society for Experiential Education, Journal March 2006, http://www.nsee.org/

Ricketts, M., & Willis, J. (2002). The power of experiential learning. http://www.teambuildingguru.com

Saddington, A. (n.d.). What is Experiential Learning? http://www.el.uct.ac.za/

Smith, M. K. (2001). David A. Kolb on experiential learning. the encyclopedia of informal education, http://www.infed.org/b-explrn.htm

Smith, M. K. (2003). Introduction to informal education. the encyclopedia of informal education,http://www.infed.org/i-intro.htm

WNBC.COM (NBC Television New York City, NY). (2005, September 7). HoursPassed Before FEMA Chief Asked For Volunteers. Retrieved September 7, 2005, from http://www.wnbc.com/news/4887230/detail.html

WOAI.COM (WOAI-AM Radio San Antonio, TX). (2005, September 5). Katrina Victim Separated From Family. Retrieved September 9, 2005, from http://www.woai.com/news/local/story.aspx?content_id=EFF63751

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

i·Society 2007

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Reducing instructor workload in online classes

Joy Colwell
Carl Jenks
Shoji Nakayama
Organizational Leadership & Supervision
Purdue University

Abstract This paper discusses techniques which instructors in online classes can use to help manage their time. Using the Quality matters rubric and an extensive syllabus, and instructor can incorporate provisions which will reduce students questions and issues. Instructors can also use a syllabus test, exemplars of student work, and policies on participation and missed work.

Keywords online learning, course management, distance education

1 INTRODUCTION

The objective of this paper is to provide field-tested techniques that improve efficiency and save the instructor valuable time in the management of online classes. These techniques can be applied to both online and traditional classes, but they are most valuable when used in managing online classes. Not only are these techniques valuable to the instructor, but they are also valuable to students in providing structure in areas that they feel are somewhat ambiguous and ill defined in many online course structures.

It is a given that online classes require much more detailed instruction on course structure and course policies than do most traditional courses. Since a course syllabus is supposed to be the primary source of course information for students, the instructor must pay particularly close attention to everything that goes into this document. All contingencies must be covered in advance and addressed in the syllabus. In terms of necessary elements, the Quality Matters (QM) project has promulgated a course rubric for quality in online courses, which includes several standards that affect the syllabus and information on course structure and policies. [1] The QM rubric encourages navigational instructions, a statement introducing the student to the course and to the structure of the student learning, netiquette expectations with regard to discussions and email communication, minimum technology requirements, minimum student skills, and, if applicable, prerequisite knowledge in the discipline. In addition, the rubric also requires learning objectives, clear and understandable grading policies, the number of graded items and their weight, email and virus information, last date to drop the course, server availability, testing information, software necessary to use the course information, FAQs, policies on withdrawal for nonparticipation, American Disability Act (ADA) information on accommodations, academic honesty policies, civility codes, campus emergency procedures, etc. The amount of information that must be front-loaded is overwhelming to many students.

The point of listing all these necessary elements of a syllabus is that, with all this detailed course information, students rarely read it all! It is too long and boring for most students, with the average online syllabus being two to three times longer than a traditional class syllabus. Based on the authors' experience, it appears that students would rather email the instructor for an answer to their questions instead of reading the posted course information. Moreover, since there is so much material, students tend to forget it soon after reading it. This creates two problems: email headaches for the instructor, and lack of familiarity with the course website and course structure on the students' part. An unwillingness to look for specific information means that many questions a student has will be emailed to the instructor, because students do not want to "waste" time looking for course information. Students have been taught to ask the instructor for answers in the traditional classes and that carries over to the online classes. Unfortunately, if students don't read the posted course information, its only value is justification for the instructor's actions later concerning grading decisions.

2 DISCUSSION

In order to overcome some of the students' resistance to obtaining information on their own, the authors offer some suggestions via the course syllabus. In designing a syllabus, the authors suggest that online instructors should pay particularly close attention to the following: a test over the syllabus (which can be an assessment delivered through the course management system), summaries of course information, exemplars of assignments, and policy statements on non-

participation, missed tests, challenged test questions, and instructor response time.

2.1 Syllabus Test

Since the authors have instituted the policy of testing on course structure at the beginning of the course, it is evident that many students do not read the syllabus until after trying and failing to pass the Syllabus Test. The questions are difficult enough that a student cannot obtain 100% (required) without putting forth some effort to learn the requirements. Although some students might guess the right answers, it is nearly impossible without studying the syllabus. Moreover, it is possible to tell by the number of times it takes a student to successfully complete the pretest whether (and how thoroughly) the student has read the course information. In one extreme case, a student took 19 tries to complete the pretest (and it only contains 20 randomly selected questions!). However, the average number of tries is probably three.

The syllabus test serves several purposes. It acquaints the students with the course; it forces them to read the syllabus; it helps them practice navigating through the course; and it accustoms them to the testing procedure which will be used throughout the class. With practice on the syllabus test, most of the startup problems with testing are alleviated because the students are totally familiar with the online testing procedure.

Typical student questions about the online course were very basic:

- Where are the discussion forums?
- When will my quiz grade be posted?
- When is the first quiz?
- What is my grade based on?
- Where are the course materials?
- When are assignments due?
- What happens if I have problems while taking a quiz?
- Where are assignments located?
- What is the university rank of the instructor?
- What is my instructor's highest degree?
- What average on tests earns an A grade?

The importance of well-constructed syllabus questions is that instructors can significantly reduce the time required for answering course management questions. Thus, the pretest can reduce wasteful administrative emailing and allow more time to be spent on the subject matter of the class.

In some classes, students are required to get 100% on the syllabus test prior to the first quiz in order to get credit for the first quiz. In other classes, students will not be allowed to proceed with the course or may even be dropped from the course if they fail to successfully complete the test within a reasonable period of time.

In all online classes, students can take the syllabus test as many times as necessary to get 100%. Of course, the more familiar a student becomes with the testing mechanics on the syllabus test; the fewer the questions and problems on subsequent tests.

2.2 Summaries of Course Information

Another helpful technique is using summaries of course information as much as possible and in as many places as possible. Regardless of how clearly the information is stated in a syllabus, it becomes lost in the maze of information presented there. As a result, students expect to have everything summarized in multiple easy to find locations. The announcements section is a good place to summarize as well as emails to each student. Redundancy of important deadline dates is an absolute necessity. In fact, if the important deadline dates are not repeatedly emphasized in one way or another, even the top students will miss some of the requirements.

2.3 Exemplars of Assignments

Another technique which the authors have found particularly helpful is the use of sample student work or exemplars of assignments. If this information is available for student to review, instructors will tend to receive far fewer questions about what is expected in an assignment. One of the authors has posted rubrics for assignments, and this does not necessarily reduce the email about assignment requirements. However, samples of student submissions showing the format and all requirements significantly reduce student questions on this issue far more effectively than any other approach.

2.4 Policies on nonparticipation

In nearly all online classes, teams are a necessity in order to make cases or projects meaningful. However, students are generally reluctant to become involved with a team unless there is penalty for nonparticipation. Therefore, the syllabus must contain a mechanism for getting the student involved from the first day of class and provide specific penalties for not doing so. Penalties seem to work better than incentives. Therefore, students are required to send a business resume to the instructor with emergency telephone numbers and email addresses as soon as the class begins. Shortly thereafter, the students must introduce themselves to their team members with the same resume. More importantly, anyone not meeting the initial requirements of a completing the syllabus test, contacting the instructor, and introducing himself or herself to the team by specific deadlines will be penalized a specific percentage of the final grade or even dropped from the course.

2.5 Missed Tests

The authors always include a policy on missing tests in their online courses. There is a 10% penalty per day on failing to make arrangements to make up an exam. This is designed to encourage students to make arrangements as soon as possible. This also helps to combat the student assumption that online courses are "work at your own pace", which seems to be a commonly held myth about online classes. This is true even when the syllabus outlines specific deadline dates for each requirement. Some students tend to view online classes

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

as having ambiguous requirements which they can interpret to their own benefit or schedule.

2.6 Challenged Test Questions

The syllabus must contain a procedure for students to challenge test questions. No matter how many times a test is given, no test is perfect. Design flaws show up with each new class, and there must be a viable appeal process that the students know and understand. If a student appeals a test question, the student should be required to present the entire question with all its alternatives within a specified period after the test. In addition, the student should be required to choose the "correct" answer and then justify that answer by some legitimate reference (which could be the text). If outside research is done, the student should get some acknowledgement of that extra effort.

2.7 Instructor Response Time

One very helpful item is to include instructor response time or turn-around time (which is also included in the QM rubric as instructor availability). For example, it is desirable for an instructor to let students know that assignment grades or test grades will not be posted until one week after the due date. This eliminates (or at least reduces) the number of emails asking "did you get my assignment?" It is also helpful for an instructor to build some flexibility into the testing procedure because of missed tests or technical difficulties. If it's possible to leave the test available slightly longer than the due date it is easier to deal with students who have issues and who are allowed to complete the test after the deadline has passed. The instructors acknowledge that this can create the potential for cheating on exams, but in these authors' experience this has not been a major issue. In any case, the authors generally structure their grading so that tests and quizzes are 50% or less of the grade, so cheating on exams will not ensure a passing grade. Multiple methods of assessment of a student's work also yields a better understanding of the student's learning.

3 CONCLUSION

All of these techniques together can go a long way in reducing instructor workload in an online course. One of the most helpful, however, involves restating summarized information in as many places as it might be relevant. For example, the discussion board guidelines can be incorporated again in every discussion board in the class. Given the amount of work involved in online classes, structuring as much efficiency as possible into the class benefits the instructor and provides more usable information to the student.

REFERENCES

1. See the QM Peer Review Course Rubric (Rubric Annotated FY0506) at www.qualitymatters.org/

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Formative assessment of the effectiveness of collaboration in GCB

Xing Hang

Computer Network Information Center (CNIC),
Chinese Academy of Sciences,
Beijing, China

David Villegas Castillo
S. Masoud Sadjadi

School of Computing and Information Sciences (SCIS),
Florida International University,
Miami, FL 33199, USA

Heidi Alvarez

Center for Internet Augmented Research and Assessment (CIARA),
Florida International University,
Miami, FL 33199, USA

Abstract With the rapid emergence of new communication software and hardware tools and the improvement of telecommunication infrastructures, a new collaboration paradigm is on the horizon that allows researchers around the globe to expand their loop of collaborators to cross geographical and cultural boundaries. However, much needs to be learned from the user experiences not only to improve the quality of the collaboration facilities, but also to develop new social protocols for distributed human interactions. In this paper, we try to analyze the usage of cyberinfrastructure in remote collaboration among researchers. For that, we draw on survey data and interviews with members from different collaborative projects, and we analyze how our current communication tools meet the needs of collaborative research activities. Then, we articulate a series of key challenges and requirements that contemporary teams are facing. In the end, we present ideas on what sorts of collaborative tools need to be built in order to fulfil the distributed and interdisciplinary collaboration projects. Our findings shed light on the factors that drive the use of cyberinfrastructure and the effectiveness in the success of cross-national and interdisciplinary research collaboration and distance learning, and suggest further research topics.

Keywords e-Science, formative assessment, group collaboration, distributed collaboration, distance learning.

1 INTRODUCTION

A majority of scientific researchers currently do not perceive the relevance of cyberinfrastructure (CI) for their own research [1,4]. Realization of such applicability often requires an in-depth understanding of both their scientific domain as well as the promise of CI. The Global CyberBridges (GCB) project [3] is designed to address this problem of inadequate adoption and use of cyberinfrastructure. GCB is a model global collaboration infrastructure for e-Science [2] between USA and international partners. The project is a multinational effort which aims at fully integrating cyberinfrastructure into the whole educational, professional, and creative process of diverse disciplines, bridging the divide between the information technology communities and the disciplines and creating a global community of scientists and researchers capable of collaborating with their counterparts through the integrated cyberinfrastructure.

Currently the project spans five research institutions spread over three regions: USA, China and Hong Kong, including faculty members and graduate students. Graduate students, 10 participants from USA and China, form four distributed teams of two to three individuals, with at least one person at each site in a specific team. Team members are playing the role of either CI researchers or disciplinary researchers. Disciplinary researchers are selected from academic areas such as biology, chemistry, meteorology and others. The CI researchers with a Computer Science or Information Technology background work with the disciplinary researchers in an attempt to satisfy their needs by making the best use of the cyberinfrastructure. Members of the four distributed teams have been working collaboratively for the past six months

on the following projects: *Computational Modeling & Simulation of Biodegradable Starch-Based Polymer Composites*, *Grid Enablement of Hurricane Simulation Application*, *On-Demand Weather Forecast Visualization via Efficient Resource Utilization in Grid Computing*, and *Collaborative Platforms*.

Our team, which is composed of two graduate students one from USA and one from China and two faculty advisors from USA, is built to work on the last project mentioned above. (The authors of this paper are members of the Collaborative Platforms project.) We believe that the success of building research collaborations across national boundaries depends very strongly on the quality of the collaborative platforms available for use by the participants and their effectiveness in using them. Therefore, our goal in this project is to provide a more convenient and efficient collaborative platform for our researchers, making global research collaboration a more productive and enjoyable experience.

To reach this goal, we have been observing, participating and studying the distributed and interdisciplinary collaboration of all four teams (including our team). We try to articulate a series of key challenges and requirements facing contemporary teams through both our observations and analysis of the survey we have conducted. Our aim in this paper is to present our research on the practice and feasibility of our existing social and technological support in providing distributed and interdisciplinary scientific researchers with an efficient and easy-to-use collaboration environment.

In current literature, similar evaluation research works using the method of survey and interview, have been done as can be seen in [7,8,9,10]. However, these works are focused on specific applications and domains different from our scenario, thus to some extent their findings does not effectively reflect the issues that we are faced with them in Global CyberBridges, and even at some points have contradictory findings compared to ours, which will be discussed later. Our work in this paper is unique in that it is the first assessment on the success of effectiveness in scientific research collaboration conducted among cross-nation/cross-culture graduate students and their respective faculty advisors. Before the research collaboration phase, these graduate students have gone through a semester of technical training on High Performance Grid Computing and Networking Research together. (This course was taught for a semester at Florida International University (Miami, Florida, USA) and the Chinese graduate students attended the class remotely through collaborative platform utilities including SAGE tile display wall, PolyCom video conferencing, and Skype and MSN audio and text chat systems.) Thus, our findings shed light on the factors that drive the use of cyberinfrastructure and the effectiveness in the success of cross-national and interdisciplinary *research* collaboration as well as *distance learning*.

The rest of the paper is organized as follows. In Section 2, we introduce the assessment activities that we chose to conduct our formative assessment. In Section 3, we discuss our findings on the current social and technological issues that have been challenging our teams. These findings are based on the result of the survey and interview as well as our observational work. In Section 4, we present detailed examination of current technologies and propose some improvements. Finally, in Section 5, we conclude our work in this paper and present the research issues needed to be addressed in the future.

2 ASSESSMENT ACTIVITIES

In traditional methods of studying the effectiveness of scientific research, publication volume is often used as the key evaluation criteria. However, in distributed collaborative scientific research such as cases like our projects, publication volume itself as the only evaluation criteria is not sufficient. Because collaboration is starting between cross-cultural strangers, it may take a while for the collaboration to mature and reach to a point where publications can be measured.

As observed in the collaboration process, the project often includes the development, integration, deployment, and testing of the technical infrastructure, as well as the coordination and building research communities needed for cross-cultural and cross-national collaboration. Collaboration activities normally include joint research projects and joint working papers. However, not all projects utilize cyberinfrastructure and collaborate in the same way.

For example, as observed in the team of *Computational Modeling & Simulation of Biodegradable Starch based polymer composites*, interaction is not much intensive since tasks are clearly divided between CI researchers and disciplinary researchers. CI researchers mainly focus on assisting disciplinary researchers in operating computer clusters and parallel computers that can meet the massive computing requirements posed by applying quantum-chemical methods to polymeric systems. In more detail, disciplinary researchers first decide which software to use in the quantum computing; Next, CI researchers install and test the software on the cluster; Then, it is again disciplinary researchers' turn to program with the software to solve disciplinary problems; Finally, CI researchers will work on how to fully utilize the cluster and improve the computation efficiency, including parallel processing, and thus enable disciplinary researchers to meet their research goals.

However, in other teams such as ours, tasks are not divided so clearly. In almost every step of this collaboration process, we had to interact with each other closely. Since our project goals was being adjusted occasionally, weekly audio meeting, together with frequent E-mails and instant message sessions, should have been well guaranteed. As a first step in this collaboration, we needed to develop a common understanding of what we would like to achieve, and how we would go about doing it. The best way we found was to exchange ideas and document notes to clarify our understanding of the project and to discuss further if something is not completely understood.

Thus, we decided to assess the effectiveness of collaboration using the following activities: attending all the distance learning sessions including both the video and audio meetings; monitoring the E-mail interactions in the project

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

group forums and the associated mailing lists; conducting a survey one month after the distance learning course has finished and the collaborative research has started; and finally, interviewing face-to-face with the team members. In the next section, we present our findings of the problems we want to further explore.

3 OUR FINDINGS

As a whole, collaborative activities we have observed in all projects include two parts: distant class participation and distributed collaborative research work. We have worked on an interim assessment of the effectiveness of the existing collaboration in the four GCB projects. As a part of the two assessing items (the technology support and the social protocols), we conducted a survey and set up interviews with respondents respectively at the two sites (Beijing and Miami) as to how satisfied they are with the progress of their respective projects, and how and what can be done to improve the effectiveness of their distance learning and research collaboration.

3.1 Survey

In the survey, we employed the extended Task-Technology-Fitness framework [5,6,8] to define the types of tasks. The survey was composed of the following three major sections

- **Main activities in research**

The different teams were asked to enumerate the activities in which they spent most of their time. Among the possible answers were: Writing documents, taking decisions with other peers, installing software used for their research, developing needed tools and researching. Figure 1 shows which tasks had a greater importance for the reviewees. The graph shows the possible answers in the X axis and the percentage of individuals who chose them as principal tasks in the Y axis. We can see that the most prominent activity is research itself. Nevertheless, this is a task that is carried mostly in solitary, and doesn't gain much from collaborative tools. The next two most common tasks, writing documents and taking decisions are much better targets for collaborative technologies, since they require more direct interactions among the team members.

Figure 1. Main Activities in research

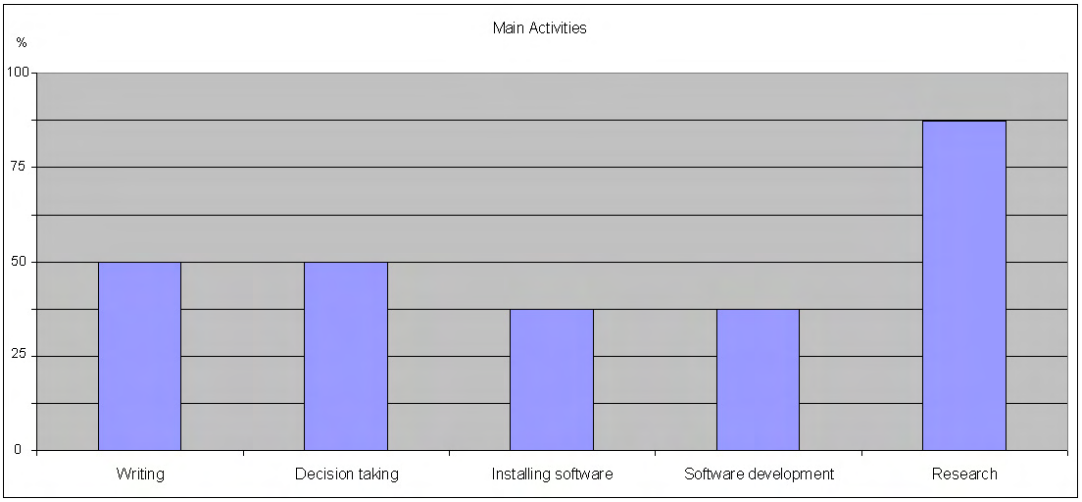
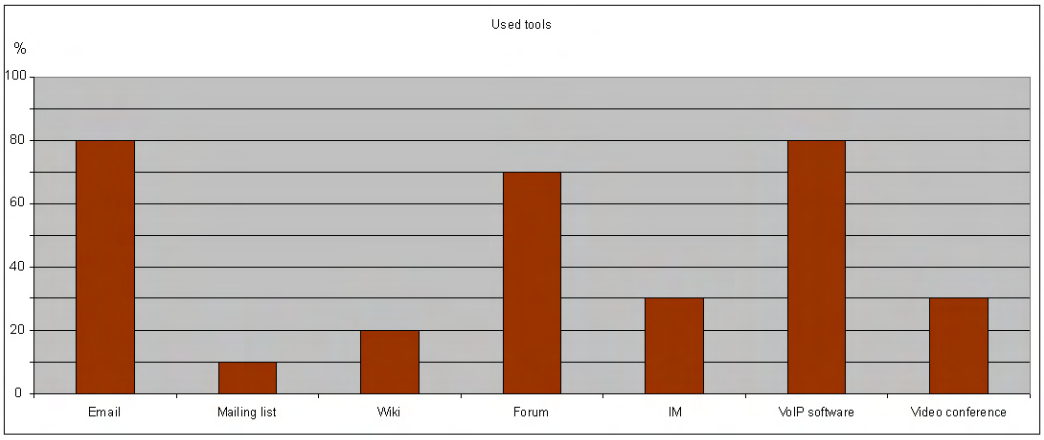


Figure 2. Tools used in the collaboration process



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

- **Used tools**

  Another question posed in the survey was the use of the existing tools during different tasks. Figure 2 shows that the two most used tools are E-mail and Voice-over-IP (VoIP) Software. E-mail and VoIP Software are existing solutions that have been available for a long time (the latter can be considered just as an extension of the regular phone line). We believe that others, like video conferencing, were not used as much due to the difficulty of using them. For example, we have Polycom video conferencing systems in USA, China, and Hong Kong, but unfortunately they were not easily accessible and there are other social issues with video conferencing that makes the users uncomfortable as discussed later.

- **Tool effectiveness**

  Finally, we asked all partners to rate the effectiveness of different types of tools for given tasks. These tool types comprise Text Systems (which includes Instant Messaging, E-mail, Wiki, Forum), Audio Systems (which includes Telephone and VoIP Conferences), Video Systems (which includes PolyCom and Skype video conferencing systems) and Face-to-Face Meetings. The low acceptance for video systems was unexpected for us, since it seems to be a more powerful tool than voice systems; providing the same service as the former plus visual appearance of partners. Our suspicion was that this characteristic of video conferencing systems, which is being seen by others, is what made some users uncomfortable with it. Many of them prefer the anonymity of text or voice systems to perform most of their tasks as can be inferred from the graph in Figure 3.

3.2 Ongoing Feedbacks and Interviews

Apart from the ongoing feedbacks and complaints that we were receiving from time to time from the partners, which were very helpful on trying quick solutions to the problems, we also set up semi-formal face-to-face interviews with the team members after analyzing the survey. We believe that there are more aspects to collaboration that cannot be demonstrated on the above histograms and ongoing feedbacks; this is where face-to-face interviews can help. We also wanted to have their ideas on some open-ended questions that could not be elaborated in our survey. Below, we will present our analysis based on informal introspection combined with more formal observational work. Our findings of current issues and problems with regards to both the technology support and the social protocols are as follows.

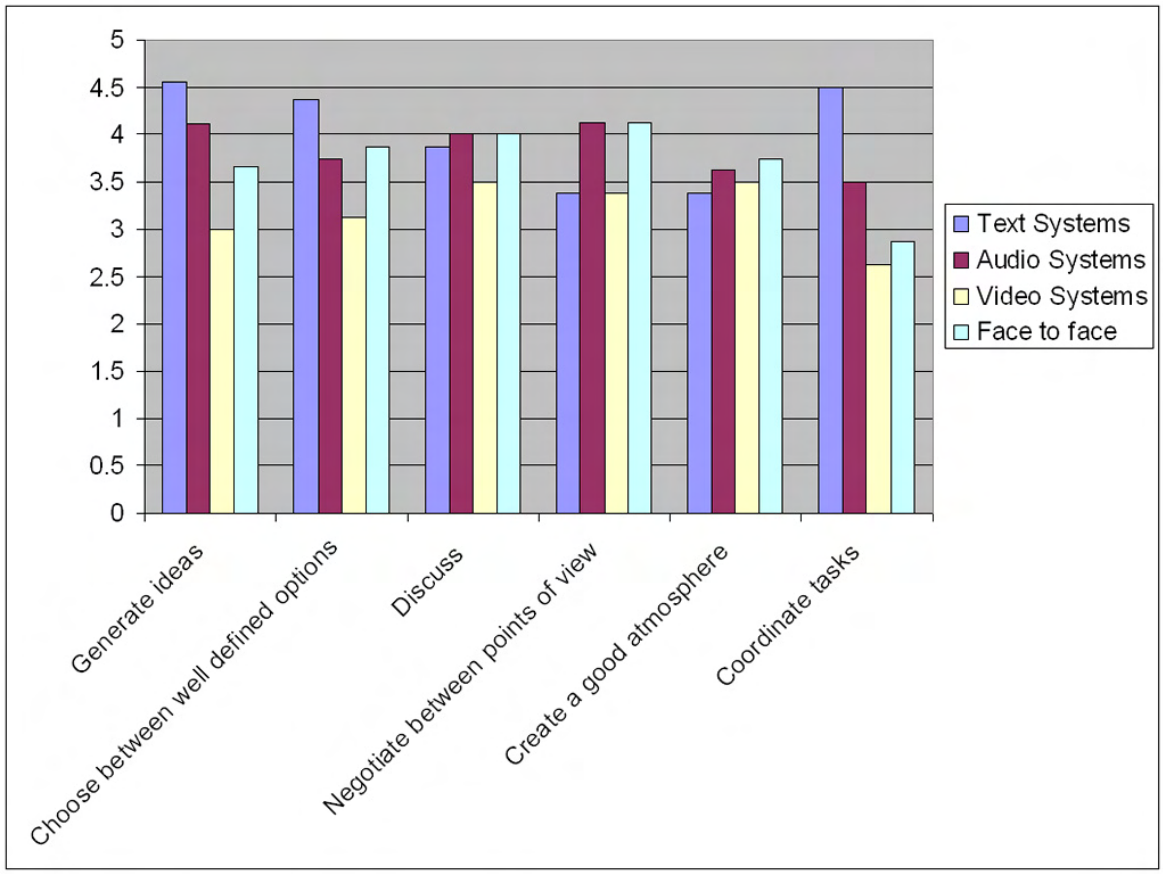
Discussion on the technological aspects

In this part, we present three of the technology requirements in the GCB project and discuss the respective issues and problems with regards to each of them.

- **We require effective communication technologies for distance learning.**

  Video Tele Conferencing (VTC) is used as the main means of distance learning. Although we admit that VTC is currently the best way that we can connect

Figure 3. Tool effectiveness according to tasks



people regardless of their location and that video plays an important role in developing a cosy atmosphere as well as building a close affinity among cross-cultural strangers, current VTC technologies are far from perfect. We believe that the distance barrier can be broken only when distributed people are communicating just as they are talking face-to-face, which requires views from various perspectives with no distraction from the technology itself, but the widely available VTC technologies today does not yet provide such an environment. The not-so-good quality and occasional delay (jitter) of sound and video that happens every now and then are among the major problems that have greatly challenged the Chinese students and have added to the language difficulty that they already had. To mention another problem, the VTC technology that we used does not support sharing of the presentation slides. This has introduced yet another problem as the projected PowerPoint slides could not be displayed simultaneously in the two sites (USA and China) together with the image of people who were participating in the classroom, not to mention other views that only the local participants can perceive. Remote participants usually find it hard to figure out who is talking about which part of the slide.

As we learned about such problems, we tried to address them by trying different solutions. For example, for each session of the class, a student volunteer would take the control of the PolyCom VTC and would change the focus of the camera to whoever was talking at the moment. Among other improvements we set up a chat system to go along with the VTC. This was very helpful as the teacher could clarify a point, send a URL, etc. using text messages without causing new confusions. Also, we developed a simple human-to-human protocol that would force us to pause several times during the lecture to make sure that remote participants can catch up with the teaching pace. Still, we admit that there is a lot to improve to make distance teaching more satisfactory and effective.

- **We require effective communication technologies for distributed collaborative research.**

First of all, based on the analysis of survey and interview, we found that our users are more willing to use the simple collaborative tools that they have been familiar with such as E-mail and voice conferencing, which shows the fact that a criterion of use for a given technology is the level of familiarity that users already have with that technology. According to our study, we found out that E-Mail was the most widely used tool in communication between peers, as well as VoIP software in the case of small groups of people; the former medium is one of the most extended online tools, and the same can be said for the VoIP technologies, which can be seen as an extension to the traditional telephone systems. Users feel more comfortable if they can keep their existing habits, and this criterion is also valid for the opposite case: the use of video conferencing or wiki software, for example, was not as well rated among users, given that these technologies require new habits and skills.

Secondly, many interviewees agreed on the point that E-mail was more effective for generating ideas and coordinating tasks, while voice conference calls were better for discussion and negotiation. Our study results also suggest that occasional video conferences are instrumental in a good atmosphere which is a crucial precursor to the effective use of distance communications technologies. However, the 12 hour time difference between Miami and Beijing has forced our team members to limit their synchronous communication (e.g., voice/video teleconferencing) to only once or at most twice a week and communicate mostly through E-mail, IM, mailing list, and group forums.

Thirdly, contrary to the point "*Instant Messaging are not seen as efficient or well suited communication channels for collaborative tasks*" as discussed in [8], our findings reveal that IM is necessary for effective collaboration. The biggest barrier in distributed collaboration is the unreachability of distant partners. Normally partners make schedules and meet regularly to coordinate tasks and make new schedules. However, when unscheduled and exigent problems arise, they usually feel helpless either because of the unawareness of their partner's phone number, or as a result of the unwillingness to get in touch with their partners at the cost of expensive international calls, which largely decrease the effectiveness in distributed collaboration. In such cases, IM has proved to be a good solution for our team members.

Last but not the least, we found that efficient collaboration necessitates a seamlessly integrated work platform where researchers have easy access to all collaboration tools. Currently, we have forums, wiki, E-mails, mailing list, video conference, and instant messaging at our disposal. As a direct consequence of all these different, but complementary technologies, we have to keep track of many usernames and passwords, and have to login and open several websites and/or collaboration software tools in order to be able to communicate with our collaborators. We believe that it would be very helpful to the researchers if we could free them from going through such tedious and distracting tasks by providing them with an integrated and single-sign-on system that would satisfy all their communication needs. A successful collaboration platform is one that allows researchers to focus on the flow and integration of information and not on the details of the underlying components. One such approach is underway in the Communication Virtual Machine project [11].

- **We require effective technology for visualizing complex phenomena.**

As another result of our study, we realized that for effective scientific collaboration, the ability to visualize and share high-resolution and complex phenomena is

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

a must. For the GCB project, we chose to use the tile display wall technology instead of expensive custom-design solutions, both because of the cost involved (using off-the-shelf equipment) and because of the ability to scale the system, if needed. A wall-mounted tile display consists of a number of regular LCD screens placed one next to the other, which forms a grid of LCD screens. In our case, we use SAGE [12], a software utility that synchronizes the output of several computers so that an image can be plotted with a very high resolution on the grid of LCDs. Although this software was originally developed to provide a solution to displaying very high resolution images, we also plan to use it as a collaborative whiteboard where distant peers can exchange ideas.

Discussion on the geographical and social aspects

Based on our observational work and the analysis of the survey and the interview, we reached to some new findings as follows.

Firstly, contrary to the common belief that it is hard for global collaboration to attain the same effectiveness as local collaboration does because of the poor technological support and incomplete social protocols, we have an interesting and exciting finding that shows this global work pattern can become even more efficient than local collaboration as a direct result of the time zone difference. Let us consider coordination, which is a common task in collaboration, as an example. Assume that the workload related to a task assigned to two remote team members is one day for each of the members and one member has to wait until the other one is done before the other one can start his/her part. As the global partners in our project have 12 hours time zone difference, when it is the beginning of the day for one person, his/her partner at the other side of the world has just finished a whole day of work and is ready to go to bed. Therefore, it actually takes this distributed team only one full day to finish the whole task (24 instead of 48 hours)!

Secondly, in order to maintain such a success in the effectiveness of distributed collaboration, we realized that maintaining a strong commitment to different tasks is absolutely necessary to its success. Keeping communication and interaction, especially quick response and having each other updated about the latest status in collaboration, is also very important when distant strangers try to build a collaborative research relationship. Interaction should include not only technical discussion but also chatting about day-to-day matters. Feedback should be provided as soon as possible to correct problems if things are not working properly. We believe that in a collaborative research work, feeling like being a part of the team is the ultimate key to success of the team and any effort in this regard cannot be overestimated.

Finally, we fount out that technology reliability has a strong affect on the building of trust and social relationships. For example, the group forums used in the collaboration did not work well for some time at the beginning. The expectation from the forum software was to send an E-mail to all the group members for each new posting to the forum. However, as the forum software was updated and the new version requires explicit mention of such setting, there was no E-mail sent to the group members (by default) to notify them of the new postings. Misunderstandings and disappointments arose in this condition and harmed the willingness to collaborate with each other for the period in which the problem persisted.

Based on the discussions presented in this section, we identified several methods that are useful and should be adopted in a distributed collaboration activity. First, you need to make sure that people at remote sites understand each other, and more importantly, are in synch with each other. For example, during a distant learning session, it may be useful to develop a protocol to signal or interrupt an ongoing conversation for more clarifications on the subject matter. In addition, according to our experience, we found out that sending the presentation slides and the other shared material in advance to all the involved parties is very helpful in light of the technological deficiencies of the current VTC technology, as well as making sure redundancy in the system if a channel fails – SAGE Tile Display Wall, Polycom, Chat, or Moodle [13].

4 POTENTIAL IMPROVEMENTS TO THE EXISTING TECHNOLOGY

The initial set of tools selected for enabling collaboration in different GCB projects were based on well known communication software tools. Since the project was not starting from any prior study of communication tools effectiveness, the choice was guided by the familiarity and availability of the tools, rather than by their effectiveness in distant collaboration. Among the basic technologies used were E-mail, instant messaging using MSN Messenger, a Web-based forum using Moodle, Wiki pages of LA Grid, video conferencing using Polycom and VoIP voice conference using Skype. A wall-mounted tile display running SAGE software for visualization was also available. These technologies proved to be enough to allow distant researchers to communicate, follow a lecture across geographical boundaries, and carry on with different research projects, but many improvements can still be done to facilitate users' experiences. Below, we first classify communication tools and then provide some improvement suggestions to the design of future communication tools.

4.1 A Classification of Communication Technologies

There are other existing technologies which were not used in the different GCB projects and we believe that they can help researchers to collaborate more effectively. Below, we introduce a classification of the available tools that we developed to better understand our options and to be able to recommend possible replacements to the currently selected tools for the future GCB projects. The categories that we came up with are the following.

- Synchronous Communication Software

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

  - Telephone and VoIP, Instant Messaging, Video Conferencing, etc.
- Asynchronous Messaging Tools
  - E-mail, Mailing Lists, Forum Software, etc.
- Collaborative Editors
  - Wiki Pages, Synchronous Word Processors (Google Docs), etc.
- Workspace Sharing Applications
  - Whiteboards, Desktop Sharing Software, etc.
- Integrated Environments
  - Some tend to be focused to a given area, like Basecamp or Novell's GroupWise, and some tend to be more general with ways to customize the communication such as CVM. These environments sometimes encompass existing technologies like video conferencing, E-mail, etc.

Although some tools can fall into more than one category, each of these categories presents some strong points and also some deficiencies. We found that *synchronous tools* for communication, like telephone, instant messaging and video conference systems are needed less frequently (e.g., once or twice a week) in order to discuss ideas, take decisions and create a consistent collaborative ambient. Nevertheless, they make necessary the presence of both interlocutors at the same time, and in the case of telephone and video conferencing systems participating peers have a higher personal exposure, which in some cases proved to be counter productive in the early stages of the projects.

Asynchronous messaging tools like E-mail, mailing list, and Web-based forum software are more suitable when researchers are situated in different time zones and cannot be present at the same time. A drawback of these tools, however, is the difficulty to maintain prolonged conversations, since responses are not necessarily immediate.

Collaborative editors are used as an important technology to accumulate knowledge. Wiki software allows partners to collaborate by putting ideas together and discussing them, although not as effectively as with other synchronous tools like phone systems or instant messaging. On the other hand, synchronous editors like Google Docs [14] allow a more immediate feedback to changes, letting peers to exchange ideas as they propose them.

Workspace sharing applications are similar to collaborative editors, but they usually allow participants to express ideas in a visual fashion. These tools are synchronous in many cases and provide the best alternative for sketching ideas, initiating discussions and showing concepts from different points of view.

Finally, *integrated environments* integrate different tools in a single package, making it easier to use them effectively. Nevertheless, these technologies are developed for specific groups of users, and bring problems when adapting them to the collectives for which they weren't addressed primarily. Although, there are some attempts to address this issue [11], there is no widely used integrated communication environment that can be adopted immediately for the GCB projects. Additionally, such environments typically require a high learning effort, which is not desirable and in most cases not practical.

4.2 Suggested Improvements to the Design of Future Communication Tools

There are many improvements that can be applied to the exiting communication tools, which were used in the GCB projects, and there are a number of new technologies that can make collaboration among researchers more efficient. There are also some features that make one communication utility more useful than others, and such features have to be considered when choosing a tool. For example, as discussed in Section 3, we found that a criterion of use for a given technology is the level of users' familiarity with the tool. This fact shows that a potentially useful communication utility is normally disregarded or falls into disuse when it conflicts with the users' daily habits or when it brings a burden associated with its high learning curve. This point indicates that new tools have to be similar to existing and established ones and new features should not require dramatic changes in the use of the existing tools.

Another improvement to the existing technology is the detachment of the software from the typical input and output devices. Bigger screens, tactile input and new interaction metaphors can enhance collaboration experiences and bridge the gap among distance and cultural frontiers. The drawbacks of such technologies, however, are primarily their cost, their low adoption rate and immaturity. In our particular case, the wall-mounted tile display running SAGE software, although being a powerful tool, has a slow adoption phase. We associate this deficiency to different factors such as the technology being in its early stages, the difficulty of deploying it, and the lack of user preparedness from our team members. SAGE is a software utility that enables high definition image and video representation, but so far proved to be inefficient for other uses such as videoconferencing or real time collaboration for our projects. We plan to address the inefficiency of SAGE in our future work.

5 CONCLUSION AND FUTURE WORK

This paper provides a formative assessment of the effectiveness of the collaboration in the GCB projects. We started from an analysis of the survey and interview about collaborative tools used by the project participants, their functionality, and their usefulness and effectiveness regarding different task types. Our assessment items cover both technological support and social protocols. Next, we examined current communication tools supporting collaborative work and in combination with the GCB communication requirements, we proposed our future work based on a currently inchoate collaborative tool, called *SAGE*.

Our work in this paper is based on both informal introspection and formal observational work combined with analysis of survey data and interview. This *Formative Assessment* is in-

strumental in that it not only helps the GCB team members make required, mid-project changes and corrections to the communication/collaboration styles and technology tools, but it also helps us identify needs for additional tools and social protocols, which should either be acquired or developed, and sheds light on future development work on our Collaborative Platforms project. In addition, as the primary aim of GCB is to foster collaboration between early-career scientists from USA and its international partners, this early feedback should increase the chances of success of GCB. We believe that continuous monitoring and refinement of the GCB collaborative projects will reveal even more issues related to cross-culture, cross-nation exchange of ideas that would require new improvements to the collaboration tools and would need new coordination, communication, and management techniques and solutions to be explored in the future.

ACKNOWLEDGEMENT

This work is part of the Global CyberBridges project. It was supported in part by the National Science Foundation (grants OCI-0636031, REU-0552555, and HRD-0317692) and in part by IBM (SUR and Student Support awards in 2005, 2006 and 2007).

REFERENCES

1. Cerf V, and the Committee on a National Collaboratory, Computer Science and Telecommunications Board (National Research Council). (1993), National Collaboratories: Applying Information Technology for Scientific Research, National Academy Press, Washington, D.C..
2. Kouzes R, Myers J, and Wulf W. (1996), Collaboratories: Doing science on the Internet, IEEE Computer, August, pp. 40-46.
3. The GCB project; http://www.cyberbridges.net/
4. Christine M. Hine (Editor). (2006), New Infrastructures for Knowledge Production, Information Science Publishing, United Kingdom, pp.143-166.
5. Joseph E. McGrath. (1993), 'Small group research, time, task and technology in work groups: The Jemco Workshop Study', International Journal of Theory and Application, August 1993, Vol 24 No. 3, s. 283-423, Sage Periodicals Press. ISSN: 1046-4964.
6. Joseph E McGrath, and Andrea B. Hollingshead, (1994), "Groups interacting with technology", Sage Library of Social Research 194, Sage Publications Inc. ISBN: 0-8039-4898-0.
7. Edward H. Shortliffe, Vimla L. Patel, James J. Cimino, G. Octo Barnett and Robert A. Greenes. (1997), A Study of Collaboration Among Medical Informatics Research Laboratories; http://smi.stanford.edu/smi-web/reports/SMI-97-0685.pdf
8. Anita Gupta, Marianne H. Asperheim, Odd Petter N. Slyngstad and Harald Ronneberg, (2006), 'An Empirical Study of Distributed Technologies Used in Collaborative Tasks at Statoil ASA', Proceedings of International Conference on CollaborateCom 2006; http://www.idi.ntnu.no/grupper/su/publ/sevo/gupta-slyngstad-TTF-framework-2006.pdf
9. Xianghua Ding, Thomas Erickson, Wendy A. Kellogg, Stephen Levy, James E. Jeremy Christensen, Jeremy Sussman, Tracee Vetting Wolf and William E. Bennett. (2007), An empirical study of the use of visually enhanced voip audio conferencing: the case of IEAC, Proceedings of the SIGCHI conference on Human factors in computing systems, Conference on Human Factors in Computing Systems, pp.1019-1028.
10. Baker, K., Greenberg, S. and Gutwin, C. (2001), Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration, Proceedings of the 8th IFIP Working Conference on Engineering for Human-Computer Interaction (EHCI'01).
11. Yi Deng, S. Masoud Sadjadi, Peter J. Clarke, Chi Zhang, Vagelis Hristidis, Raju Rangaswami, and Nagarajan Prabakar. A communication virtual machine. In Proceedings of the 30th Annual International Computer Software and Applications Conference (COMPSAC 2006), Chicago, U.S.A., September 2006.
12. Renambot, L., Rao, A., Singh, R., Jeong, B., Krishnaprasad, N., Vishwanath, V., Chandrasekhar, V., Schwarz, N., Spale, A., Zhang, C., Goldman, G., Leigh, J., Johnson, A. (2004), SAGE: the Scalable Adaptive Graphics Environment, Proceedings of WACE 2004.
13. The Moodle Project; www.moodle.org
14. Google Docs & Spreadsheets; http://docs.google.com

i·Society 2007

Intellectual scrutinizer for compute, storage, network & system characteristics of Linux system

A. Balamurugan
M. Savithashree
H. Sriram
G. Vidya
Department of Information Technology
V.L.B.Janakiammal College of Engineering and Technology
Kovaipudur, Coimbatore-641042, Tamilnadu, India.
csebala@rediffmail.com, savithashree@gmail.com, srisum1986@gmail.com, gopvidya@gmail.com

Abstract Enterprise class servers are one of the important components in an enterprise. Management of these servers by understanding their behavior, load conditions, and monitoring any failures is very important. The proposed paper is towards observing and reporting of the Compute, Storage, Network and System Characteristics in a simulated environment of server running in Linux operating system software. The mechanism used for observing and reporting can be achieved by retrieving relevant data through operating system provided system calls, library APIs or by accessing kernel data and by developing an Intelligent Agent in User Space following SNMP protocol.

Thus the agent is monitored from the manager by measuring the parameters such as the number of CPUs, Type of CPUs, and Total CPU utilization for observing the CPU characteristics. Storage related characteristics for observation and analysis are type of storage devices, capacity of storage devices. Using these characteristics the performance, security, configuration, accounting and fault occurring in the network can be monitored.

Keywords Enterprise class servers, compute, storage, network and system characteristics, system call, library APIs, Intelligent Agent.

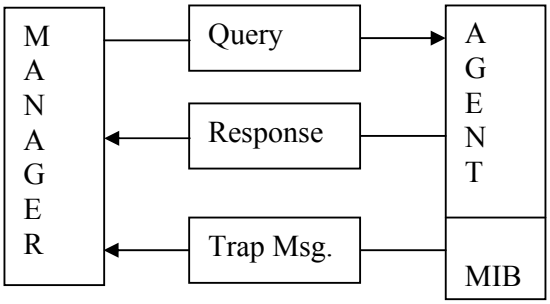
1. INTRODUCTION

The key issue in this concept is that this has been implemented only in the Windows server till date. Most of the enterprises nowadays are switching over to Linux domain because of its advanced features and reliability when compared to Microsoft Windows. Therefore the performances of these Linux servers have

to be monitored so that they are available all the time. As the size of the networks grew, they became harder to manage and maintain, thus the need for network management was realized. The change in the type of protocol used for monitoring had a greater influence in response time and availability. SNMP is a management protocol used for monitoring network activity which can be used to monitor a wide range of devices in real time. SNMP has become interoperable on account of its widespread popularity. The Real Need for Network Management is for

- Reducing Mean Time to Repair
- Improving Service Availability
- Providing Critical Reporting
- Capturing Performance Metrics

Figure 1. Interaction between agent and manager



The Figure1 depicts the interaction between a manager and an agent. The interaction between a manager and an agent is similar to the interaction between a master and a slave device. The manager can initiate a poll to the agent requesting information or directing an action. The agent, in turn, generates a response to the query. This is how a remote I/O protocol works. However, the manager can request that a trap be set by the agent. A trap is simply a report to be issued in the future which is triggered when a set of conditions are met, similar to an alarm. The trap is triggered upon an event and, once it occurs, the agent immediately reports the occurrence without a poll from the manager. The NMS receiving

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007

the trap can then take appropriate action such as notifying personnel of the event. In this situation, the NMS is acting as a server by gathering data from agents and providing information on the state of devices to clients. Management applications can monitor, configure and control managed elements. A management protocol is used to communicate management information between the management stations and agent.

2 SNMP – SIMPLE NETWORK MANAGEMENT PROTOCOL

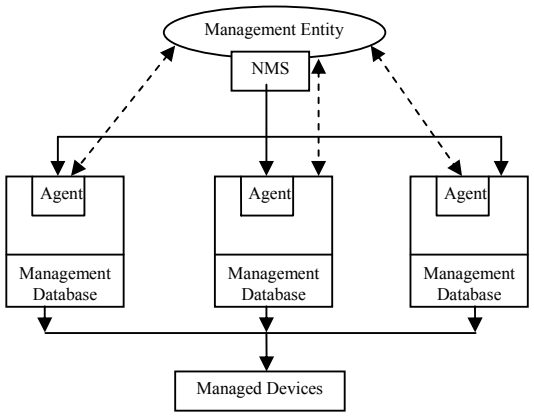
SNMP is the most popular network management protocol in the TCP/IP protocol suite. SNMP is the prevailing standard for management of TCP/IP networks. SNMP is layered on top of UDP, the User Datagram Protocol. All SNMP transactions take place using PDUs (Protocol Data Units). It is a simple request/response protocol that communicates management information between two types of SNMP software entities, SNMP applications (also called SNMP managers) and SNMP agents.

2.1 BASIC COMPONENTS OF SNMP

The SNMP-managed network (Figure2) consists of three key components: managed devices, agents, and network-management systems (NMSs).

A managed device is a network node that contains an SNMP agent and that resides on a managed network. Managed devices collect and store management information and make this information available to NMSs using SNMP. An agent is a network-management software module that resides in a managed device. An agent has local knowledge of management information and translates that information into a form compatible with SNMP. A Network Management System (NMS) executes applications that monitor and control managed devices. NMSs provide the bulk of the processing and memory resources required for network management. One or more NMSs must exist on any managed network.

Figure 2. SNMP Managed Network



2.2 SNMP Commands

Managed devices are monitored and controlled using four basic SNMP commands: read, write, trap, and traversal operations.

The read command is used by an NMS to monitor managed devices. The NMS examines different variables that are maintained by managed devices.

The write command is used by an NMS to control managed devices. The NMS changes the values of variables stored within managed devices.

The trap command is used by managed devices to asynchronously report events to the NMS. When certain types of events occur, a managed device sends a trap to the NMS.

Traversal operations are used by the NMS to determine which variables a managed device supports and to sequentially gather information in variable tables, such as a routing table.

2.3 SNMP Protocol Operations

An SNMP application can read values for the SNMP objects (for monitoring of devices) and some applications can also change the variables (to provide remote management of devices). The SNMP agent software on a device listens on port 161 for requests from an SNMP application. The SNMP agent and application communicate using User Datagram Protocol (UDP). Trap messages, which are unsolicited messages from a device, are sent to port 162.

Get: Retrieves the value of a MIB variable stored on the agent machine.

GetNext: Retrieves the value of the next MIB variable.

Set: Changes the value of a MIB variable

Trap: A notification of something unexpected, like an error is sent by an agent to a manager.

2.4 Working of SNMP in a Network

A network consists of managers (clients) and agents (servers) as shown in Figure 3. Communication between the agents and clients can be established through a protocol SNMP. Each agent consists of a collection of information which is organized hierarchically called MIB. MIB can be accessed using the network management protocol SNMP. Using SNMP agents can be monitored by the managers.

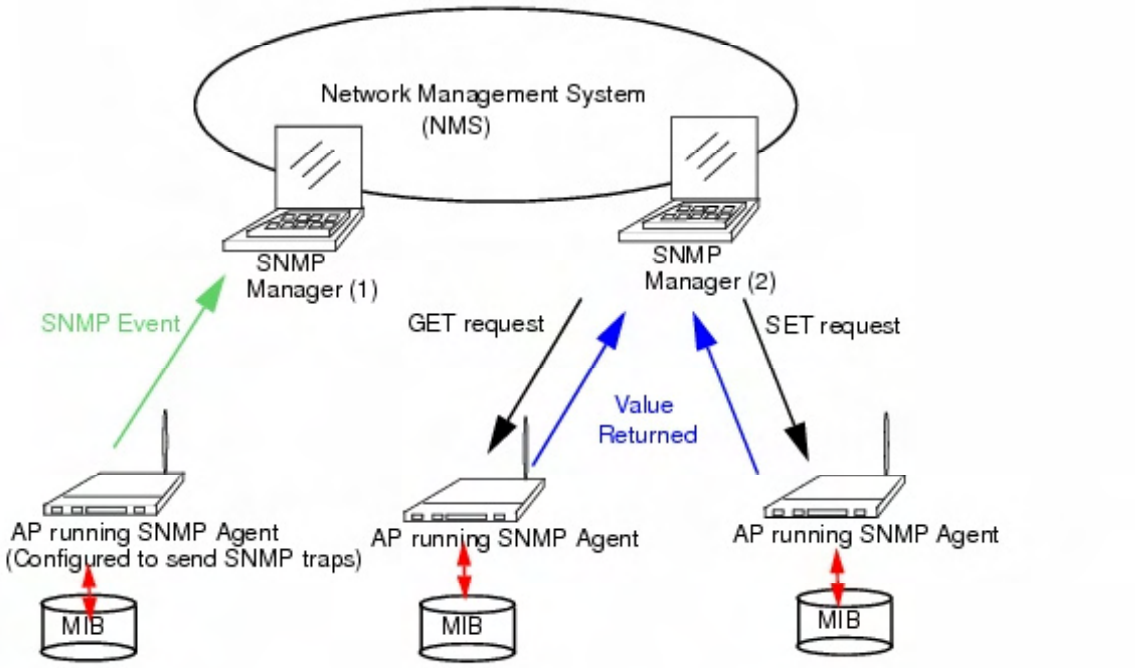
When a manager needs to retrieve values regarding the agent it sends GET request to the agent and the agent in turn returns the value queried by the manager.

A SET request is used by the manager when it needs to change a value present in the MIB of a particular agent. This

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. Working of SNMP in a Network



event is invoked when the manager specifies the OID of the object it want to change.

3 NET-SNMP

NET-SNMP provides tools and libraries relating to the Simple Network Management Protocol including an extensible agent, a SNMP library, tools to request or set information from SNMP agents, tools to generate and handle SNMP traps, etc. The SNMP implementation is done using NET-SNMP with Linux as a platform. The NET-SNMP project called in the past as UCD-SNMP was historically developed by the American university. Carnegie Mellon University (CMU) and then improved and maintained now by the American university, University of California Davis (UCD). NET-SNMP supports SNMPv1, SNMPv2 and SNMPv3 from the SNMP agent side and from the SNMP manager side by using on line commands.

3.1 NET-SNMP Commands

The NET-SNMP toolkit provides a suite of command line applications that can be used to query and act on remote SNMP agents. The following are the commands used in NET-SNMP. The Table 1. exemplifies the set of commands that are used to query a SNMP agent by the manager.

Table 1. NET-SNMP Command Set

Command Set
SNMPGET
SNMPWALK
SNMPGETNEXT
SNMPTRANSLATE
SNMPSET
SNMPTABLE
SNMPTRAP

SNMPGET Command

Snmpget communicates with a network entity using SNMP GET requests. Snmpget is an SNMP application that uses the SNMP GET request to query for information on a network entity.

One or more object identifiers (OIDs) may be given as arguments on the command line. Each variable name is given in its own format.

Syntax: snmpget options hostname community object

SNMPWALK Command Snmpwalk - retrieves a sub tree of management values using SNMP GETNEXT requests. Snmpwalk is an SNMP application that uses SNMP GETNEXT requests to query a network entity for a tree of information. An object identifier (OID) may be given on the command line. This OID specifies which portion of the object identifier space will be searched using GETNEXT requests.

Syntax: snmpwalk application options common options OID

SNMPGETNEXT Command Snmpgetnext - communicates with a network entity using SNMP GETNEXT requests. Snmpget is an SNMP application that uses the SNMP GETNEXT request to query for information on a network entity. One or more object identifiers (OIDs) may be given as arguments on the command line. For each one, the variable that is lexicographically "next" in the remote entity's MIB will be returned.

Syntax: snmpgetnext options hostname community objected

SNMPTRANSLATE Command Snmptranslate - translate MIB OID names between numeric and textual forms. Snmptranslate is an application that translates one or more SNMP object identifier values from their symbolic (textual) forms into their numerical forms (or vice versa).OID is either a numeric or textual object identifier.

Syntax: snmptranslate options OID

SNMPSET Command Snmpset - communicates with a network entity using SNMP SET requests. Snmpset is an SNMP application that uses the SNMP SET request to set information on a network entity. One or more object identifiers (OIDs) must be given as arguments on the command line. A type and a value to be set must accompany each object identifier.

Syntax: snmpset options hostname community objectID type value

SNMPTABLE Command Snmptable - retrieves an SNMP table and display it in tabular form. Snmptable is an SNMP application that repeatedly uses the SNMP GETNEXT or GETBULK requests to query for information on a network entity. The parameter TABLE-OID must specify an SNMP table. Snmptable is an SNMP application that repeatedly uses the SNMP GETNEXT or GETBULK requests to query for information on a network entity.

Syntax: snmptable options hostname community objected

SNMPTRAP Command Snmptrap, snmpinform - sends an SNMP notification to a manager. Snmptrap is an SNMP application that uses the SNMP TRAP operation to send information to a network manager. One or more object identifiers (OIDs) can be given as arguments on the command line. A type and a value must accompany each object identifier. When invoked as snmpinform, or when -Ci is added to the command line flags of snmptrap, it sends an INFORM-PDU, expecting a response from the trap receiver, retransmitting if required. Otherwise it sends a TRAP-PDU or TRAP2-PDU.

Syntax: snmptrap options hostname community trap parameters

3.2 Management Information Base

A MIB is a collection of information that is organized hierarchically. The data is structured in a tree form, and there is a unique path to reach each variable. This structured tree is called the MIB. MIBs are accessed using a network-management protocol such as SNMP. They are comprised of managed objects and are identified by object identifiers.

MIB's, or Management Information Bases, provide a map between numeric OID's and a textual human readable form. The structure of a MIB comes from the Structure of Management Information (SMI) standard detailed in IETF RFC 1155 and 2578. SNMP defines a separate standard for the data managed by the protocol. The data is structured in a tree form, and there is a unique path to reach each variable. This structured tree is called the Management Information Base (MIB).

The subset of managed objects that make up the TCP/IP portion of the MIB is maintained by each TCP/IP node. Objects in the MIB are defined using a subset of Abstract Syntax Notation One (ASN.1) called "Structure of Management Information Version 2.

In SNMP, MIB objects are given a unique object identifier consisting of a sequence of numbers separated by a period (.). These sequences of numbers are read from left to right and correspond to nodes of the object name tree.

Each device in an SNMP network is defined by a data file called the Management Information Base, or MIB. The MIB defines the device as a set of managed objects - values that can be read or changed by the SNMP manager. Managed objects can include alarm elements, controls, device uptime, or other aspects of the device.

Device elements must be listed in the MIB for them to be visible to the SNMP manager. All SNMP devices support a generic MIB that defines generic traps. But to use advanced network management features, granular traps are used which contains device-specific information. Granular traps require detailed MIBs that fully describe the managed objects of the devices.

3.3 Object and ObjectID

Object name is given by its name in the tree. All child nodes are given unique integer values within that new sub-tree .Children can be parents of further child sub-tree (i.e. they have subordinates) where the numbering scheme is recursively applied.

The Object Identifier (or name) of an object is the sequence of non-negative Integer values traversing the tree to the node required. Allocation of an integer value for a node in the tree is an act of registration by whoever has delegated authority for that sub tree.

3.4 MIB Tree

Network management is defined under the iso(1) tree branch. Under this tree branch are a number of subordinate organization definitions. Network Management falls under the org(3) node.

Under the org(3) node are a number of subordinate organizations. Network Management falls under the dod(6) node for the Department of Defense (DoD).

Under the dod(6) node are a number of subordinate networks. Network Management falls under the internet(1) node for the Internet.

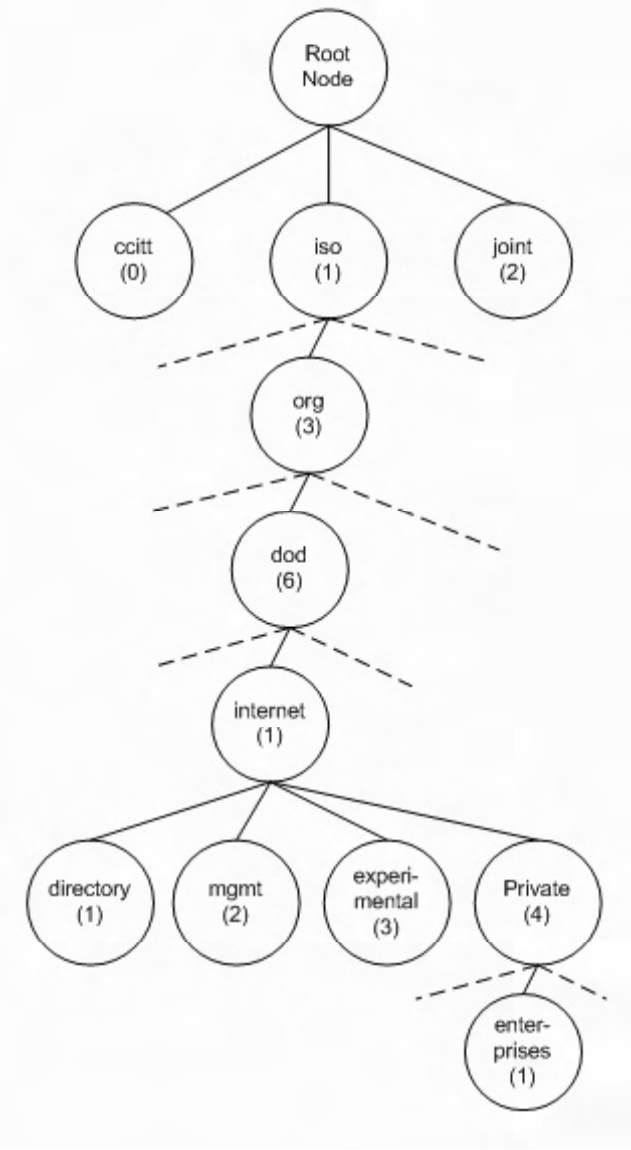
Under the internet(1) node are a number of subordinate nodes representing different standardization efforts in the area of directory, management, experimental and private MIBs. Network Management objects that have been standardized fall under the mgmt(2) node.

Under the mgmt(2) node are a number of subordinate nodes representing different standardization efforts. Network Management objects that have been standardized fall under the mib-2(1) node.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Under the mib-2(1) node are a number of subordinate nodes representing groupings of MIB variables. Figure 4. shows the system(1) and interfaces(2) MIB variable groupings. The MIB tree illustrates the various hierarchies assigned by different enterprises.

Figure 4. MIB Tree



4 RETRIEVING THE CHARACTERISTICS USING NET-SNMP COMMANDS

4.1 Installing and Configuring NET-SNMP Agent

NET-SNMP is a suite of software for using and deploying SNMP protocol. It supports IPv4, IPv6, IPX, AAL5, UNIX domain sockets and other transports. It contains a generic client library, a suite of command line applications, a highly extensible SNMP agent, perl modules and python modules. The suite includes:

- A graphical MIB browser (tkmib), using Tk/perl.
- A daemon application for receiving SNMP notifications (snmptrapd).
- An extensible agent for responding to SNMP queries for management information (snmpd).
- A library for developing new SNMP applications,

NET-SNMP is available for many Unix and Unix-like operating systems and also for Microsoft Windows.

In the following section we will discuss about installing and configuring the SNMP agent. The version NET-SNMP 5.3.1 will be used for this purpose.

SNMP packages to be installed on FC6 system:

- net-snmp-5.3.1-11.fc6.i386.rpm
- net-snmp-perl-5.3.1-11.fc6.i386.rpm
- net-snmp-devel-5.3.1-11.fc6.i386.rpm
- net-snmp-utils-5.3.1-11.fc6.i386.rpm
- net-snmp-libs-5.3.1-11.fc6.i386.rpm
- php-snmp-5.1.6-3.i386.rpm
- net-snmp-5.3.1-11.fc6.src.rpm

4.2 Generation of snmp.conf and snmpd.conf files

Automatic generation of snmp.conf and snmpd.conf are possible through the utility called snmpconf. Execute snmpconf which prompts an appropriate menu to create the configuration files.

The snmp.conf and snmpd.conf files define the behavior of:

- Agent
- Client
- SNMPv3 security settings
- MIB handling

Depending on the need, the above features can be set-up for use.

4.3 MIB Definition File

The new MIB definition file will be hand-coded by an individual for a particular feature. For all the SNMP projects that are being executed, this file will contain a specific OID which will be private in nature.

The generation of a new MIB module involves:

1. Creation of New MIB definition file - defines the elements to be managed,
2. Creation of a C file that contains the snmp related code and also the code for retrieving/setting the appropriate parameters as defined in the MIB,
3. A H file that contains the header related information

4.4 OID Definition

The OID definition is part of the MIB tree defined universally. Now, most of the enterprises who have their Private features - define their MIBS under the following tree:

.iso(1).org(3).dod(6).internet(1).private(4).enterprises(1).

.iso(1).org(3).dod(6).internet(1).private(4).enterprises(1).netSnmpTutorialMIB(4).nstMIBObjects(1).nstAgentModules(1)

4.5 Writing a Dynamically Loadable Object

1. Steps to build the shared object
   a. Get the Makefile file, the nstAgentPluginObject.h file, and the nstAgentPluginObject.c file.
   b. make nstAgentPluginObject.so
2. Steps to test the shared object via runtime MIB configuration
   a. Start the snmpd and observe the dlmod and nstAgentPluginObject modules interact using the debugging flag
3. Load the shared object into the running agent
4. Access the shared object's data using the snmpget command

4.6 TEMPLATE .c and .h file Generation

The TEMPLATE .c and .h file will be initially created using the mib2c tool. The mib2c tool will be used in such a way that the code generated will be of ucd-snmp style. The mib2c tool has to be executed giving the appropriate module-identity or the node definition in the MIB file.

Once the .c and .h file modifications are complete, a separate library (*.so) should be generated. The generated dynamically linkable library (*.so) should be included in the snmpd.conf file for the snmp agent to load this library *.so during startup. Re-start the snmpd with these changes, and check whether the *.so will be loaded by snmpd or not. The library will be loaded and the snmp agent is ready to accept inputs on the new MIB henceforth.

Installation of SNMP agent is the most important part of this paper.The installation process is described step by step. Once the installation is done the NET-SNMP package is configured. When the installation and configuration of SNMP agent is done the agent is ready to monitor the network and the system characteristics.

5 CONCLUSION

In this paper significance of the network management is illustrated by using the SNMP protocol. Thus the SNMP protocol was developed to facilitate the network management. Hence with simple SNMP requests (GET, SET, GETNEXT) and SNMP TRAPs, it is possible to manage a network. The SNMP implementation under Linux is done by using software called NET-SNMP package. NET-SNMP supports SNMPv1, SNMPv2 and SNMPv3 from the SNMP agent side and as a part of SNMP manager it makes use of on line commands.

NET-SNMP consists of tools and functionalities:

- SNMP API (Programming Application Interface).
- An extensible SNMP agent.
- On line commands to query a SNMP agent.
- On line commands to manage and generate SNMP Traps.

6 REFERENCES

1. SNMP – Simple Network Management Protocol. SNMP V1, SNMP V2C & SNMP V3. http://net-snmp.sourceforge.net/
2. "Simple Web" – Links for information and Network Management. http://www.simpleweb.org/
3. NETCOM System – Enterprise Management. http://www.netcom-sys.com
4. SNMP Info: Current SNMP Standards – Understanding SNMP MIBs. http://www.snmpinfo.com/
5. Linux SNMP Network Management Tools. http://linas.org/linux/NMS.html

6NET SNMP – An SNMP Library, tools. http://www.sourceforge.net

7. NET-SNMP: MIB OIDs - http://www.net-snmp.org
8. Network Management and Monitoring with Linux. http://www.snmplinux.com

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Asynchronous network for QAE: community schools of African developing countries

Kenedy Greyson, Mussa Kisaka, Damian Haule

University of Dar es Salaam, College of Engineering and Technology,
P.O Box 35131, Dar es Salaam.
kenedy@yahoo.com, kissaka@ee.udsm.ac.tz, haule@ee.udsm.ac.tz

Abstract There are three main enemies affecting development strategies in developing countries: ignorance, poverty and diseases. These problems are closely related. In order to deal with them, efforts should be directed equally to all of them. Although most developing countries have several strategies in poverty alleviation and eradication, fighting against diseases such as HIV/AIDS, malaria, etc., yet, fighting against ignorance is not taking place accordingly. Lack of access to quality education in the rural areas of the developing countries is in high level compared to that of the cities and big towns. Problems in education include the lack of qualified teachers, academic resources such as libraries, laboratories, etc. This paper address the method in which Information Technology can be used to deal with the problem in the rural areas where Information and Communication Technology (ICT) infrastructure is limited. This paper presents the proposed method of asynchronous network (asynch-Net) technology to be used for implementation and Quality assurance of Education (QAE) Rural Education Development project (RED-p) in the rural area community of rural areas of developing countries.

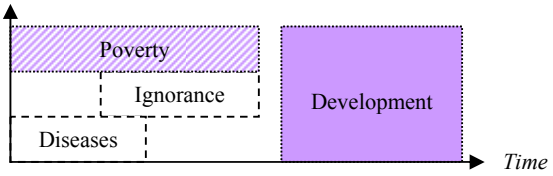
Keywords RED-p, Asynch-Net, QAE, ICT Keywords

1 INTRODUCTION

The poverty level in African developing countries is very high. It is prevailing in the rural areas that in the urban areas. The best way to deal with this problem is to deal with its related components, such as, ignorance and diseases. Knowledge (education) and health are main elements that give a hand to eradicate poverty. Basic step in fighting against ignorance could start with providing better education in the schools of rural areas. Research reports shows that, more than 80 percent of rural inhabitants have only primary education level. Tanzania Government education policy clearly explains that, I quote… "It is through quality education Tanzania will be able to create a strong and competitive economy which can effectively cope with the challenges of development and which can also easily and confidently adapt to the changing market and technological conditions in the region and global economy". This fact should be focused and equally distributed in all areas, rural and urban; otherwise, the poverty in the rural areas will never be reduces and hence it will stay for a long time.

Over 80 percent of Tanzania population lives in the rural areas [1]. This paper concentrates in Tarime district situated in Lake Zone, as a pilot of rural areas of developing countries. Tarime District has a population of 492,798 according to the 2002 population census. This district is divided in has 41 wards [1]. The 90 percent of the district inhabitants live in the rural areas at the average income of less than a dollar per household a day, which is the case of most rural areas in developing countries. Economical activities in Tarime include small scale agro-pastoral activities and fishing. This low income per household, results in lack of ability to afford good schools, mainly private schools. Figure 1, is a block structure scenario of three problems in the rural areas of developing countries. The block diagram shows that diseases and ignorance are the foundation of poverty. If either of the foundation blocks is removed, the level of poverty is reduced but still gets the support from the other block. Both blocks are supposed to be reduced so as to eradicate the poverty, and therefore development comes.

Figure 1. Block foundation of poverty in the rural areas of developing countries



1.1 Education Status in Tarime District

Government, through the Ministry of Education and Vocation Training (MoEVT) is in the progress of increasing the number of secondary schools to the level of at least one secondary in every ward. This is done through the local govern-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

ment strategies and at least the result is positive and Tarime is among the district of at least 80 percent has managed to reach the ratio of at least two secondary schools in each ward up-to-date. The main issue discussed in this paper is how to make this succeed in terms of Quality of Assurance of Education (QAE) where educational resources such as teachers in terms of lectures, libraries, laboratories are limited and some of them are not available at all. For example, in Zone A, that has four secondary schools, i.e., Bukwe, Nyanduga, Manga and Tarime, the National Examination Council of Tanzania (NECTA) results in the year 2004 is shown in Table 1. Pass is denoted by division I to III, where student go for further trainings such as high schools or technical colleges.

Table 1. General academic performance of secondary schools in the rural areas

School	Div. I	Div. II	Div. III	Div. IV	Div. 0	PASS (%)	FAIL (%)
Bukwe	0	0	0	58	42	0	100
Nyanduga	4	0	4	60	31	8	92
Manga	3	0	28	66	3	31	69
Tarime	0	5	30	56	9	35	65

1.2 ICT Status in Rural Areas of Tarime District

Although the penetration of ICT services in the rural area is limited by the lack of infrastructure in those areas, several efforts have been made to rectify the situation. Telephone operators, wired and cellular network have done significant efforts towards implementation of theirs services across the rural areas. Cellular networks, such as Celtel, TiGo, Zantel, and Vodacom, have their services available along the highways across the country. This has managed many villages in the rural areas to get cellular services in their areas. Tanzania Telecommunication Company Limited (TTCL) on the other hand, has made an effort of providing to over 75 percent of district an Internet access point. This access of Internet at the district level in the country has contributed in ICT access in the rural areas.

Despite of the significant efforts mentioned above, the use of ICT which has been identified as a key role in development is at low level, therefore, the effort to eliminate or eradicate poverty is still facing problems. Moreover, other social and economical services are left behind in this globalization scenario.

2 OBJECTIVES

The main objective of this research is to strengthen and expand opportunities for rural area populace in academic development in their own community without migrating to the urban areas. This is accomplished by opening academic doors for education development (continuing education) by making education resources (such as educational materials and accessibility) to be available in the rural schools of developing countries.

In order to achieve the main objective, several sub-objectives are to be considered:

i. To establish a cost effective network for ICT services in the rural areas,
ii. To develop relevant contents for people in the rural area, and
iii. To distribute contents at the respective areas and pre evaluate the result.

3 TECHNOLOGY

The choice of technology is the most important key in establishing a sustainable ICT service in a poor community. The costs of the technology include the purchasing of the equipments and running costs. The technology discussed in this section will cover, the network, contents development, and hence the distribution.

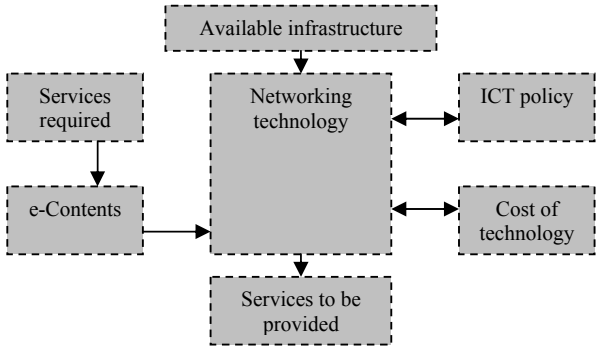
3.1 Network

For the ICT service to be made available, network infrastructure should be established. There are so many technologies available for establishing network. These are categorized in wired or wireless. Wireless technology has been identified the most cost effective technology in the rural areas of developing countries [4]. Wireless technologies standards such as Very Small Aperture Terminal (VSAT), Wireless Fidelity (WiFi), Wi-MAX, etc., have limitations in the rural areas of developing countries in terms of type of services, implementation costs, and running costs. Type of services to be provided and costs related to its implementation and running together will help to choose the type of network to be used in the low-income areas. Network can be real-time (RT) network to provide real-time services such as telephone services, or non real-time (NRT), sometimes termed as asynchronous network (Asynch-Net) for asynchronous services such as electronic libraries (e-Library), e-Lectures, etc. In this paper we are limited with asynchronous services required for academic purposes. The current project that works on the development of academic e-Resources (RED-p) uses ICT technology to support and enhance educational resources such as e-Library, teaching materials, etc.

Figure 2, shows steps and components to be considered when implementing ICT network and its services in the rural areas. Availability of other infrastructure is one of the components that supports in choosing the technology. For example, if there is no electricity, equipments to be used must have power management. Also, if there is a public transportation, then a mobile access point (MAP) technology may be used, etc [3]. Services required also another factor for choosing network technology. It should be noted that, the cost of technology and ICT policy may be relaxed.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Components for providing ICT services in the rural areas



Asynchronous Link- Access (ALA)

Due to the high cost of implementing ICT network infrastructure in the rural areas, the cost effective design, MAP is proposed. For the services to be sustainable in the rural areas, rural areas inhabitants must be willing and able to pay for services provided. Due to the low income situation in those areas, the cost effective infrastructure is the key element in its success. Real-time networks are not cost effective at the starting point. Due to the nature of the required services, asynchronous network can provide most of the services. For example, electronic email (e-mail), electronic library (e-library), electronic lectures (e-lectures), etc., can be recorded and played back when required. This means that, they can tolerate delays. Through Store-and Forward (SaF) technology, most required services can be delivered. Using Mobile Access Point (MAP) network technology, e-contents can be stores and taken to the respective destination using hybrid technology using wireless and public transportation means [4].

3.2 Contents Development

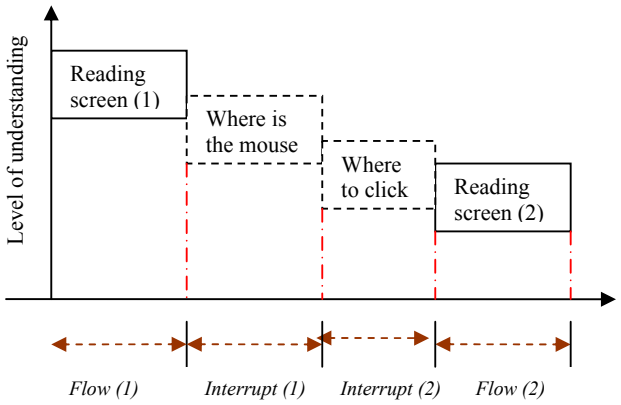
The education curriculum of secondary schools in Tanzania require finalists, form four students, to seat for at least seven (7) subject including compulsory subjects which are Basic Mathematics, Biology and Civics. Other subjects may include Chemistry, Physics, Geography and English. Alternatively, History, Kiswahili, Religions Education and others may be included.

Presentation is a key factor that may invite understanding environment of the class. Electronic contents (e-content) may be in three basic formats: text-, audio-, or video-.

i. Text Format

Through survey, it is observed that, information made in text-format is widely used when in hardcopy (printed books, handouts, etc) than when made in softcopy (acrobat files, Microsoft word, etc) by the secondary students. When using computer, the flow of understanding is interrupted by other operations such as: finding the position of mouse, finding (using eye) a place to click so as to proceed in the next page, moving the cursor to the point to click, to connect the last flow of understanding in the reading screen (1) and the present line in the reading screen (2). Figures 3, shows the effect of number of interrupt event on the level of understanding, neglect the operations of events done by eyes, and consider the operations using the printed book as a standard.

Figure 3. Level of understanding in Text-format



If lecture has an average of 15 pages the effect of interruption due to operation of computer depends on the speed of the computer, computer literacy level of the user, human interface form, etc. Assuming there is no other effects on the Interrupt effect, the interrupt number, is the number of events to take place before the next screen to appear. Interrupt(.) is the time taken in seconds while event of the interrupt is taking place. Note that, there is no any event taking place at less than 1 second events and its average duration. Assuming the fast event take 1 second. Considering the time used for the study to be T (minutes), the percentage of understanding level (PUL) is given by equation (i)

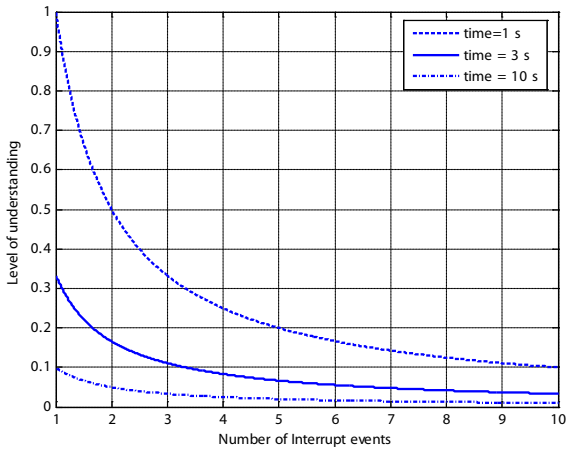
$$PUL = 100 - \frac{n\sum' \text{interrupt}(.)}{T} \times 1.67 \qquad \text{(i)}$$

This equation is derived with an assumption that the interrupt effect (IE) is directly proportional to the number of interrupt n and the interrupt (.).

$$IE = kn\sum' \text{interrupt}(.) \qquad \text{(ii)}$$

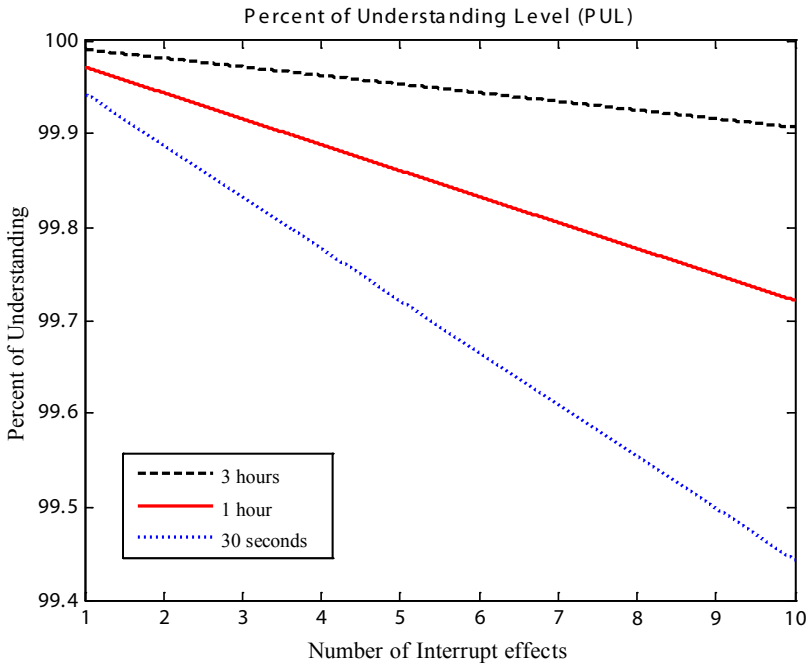
Where

Figure 4. Effect of interrupt number and their duration on the understanding level



In the results provided in Figure 4, assumes the minimal operations using acrobat reader (PDF) files. If the program to

Figure 5. Percent of understanding level (PDF file fromat)



be used takes longer period in one event and if the number of events is increased, then the level of understanding is more reduced. However, the percentage of understanding is shown in Figure 5. It is observed that, the for an hour of study where at least 10 number of interrupts and interrupts(.) of 1 second in average are applied, only maximum of 0.25 percent is lost. This being a linear function of number of interrupts, the loss of understanding is never beyond 1 percent. In that case, we conclude that, due to the cost of the text books the use of digital library can be used to assist secondary schools in the rural areas of poor communities.

ii. Audio and Video Format

The audio and video format have high acceptance compared to text format. Also, due to the less interrupt events required, level of understanding is improved. In fact, the cost of developing contents in audio and video format is more expensive and takes more time during editing. Survey made in Tarime district shows that, the use of video contents has more power in level of understanding than any other format. It is also observed that video format invites the class-interest among the listeners, thus and flow of understanding is maintained. Due to the bigger size of video files, the real-time required for these would be very expensive since they require high bandwidth, hence, the use of asynch-Net is proposed in the rural areas.

iii. Distributions of Contents

Unfortunately, due to the lack of electric power and paucity of computers, televisions, and video-players, rural learners are disadvantaged [2]. In order to provide e-Resources in the rural areas teachers and learners depend on rural community centers known as Multi-purposes Community Telecenter (MCT). Electronic distributions still remain in the form of SaF, where contents are brought to the MCT and accessed locally. Other means can be though radio broadcastings, TV stations, etc. However, the limited community radios and TV stations cost of airing the program in national radio and TV broadcasting is very expensive.

The RED-p network provides an asynchronous link to the schools through MAP where public transportation and wireless technology to provide SaF services. [4]. Multimedia Access to Resources for Teaching-board (smart-board), a web-based platform is the core technology for the training provision. RED-p network connectivity is a proprietary technology capable of transmitting video, sound at the reasonable Quality of Service (QoS) from the main center to the networked schools. Schools will have the access to the digital library owned by RED-p and receive recorded lectures from the main center database. Through RED-p network, you can retrieve and submit assignments, participate in discussion, take tests, etc.

4 METHODOLOGY

Endeavour effort on searching the best way to bridge quality education in the rural areas started at Dar es Salaam Institute of Technology when team of Engineers (RED-p) joined their hands to propose the solution.

4.1 Training teachers and students how to use computer to create and access educational resources

This project was intended for secondary schools where at the surveyed area of Tarime district Zone A, at least 80 percent of teachers has idea and basic knowledge of computer. The job ahead was to train how they can digitize their material using word processing, database and PowerPoint for being used in their e-Libraries. Few teachers from each school will be trained to be trainers for others and for students. Training will also cover the use of multimedia equipments. The pilot

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

area has four secondary schools and one Teachers Training College (TTC).

4.2 To develop an e-library in sound- and/or video-format resource center for teachers and students to access academic resources

The proposed method is in three phases, phase one starts from video format development where a professional digital video camera is used to record presentation. Qualified teachers (regardless of their location) are invited and asked to prepare topic presentation in the specified subject. Using the secondary school curriculum, the presentations are collected, edited and sent to the server database using MySQL that handle queries for future use. Phase two and three deals with text format where several e-library service providers are contacted and e-books are presented to the RED-p center.

Surveying the area and locating the schools cluster was the first step. During this survey, several factors were considered such as, access to the national grid (electricity), how the areas has been affected by the lack of educational resources based on the students' performance and total RED-p services demand. The geographic condition of the area was used to determine the link to be used for the network.

4.3 To evaluate the result and modify the model to be multiplied in the rural schools of Tanzania

This is the last step where the model is tested and output is analyzed for the decision in how to modify and correct stages taken.

5 CONCLUSION

In this paper, the use of extending wireless technology asynchronously has been identified as a cost effective network for a first step in introducing ICT services in the rural areas of developing countries. QAE can be improved when these services will be integrated in provision of academic materials so as to bridge the gap exists in access of academic resources in the secondary schools in the rural areas and those in urban areas. The similar network is expected to be used in different areas so as to cover other rural areas.

REFERENCES

1. Census (2002), 'Population and Housing Census', General Report, 2002.
2. APSCE, (2006), Research and Practice in Technology Enhanced Learning, World Scientific Publishing Company & Asia-Pacific Society for computers in Education, Vol. 1 No. 3 November 2006, pp. 297-308
3. Greyson, et al, (2006), 'Asynch-NET: Footstep for Always-on e-Services in the rural Areas of Developing Countries: Case Study Tanzania', Proceedings, 4th IEEE International Workshop on Technology for Education in Developing Countries (TEDC '06), Iringa, Tanzania July 12-14, 2006, pp. 5-6.
4. Greyson, et al, (2006), 'Extension of Asynch-NET Architecture for Telemedicine: Network for Clinic Centers in the Rural Areas of Developing Countries', Exploiting the Knowledge Economy: Issues, Applications, Case Studies Applications, Paul Cunningham and Miriam Cunningham (Eds), IOS Press, Vol. 2, 2006 Amsterdam, ISBN: 1-58603-682-3.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Designing web-based business application with multimedia data

Mohammed Hassouna

University of East London, UK
mhassonas@gmail.com

Abstract Although there are paradigm shifts in web application especially applications with multimedia content, developing methodologies of this kind of application are still in the initial stage of development. This essay discusses four key issues challenging design web based business applications with multimedia data. These issues are: design methodologies, Interactivity, Identifying virtual user requirements and heterogeneous composition of the designing team. As well, this essay suggests some guide lines and good practices to help practitioners and designers to face these challenges.

Keywords Multimedia methodologies, Interactivity, virtual user requirements, designing team

1 INTRODUCTION

Although there is a tremendous evolution in technology and applications of multimedia, there was no acceptable approach to develop this kind of applications [1]. According to a study conducted in Ireland by Barry and Lang, 61% of the surveyed said their thought about the future trends of developing multimedia applications encouraged them to develop business applications with multimedia data. The weakness of development practice of this type of applications could lead to lower productivity and higher maintenance costs or could lead to a complete failure of the application [2]. This essay discusses the design of this type of applications in particular designing the applications which delivered and used over the web.

This essay critically examines some key issues challenging designing web based business applications with multimedia data. These issues are design methodologies, interactivity, virtual user requirements and design team. The outline of this essay is as follows: section 1 calls to develop and improve the practices of multimedia application to meet the growing need for this type of applications at the same time this section shows the layout of this essay. Section 2 argues the adaptation process of the tradition design methodologies to match the special characteristics of web-based business application after that this part discusses some of the efforts to adopt the rational unified process (RUP). Section 3 discusses the interactivity issues and why it is important for the web based business applications at the end of this section there are some guidelines to the practitioners and the designers to utilize interactivity effectively in their applications. Section 4 talks about the difficulty of identifying the requirements of the virtual users then some suggestions presented at the end of this section to solve this problem. Section 5 discusses the heterogeneous composition of the designing team, which includes programmers, art designers, marketing specialists and others and suggests some ideas to overcome this problem. Section 6 concludes the essay with some guidelines and suggestions to meet the challenges facing design web-base business applications.

2 DESIGNING METHODOLOGIES

There are several methodologies, tools and techniques for developing software in general. This part of the essay focuses on finding an appropriate methodology for developing business applications with multimedia data. The methodology in software development is a general description of how to get something done. Avison and Fitzgerald [3] suggest some attributes for the methodology these attributes are:

- Consisting of a series of stages
- Consisting of a series of tools
- Having A philosophy
- Having a training scheme
- Including some techniques

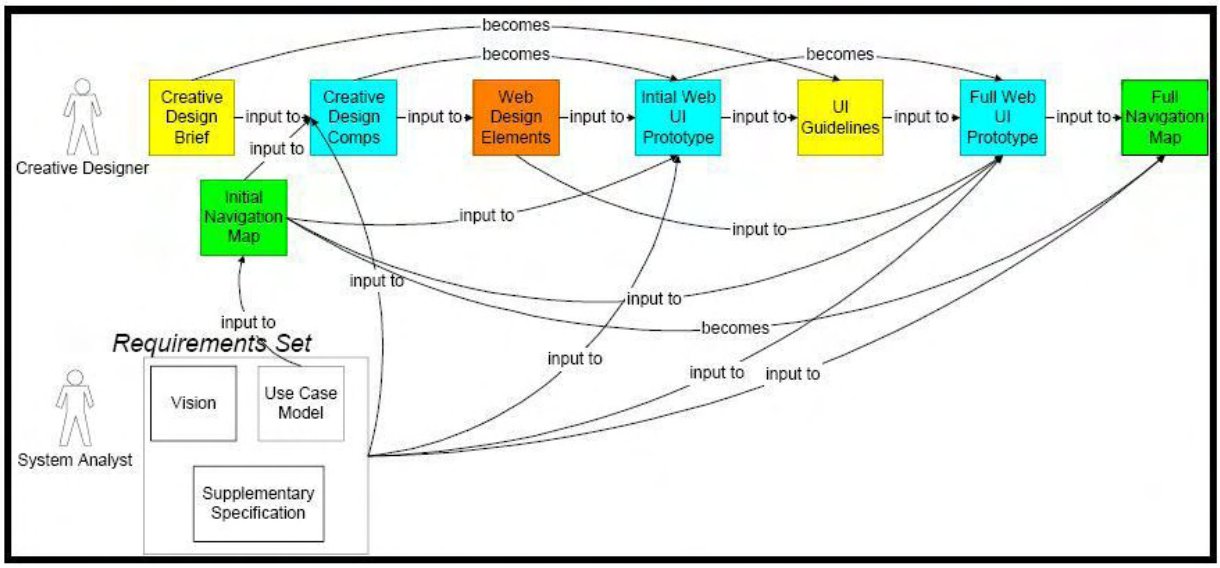
Avgerou and Cornford [4] have a different opinion. They put a list of questions that must be addressed by any methodology; these questions are:

- Why we need to develop the system and how we can do that.
- How the designing process can be managed.
- How we can teach this methodology.

Some of the general methodologies used on developing information system are listed below:

- Object Oriented Analysis and Design (OOAD)
- Dynamic Systems Development Method (DSDM)
- Structured Systems Analysis and Design (SADM)

Figure 1. Integrating Rational Unified Process and Creative Design Process.



- Information Systems work and Analysis of Changes (ISAC)
- Rational Unified Process and Unified Modeling Language (RUP), (UML)

Unfortunately all the traditional software developing methodologies are inappropriate for multimedia designing [5]. Business applications with multimedia data have special attributes and delivering them over the web adds another dimension to the scene. Designing this type of application requires addressing many issues like software development, media production, interface design and project management [6], [7], not that only but also caring with identifying user requirement. Researchers already try to adapt the traditional software methodologies to be adequate to developing multimedia applications [6]. Some of the methodologies used to develop multimedia application are [1]:

- Hypertext Design Model (HDM)
- Relationship Management Methodology (RMM)
- Object Oriented Hypertext Design Model (OO-HDM)

Studying and analyzing the methodologies listed above is out of scope of this essay. Moen Design Center application is one example of business application with multimedia data (www.moen.com) [8]; Moen gives the user the opportunity to design their own bathrooms and kitchens. To my knowledge, there is no fully adequate methodology suitable for designing this sort of application but there are tries to adopting existing methodology.

Nielsen said that "Software design is a complex craft and we some times arrogantly think that all its problems are new and unique" [5],[9]. Lang points to the fact that methodologies should be inherited from the root disciplines like software engineering, human computer interaction, graphic design, marketing and other disciplines. In this instance, I share Nielsen and Lang in their thoughts and from all the argumentation we can realize how it is important to learn from previous experiences in developing and designing application. For that I think what we need is to start where others stop. Rational unified process (RUP) is one of the methodologies adapted by the researcher and practitioner. Next section presents an adapting process to adapt Rational unified process methodology to make it suitable for multimedia applications.

Word and Kroll adapt the rational unified process (RUP) to combine creative design and software engineering [10]. They propose a way to reach a common language that anyone involved in developing – manager, programmer, artiste and architecture- can understand what the application will do (see Figure 1.) On the other hand, Bygstad investigates the same methodology (RUP) but he found that there is no sufficient support to the internal and external technical integration [11], which makes integrating business process with developing process a complex challenge. Bygstad suggests a theory extension to (RUP) to overcome this shortness. The two attempts of adaptation are a contribution toward finding a suitable methodology for designing the business application with multimedia data and we still need more research to cover all the issues which impact designing this type of application such as software development, interface design media production, and project management

3 INTERACTIVITY

According to the Longman Interactive English Dictionary, "interactivity "is the noun of the adjective interactive that refers to

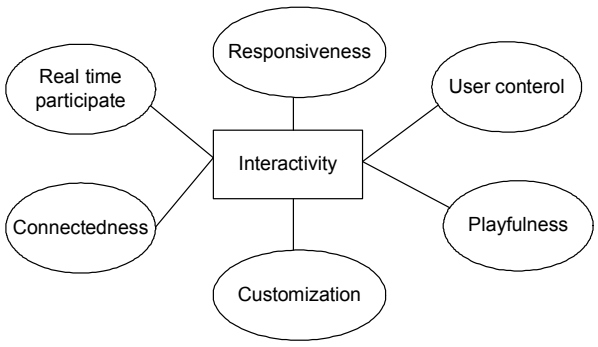
1. Teaching methods, processes involve people to work together
2. An adjective for a computer program linking the computer and the person who use it.

The online version of this dictionary adds "and does things in reaction to your action". El-Gayar and others argue the definitions of the term interactive and they conclude that there is no agreement on what interactivity is, and this term may have different meanings to different people [12]. Iuppa [13] shares El-Gayar's opinion and he demonstrates that by comparing interactivity in different fields like gaming, learning and internet for example when he talks about interactiv-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

ity on the internet application he points to the existence of the interactivity on all internet application. Iuppa [13] said we could not imagine that buying a book from Amazon.com for instance without interactivity with the Amazon's web application. Gassaway [14] said interactivity is a dramatic demonstration of user control [15]. The definition of Dholakia et al. [16], more appropriate to describe interactivity on internet applications. They suggest six components must be taken into consideration when we talk about interactivity these components are: real time interactions, user control, customization, connectedness, playfulness and responsiveness, see Fig 2.

Figure 2. Interactivity components



The issue here is what is important of the interactivity for designing business applications with multimedia data. Interactivity is utilized on these applications to encourage the visitors to revisit the web site [16], to increase the satisfaction of the user [17],[12], and to transform the application to effective marketing and sales tool [18]. Researchers and practitioners in the interactivity field focus essentially on the definitions, evaluation, and application of the interactivity besides designing issues [12]. This essay mainly is concerned with the designing issues of the interactivity and the kind of interactivity to be considered while designing the business application with multimedia data. The designer of the multimedia application and especially web based business application must take into consideration interactivity by utilizing interactive elements in the application. Dysart [18] lists some of the interactive elements which can be used on the web site; some of these elements are: Chat rooms, Ask the expert, Search engines, Auto-responders, online newsletters, Members only area ,Online product demo, Online bulletin bored, Online video conferencing ,Interactive forms and tours and Request-for-information forms

El-Gayar [12] propose a comprehensive framework addressing the interactivity elements based on the six components which presented by [16], see Figure 3.

In this instance I share El-Gayar [12] in their thought about the important role of interactive multimedia system in the marketing and advertising. At the same time I consider El-Gayar's framework a practical guideline for the designer of the multimedia application to make their application compete in today's competitive environment. Many industries are already dominated by the interactivity for example employing virtual real estate tours on the real estate industry El-Gayar et al [12]. Consequently we need efficient utilizing of the interactivity in the business application.

4 IDENTIFYING VIRTUAL USER REQUIREMENTS

In the early days of computing, the contribution of the user was highly insignificant. Change in times has resulted in a move from the basic functionality of a program to the usability which has resulted in a 'power change' thus allowing for a greater need for user involvement [19]. The introduction of the user into the design process has resulted in some

Figure 3. El-Gayars' framework of mapping multimedia and web features.

Interactivity dimensions	Multimedia/Web features	
User control	• Alternative options for site navigation • Linear interactivity, where the user is able to move (forward or backwards) through a sequence of contents	• Object interactivity (proactive inquiry) where objects (buttons, people or things) are activated by using a pointing device.
Responsiveness	• Context-sensitive help • Search engine within the site	• Dynamic Q&A (questions and responses adapt to user inputs)
Real-time participation	• Chat rooms • Video conferencing	• E-mail • Toll-free number
Connectedness	• Video clips • Site tour	• Audio clips • Product demonstration
Personalization/ Customization	• Site customization • Bilingual site design	• Customization to accommodate browser differences
Playfulness	• Games • Software downloads • Visual simulation	• Online Q&A • Browser plug-ins (e.g., flash, macromedia, etc.)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

complication for the design team this is because the need for effective communication during the design stage. This is to ensure that neither functionality nor usability of the multimedia product is compromised [20].The inclusion of the perspective of the user is of great importance to multimedia applications since the user will not invest in products if they are perceived as low quality [21]. The designer must take into consideration needs of the users to design effective web based business application. User requirements' engineering is a software engineering issue that tries to identifying user requirements and ensuring that a reliable requirement specification can be developed [22]. The Usable system is the system that fulfills the requirements of all users [7]; accordingly, there is an intersection between the user requirements and the usability.

The issue that must be raised here is how to identify and collect the virtual user requirements of the web based application. Collecting requirement from this kind of user is most difficult [5]. The main cause of the complexity of requirements capturing is combining requirement capturing and designing in one phase at the beginning of the project lifecycle [23]. I agree with MacDonald and Welland in their thoughts because they are based on the results of their interviews with a number of people involved in developing web application. Smith et al. suggest criteria for user requirements in multimedia application to ensure the user can easily achieve the following:

- Acquire needed information when needed.
- Locate information that is needed in the expected locations.
- Recognize the meaning of a link so as to ensure that results provided

match the required need.

- Obtain a summary of the information available [24]

The three main factors that can affect the interaction level between the user and the design team are usability, accessibility and learnabilty [20]. "According to ISO 9241 part 11," usability is defined "as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241-11). The key things to look out for in usability are: effectiveness, efficiency and satisfaction [25]. Other characteristics to look out for in effective usability are simplicity in design, consistency, reduced memory load, constructive feedback mechanism, good troubleshooting support, effective shortcuts and well designated exits [26]. Accessibility deals with the ease and transparency to get the user to interact effectively with the design system. In order for accessibility to be truly achieved, alternate versions need to be design to cater for the need of a broad spectrum of users [27]. Learnability deals with the interaction of the user without need for additional learning materials [20].

To contribute to solving the problem of defining virtual user requirements, Lang contrasts between virtual users and mass-marketing [5]. Therefore he points to the importance of combining marketing research techniques and usercenter requirements definitions techniques. Marrenbach adds to that the contribution of the domain experts and technology experts [25]. In this instance. I think, to deal with the virtual user requirements, we need a cooperative effort not only from the designers and the developers but also from the marketing team and the domain expert besides taking into consideration usability and accessibility and learnability.

5 HETEROGENEOUS COMPOSITION OF THE DESIGNING TEAM

One of the major issues that affect multimedia application is the multidisciplinary nature of the design team. In most cases, there are several design teams involved in a project each with their own unique skills, ways of communicating as well as interpretation of a given task. It is, therefore, of great importance that the issue of diversity within the design group is dealt with appropriately in order for it not to become a deterrent to the accomplishing the main aim of the project. The importance of effective communication and sharing information is, hence, of great importance within the various design teams. Once effective communication is set in place, a whole range of management problems is avoided which might have reduced productivity [28]. It should be noted that another issue facing a multi-disciplined team is actually getting the various groups to work as a team keeping the main goal in focus.

The cultural diversity of the design teams is something that should also be considered. A method of dealing with the issue of diversity within the team is the agile approach. Agile approach focuses on:

- The interaction between people rather than on processes
- Functioning software rather than an all inclusive document
- Relationship with the customer rather than on dealing with the contract

Agile approach focuses on people and interaction and that is of great benefit to multimedia design application [29]. The diverse nature of the design teams involved in the multimedia application has to be considered when improving methods to deal with the various challenges that will be encountered [5].

The person that has the most tasks to do within a team like this is the project manager. The project manager has to have an understanding of each individual's function within the group, the impact their contribution makes to the bigger picture of the project, as well as what keeps them inspired. The project manager has to play the role of both the coach and the manager to effectively accomplish the task of the project. A typical multimedia project team consists of three major groups: content, graphic design and technology groups. It is the job of the project manager to manage activities across these three major groups effectively as well as managing communication effectively between the team and the client [30]. The importance of communication in successfully achieving the goal of getting the project accomplished with a multidisciplinary team cannot be overemphasised enough. It should be a two-way process which also for feedback from the team. It should be noted that the translation and relaying

of messages effectively by the project manager is a core part of this process [30].

6 CONCLUSION

Business applications with multimedia data have special attributes and delivering them over the web adds another dimension to the scene. Designing this type of application requires addressing many issues like software development, media production, interface design and project management [7]. We need to start from where others stopped and we do not need to reinvent the wheel. In this instance we can use the adapted version of the rational unified process (RUP) methodology which I mentioned earlier to develop the multimedia applications. Interactivity plays pivotal role in multimedia applications for that we should look at Interactivity in the wider sense.The comprehensive framework proposed by El- Gayar et al [12], can be considered as a practical guide lines to designers and practitioners. The designer of the multimedia application and especially web based business application must take into consideration interactivity by utilizing interactive elements in the application.

Identify and collect the virtual user requirements of web based applications are most difficult. Concerted efforts from all people involved in designing and developing multimedia applications are required to overcome this difficulty. These efforts should go beyond design team to include marketing team and domain experts. Marketing research and user-center requirements definitions techniques beside usability, accessibility and learnability requirements are essential tools and techniques to Identify and collect the virtual user requirements.

Multidisciplinary nature of the design team is one of the major issues that affect multimedia application. Effective communications are necessary to face the management problems. These communications should be a two-way process and the project manager is a core part of this process and he/she should take the responsibility.

More researches are required to adapt the traditional design methodology and techniques. Analyzing the existing methodologies and design techniques are essential to reach to suitable sets of methodologies and techniques that can meet the special characteristics of multimedia application.

REFERENCES

1. Barry, C. & Lang, M. (2001) "A Survey of Multimedia and Web Development Techniques and Methodology Usage". IEEE Multimedia. 8(3), 52-60. (Barry and Lang 2001)
2. Ginige, A. and Murugesan, S. (2001) " Guest Editors' Introduction: The Essence of Web Engineering— Managing the Diversity and Complexity of Web Application Development ". IEEE MultiMedia 8, 2 (Apr. 2001), 22-25.
3. Avison D, Fitzgerald G, (2003), "Information Systems Development: Methodologies, Techniques, and Tools" Third Edition, McGraw Hill
4. Avgerou C and Cornford T, (1998) "Developing Information Systems: Concepts, Issues and Practice", 2nd Edition,Macmillan
5. Lang, M. , (2005) " Issues and Challenges in the Design of Web-based Hypermedia Systems." In Pagani, M. (ed), Encyclopaedia of Multimedia Technology and Networking. Hershey, PA: Idea Group Publishing.
6. Engels, G.,S.,Neu,B.,(2003)," Integrating Software Engineering and User-centred Design for Multimedia Software Development", IEEE
7. Martin S., Bolissian J. M., Pimenidis E. (2003) "PURE and SIMPLE: A framework for the evaluation of multimedia products ", in proceedings of the 10th European Conference on Information Technology Evaluation, 25-26 September 2003, Instituto de Empresa, Madrid, Spain, pp. 431-440
8. Perfetti C. and Spool J. (2002)"Macromedia Flash™: A New Hope for Web Applications", User Experience White Paper, http://www.adobe.com/resources/ business/solutions /user_centric/flash_web_apps.pdf accessed on 28/4/2007
9. Nielsen, J. (1997), "learning from the real word", IEEE Software,14(4),98-99
10. S. Ward, P. Kroll, 1999, 'Building Web Solutions with the Rational Unified Process: Unifying the Creative Design Process and the Software Engineering Process', Rational Software Corporation, July1999. http://www.rational.com/
11. Bygstad, B. (2004). "Controlling Iterative Software Development Projects: The Challenge of Stakeholder and Technical Integration". Proceedings of the 37th Annual Hawaii international Conference on System Sciences, IEEE
12. El-Gayar, O.F., Kuanchin Chen, and Kanchana Tandekar (2005). "Multimedia Interactivity on the Internet ". In: Encyclopaedia of Multimedia Technology and Networking (Ed.) Margherita Pagani Idea Group Inc. Hershey, PA.
13. Iuppa, Nicholas.(2001)," Interactive Design for the New Media and the Web". London: Focal, 2001.
14. Gassaway, Stella et.al.(1996), "Designing Multimedia Web Sites". Indianapolis: Hayden Books.
15. Dysart, J. (1998). "Interactivity: The Web's new standard" . NetWorker: The Craft of Network Computing, 2(5), 30-37.
16. Dholakia, R.R., Zhao, M., Dholakia, N., & Fortin, D.R. (2000). "Interactivity and revisits to Web sites: A theoretical framework" , Research institute for telecommunications and marketing.http://ritim.cba.uri.edu/wp2001/wpdone3/Interactivity.pdf accessed 26/4/2007
17. Rafaeli, S., & Sudweeks, F. (1997). "Networked interactivity". Journal of Computer- Mediated Communication, 2(4). http://jcmc.indiana.edu/vol2/issue4/ rafaeli.sudweeks.html
18. Dysart, J. (1998). "Interactivity: The Web's new standard" . NetWorker: The Craft of Network Computing, 2(5), 30-37.
19. Grudin, J. (1991). "Interactive systems: Bridging the gaps between developers and users". IEEE Computer, 24, 4, 59-69.
20. Elsom-Cook, Mark.(2000). "Principles of Interactive Multimedia". New York: McGraw-Hill Publishing Co.
21. Ghinea, Georghita and Sherry Y. Chen (2006). "Digital Multimedia Perception and Design". Hershey: Idea Group Publishing,.
22. Van der Poll, J. A., Kotzé, P., Seffah, A., Radhakrishnan, T., and Alsumait, A. (2003). "Combining UCMs and formal methods for representing and checking the validity of scenarios as user requirements". In Proceedings of the 2003 Annual Research Conference of the South African institute of Computer Scientists and information Technologists on Enablement Through Technology
23. McDonald A. and Welland R.(2001), 'Web Engineering in Practice', Proceedings of the Fourth WWW10 Workshop on Web Engineering, Page(s): 21-30. http://www.dcs.gla.ac.uk/ ~andrew/webe2001.pdf accessed on 24/4/2007
24. EMMUS a, (European MultiMedia Usability Services), "Multimedia Design: The role of Guidelines" http://www.ucc.ie/hfrg/emmus/MCGDoc/ guidelines.html accessed on 26th April 2007
25. Marrenbach, J. (1999)," Rapid Development and Evaluation of Interactive Systems". Proceedings of the 5th ERCIM Workshop User Interfaces for All , Volume Report 74, pp. 81-86, Dagstuhl

26. EMMUS b, (European MultiMedia Usability Services)."Multimedia & Usability Principles: An introduction", http://www.ucc.ie/hfrg/ emmus/guidelines/ d11what.html accessed on the 26th of April 2007.

27. Lynch, Patrick and Sarah Horton.(2001), "Web Style Guide". New Haven: Yale University Press, 2001.http://webstyleguide.com/multimedia/access.html accessed on the 26th of April2007.

28. Phillips, R. A. (1996). A Methodology for Developing Educational Applications of Interactive Multimedia, Paper presented at the Third International Interactive Multimedia Symposium, Perth, Western Australia, 309-315 http://www.ascilite.org.au/aset-archives/ confs/ iims/1996/lp/phillips.html accessed on 27th April 2007.

29. McDonald A. and Welland R., b (2001) 'Agile Web Engineering (AWE) Process', Department of Computing Science Technical Report TR-2001-98, University of Glasgow,Scotland,2 December 2001.

30. Shelford, J., Thomas and Gregory a. Remillard. (2003), "Real Web Project Management", Boston: Addison-Wesley,.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007

A Survey of DRM in digital video

John F. Duncan

Indiana University, Bloomington, IN
johfdunc@indiana.edu

Abstract With the rapid expansion of video technology, intellectual property (IP) holders are increasingly concerned with controlling the distribution of copyrighted video. These IP holders, predominantly large industry groups, use digital rights management (DRM) techniques designed to restrict use of their copyrighted works. However, these DRM systems have many potential pitfalls, from unintentional consequences to their restriction of legal consumer practices.

This paper illustrates the ways in which the techniques used in modern video DRM are inherently incapable of addressing large-scale commercial piracy. This result, along with other observations, suggests that most modern video DRM is designed to control the behavior of consumers, not pirates.

Consumers exposed to DRM find themselves increasingly able to do less with legitimately purchased video content. The resulting gap in expectations only strengthens commercial piracy, by creating a viable business model for pirates – providing the same content in a more usable format.

Keywords DRM, video, fair use

1 INTRODUCTION

DRM (Digital Rights Management) is a broad term for techniques and systems designed to manage access to copyrighted information. Entertainment content, including videos, has been the primary focus of DRM systems due to the perfection of digital copies, as opposed to the decay in analog copies. Exact definitions of DRM are not standardized. The systems themselves and claims as to the goals and the means of these systems vary widely. The current spectrum of video DRM technology is the result of a continuing effort by industry groups to retain control over all aspects of video consumption. The increasing restrictions these DRM systems impose on consumer use are justified as anti-piracy measures. Frequently it is also suggested that without these controls, the market would be severely impacted, leading to the halting or slowing of new content production due to commercial infeasibility. [1]

The industry uses the term 'piracy' to mean copyright infringement, as demonstrated by the MPAA's definition: "Anyone who sells, acquires, copies or distributes copyrighted materials without permission is called a pirate." [2] Industry groups also often conflate piracy with theft of physical property. The MPAA claims: "Downloading a movie without paying for it is no different than walking into a store and stealing a DVD off the shelf." [2] However, the Supreme Court noted in Dowling v. United States, 473 U.S. 207 (1985), that "interference with copyright does not easily equate with theft." [3] In this paper, the term 'commercial pirate' will be used to refer to someone who engages in large-scale copyright infringement for profit, to differentiate them from consumers who often engage in small-scale copyright infringement for personal uses. The degree to which this small-scale copying is covered by fair use provisions is an open legal question.

Online file trading, also distinct from commercial piracy, has been one of the driving forces behind the industry's increased deployment of DRM. Currently, the primary types of file traded online are audio files, but with increasing home bandwidth and improvements in compression technology, a rise in video file trading is likely. [4] In 2006, for example, the NPD group estimated that eight percent of American households had downloaded a video from P2P services, as opposed to two percent that had paid to download a video file. [5]

Section 2 provides an overview of the major organizations involved in developing and implementing DRM systems and standards. Section 3 discusses the major video DRM technologies, including specific ones like AACS and broad categories like digital watermarking. Section 4 examines the major types of commercial video formats and the particular DRM systems commonly used in each format. Section 5 details how the aforementioned DRM systems affect consumer rights. Finally, Section 6 proposes a new direction for the industry to pursue in order to more effectively serve both their goals and those of the consumer.

2 ORGANIZATIONS INVOLVED IN DRM

DRM has primarily been developed by large corporate IP holders, who have had the most to lose from unauthorized reproduction and distribution of their IP. In 2003 alone, the IIPA estimated losses due to unauthorized reproduction and distribution at $22 billion. [6] The precise way in which this figure and other similar ones were determined is unknown and subject to debate about size as well as direction. [7] Certainly, the organizations involved in DRM have clear motivations for protecting their revenue streams and business models. In particular, there is increasing concern with illegal distribution over the internet. The Motion Picture Association of America (MPAA) in particular has argued that the availability of video files on the Internet directly prevents the content owners from recovering the production costs through secondary market sales. [4]

The MPAA has been the one of the strongest proponents of video DRM. Representing the major studios in Hollywood, they have been a major force behind the broadcast flag, DRM systems in physical media, and watermarking to protect against copying by awards screeners. The MPAA has stated that without video DRM in digital broadcasts, its members are unwilling to release their content to this media. [4]

The analysis of the broadcast flag in [4] concludes three things are protected in theory: the content being broadcast, the corporate business model tied to scheduled broadcasting, and the larger IP paradigm. In practice, the broadcast flag fails to protect the actual content from piracy, but succeeds in protecting the MPAA's business model and ideas about IP. This is consistent with the MPAA's drive to reverse social norms regarding file sharing to place more power in the hands of content owners. [4] As for fair use, Michael Lesk notes that publishers have repeatedly asserted that they have no responsibility to protect fair use rights either in theory or practice. [8] Their stance toward DRM technologies that prevent fair use rights is consistent with this position.

While discussing the specifics of the proposed Video Home Recording Act of 1996, a forum of the industries involved (the MPAA, CEMA, BSA, ITIC, and RIAA) evolved into an industry group known as the Copy Protection Technical Working Group (CPTWG). The CPTWG in turn spawned working groups such as the Digital Transmission Discussion Group (DTDG) and the Data Hiding Subgroup (DHSG). [9] The CPTWG also formed the Broadcast Protection Discussion Group (BPDG), one of the main designers of the broadcast flag. [4] When the Consumer Electronics Manufacturers Association (CEMA) presented their list of desirable attributes for a DRM system to the CPTWG and the MPAA, five major areas of consumer rights were outlined. These included fair use rights such as time shifting and place shifting, free copying of non-premium broadcasts, equipment and content interoperability and the ability to upgrade systems without rendering previously recorded material unviewable. [9] When the MPAA later issued its list of twelve points, none focused on customer rights, and the original CEMA recommendations in this area were rejected. [9]

3 DRM TECHNOLOGY

Different technologies are employed based on which of the major video distribution channels is being targeted and which organization owns the content. The efficacy of these technologies depends not only on their actual implementations but also on the social and societal norms of their regions of deployment. The common varieties of video DRM and several specific well-known systems are discussed below.

While the MPAA states that it fights piracy by "encouraging the development of new technologies that ensure movies can be made available legally over the Internet and other digital media," [10] none of these DRM systems are legally required for content distribution. Neither do they target commercial piracy. Rather, these technologies are designed to prevent consumer-level copying.

Many of the DRM systems discussed here rely on secret keys, at the player level, the disk level, or even for individual sectors of a disk. If one of these keys is compromised, a desired feature of most DRM systems is that the key be revocable. Should a key compromise be detected, compromised devices can be revoked by the manufacturer. The revocation consists of ceasing to distribute content that can be decrypted by that compromised key. [11] Revocation often targets a set of devices, rather than a specific compromised device. The next time a legitimate user of one of these devices loads an updated DVD or connects to the internet, the device may no longer play their new media. [12] However, revocation alone may not always be sufficient to prevent compromises, as noted specifically in the discussion of AACS.

3.1 CSS

The Content Scrambling System (CSS) marked the first wide-scale deployment of encryption for DRM. [11] CSS was developed for DVDs by Matsushita in 1996 - it was designed to prevent byte-for-byte copies of a disk from being made, as well as encouraging manufacturers to make DVD players that supported the full range of DRM features desired by the industry. This was done by requiring hardware manufacturers that had a license to use CSS to also certify their hardware as compliant with other DRM measures. Since DVDs encrypted using CSS could (in theory) not be read by players without a license, the industry was in a position to force hardware manufacturers to comply with CSS in order to appeal to purchasers of CSS-encrypted DVDs. [13]

On a CSS-protected disk, there is a visible block of encrypted content (the digital video content the consumer wants to view) and a hidden portion that includes encrypted key information. When the disk is inserted into a CSS-compatible player, the host system or computer and the DVD drive use a challenge-response system based on shared secrets to establish that they are both in compliance with CSS. During

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

this process they agree on a session key, which will be used to encrypt the transmission of other keys over the bus between them. The player then uses its player key to decode the disk key. With the disk key decoded, the player sends the title and disk keys to the host, encrypting this message with the session key. To play the DVD, the player then sends each sector to the host. The host first uses the disk key to decrypt the title key, and then uses the title key to decrypt the sector key, which allows it to decrypt the sector and play it. [14]

CSS is relatively simple, and its key length is only 40 bits, which makes it relatively easy to brute-force. [15] In addition to these problems, CSS suffered from a limited number of possible player keys and the fact that all players of the same type used the same key, making revocation difficult. These weaknesses in the CSS algorithm lead to the development of DeCSS, a program developed in 1999 by Jon Johansen and two other anonymous programmers. [16] DeCSS was the first program to defeat the CSS algorithm – since then, many others have been developed, and the original DeCSS is somewhat obsolete. [12] DeCSS and its successors allow users to play DVDs under linux, skip commercials, circumvent DVD region coding, and make copies of DVDs. [17]

Possessing or distributing DeCSS or similar programs is illegal in nations that implement many of the World Intellectual Property Organization (WIPO) anti-circumvention recommendations. Even linking to pages that host the software has brought legal action. The most famous of these lawsuits was an action against Eric Corley, a maintainer of www.2600.com, by a number of movie studios. [17] While industry organizations have aggressively worked to portray tools like DeCSS as the main source of illegally distributed movies, these movies primarily come from illegal in-cinema recordings or movie screeners. [4]

3.2 AACS

AACS, or the Advanced Access Content System, was developed in 2005 by a multi-company consortium including Disney, Intel, Microsoft, Panasonic, Warner Brothers, IBM, Toshiba, and Sony. [18] AACS was adopted as the major DRM technology by both Blu-ray and HD DVD. AACS uses AES, the Advanced Encryption Standard, to encrypt content using a variety of keys. AACS differs from CSS in the way that it manages keys. Unlike CSS, which used a single key for all players of the same model, AACS allocates a unique set of decryption keys to each player. [19] This was done in an effort to increase revocability.

Devices that use AACS are given a set of secret keys known as device keys. Each device key enables the player to calculate a number of processing keys. On a disk that uses AACS, the video content is encrypted using a title key unique to a given printing of a movie. Several copies of this title key are stored on the disc, each encrypted with a different processing key. When the disk is placed into an AACS-compliant player, the player tries all of its processing keys until it is able to decrypt the title key. The title key is then used to decrypt the movie. [20] The multiple copies of the title key are stored in an area known as the Media Key Block (MKB). The term 'subset difference set' is used to refer to the group of players able to decrypt a particular copy of the key in the MKB. By altering the entries in the MKB, it is possible to define these groups differently, allowing compromised players to be disabled for discs with new versions of the MKB. [21]

The AACS system contains an inherent flaw when used on personal computers. To play the content encrypted on a disk using AACS, the various keys involved (the device key, the title key, etc.) must all reside in main memory at some point. While steps can be taken to protect these values through obfuscation, a determined seeker can find them. Because of this, the device key for a particular device can be retrieved and used to create non-AACS compliant systems for playback and duplication.

By December of 2006, exploits using this flaw in AACS became widely known. These began with the publication of BackupHDDVD, a program designed to decrypt AACS-protected disks, either to create backups or illegal copies for distribution. [22] This lead to the circulation in the months between then and April of 2007 of various processing keys for the AACS system, the most famous of which is a 128-bit hexadecimal number known as the 09 F9 key for its beginning hex digits. [23] The widespread publication of this key on major user-submission news sites like SlashDot and Digg led to DMCA takedown notices from the MPAA and the AACS LA. These notices were also sent to Google for indexing these sites. [24]

One direct result of this widespread key distribution was that the AACS LA decided to officially revoke the 09 F9 key, removing the ability of players that used that key to play content encrypted with AACS after April 23, 2007. However, in May of 2007, a new key known as 45 5F began to circulate, amid claims that it was the replacement for the 09 F9 key. [25]

The AACS licensing agreement requires that hardware manufacturers be given ninety days notice prior to one of their keys being blacklisted. This advance notice is required so that the manufacturers have enough time to properly update their products with new keys. This suggests that key revocation in the AACS system is ineffective. The time required by attackers to obtain a new key is much shorter than the notification period guaranteed to hardware manufacturers.

3.3 Digital Watermarking

The idea behind watermarking is simple – data can be hidden within a larger block of data in such a way that it minimally distorts the original information. That data can then be recovered during playback of the larger data file. This hidden information can be used for many purposes, including identifying the copyright holder, determining the ways in which the content can be played or copied, or establishing authentication. Watermarks have been extensively used to mark ownership and implement DRM. [26]

Watermarking was first introduced to digital video by the Data Hiding Subgroup (DHSG) of the CPTWG in 1997.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Their original proposal relied on drive detection of pressed disks vs. burned disks – with watermarks on burned disks indicating illegal copies. [11] Watermarking was seen as a simultaneous solution to the illegal copying of DVDs, digital cinema reels and movie screeners. Indeed, watermarks have been used in legal action in regards to such copying. [12]

The main difficulty in implementing watermarks comes from the opposing requirements that they cause minimal perceptible distortion to the viewer and that they survive attacks such as compression, which removes small details unlikely to be noticeable to the viewer. [26] The goal for modern video watermarking systems is to be so robust to distortion that removal results in undesirable degradation of the original video source. Additionally, detection of the watermark must remain rapid (the declared DVD detection time is 10 seconds) in order for the player to be able to respond to the consumer. All of this must be implemented in the limited space available for additional hardware inside the physical drive. Watermarks must have enough payload capacity to implement all the desired modes of operation for consumer electronics, and to contain unique identifiers for user tracking. [27]

Defending watermarks against attacks is difficult because legitimate users of video streams are likely to process the video in ways (such as compression) that damage the watermark in the course of their normal use of the content. [12] Other difficulties are introduced by different distribution channels, such as broadcast media. While the content distributor would like to send a unique id to each recipient, a broadcast signal by definition involves a single version being sent to all broadcast receivers. [28] Watermarks are also notoriously difficult to evaluate. Each industry group uses different test cases and methods. [29]

Watermarks are often used to implement copy control. Three main states for media are used: Copy-never, copy-once, and copy-always, of which content owners traditionally use only the first two. Of these, copy-once is the most interesting, as it is superficially designed to preserve at least some of the consumer's rights by allowing backup copies to be made, while also being designed to prevent further copying of the backups (usually by watermarking the backups as copy-never). The two main methods of implementing this principle with watermarks are secondary watermarking and ticketing.

In the first system, a secondary watermark is added to the copy – the presence of two watermarks becomes the identifying feature of the copy-never state. In ticketing, the media features a unique identifier in the watermark. During recording, this identifier is fed through a one-way function to generate a new identifier, in such a way that the number of times this has been done is detectible. Thus, a given id can be issued so that only one copy can be made by virtue of the function used. The failure modes for these two systems are different. For the first, failure to insert the second watermark results in a copy that is also copy-once. In the second, mishandling of the ticket may prevent copying even where it should be permissible. Thus content providers must decide whether a failure mode in their favor or that of the consumer is preferable. [27]

Failures of the larger watermarking process are also a concern. For example, if watermarks are used to uniquely identify recipients of copyrighted works so that any subsequent illegal distribution of the works can be traced back to the original source, a possible danger could be the content provider issuing the same unique identifier to multiple recipients. [28] False identification (detecting a watermark when none is actually present) must also remain small (less than 10^-12 percent for DVDs), or legitimate users will be too frustrated with the system to use compliant players. The other type of false positive involves detecting a watermark, but incorrectly extracting the data from it, leading to non-desired operation (either from the point of view of the consumer denied lawful access or the studio whose content is more accessible than intended). [27]

3.4 The Broadcast Flag

The broadcast flag is a proposed system for controlling digital broadcast video. Like watermarks in physical media distribution, the broadcast flag (physically realized by a stream of bits sent with the broadcast program) is intended to prevent unauthorized copying by encoding information about copyright status in the actual broadcast. In this way, the same ideas of copy-never, copy-once, and copy-always can be implemented.

The MPAA and the major television broadcasters have argued that the broadcast flag increases the quality of programming without compromising the consumer rights. [6] While the stated goal of the broadcast flag is to prevent piracy, its actual goal is to alter the public's perception of intellectual property to one where file sharing is not seen as a given, fair use rights no longer apply, and control over a given piece of content remains solely in the hands of its creator. As one major IP lawyer puts it, "so long as the general public believes that private copying for non-commercial use is not wrong in the digital environment, it is simply a given that we will see the immediate uploading and free downloading of best-selling novels, music, and – once bandwidth is available – theatrical motion pictures by millions of people." [4] The broadcast flag is intended to change this attitude, by implementing protections in a way that causes users to believe that content is not intended to be copied or shared. [4]

Implementing the broadcast flag requires investments at every step of the broadcast process. Broadcasters must support this additional information. Consumers must purchase or be given equipment that enforces the DRM system. The content itself must be altered prior to transmission to include the broadcast flag. These requirements prevent the creation of a simple add-on device that would be inexpensive to manufacture. To meet all of the requirements, it would be more likely that the consumer might simply have to buy a new television. [4] The broadcast flag is thus tied into the push for DTV and HDTVs.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Adoption is made more difficult in that consumers appear aware of the consequences and expense involved with the broadcast flag. When the FCC solicited comments about the need for federal regulation in regards to the broadcast flag, it received over 6,000 comments, primarily from consumers. [6] Friedman et. al. echo this concern, stating that purchasers of these systems were likely to lose freedoms and pay more. Additionally, Friedman et. al. note with concern that the regulations on data transmission required for the broadcast flag are likely to stifle future innovation. [4]

The right of the FCC to issue mandates over devices capable of recording has also been questioned, as the FCC has traditionally only had domain over broadcasting itself. [8] Indeed, the US Court of Appeals, DC Circuit, ruled in December of 2005 that the FCC had overstepped its authority in regards to the broadcast flag. [30] Legislation to give the FCC authority to implement the broadcast flag was pending before Congress in 2006, but the bill died without receiving floor time. [31]

3.5 DRM in Other Video Systems

DRM is increasingly appearing in modern operating systems, with Windows Vista being a good example. Vista implements a technology known as the Protected Media Path (PMP) to protect video content being played in software. This DRM system was subverted in January of 2007, although the Canadian programmer who developed the exploit has been too concerned with prosecution under the DMCA should he travel to America to publish his results. [32] Problems with Vista's DRM and the relationship between Microsoft and the major studios are believed to be responsible for a lack of support for HD content in certain 32-bit versions of Vista. [33]

Peter Gutmann, a researcher at the University of Auckland in New Zealand, observes that Windows Vista includes many changes at the operating system level, all of which are designed to protect content such as Blu-ray and HD-DVD video. However, "providing this protection incurs considerable costs in terms of system performance, system stability, technical support overhead, and hardware and software cost." [34]

DRM is also appearing in video game systems. Many new video game consoles, such as Sony's Playstation 3 and Microsoft's XBox 360, are capable of video output specifically designed for HDTVs. Most of these systems use the HDMI (high-definition multimedia interface) standard for output. HDCP (high-bandwidth digital content protection) is a DRM system that is part of the HDMI standard, requiring devices that output in HDMI to implement DRM. HDCP has been found to cause device operability problems in certain game consoles, even during legitimate use. [35] HDMI also contains provisions for downgrading content when connected to non-HDMI compliant systems, using a system called the Image Constraint Token (ICT). While this system is not currently in use, manufacturers might simply be waiting for more adoption of HDMI before making use of this provision. [36]

Another increasingly popular venue for digital video is the digital video recorder (DVR). The first commercial DVR systems were ReplayTV and TiVo, which launched in 1999. These devices embody the concept of time-shifting: they are explicitly designed to record broadcast television for later viewing. Consumers commonly use DVRs and portable media devices for three primary functions. First, they time shift by watching broadcast programs at different times from when they aired. Second, they space shift by watching these programs in a location where direct access to the broadcasts is not normally available. Third, they format shift by taking this programming with them on mobile devices, for consumption elsewhere.

Time shifting was explicitly stated as legal in the AHRA, [8] and has been upheld by the US Supreme Court. [13] Additionally, the rights to make backup copies and to transform content between different formats have been recognized as well. [37] These devices have also begun to offer more options for place-shifting, with services such as TiVo ToGo, which allows the transfer of digital video recorded on a TiVo to a personal computer. DRM is used by the TiVo box to encrypt recorded content. Currently, the TiVo ToGo DRM system has been broken, and various programs exist to remove it. [38]

4 MODERN COMMERCIAL VIDEO

4.1 Video Types

The modern video experience has undergone qualitative and quantitative changes since the development of purchasable media and broadcast video. Since the landmark Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984) [39], home consumption has become increasingly important. Packaged physical media sales alone can easily account for more revenue than box office sales. [40] Three primary formats have emerged for modern consumer-oriented video:

- Packaged Physical Media (DVDs, etc.)
- Mass Broadcast Media (DTV, etc.)
- Streaming / Downloadable Media

These formats are the result of simultaneous development by the motion picture studios, the television studios, and the consumer electronics industry.

4.2 Physical Media

Display resolution is one of the factors commonly used to discuss differences in picture quality between different video formats. Display resolution is expressed as the number of pixels that can be displayed in both the horizontal and vertical dimensions, with the horizontal listed first by convention. Multiplying these two numbers together yields the total number of pixels in the display.

The NTSC broadcast television format uses a resolution of 270 x 486. In comparison, the NTSC S-VHS standard is 400

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

x 480 and the NTSC DVD standard is 720 x 480. HDTV, Blu-ray, and HD DVDs support resolutions of 1280 x 720 and 1920 x 1080. Digital theater standards run as high as 4096 x 3072. [41]

The DVD format was originally developed in 1995 as a compromise between formats proposed by several different companies. At the same time, a group known as the DVD Consortium was formed from these original companies. From the 10 founding members, it quickly grew to include over 100 companies, and was renamed the DVD Forum. The DVD Forum, which controls the licensing for CSS, is also the organization responsible for its adoption in the DVD format. [9] The DVD format supports region coding over 6 regions. Currently, consumers are beginning to see the next generation of DVD competitors, including Blu-ray Disc and HD DVD.

The Blu-ray format was developed by Sony in 2002 as its next-generation replacement for DVDs. Sony is joined by Apple, HP, Dell, Panasonic, Disney, Fox, and many others in supporting Blu-ray. [42] Development of the format has continued in recent years, resulting in several different disk standards with different storage capacities. Blu-ray disks use several different types of consumer-level DRM: AACS, BD+, and MMC. BD+ is a virtual machine designed to allow DRM-enforcing executables to be embedded in Blu-ray discs. Mandatory Managed Copy (MMC) is a system designed to allow limited copying by consumers. MMC was added at the request of HP, who made their participation in the Blu-ray format contingent on it. [43] Blu-ray disks also include a DRM system known as BD-ROM Mark, which consists of a small piece of cryptographic data written to the disk using dedicated hardware. BD-ROM Mark is designed to deter commercial piracy by making it difficult to burn professional-quality Blu-ray disks. The Blu-ray standard supports region coding over 3 regions.

HD DVD, or High-Definition DVD, was jointly developed by Toshiba and NEC in 2004. Companies supporting the format include Sanyo, RCA, HP, Acer, Microsoft, Universal, Paramount Pictures, Warner Brothers, and others. [44] In November of 2003, the DVD Forum selected HD DVD as the successor to the DVD standard. However, customer reaction has been mixed, with HD DVD and Blu-ray currently competing for dominance in the market. Like Blu-ray, HD DVD also uses the AACS and MMC DRM systems. However, HD DVD does not use region coding.

4.3 Mass Broadcast Media

Digital television is a two-pronged approach for reinventing the way consumers receive and use mass media broadcasts. The first part involves the actual transmission – while many signals are still broadcast in the USA in analog, there is an increasing move to digital, especially in cable users. At some point (originally scheduled for 2006, but now moved back to February 2009), all American analog broadcasting is due to switch over to digital signals (DTV). So far, only the Netherlands has made the full transition to digital broadcasting. [45]

The second part involves new television hardware. High-definition televisions (HDTV) have been introduced to consumer markets which are capable of better using DTV signals and with overall improvement in picture and sound quality. [4] HDTVs, however, remain expensive and have not achieved the household penetration rates hoped for in the industry. Estimates of HDTV penetration in American households are as low as 36% in 2007 and only half of those sets receive HD programming. [46] This lack of adoption is one of the direct reasons for the delay in the digital switchover, as older televisions may not be able to accept the new digital signals. [4]

The switchover to DTV is also one of the major forces driving the development of DRM watermarking techniques, as the FCC requires terrestrial broadcasts to be sent unencrypted (whereas current digital signals to cable subscribers can be encrypted). The various content providers are therefore concerned about sending video content without some form of protection. [4] Current cable DTV systems use set-top boxes which support specific DRM systems from specific content providers. This makes migration difficult for end users as well as content providers interested in upgrading or changing systems. Users are also prevented from receiving services from multiple content providers at once.

4.4 Streaming / Downloadable Media

An increasing number of sites offer digital video in streaming format. YouTube, created in 2005 and acquired in 2006 by Google, is among the most popular. YouTube allows users to upload videos that can then be watched, but not downloaded, by other users. Instead of attempting to deploy active DRM technologies, YouTube has chosen a highly responsive removal policy. As users post videos whose copyrights they do not own, YouTube removes these infringing videos at the request of the actual content owners. Currently, YouTube plans to offer more advanced DRM technologies and upload filtering to those companies willing to sign distribution deals. [47]

Downloadable video stores are slightly less popular. These include the portion of Apple's iTunes store that sells videos and competitors such as Google Video store. Both of these implement DRM systems (Apple's is known as FairPlay) to protect the videos they sell. Apple's popular portable digital media player, the iPod, also uses the FairPlay system to protect its content. Various programs, such as PlayFair, exist to remove FairPlay DRM.

Microsoft, which recently entered the portable digital video hardware market with its iPod competitor, the Zune media player, has been criticized for the lack of compatibility between the new player and DRM-protected video previously sold through other online retailers under the 'PlaysForSure' system, also developed by Microsoft. [48] Vendors and customers who invested in products containing PlaysForSure DRM have found Microsoft willing to abandon them in favor of its newest DRM technologies. Neither is there any guarantee that this abandonment will not occur again in the future. DRM systems are rarely backwards-compatible.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

5 CONSUMER RIGHTS AND CONCERNS

Many of the DRM systems discussed here and the laws which support them threaten basic practices that have long been afforded to consumers – fair use, time shifting, and making backups. These measures rightfully inspire concern in consumers of modern digital video content. DRM systems inherently threaten fair use because they have no way of distinguishing between legal infringement for fair use and illegal infringement for wholesale copying. DRM systems, for example, cannot detect satire.

Additionally, companies that use DRM have even more to gain than preventing copying. Activities such as unfavorable news reporting, research disliked by the content owner, and criticism of the content can all be incidentally prevented by DRM technologies that disallow the needed access. [49] The DMCA and other legislation has been used to attempt to suppress unfavorable discussion of DRM systems in the past, and industry groups have been notoriously unfriendly to the concept of fair use – the RIAA in particular has stated its members do not believe fair use is ever valid in relation to music. [8]

Material that has passed out of copyright and into the public domain is free for use by the public. However, editions of these works in formats protected by DRM may remain inaccessible for all time. [8] Similarly, public domain content broadcast over systems that implement watermarking might prevent users from recording programming they are legally entitled to copy.

DRM technologies reduce competition and lock users into given vendors by preventing interoperability. Players that use DRM systems are more expensive, in addition to being more complex and thus, more prone to failure. DRM systems also potentially force users to violate their own privacy by releasing identifying information. The loss of privacy is especially concerning given the increasing value of identity in the digital age. [49] In a recent example of how this may affect customers, Apple was found to be embedding customer names and email addresses in its DRM-free audio files, as a means to trace the source of unauthorized distribution. [50]

6 CONCLUSION

Large-scale commercial piracy is the true source for most unauthorized copies of DVD movies and other copyrighted digital video files. [4, 8] Several fundamental aspects of the technology involved in digital video content distribution make securing it against such piracy impossible. It is simply a question of the amount of time the pirates will require to break a system. [6] In addition, only a single person or group need defeat a DRM system for it to be rendered useless. Some of these systems need not even be broken – exemptions in DRM law for professional equipment mean a pirate operation with enough resources can completely circumvent them. [4]

The MPAA and other groups are largely focused on targeting non-commercial internet sharing, but this is not relevant to piracy – pirates want a financial return for their work. [4] In the name of combating piracy, the major industry groups responsible for controlling most commercial digital video content have introduced legislation and technological means that implement and protect DRM systems. However, none of these systems are sufficiently 'hacker-proof' to be able to keep dedicated commercial pirates out – and indeed, they are targeted at the consumer. [6] Rather than preventing piracy, these systems merely grant more and more control over home viewing to industry groups. At the same time, these systems increase the price of technology and the complexity and chance of failure of video players.

Consumers, frustrated with DRM, turn to commercial pirates. These pirates provide DRM-free, high-quality content, which is the only thing consumers wanted in the first place. By failing to provide sufficient added value to counteract the added penalty of DRM systems, IP owners have taken a small problem and created their own worst enemy – commercial piracy.

As Eric Flint of Baen Books puts it:

"Electronic copyright infringement is something that can only become an 'economic epidemic' under certain conditions. Any one of the following: 1) The products they want... are hard to find, and thus valuable. 2) The products they want are high-priced, so there's a fair amount of money to be saved by stealing them. 3) The legal products come with so many added-on nuisances that the illegal version is better to begin with. Those

Table 1. Summary of Major DRM Systems

	CSS	AACS	Digital Watermarks	Broadcast Flag
Designed to Protect	DVDs	Blu-ray, HD DVD discs	Any digital video	Any broadcast video
Designed to Prevent	Unauthorized disk copying	Unauthorized disk copying	Unauthorized copying / Distribution	Unauthorized copying / Distribution
Revocation	Model – level	Player – level	-	-
Common Cryptographic Elements	Single player key, title key (on the disk)	Set of player keys, title key (on the disk)	-	-
Dissimilar Cryptographic Elements	Session key hides communication, disk key	Set of processing keys to decrypt the title key	Steganography conceals an identifier	-
Requirements	Player Hardware	Player Hardware	Receiver / Player Hardware	Receiver Hardware

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

are the three conditions that will create widespread electronic copyright infringement, especially in combination. Why? Because they're the same three general conditions that create all large-scale smuggling enterprises. And... Guess what? It's precisely those three conditions that DRM creates in the first place. So far from being an impediment to so-called 'online piracy,' it's DRM itself that keeps fueling it and driving it forward." [51]

REFERENCES

1. MPAA (2007), 'Who piracy hurts – consumers'; www.mpaa.org/piracy_Consumers.asp (June 2007)
2. MPAA (2007), 'Who are movie pirates?'; www.mpaa.org/piracy_whoAre.asp (June 2007)
3. FindLaw (2007), 'Dowling v. United States, 473 U.S. 207 (1985)'; caselaw.lp.findlaw.com/scripts/getcase.pl?navby

=search&court=US&case=/us/473/207.html (June 2007)

4. A. Friedman, R. Baliga, D. Dasgupta, and A. Dreyer (2004), 'Understanding the broadcast flag: a threat analysis model', Telecom. Policy, vol. 28, pp. 503-521
5. J. Palenchar (2007), 'NPD: illegal downloads outpacing legal downloads', This Week In Consumer Electronics; www.twice.com/article/CA6424429.html?rssid=84

(March 2007)

6. A. Eskicioglu (2003), 'Protecting intellectual property in digital multimedia networks', IEEE Computing, vol. 36, pp. 39-45
7. K. Fisher (2006), 'The problem with MPAA's shocking piracy numbers', Ars Technica; arstechnica.com/news.ars

/post/20060505-6761.html (May 2007)

8. M. Lesk (2005), 'Salute the broadcast flag', IEEE Security and Privacy, vol. 3, no. 3, pp. 84-87
9. A. Eskicioglu and E. Delp (2001), 'An overview of multimedia content protection in consumer electronics devices', Signal Processing: Image Communication, vol. 16, no. 7, pp. 681–699
10. MPAA (2007), 'Piracy & the law'; www.mpaa.org/piracy_AndLaw.asp (June 2007)
11. W. Jonker and J. Linnartz (2004), 'Digital rights management in consumer electronics products', IEEE Signal Processing Magazine, vol. 21, no. 2, pp. 82-91
12. E. Lin, A. Eskicioglu, R. Lagendijk, and E. Delp (2005), 'Advances in digital video content protection', Proceedings of the IEEE, vol. 93, no. 1, pp. 171-183
13. J. Bloom, I. Cox, T. Kalker, J. Linnartz, M. Miller, and C. Traw (1999), 'Copy protection for DVD video', Proceedings of the IEEE, vol. 87, no. 7, pp. 1267-1276
14. G. Kesden (2000), 'Content scrambling system (CSS): introduction', G. Kesden; www.cs.cmu.edu/~dst/DeCSS/

Kesden/index.html (December 2006)

15. E. Edwards (2006), 'The content scrambling system: a technical description', E. Edwards; www.tinyted.net/eddie/css.html (May 2007)
16. A. Hawkesworth (2001), 'DivX: DVD quality movies on a CD-R?', Department of Electronics & Computer Science, University of Southampton; citeseer.ist.psu.edu/559183.html (October 2006)
17. K. Eschenfelder (2005), 'Chasing down the social meaning of DeCSS: investigating the internet posting of DVD circumvention software', Bulletin of the American Society for Information Science and Technology, vol. 31, no. 5; www.asis.org/Bulletin/Jun-05/rschenfelder.html (November 2006)
18. AACS Licensing Admin (2007), 'Advanced access content system'; www.aacsla.com/home (May 2007)
19. AACS Licensing Administrator (2006), 'AACS, pre-recorded video book'; www.aacsla.com/specifications/

specs091/AACS_Spec_Prerecorded_0.91.pdf (May 2007)

20. E. Felten (2007), 'AACS: a tale of three keys', Freedom To Tinker; www.freedom-to-tinker.com/?p=1121 (June 2007)
21. FoxDisc [alias] (2007) , Reply to 'Clarification on the state of AACS', Doom9; forum.doom9.org/showthread.php?t=124505 (May 2007)
22. Muslix64 [alias] (2006), 'BackupHDDVD, a tool to decrypt AACS protected movies'; forum.doom9.org/showthread.php?t=119871 (May 2007)
23. F. Lane (2007), '09 F9: an unlikely star is born thanks to digg.com', Sci-Tech Today; www.sci-tech-today.com/news/09-F9--An-Unlikely-Star-Is-Born/story.xhtml?story_id=011001CEELPZ (June 2007)
24. Chilling Effects(2007), 'AACS licensor complains of posted key'; www.chillingeffects.org/notice.cgi?

sID=03218 (May 2007)

25. R. Paul (2007), 'Latest AACS revision defeated a week before release', Ars Technica; arstechnica.com/news.ars/post/20070517-latest-aacs-revision-defeated-a-week-before-release.html (May 2007)
26. F. Duan, and I. King (1997), 'A short summary of digital watermarking techniques for multimedia data', Department of Computer Science and Engineering, Chinese University of Hong Kong; citeseer.ist.psu.edu/468553.html (October 1996)
27. M. Maes, T. Kalker, J. Linnartz, J. Talstra, G. Depovere, and J. Haitsma (2000), 'Digital watermarking for DVD video copy protection', IEEE Signal Processing Magazine, vol. 17, no. 5, pp. 47-57
28. S. Emmanuel and M. Kankanhalli (2003), 'A digital rights management scheme for broadcast video', ACM, Springer Verlag Multimedia Systems Journal, vol. 8, pp. 444-458
29. F. Petitcolas and R. Anderson (1999), 'Evaluation of copyright marking systems', Proceedings of IEEE Multimedia Systems, vol. 1, pp. 574-579
30. Public Knowledge (2005), 'ALA v. FCC, MPAA, Docket No 04-1037'; www.publicknowledge.org/pdf/bfcase-decision-20050506.pdf (December 2006)
31. GovTrack.us (2007),'"S. 2686 [109th]: communications, consumer's choice, and broadband deployment act of 2006'; www.govtrack.us/congress/bill.xpd?bill=s109-2686 (June 2007)
32. A. Ionescu, 'Update on driver signing bypass', A. Ionescu; www.alexionescu.com/?p=24 (May 2007)
33. D. Warne (2006), 'We were wrong about HD playback in Vista', APC; apcmag.com/3111/we_were_wrong_about_

hd_playback_in_vista_microsoft (June 2007)

34. P. Gutmann (2007), 'A cost analysis of Windows Vista content protection", P. Gutmann; www.cs.auckland.ac.nz/~pgut001/pubs/vista_cost.html#thoughts (May 2007)
35. Popular Mechanics (2007), 'The mystery of the blinking Playstation 3'; www.popularmechanics.com/blogs/

technology_news/4212161.html (July 2007)

36. K. Fisher (2006), 'Hollywood reportedly in agreement to delay forced quality downgrades for Blu-ray, HD DVD', Ars Technica; arstechnica.com/news.ars/post/20060521-6880.html (May 2007)
37. P. Samuelson (2003), 'DRM {and, or, vs.} the law', Communications of the ACM: Special Issue Digital Rights Management and Fair use by Design, vol. 46, no. 4, pp. 41-45
38. C. Doctorow (2006), 'TiVoToGo DRM cracked', Boing Boing; www.boingboing.net/2006/12/04/tivotogo_drm_

cracked.html (July 2007)

39. FindLaw (2007), 'Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984)'; caselaw.lp.findlaw.com/scripts/getcase.pl?navby=CASE&court=US&vol=464&page=417 (May 2007)
40. Firestone Entertainment (2007), 'Theatrical revenues & expenditures'; firestoneentertainment.com/docs/

Theatrical_Revenues_&_Expenditures.htm (June 2007)

41. J. Sokol (2004), 'Video format resolutions', Video Technology Magazine; www.videotechnology.com/

0904/formats.html (July 2007)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

42. Blue-ray Disc (2007), 'Supporting companies'; www.blu-raydisc.com/general_information/Section-14009/Index.html (May 2007)

43. Hewlett-Packard (2005), 'HP to support HD-DVD high-definition DVD format and join HD-DVD promotions group'; www.hp.com/hpinfo/newsroom/press/2005/

051216a.html (May 2007)

44. Time For DVD (2007), 'High definition-capable DVD'; www.time-fordvd.com/tutorial/HDDVDTutorial.shtml (June 2007)

45. T. Sterling (2006), 'Dutch pull plug on analog television', ABC News; abcnews.go.com/Entertainment/wireStory?id=2716983&CMP=OTC-RSSFeeds0312 (April 2007)

46. E. Sass (2007), 'HDTV Penetration Will Hit 36% in 2007', Media Post Publications; publications.mediapost.com/index.cfm?fuseaction=Articles.showArticle&art_aid=63061 (June 2007)

47. Reuters (2007), 'Google reserves YouTube DRM for partners only', PC Pro; www.pcpro.co.uk/news/105118/google-reserves-youtube-drm-for-partners-only.html (July 2007)

48. D. Slater (2006), 'Microsoft's Zune won't play protected Windows media', Electronic Frontier Foundation; www.eff.org/deeplinks/archives/004910.php (June 2007)

49. J. Erickson (2003) , 'Fair use, DRM, and trusted computing', Comm. of the ACM, vol.46, no.4, pp.34–39

50. P. Eckersley (2007), 'Apple's DRM-Free AAC files contain more than just names and email addresses', Electronic Frontier Foundation; www.eff.org/deeplinks/archives/005282.php (June 2007)

51. E. Flint (2007), 'There ain't no such thing as a free lunch', Salvos Against Big Brother, Baen Books; baens-universe.com/articles/salvos6 (June 2007)

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

CRM data grid services

Yongmin Tang

Technical College of Xi'an.
Shaanxi University of Science & Technology, 710016,China
email:tymanna@126.com

Abstract grid is a distributed component system, like service discovery, service creation, lifetime management and notification. CRM: Customer Relationship Management information at knowledge level and construct high-level expert system such as CRM tutor expert system. Urged us to provide basic infrastructure supporting Customer Knowledge (CK) service. Therefore, we introduce a set of Database Grid Service to support CRM data resource sharing. Our ultimate goal is to aid the development of distributed systems that help Customer to retrieve, integrate and share CRM information and knowledge from geographically.

Keywords CRM, CK, data resource space, Database Grid Service, Grid computing

1 INTRODUCTION

The failure cases of implementing CRM systems revealed its weakness of adaptability such as the difficulty of customer data migration, hard integrating with demand chain partner and can not seamless access to internal and external system Second, CRM does not provide proper way of database resource registration and discovery. Adopting Grid Services technique in the evolution of CRM system would be a good solution.

The CRM database grid should provide Customers with an efficient mechanism for coordinative customers of these related databases. And providing the integration access service by constructing a virtual database so that customers can access distributed data sources with single access point. A data producer might need to get its local data shared among its partners. Both data productcers and data consumers might use a local database as data management tool. How to interconnect this database to achieve data communication and sharing is a basic function that should be supported by Database Grid. Grid Services are created, managed, and destroyed within any particular hosting environments. Thus, Services that conforming to this specification. Grid is a distributed component system, like service discovery, service creation, lifetime management and notification. CRM: Customer Relationship Management information at knowledge level and construct high-level expert system such as CRM tutor expert system. Urged us to provide basic infrastructure supporting Customer Knowledge (CK) service. Therefore, we introduce a set of Database Grid Service to support data resource sharing

In the following, we first describe the notions of two backbone of CRM-Grid: Database Grid Service and Knowledge Service; then, we propose the CRM-Grid infrastructure in section 3

2 CRM-GRID DATABASE SERVICE

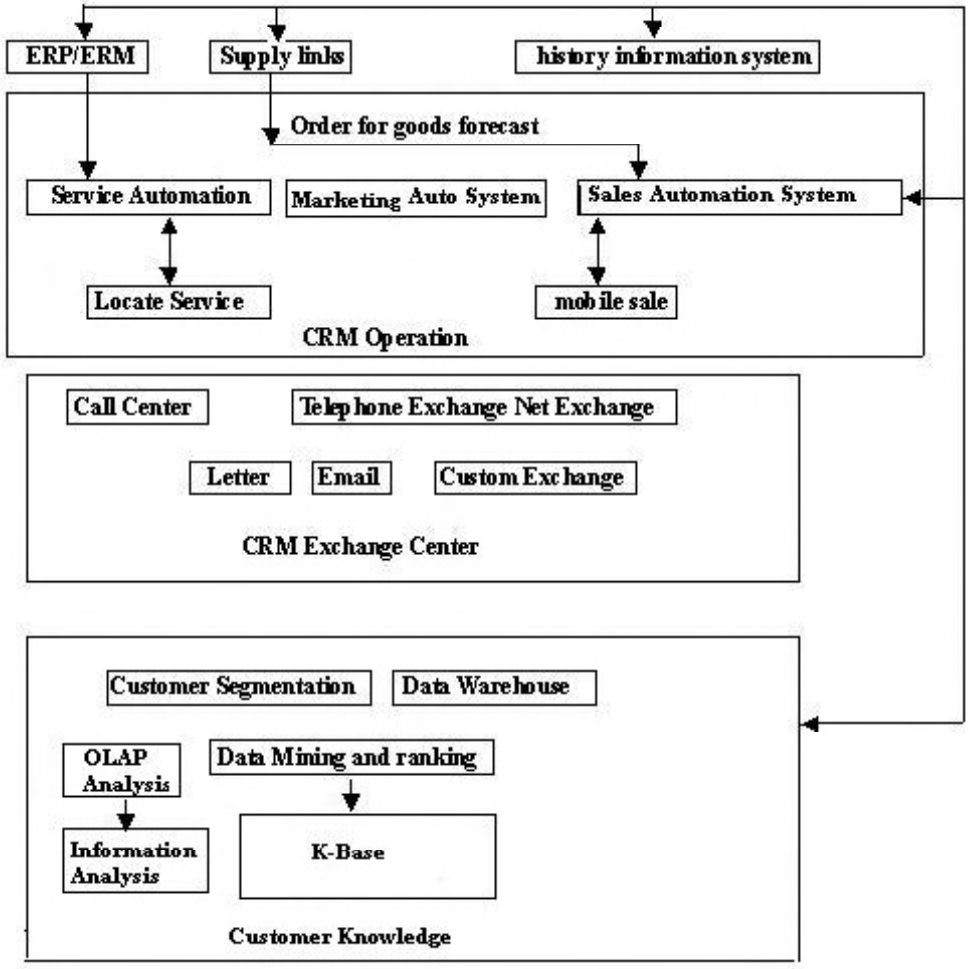
2.1 Elements of CRM System

CRM: Customer Relationship Management focuses on automating and improving the institutional processes associated with managing customer relationships in the areas of recruitment, marketing, communication management, event management, service, and support. CRM is both a business strategy and a set of discrete software and technologies whose goal is to reduce costs, In large-scale e-business and e-learn activities, These entities include various kinds of applications, computing appliances storage appliances and databases.

A complete CRM system consists of the following components (www.syntelinc.com):

- Marketing Automation System: to manage the marketing campaign and to run e-mail and fax management
- Product Configuration: the CRM part dedicated to the business products;
- Outbound Call Center: call center and telemarketing automation software;
- Sales Automation System: to manage all the items about the Sales Force of the company (sales forecasts, contact management, etc.);
- Inbound Call Center: call-tacking management;
- Customer Service: service and direct contact management;

Figure 1. CRM System and Customer Knowledge (CK) Service



- Customer Analysis: Customer Segmentation, Data Warehouse, Data Mining and ranking

These 7 parts that compose a CRM suite can be grouped in 4 categories by which CRM is completely defined (source IDC and Cap Gemini, 1999)

2.2 Customer knowledge service

Customer Knowledge (CK) Service can obtain implicit, unknown and useful CK from abundant data through the inlet and usage of intelligent data processing and analyzing tools such as OLAP and DM. Part of the data, information and knowledge about the customer obtained from database such as the call center, marketing, sales and customer service departments are transmitted directly to the knowledge base and part of those are processed with intelligent tools to turn them into more valuable CK and save them in the enterprise's knowledge base. OLAP is a tool for data analysis and inquiry. It provides the decision-maker with necessary information by using CRM Resource Space to analyze, inquire and generate statements. However, it only satisfies the most primary demand. Data mining technique searches for the relevant data according to the defined business object, and then processes them. After that, it selects the corresponding arithmetic to mine them, and eventually the interpreted results enter into the knowledge base. All kinds of CK obtained by data mining can be integrated in the enterprise and then are shared in the decision in sales, segmentation, high-valued customer identification, customer winning, analysis of customer retention and turnover, fraud examination, new product development and analysis of customer benefit. In this way, it is possible to optimize customer structure and customer relationship and deploy the enterprise resources efficiently.

Most exciting of all is CRM's ability to promote and enable e-Business, which is the seamless, web-based collaboration between an institution and its customers, suppliers, and partners. CRM applications track and manage interactions and transactions with various customers across multiple channels, including the Web. For institutions with a high degree of personal interaction, such as admissions recruiters or development officers, CRM can extend these channels to the Web by providing a framework for managing the interactions and transactions. CRM can also enable purchase of products or services online, and provide Web-based services and support, all personalized for the individual customer.

2.3 Grid data resources of the CRM Space

The expectation is that data resources in a grid will still generally be managed using existing systems Such as Customer Relationship Management Database Systems or file system. In this case, an existing system will already provide consumers with mechanism for accessing data resources, and such data resources are called service managed data resources.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

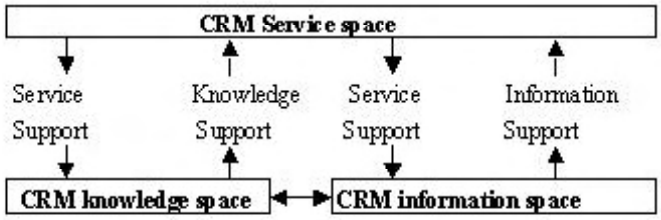
The CRM Space of Database Grid Information includes: information space, Service: space and knowledge space

The CRM information space is a resource space (information-category, information-level, location), where the former two identify information content, and the third identifies the locations that store information.

The CRM Service space is a three resources space (the service category, service level, location) service level contains four parts from low (close to hardware) to high (close to users): system level, middleware level, application interface (API) level, and application level. Each point in the space represents a set of services at a certain service level of a service category and is stored at a certain location.

The CRM knowledge space is a three-resource space (knowledge-category, knowledge-level, location), where the knowledge-category and the knowledge-level of a knowledge space identify knowledge content at a certain knowledge level of a certain knowledge category, the location identifies the locations that store knowledge. With reference to the knowledge levels of the axiom system, we can classify the knowledge space into four knowledge levels from low to high: conceptual level, axiom level, rule level, and method level. The conceptual level contains the basic concepts in the form of noun or noun phrases together with their explanations or definitions like dictionaries. The axiom level contains the commonsense knowledge of knowledge categories. A knowledge category together with its all-level sub-categories constitutes a knowledge category hierarchy.

Figure 2. Grid data resources of the CRM space



3 CRM GRID SERVICE MODEL

Based on the Internet, Grid computing seeks to extend the scope of distributed computing to encompass large-scale resource sharing including massive data-stores, high-performance networking and powerful computers, be they supercomputers or networks of workstations.

To explain the idea of Grid computing, one often uses the metaphor of a power distribution grid. This metaphor directly relates the Grid computing paradigm to the Utility computing business models. However for this metaphor to be realistic, the following issues relating to Grid computing need to be addressed: Different kinds of resources, Different kinds of interaction, Dynamic resource allocation and integration.

. A CRM compliant grid is a distributed component system that utilizes basic elements and patterns of a distributed system, like service discovery, service creation, lifetime management and notification.

•CRM Service discovery requires the ability of a service to describe itself. This is covered inside the mandatory GridService port type, which contains a FindServiceData function. Additionally a service can act as a registry if it implements the optional Registry port type.

•CRM Dynamic Service creation is a central element of OGSA. It is supported by the idea of factory services, which implement the optional Factory port type. A factory service is responsible for creating a new Grid Service instance and providing the client with a handle for accessing the service.

•CRM Lifetime management is a basic issue in distributed systems and is therefore addressed in the mandatory GridService port type. The basic idea is a Soft-State approach. This means that each service has a limited lifetime, therefore no distributed garbage collection is needed that tracks the necessity of services within the Grid and destroys them. An interested client may extend the lifetime if the policy of the service allows it. If the lifetime is expired, the service may terminate.

•CRM Notification is important because of the dynamic nature of the collaborating services. There are many scenarios imaginable, where a service creates another service and needs to be informed about specific events or changes in the state of the created service. Therefore a service can implement one or both of the optional NotificationSource and NotificationSink port types, which allow message delivery.

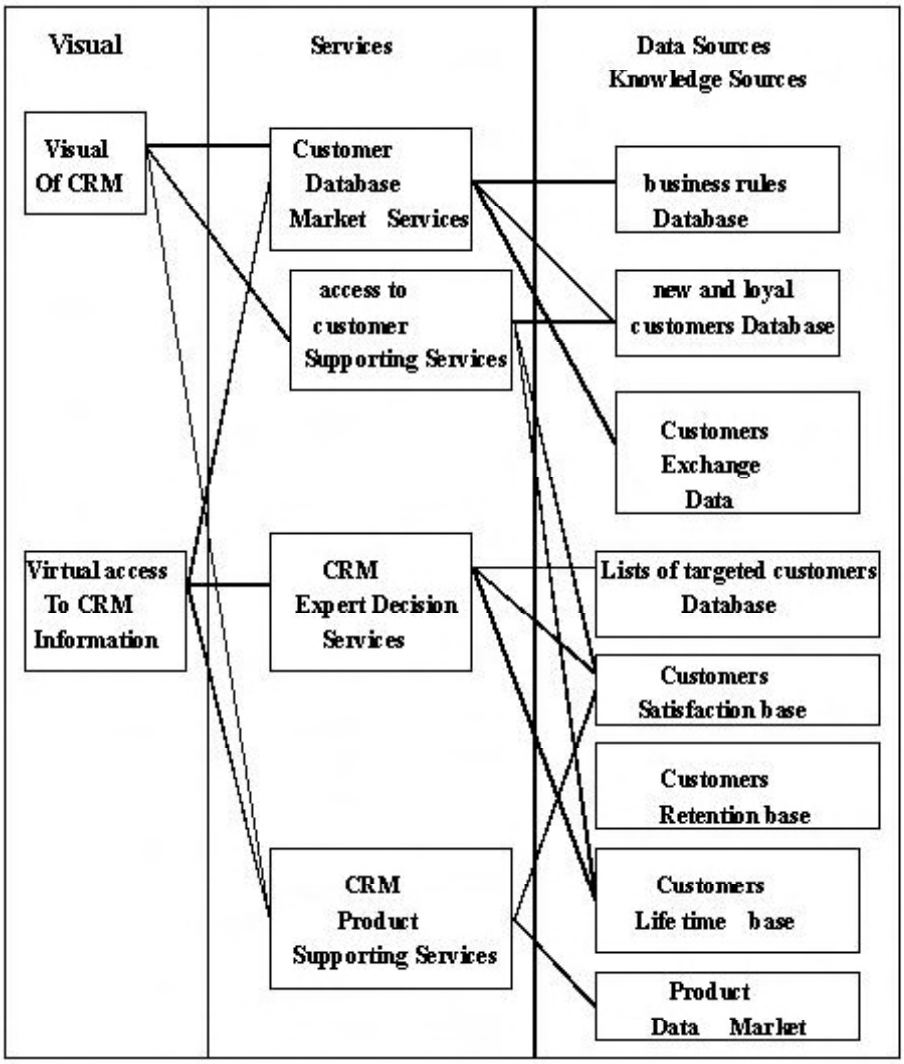
4 ESTABLISHING A COMPREHENSIVE CK SPACE GRID SYSTEM

Database Grid Service is used to publish database resources in a CRM.In this case; there are a number of basic databases, including Marketing, Product Configuration, Sales, and customer service departments. Those databases serve as public information source, which are geographically distributed and owned by different institutions or individuals. For example, the Customer databases belonging to different data resource space

Just as Figure 3 has illustrated, we divided the CRM Grid infrastructure into three layers. The backbone consists of a variety of data sources and knowledge sources. Above those sources, we could construct high-level services including some high-level database grid services. Above those services, a Virtual Database service is constructed to integrate those basic CRM databases and provide data access service, CRM Knowledge services. Relying on these basic data and knowledge sources, we could construct high-level services.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. The CRM grid infrastructure



5 CONCLUSIONS

This paper describes our experience with building a Grid for CRM.Our ultimate goal is to aid the development of distributed systems that help Customer to retrieve, integrate and share CRM information and knowledge from geographically decentralized CRM database resources and knowledge based services globally.

REFERENCES

1. http: //www .crmassist.com/documents/document.asp?i=430, Mandeep Khera, "Customer Relationship Management -Beyond the BUZZ." ITtoolbox CRM. Downloaded (02/05/2002).
2. http://houns54.clearlake.ibm.com/solutions/crm/crmpub.nsf/detailcontacts/Home? , OpenDocumet, "Build Lasting Customer Relationship, and Build Your Profits", (IBM-CRM). Downloaded (02/05/2002).
3. http://www.cio.com/research/crm/edit/crmabc.html,Customer Relationship Management Research Center, 200 1. Downloaded (23/04/2002).
4. Abdulreza Salahi (Y. El Saffar) Hussein H. Fakhry "Study and Development of CRM Information System for Small and Medium Businesses" 0-7803-8482-21041 02004 IEEE. 503
5. Davids, M. (1999) "How to avoid the 10 biggest mistakes in CRM", Journal of Business Strategy, Vol. 20, No. 6, pp. 22-26.
6. Hai Zhuge (2004) Resource Space Grid: model, method and platform 2004; 16:1385–1413 Published online 14 September 2004 in Wiley InterScience (www.interscience.wiley.com).
7. Yongmin tang, "Data mining based on WaveCluster", the Proceedings of THE 1ST International Conference on Digital Communications and Computer Applications.Jordan,2007,pp276-283
8. Henning Gebert, Malte Geib, "Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts", Journal of Knowledge Management, Vol 7, No. 5, pp. 107-123, 2003.
9. Marcus Blosch, "Customer Knowledge", Knowledge and Process Management, Vol 7, No. 4, pp. 265-268, 2000.
10. Ye Naiyi, "A Model of Customer Knowledge in the Information Age", Operations Research and Management science, Vol 11, No. 4, pp. 121-127, 2002.
11. Yang Jinyue, Yang Lin, Sun Yin, "Customer intelligence, core of CRM", Logistics Technology, No. 7, pp. 22-24,2002.
12. Mario Antonioletti "Web Services Data Access and Integration – The Core (WS-DAI) Specification, Version 1.0 ",Copyright © Open Grid Forum (2006). All Rights Reserved
13. Eppler, M., Seifried, P. and Röpnack, A., "Improving Knowledge Intensive Processes through an Enterprise Knowledge Medium", ACM SIGCPR Conference on Computer Personnel Research, New Orleans, USA, 1999, pp. 222-230.
14. Riempp, G., "Von den Grundlagen zu einer Architektur für Customer Knowledge Management", in: Kolbe, L. M., Brenner, W. and Österle, H. (Eds.), Customer Knowledge Management, Springer, Berlin, 2003, pp. 23-57.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

A switch interaction solution for detecting and isolating ARP spoofing

Dengke He

Chongiqng iSoft Technology CO., LTD.
Chongqing, P.R. China
400039

Jiashang Yan
Jiang Du

Chongqing University of Posts and Telecommunications
Chongqing, P.R. China
400065

Abstract ARP (Ethernet Address Resolution Protocol) spoofing is omnipresent at present. The existing technologies (personal firewalls, middleware, watch ARP changes, etc.) are incapable of finding out exactly where the spoofing host is, and separating it completely from local network. So, a switch interaction solution is presented. Through a special deployment of Monitors and Center, this solution could detect ARP spoofing, locate the malicious host' s exact position in physical access layer, as well as isolate it completely from local network.

Keywords ARP spoofing, Detection, Isolation, Switch interaction, Physical position

1 INTRODUCTION

ARP spoofing (also known as ARP poisoning) is an old problem. It was being aware since 1997 [1]. Mostly, ARP spoofing is used for Man-in-the-Middle (MITM) attacks, DoS (Denial of Service) Attacks, and MAC Flooding [2].

This is a nightmare for network administrators, as they have to locate where the spoofing attack issues. This is not an easy work (discusses below). It is increasingly true since easily-used tools are developed, such as Ettercap [3], Arpoison [4], etc.

2 SECURITY IN ARP

ARP is used for converting protocol addresses (e.g., IP addresses) to local network addresses (e.g., Ethernet addresses) [5]. Since it is stateless, and it is believed that all hosts in local network can be trusted, so this protocol has its innate deficiency in security. And this deficiency provides a favorable condition for ARP spoofing.

ARP spoofing could be implemented in two ways: sending a spurious address request; and sending an unsolicited address response.

The first way to spoof ARP cache is sending a spurious address request. The attacker could issue a legitimate request, by broadcasting or unicasting, to poison the ARP cache of requested host. This is tested to be true in RedHat Linux 9.0, Fedora Core 5.0, Windows 2000, XP, and 2003 system.

Sending an unsolicited address response is the second way. As ARP is stateless, local host will update its ARP cache when it receives an address response, no matter whether it sent this request or not. So the attacker could broadcast or unicast its forged response to local network. Even if local ARP cache is statically bound, its static entries will be changed in platforms such as Windows 95, 98 [6], and 2000.

The biggest problem to locate the spoofing host (the attacker) is that the source MAC (Media Access Control) address and sender's hardware address in ARP packet could be spurious. Although network administrators could be alerted from monitoring applications (e.g. firewall), what they cannot do is to find out exactly where the malicious host is and isolate it completely from local network.

3 RELATED WORK AND MOTIVATION

The existing solutions for ARP spoofing detection and prevention can be classified into two categories: host-based and network-based.

3.1 Host-based solutions

Static ARP entry should be the first solution. But it is not easy to get this work done in a large network.

Mahesh Tripunitara and Partha Dutta have proposed a middleware method to detect and prevent ARP spoofing [7]. In this solution, a major part - Cache Poisoning Checker (CPC) streams module is put into protocol stack of the operating system. The limitation is that this approach can only be implemented in the stream-based system, such as Solaris [8]. While in Linux, kernel module should be changed to implement the similar function [8].

The request-reply mismatch method [6] is just checking if a reply matches a request in the table. As discussed above, an ARP request could also poison cache.

The duplicate packet detection method [6] is to hash IP packets as its index, and see if packet with the same hash index appears. If so, ARP spoofing happens. The biggest problem of this method is that how to detect MITM attack if the attack already exists.

Personal firewall is another solution. Ebtables [9] and Outpost firewall pro [10] are commonly used.

3.2 Network-based solutions

Detection on switches via SNMP is a typical solution [6]. Information about ingoing and outgoing packets and bytes in ifTable is used. If ingoing packets are more than outgoing packets in a certain switch port, as well as the average packet size of this port is less than 65, this port is suspicious (i.e. attacker is likely linked to this port). But if the malicious host sends ARP request to poison, the ingoing packets are probably equal to the outgoing packets. Moreover, the final detection result is not satisfying.

APRDefender [11] is another way. It is the open source program ARPWatch. It is, however, only detecting changes of MAC addresses in local network, as well as finding new announced hosts.

3.3 Motivation

As discussed above, host-based methods can, to some extent, detect and prevent attacks, but are lack of locating the spoofing host accurately; while network-based methods could detect some attacks, but with higher false negatives. Besides, they could not isolate malicious hosts from local network.

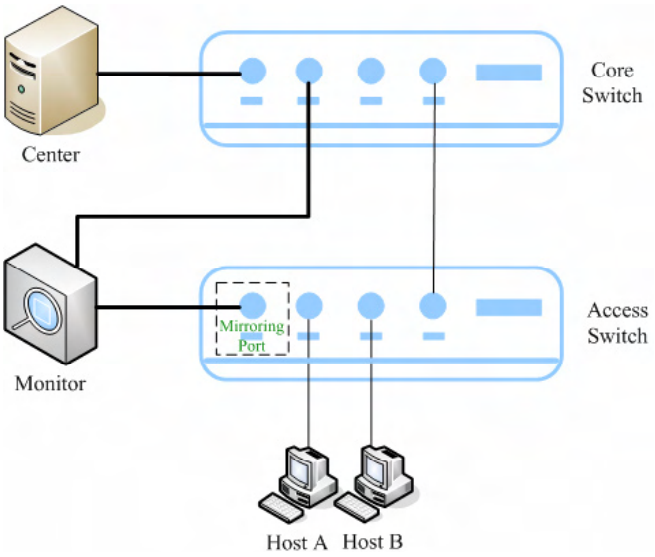
So our aims are:

A. locating the spoofing host accurately;
B. isolating the spoofing host completely from local network;
C. recovering the poisoned ARP cache;

4 ARCHITECTURE

The architecture of our solution is demonstrated below:

Figure 1. Solution architecture



4.1 Monitor

Monitor takes charge of collecting and analyzing network information. It has the monitoring interface (i.e. network interface card) linking to the mirroring port of the access switch.

Monitor communicates with Center via the communication link, between another network interface card and a core switch port.

4.2 Center

Center is in charge of storing network information and alarms, as well as taking actions by interacting with the access switch (e.g. close switch ports) according to policies made by administrators.

In order to interact with the access switch, Center should be able to log into the access switch remotely. In other words, the access switch should support remote access, and there is a way between Center and the access switch (i.e. routing). Besides, Center should know IP address, user name and password of the access switch.

In fact, Center has a management GUI (Graphical User Interface) to operate all the access switches in local network. Administrators could close and open a certain port of a specific access switch, query port status – open or close, and query switch port by host MAC through Center GUI, manually. Or, these operations could be done automatically by setting policies in management GUI. Ten minutes close for ARP spoofing port, for example. If a malicious host issues ARP spoofing, an alarm will be triggered and sent to Center. Center will interact with the relevant access switch, and close the switch port for ten minutes. After time is expired, this switch port will be open again.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4.3 Core switch

Used for communicating between Center and Monitor, between Center and access switches.

4.4 Access switch

Each port on the access switch should be connected by terminal host (i.e. no switch cascading). For we have to isolate malicious hosts completely from local network, we should make sure that if a certain switch port is closed, there will be only one terminal host affected. Another consideration is to make sure all network activities within local network could be collected. Monitor will only see outgoing and ingoing network information if it is linked to the mirroring port of the core switch. Under that circumstance, we cannot guarantee the inner network security.

As discussed above, the major difficulty against ARP spoofing attack is that it is hard to locate where the spoofing host is. In this architecture, this problem could be easily solved.

Inside every access switch, there is a MAC address table, containing the MAC address and the switch port this MAC address is from, used for forwarding network packets. Otherwise, the access switch will broadcast the receiving packet to every port except the port that received this packet, and learn to add entries to this MAC address table.

Even if all parts in an ARP packet are forged, the source MAC address of this ARP packet will still be remembered in switch MAC table, once this forged ARP packet is squeezed into local network.

So, through querying switch MAC table, it is easy to locate the switch port this spoofing host linked.

There is, however, aging time for each MAC entry in this table. Normally, the aging time is about 2 - 5 minutes. It varies in different switches.

5 IMPLEMENTATION

The system works in following steps:

5.1 Host information collection

Monitor is collecting network packets, especially IP packets, to build a <MAC, IP> table during the learning period setting by administrators.

At times, we may find Man-in-the-Middle attack already happens when the system is in learning period. In this case, the system will collect the wrong data. To ensure the collecting <MAC, IP> data are correct, to the utmost extent, ARP request is used.

Every time a new IP address is learnt, an ARP request for this IP address will be sent. For the sake of distinguishing an ARP reply responded to our request from other specious reply, we fill a nonexistent sender's protocol address (e.g. 1.1.1.1) in our ARP requests. If the response is right, the target's protocol address in the reply would have to be the same one.

By doing this, the collecting results will be guaranteed to be right to the largest extent, but not 100 percent.

5.2 Host information validation

When the learning time is over, data in this table are sent to Center. Valid data will be sent back to Monitor after administrators confirm these hosts to be legitimate in Center.

If administrators are not sure about specific host data, they could check them via tools like Arping.

5.3 ARP spoofing monitoring

Monitor will then judge if an ARP request or a response is legitimate by the confirmed data. Any sender's hardware address (i.e. MAC) and its corresponding protocol address (i.e. IP) contained in ARP packet that is different from confirmed <MAC, IP> data will be reported to Center as abnormal. The alarms contain information about relevant access switch.

Meanwhile, a right ARP reply is sent to destination to recover the poisoned entry.

5.4 Spoofing host isolation

Receiving alarms from Monitor, Center will interact with the relevant access switch: log into that switch, query switch port by MAC address, and close this switch port temporarily or permanently, according to policies.

6 EXPERIMENTATION

Experiment one – Send 100 ARP spoofing packets within 1 second to see alarm reports:

Table 1. Detection experiment

Sent	Detected	False positive rate	False negative rate
100	101	1%	0%
100	115	15%	0%
100	112	12%	0%
100	100	0%	0%

Result analysis – 100 attacks could all be detected, but false alarm rate is a little higher. Because in the test environment, not all the hosts in local network are connected to the test access switch, Monitor is not able to collect all the host information during the learning period. So in the ARP spoofing monitor stage, if some other hosts broadcast ARP request, false alarms are generated.

Experiment two – Send ARP spoofing packets continually to see if switch port this host linked will be closed. If one port is closed, connect this host to another port:

Table 2. Switch interaction experiment

Port numbers connected	Port numbers closed
5	5
5	5

Result Analysis – It could close physical port connected by the malicious host, and the effect is satisfying.

7 CONCLUSIONS AND FUTURE WORK

This solution could not only find out where the ARP spoofing hosts are, and separate them in the physical access level, it could also be used for isolating other abnormal activities, such as scanning local network, transferring data through Point-to-Point protocol, etc. These activities might not, as far as people are concerned, be malicious, but they are of influence. Numerous network packets will be generated, and a great deal of network bandwidth will be taken. Thus, normal activities will be greatly affected.

Moreover, this solution could monitor inner communication to the greatest extent, as well as outer communication, so that administrators could clearly understand their network usage.

However, improvement should also be made for this solution:

The accuracy of host information collected in the first step of the system should be improved, because later alarms are reported by this information. Although host information will be confirmed by administrators in the second step, we could not rely on this. It is such a tedious work to validate MAC addresses so that few people will do it carefully.

Detecting accuracy should be improved. From experiment results, we can see that the false positive rate is a little too high.

REFERENCES

1. Yuri Volobuev, ARP and ICMP redirection games, http://insecure.org/sploits/arp.games.html, 1997.
2. Corey Nachreiner, Anatomy of an ARP Poisoning Attack, http://www.watchguard.com/infocenter/editorial/135324.asp, 2001.
3. Ettercap, http://ettercap.sourceforge.net/index.php.
4. Arpoison, http://www.arpoison.net/.
5. David C. Plummer, RFC 826: An Ethernet Address Resolution Protocol, www.ietf.org/rfc/rfc826.txt, 1982.
6. Marco Antônio Carnut, and João José Costa Gondim, ARP spoofing detection on switched ethernet networks: A feasibility study", http://www.linorg.cirp.usp.br/SSI/SSI2003/Artigos/A25.pdf, 2003.
7. Mahesh Tripunitara, and Partha Dutta, Middle approach to asynchronous and backward-compatible detection and prevention of ARP cache poisoning, http://www.freepatentsonline.com/6771649.html, 1999.
8. Silky Manwani, ARP Cache Poisoning Detection and Prevention, http://www.cs.sjsu.edu/faculty/stamp/students/Silky_report.pdf, 2003.
9. ebtables, http://ebtables.sourceforge.net/.
10. Outpost firewall pro, http://www.agnitum.com/download/Ethernet_Attacks_Protection.pdf.
11. ARPDefender, http://www.arpdefender.com/.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Enterprise content management: bridging the academia-industry gap

Sergey V. Zykov

TEKAMA LLC
Russian Federation

Abstract The paper considers content management in web portals, embracing heterogeneous enterprise information systems (IS). Within the present-day information society, huge enterprises have acquired quite a heavy data bulk, which tends to be ever growing. Global distribution and weak-structured character of the heterogeneous enterprise data complicate its uniform management. Therefore, a problem-oriented approach for enterprise content management is suggested, which combines a formal model and a software development toolkit. An industry-level implementation proves significant terms-and-cost reduction.

Keywords enterprise content management, data model, abstract machine, enterprise portal

1 INTRODUCTION

Present-day enterprise IS currently process terabytes of heterogeneous information sources (both of data and metadata nature), which tend to change for petabytes shortly. The tremendous data bulk demands new models, methods and tools to manage them uniformly. A positive way of the solution is portal-based enterprise content management (ECM). The major software producers (Microsoft, Oracle, BEA etc.) lack adequate academia-based formal models in their ECM schemes, which results in model-level methodological "gaps" and makes the process vendor-dependent and non-uniform. On the other hand, the theoretical approaches known as yet lack efficient bridging with industry CASE and RAD standards, and therefore they generally do not result in enterprise-level solutions with practically applicable implementation features (scalability, expandability, availability, fault tolerance etc.).

The proposed ECM methodology is a part of the integrated IS lifecycle support for heterogeneous portal-based environment [12, 13].

2 THEORETICAL BACKGROUND AND RELATED WORKS

Research methods used for ECM modelingare based on an innovative synthesis of carefully selected theoretical areas of finite sequences [1], categories [2], semantic networks [3] and variable domains [4].

Finite sequences (in the form of typed lambda calculus) and categories are used for data and metadata object modelling. Categories are used to model virtual ECM machine by means of an abstract machine. Computations theory [4] enables formal content management description in terms of denotation semantics. Semantic networks are used to enable visual, rapid and rigorous bridging between enterprise-scale CASE tools and the formal model methodology level.

Only a few of the major software producers incorporate theoretical components in their software development methodologies. One of such positive examples is Microsoft .Net SDK platform with formal models based on state abstract machines [11].

Getting together heterogeneous software components is among the most challenging tasks of the enterprise-level integrated system development. To achieve the integration success, the suggested methodology uses a combination of formal model and SDK for building class-level association-based relationships.

The problem domain features high complexity of the object classes and incomplete information on the structure of certain instantiations of these classes. However, both the set of class attributes the set of operations over class objects can be determined rigorously. Thus, the suggested frame-based methodology appears to be applicable due to the following reasons: variety of heterogeneous classes, importance of association-based inter-class relationships, and class inference (the latter is possible even under weak-structured character of some of the class instances).

Ontology-based approaches (such as Cyc project and its extensions [6], [9]) even being aimed at large knowledgebase integration (as in [7], [8]) tend to be comparable in efficiency to the approach suggested only under a total class-level uncertainty. However, such total class-level uncertainties ac-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

tually belong to quite different problem domains than the ECM. Such domains usually involve thesaurus to meet the relevance levels required (as in the task of building a web-page parser for semantic web [7], [10]).

The approach suggested uses similar foundations compared to the ontology-based approaches (e.g. Cyc project uses predicate calculus-based CycL language, quite similar to Lisp [6], "conceptual model" [10] etc.). Also, the ontology-based approaches use certain tools to simplify the data modeling and integration processes (e.g., UML and XML-based tools are used in [9]). However, the ontology-based approaches lack a balanced combination of formal models and industry-level SDKs (including visualization tools) for the entire ECM lifecycle, which results in low scalability and non-suitability for the majority of enterprise-level tasks [8], [9], [10].

3 THE ECM METHODOLOGY

3.1 Overview

The ECM methodology data model is dynamical and state-based to work adequately in heterogeneous weak-structured environments. The model supports personalization-based front-end and back-end (meta)data processing. Component and event-based scripting technologies support extendable, distributed, and interoperable environments.

The variable domain-based model features event-driven (meta)data object management of heterogeneous problem domains. Therewith, the range of possible (meta)data sources can be extended up to a vast range of data warehouses, which support both front-end globally distributed enterprise IS and legacy systems. The methodology features content oriented data and metadata model based on an abstract machine.

A multi-level integrated IS design and implementation scheme is suggested (see fig. 1), which provides fast component-based ECM. This general scheme is used to continuously maintain the enterprise content adequacy and to provide content integrity control. During the ECM lifecycle, the content (i.e. data and metadata objects) specification is transformed from problem domain notations to formal computational data model entities, further, to object-relational (meta)database scheme by means of a CMS featuring CASE toolkit and, finally, to the target Web-based representation. A formal language is suggested to represent both the content and the environment during their transformations within the enterprise data lifecycle.

3.2 Data and metadata model

An object-based data model is suggested.

Within the model framework, a (meta)data unit is generally represented as follows:

$$\text{Data} = \langle \text{class, object, value} \rangle, \quad (1)$$

where under a *class* a collection of objects of the integrated enterprise database is implied. An *object* means an element of ECM IS portal page template. A *value* means a template data object instantiation and represents the problem domain dynamics.

Compared to results known as yet, principal benefits of the model suggested are more adequate mapping of heterogeneous weak-structured problem domain dynamics and statics as well as event-driven (meta)data control in a global computational environment.

The data model is based on two-level *conceptualization* [5], i.e., the process of establishing relations between problem domain concepts.

Objects, according to assigned types, are assembled into assignment-dependent collections, thus forming variable domains, which model problem domain dynamics and statics.

(Meta)data and state semantics is adequately formalized by the model based on multi-sort typed lambda calculus, combinatory logic, semantic network scenarios and categorical abstract machine with states.

The data model compression principle allows model application to concepts, individuals and states separately and to the data on the whole.

The integrated object model for data, metadata and states is characterized by structural hierarchical organization, scalability, metadata encapsulation and readability. Extendibility, adequacy, neutrality and semantic soundness of the formalization provide problem-oriented IS development with (meta)data object adequacy maintenance throughout the entire lifecycle.

On the basis of multi-parameter functional

$$F = F((v), (e), \ldots)(s)(p), \quad (2)$$

where assignment values represent:

s – IS user personal preferences;

p – IS user registration status;

v – IS client interface parameters;

e – IS data access device parameters,

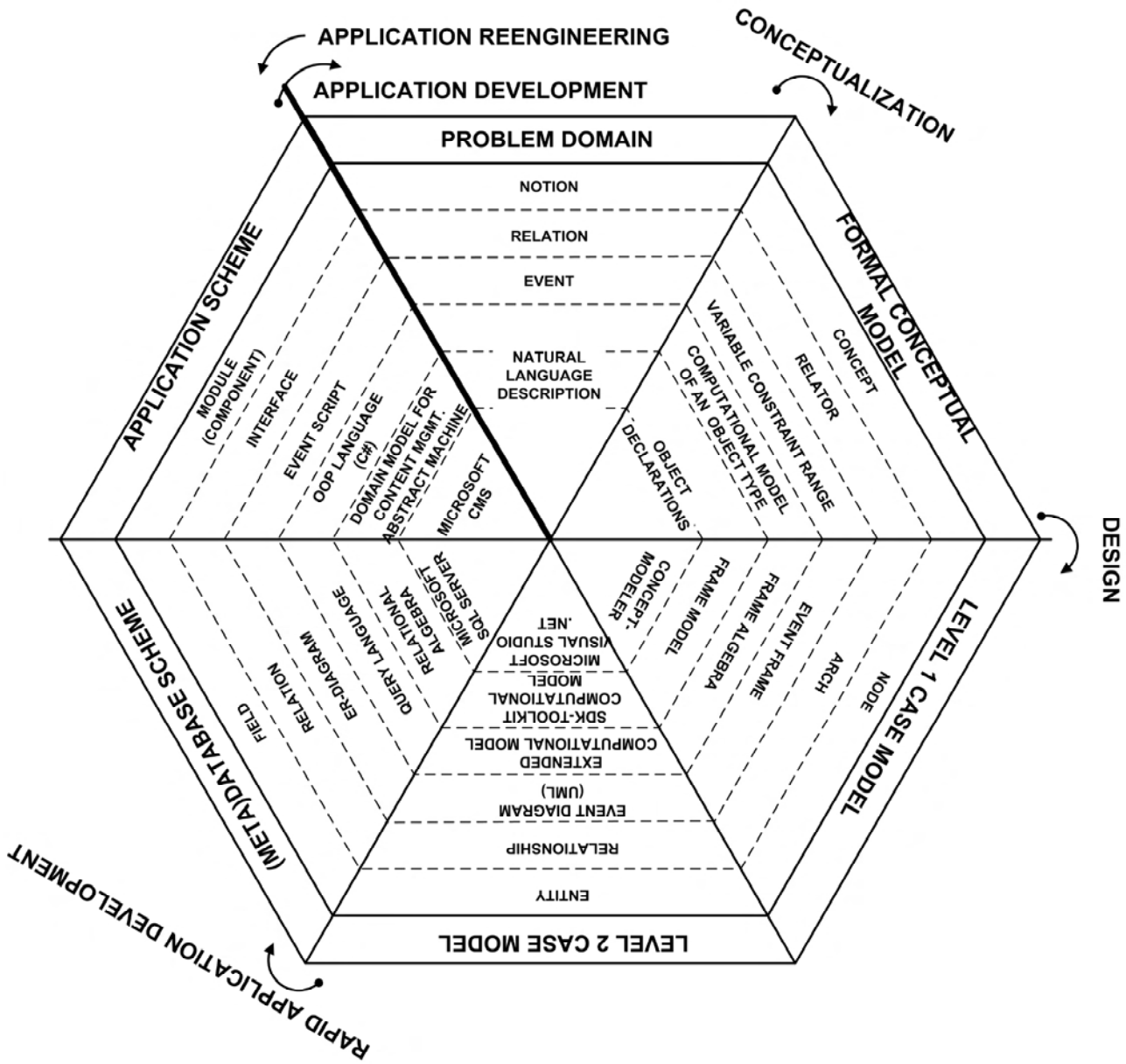
a problem-oriented object model for portal personalization has been built, which is based on functional $||F||$ evaluation function [7].

3.3 Content management model

Abstract machine for content management (AMCM) [7,10] is suggested as an ECM IS model, which is an improved version of categorical abstract machine (CAM) [2]. At any given moment AMCM is determined by its *state*. AMCM

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 1. The ECM methodology outline



work *cycle* can be formalized by explicit enumeration of possible state changes, which define the procedure of AMCM state *dynamics* modelling.

From the formal model viewpoint, when portal page templates are mapped into the pages, variable *binding* evaluates the variables that represent template elements and their values, i.e. portal page elements.

AMCM semantics can be described on the basis of D.Scott's variable domain theory [4]. Therewith, atomic template types are selected from standard domains, while more complex template types are built using domain constructors.

AMCM formal semantics is built as follows:

- *Standard* (most commonly used within the model framework) domain enumeration;
- *Finite* (containing explicitly enumerable elements) domain definition;
- Domain *constructor* (operations of building new domains out of the existing ones) definition;
- *Composite* domain formalization using standard domains and domain constructors.

Domain constructors include functional space $[D_1 \rightarrow D_2]$, Cartesian product $[D_1 \times D_2 \times \ldots \times D_n]$, disjunctive sum $[D_1 + D_2 + \ldots + D_n]$ and sequence D^*.

Let the AMCM language contain expression set E (including constant set, identifier set I, assignment operation (content "write operation" to template "slot") etc.), and command set C (comparison, command sequence etc.).

AMCM syntax is completely defined by the following syntax domain description:

$$\text{Ide} = \{I \mid I - \text{identifier}\}; \quad (3)$$

$$\text{Com} = \{C \mid C - \text{command}\}; \quad (4)$$

$$\text{Exp} = \{E \mid E - \text{expression}\}. \quad (5)$$

Let us collect all possible language identifiers into *Ide* domain, commands – into *Com* domain, and expressions – into *Exp* domain.

The state-based AMCM environment model can be represented as follows:

$$St = Mem \times In \times Out; \quad (6)$$

$$Mem = Ide \rightarrow [Val + \{unbound\}]; \quad (7)$$

$$In = Val^*; \quad (8)$$

$$Out = Val^*; \quad (9)$$

$$Value = T_1 + T_2 + \dots . \quad (10)$$

AMCM state is defined by "memory" state, which contains input values (i.e. content) and output values (i.e. web pages) of the abstract machine. Therewith, under memory a mapping from identifier domain into value domain is implied, which is similar to lambda calculus variable binding. For correct exception handling, unbound element should be added to the domains. Value domain is formed by disjunctive sum of domains, which contain content types of AMCM language.

Semantic statements describe denotates (i.e. correctly built values) of AMCM (meta)data object manipulation language.

Semantic statements for basic AMCM language commands and expressions are presented in [12].

Constant denotates are their respective values in a form of ordered pair of <*variable, value*>, while program state remains unchanged.

Identifier denotates are identifiers bound with their values (if binding is possible) in a form of ordered tuples of <*variable_in_memory, identifier, state*>, while the state remains unchanged (an error message is generated if the binding is impossible).

Thus, ECM IS template binding with the content may result in AMCM state change and in a number of limited, predefined cases (particularly, under template and content type incompatibility) – in error generation.

Semantic statement for an AMCM command, which assigns content to template element, results in state change with substitution of content value by the identifier in memory.

4 THE ECM METHODOLOGY CUSTOMIZATION

Let us apply the computational models introduced to the target ECM IS and the portal.

The problem domain model is based on two-level compression [5], which is interpreted here as establishing relationships between data object classes *C* of the integrated problem domain *D*, which, in general, are modelled by the domains:

$$C = Iw{:}[D] \ \forall \ v{:}D \ (w(v) \leftrightarrow \Delta) = \{v{:}D \mid \Delta\}, \quad (11)$$

where:

C and *D* are in a relation of partial order with each other (*C ISA D*);

is a condition of data object *w* belonging to class *C* (according to a problem domain expert).

Complex, "multi-dimensional" data objects are modelled as *n*-arity data object relationship (frame representation is given in [12]):

$$R^n = Iw{:}\ [V_1,..,V_n] \ \forall \ v_1{:}V_1 \ \dots \ \forall \ v_n{:}V_n$$
$$(w\ [v_1,\dots,v_n] \leftrightarrow \Gamma) = \{[\ v_1{:}V_1,\dots \quad ,v_n{:}V_n] \mid \quad \Gamma\}. \quad (12)$$

Thus, any class of data objects is a collection of ordered pairs $(v_i{:}V_i)$, where v_i is *i*-th attribute of the class, and V_i is type of the class.

However, class attributes contain not only data, but also metadata (e.g., object dimensions, and integrity constraints). Another metadata example is a number (bit mask), which defines effective rights of class objects usage in the templates of ECM IS.

Under assignment a_1, class *C* is instantiated with Δ_k template of ECM IS web page. Therewith, evaluation of template collection *M* assigns the value "true" to its only element m_i, the index of which equals to the template number (*k*):

$$M = (m_1,\dots, m_k,\dots, m_N), \quad (13)$$

where $\forall\ i{=}1,\dots,N\ m_i \in \{0,1\}$;

$$[M|\Delta_k] = (m_1^*,\dots, m_k^*,\dots, m_N^*), \quad (14)$$

where $m_i^* = 1$, $i = k \cup m_i^* = 0$, $i \neq k$.

Besides, the metadata attributes $v_1,\dots,v_n$ are instantiated with metadata objects, according to integrity constraints t_i of template Γ:

$$[(v_1{:}V_1,\dots,v_n{:}V_n)]t_i = ([v_1]|\Gamma(t_1),\dots,$$
$$[v_n]|\Gamma(t_n)) = (v_1'{:}V_1',\dots,v_n'{:}V_n'). \quad (15)$$

Therewith,

$$V_1' \text{ ISA } V_1,\dots, V_n' \text{ ISA } V_n. \quad (16)$$

The second assignment a_2 results in $(v_1',\dots,v_n')$ template instantiation of the ECM IS web page with content values of $(c_1,\dots, c_n)$:

$$[(v_1'{:}V_1',\dots,v_n'{:}V_n')]c = (v_1'/c_1,\dots, v_n'/c_n), \quad (17)$$

where $c_1{:}C_1,\dots, c_n{:}Cn$, and C_1 *ISA* $V_1',\dots, C_n$ *ISA* V_n'.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Note that the abstraction level of the model elements decreases during transitions from classes to objects, and, further, to their values. Data object expandability maintenance during the meta-level transitions provides extendable software development. According to the types assigned, the objects are aggregated into assignment-dependent collections and they form variable domains. Adequacy, neutrality and semantic integrity of data objects and their components provide problem-oriented ECM with continuous model-level lifecycle support.

Object classes u are defined by means of description $Iu\ \Delta(u)$ with the values of $[Iu\ \Delta(u)]$, where Δ is the selection criterion. Applying assignments $a_1 \in A$ and $a_2 \in A$ from domain A transforms the classes first into objects

$$o = [Iu\ \Delta(u)]\ a_1, \qquad (18)$$

and then into their values: $c = o\ a_2$.

Two-way assignment character (from classes to values and backwards) provides an adequate re-engineering model; description mechanism simplifies modelingin both directions.

Variable domains

$$O_T(A) = \{o \mid o : A \circledast T\} \qquad (19)$$

are constructed as collections of objects o with types T, extracted from problem domain D by selection criteria predicates Δ, while the collection of possible data objects o is contained in D, and the collection of actual ones, $O_T(A)$ is contained in T.

Thus, the basic modelingprinciple can be summarized as follows:

$$[\text{ class of objects }] : \text{assignment} \rightarrow \text{object}, \qquad (20)$$

where the bracketed part refers to the language level of class description, and the rest refers to the software level. The principle means that to declare a class we use such a selection criteria, which identifies functions from assignments into objects, i.e. class is treated like a process.

Data objects are identified as follows:

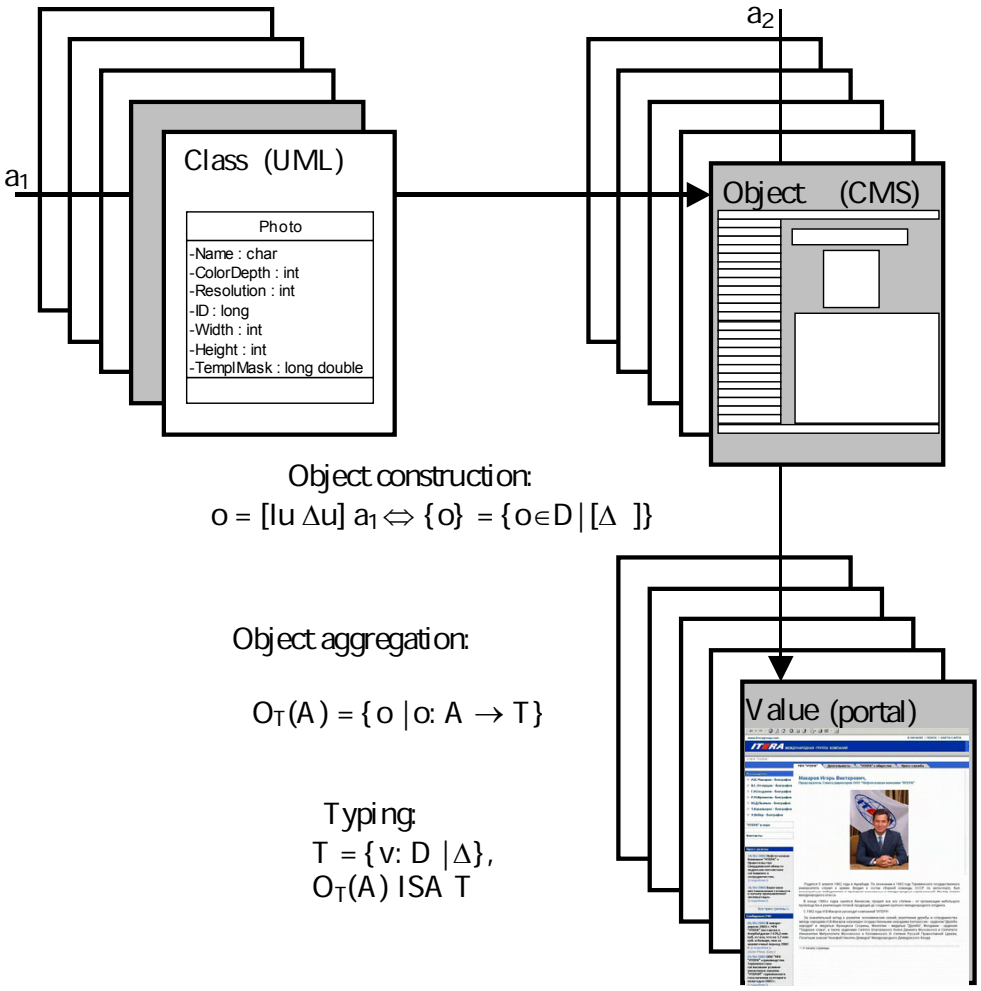
$$[\text{ object class }] : \text{assignment} \rightarrow \text{object} \blacktriangleright \text{object} \blacktriangleright \text{assignment} \rightarrow \text{value} \blacktriangleright \text{value}, \qquad (21)$$

where "▶" means abstraction level decrease.

Thus, the diagram (see fig. 2) illustrates the compression principle

$$o = [Iu\ \Delta u]a_1 \Leftrightarrow \{o\} = \{o \in D \mid [\Delta\ (\bar{o})\ ]\}, \qquad (22)$$

Figure 2. Application of the object-based CMS model to the enterprise portal



and the essence of problem domain modelingprocess, which transfers language-level classes to problem-domain ones by the evaluation function [•]. An object class is modelled by its selection criteria Δ and its description *I*. Evaluation function maps problem domain objects to data definition language ones.

Figure 2 also gives an example of ECM modelling.

The example starts from a data class (a digital photo image) representation in UML language. The class has the following attributes:

ID: char;

Name: int;

ColorDepth: int;

Resolution: long;

Width: int;

Heght: int;

TemplMask: long double;

Let us describe the ECM formal evaluation procedure in a more detailed way. First, let us define content classes as a data warehouse superstructure. They contain (meta)data description formats represented by ordered pairs of <*attribute, type*>. Derived classes are built using a predicate criterion, which collects a certain subset of the basic class objects. Let us establish class hierarchy based on *ISA* partial order relationship.

The latter attribute refers to the bit mask of the ECM template.

The first assignment a1 maps certain template attributes (such as relative position of the digital photo image, etc.) to the resulting enterprise portal web page elements. The second assignment a2 results in evaluation of the ECM template web page by the content values.

IS development methodology has been customized for enterprise portal management, including information and interface parts (i.e. data and metadata) of Internet and Intranet sites. A detailed design scheme is given in [12,15].

According to the scheme, a formal procedure for heterogeneous data warehouse processing is suggested that allows users to interact with the integrated distributed (meta)database in a certain state, depending upon dynamical script-activated assignments. Therewith, scripts depend on user-triggered events and provide transparent intellectual client-server front-end interaction. Dynamic profiles for (meta)database access provide reliable and flexible personalization, high fault tolerance and data security for enterprise information system users in heterogeneous environment.

Further, let us form a user role scheme for each class. The role scheme is defined on the Cartesian product of limited list of user types (such as "author", "editor", "content manager", "editor-in-chief" etc.) over the list of possible operations on the class objects (such as "create", "delete", "edit", "read", "publish", etc.). Roles are represented as a two-dimensional matrix of <*role_name, operation*>.

Let us define a relationship (hierarchy) for object relation modelingand build a template for portal web page publishing. A template is an ordered list of data object classes (it can be a nested one), which specifies the instantiated class metadata elements and their assigned values:

$$<\ldots<\text{class\_name, attribute\_name, value}>,\ldots,<\text{attribute\_name, value}>>\ldots>, \quad (23)$$

where certain attributes may get the undefined value of "⊥".

To formulate selection criteria of the attributes into the template classes, let us use predicate-based assignments.

The procedure of methodology application also involves a warehouse query, which is a low-level embedded procedure. The query is modelled by a logical predicate, which involves all the actual data object classes. The predicate may contain quantifiers (such as "any", "some", "exactly *N*", "at least *N*", "at most *N*", etc.), logical operations (including "and", "or", and "not"), and cause-and-effect relations ("causes", "results in").

Finally, let us form portal web page by evaluating its elements. At this stage, users control object state changes depending on their functional roles and allowed operations. Therewith, attention is paid to object hierarchies and user roles, as well as to data object types. The data object choice criteria for the portal web pages are treated as predicate-based assignments.

5 THE ECM SYSTEM IMPLEMENTATION

The ECM system development approach has been practically approved by constructing Internet and Intranet portals for ITERA International Group of Companies (http://www.iteragroup.com).

The *Menu* module of CMS is aimed at portal navigation data storage and processing. *Pages* module is related to *Menu* and tracks events of assignment, migration and deletion of portal pages, (meta)data and navigation menu items. *Images* module provides storage, retrieval and portal web publication of digital photos and graphics. *News Columns* module supports periodical portal publications (press releases, media news, etc.) including data from related third-party IS modules. *Special sections* module organizes visual management of portal content (i.e. data) and design (i.e. metadata) by given criteria set. *Administration* module implements data personalization, profiling, access control policies and (meta)database synchronization.

In terms of system architecture, the ECM solution provides assignments (depending on front-end position in data access hierarchy) with assignments for (meta)data entry, modification, analysis and report generation (from administrator

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

level down to reader one). Problem-oriented form designer, report writer, online documentation and administration tools make an interactive interface toolkit. (Meta)database supports integrated storage online access to data and metadata (e.g., object dimensions, integrity constraints, representation formats, etc.).

Implementation process included fast prototyping (of both DBMS and scripting) and full-scale integrated ERP-based implementation. The fast prototype proved adequacy of the (meta)data model and of the approach on the whole.

Upon prototype testing, the full-scale ECM solution has been developed, which manages Intranet and Internet portals (http://www.iteragroup.com).

Implementation proved a substantial term-and-cost reduction (over 30% on the average) as compared to commercial software available. Functional features have been essentially improved and include better ERP and legacy IS integration, easier handling complicated objects and smarter report generation. Advanced personalization and access control have substantially reduced risks of (meta)data damage or loss [12,13].

During the ECM, IS specification is transformed from problem domain concepts to data model entities, then to DBMS scheme, and, finally, to target IS description with the required architecture and interface. The innovative semantically oriented visual ConceptModeller ECM tool [12,15] was used to support data model customization. At the final ECM stage, the content management system and the portal were gateways between the data warehouse, Internet and Intranet sites. The original problem-oriented CMS and ConceptModeller [14,15] completed the heterogeneous data sources ECM. In case of content-critical warehouse updates, the content is automatically updated accordingly. Other options include scheduled and manual data updates and retrievals.

As a part of the ECM methodology, the heterogeneous repository processing scheme allows users to interact with the distributed data warehouse in a certain state depending on dynamically activated scripts. Thus, scripts (as data access profiles and object-oriented procedures) are initiated depending on user-triggered events. Scripts provide transparent and intellectual front-end warehousing. Dynamically adjustable content access profiles are implemented by the problem-oriented CMS, which is an intellectual media between users and the warehouse; they provide high fault tolerance and information security [14,15]. Depending on semantically oriented user profile, certain database connection and access levels are dynamically assigned; they remain valid only until the end of data exchange session. Access is granted to the content, i.e., data and metadata (object dimensions, integrity constraints, access rights, etc.).

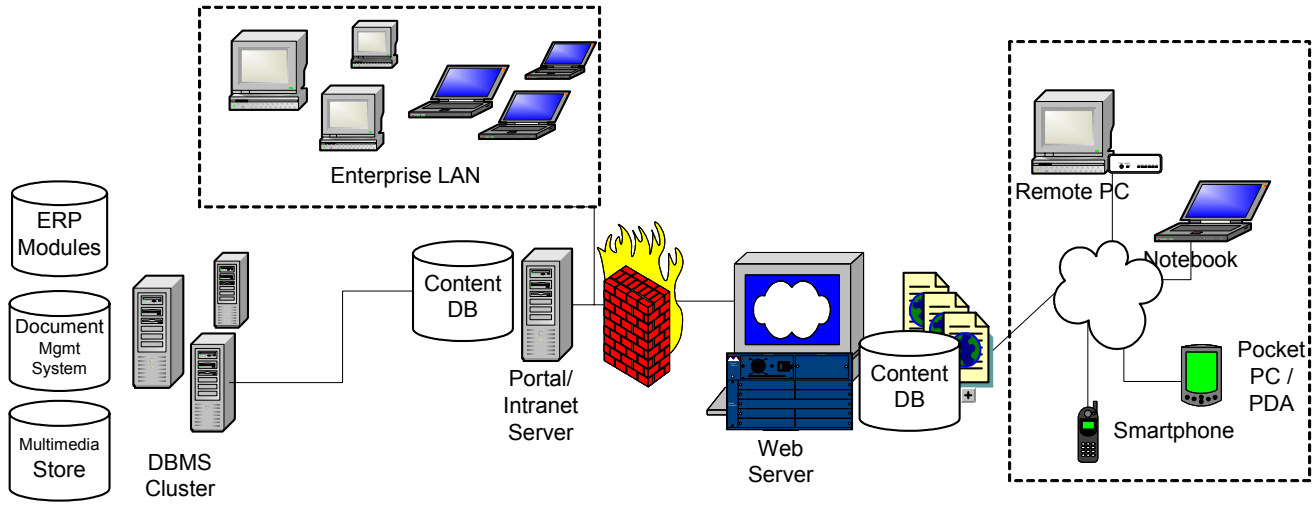
The ECM methodology treats data and metadata objects uniformly. This makes portal interface a problem-oriented, straightforward and uniform one and boosts system performance. Event-driven portal-based warehouse architecture is presented in fig.3. Client-side web page object states change depending on event script execution. Thus, while the warehouse data remains unchanged, users can request a content update or initiate a query. Moreover, the front-end interface itself is also client profile dependent. Options include personal preferences, data access device and browser settings.

Warehouse data access is also profile-dependent. At the upper level of data access hierarchy, clients can be divided into administrators, managers and ordinary users. Judging by the profile, content object state (i.e., system interface) is changed.

Scenario-based interface results in higher degree of interactivity, user-friendliness and security. User profiles (i.e. assignments) are stored in metadata base of visitors, and, depending on their properties, content access and representation level are customized.

ERP legacy system modules have been integrated by means of the portal interface.

Figure 3. Portal-based ECM system architecture



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

The interactive interface is represented by portal-based problem-oriented CMS, which includes form designer, report generator, on-line help and administration tools. The improved enterprise warehouse supports integrated content storage. During the ECM process, semantic network data model specification generated by ConceptModeller is transformed into UML diagrams, then, by means of CASE toolkit - into ERD and, finally, into the integrated warehouse. The innovative semantically oriented CMS and ConceptModeller tool have been used to transform heterogeneous content into a uniform enterprise portal-based warehouse.

The problem-oriented portal manages content of both ERP and legacy system modules. For example, the HR legacy system would provide a number of significant data items for corporation profile portal pages including total staff number, number of countries and companies in the corporation. In case integrated implementation includes vacancy module of the HR legacy system, dynamical updates of HTML page vacancy data become easier.

Similarly, financial components would provide content for a number of periodical or user-triggered financial reports. Content examples may include revenues, profits, production dynamics, stock values, etc. Production manufacturing module would provide productivity and capacity data for executive summaries and company profile. Address book from document management system would serve for contact information and provides automatic feedback routing through corporate organizational structure. To enhance the enterprise portal performance and interface, an event-driven software agent of the innovative CMS could dynamically update data published in the HTML pages.

The integrated enterprise warehouse is stored in the data center. The warehouse is based on original CMS and ConceptModeller tools and it integrates ERP IS modules with legacy systems. The ECM portal solution has successfully passed a three-year test.

6 RESULTS, CONCLUSION AND PROSPECTS

The new methodology of ECM system development supporting the entire lifecycle has been introduced. The methodology is a part of the integrated approach to enterprise IS development [12], which provides adequate, consistent and integrate (meta)data manipulation during its entire lifecycle.

A set of (meta)data object models have been built. It includes state-based dynamical models for problem domain and development tools. The models provide integrated (meta)data object manipulation in the environment of weak-structured heterogeneous problem domains.

An ECM IS for (meta)data object management software development toolkit has been implemented. The IS features content-based architecture with front-end and back-end interfaces.

A fast event-driven prototype and the full-scale ECM IS have been implemented. The IS is capable of managing Internet and Intranet portals for *ITERA International Group of Companies* employing around 10,000 people in nearly 150 companies of more than 20 countries (http://www.iteragroup.com). Implementation results proved substantial terms and costs reduction as compared to commercially available software. The ECM information system is based on a model, which integrates objects management methods for data and metadata.

The author is going to continue his studies of academia-based formal models and enterprise-level SDKs that provide a well-balanced approach to ECM systems modeling, development and implementation.

REFERENCES

1. Barendregt H.P. The lambda calculus (revised edition), Studies in Logic, No.103, North Holland, Amsterdam, 1984.
2. Cousineau G., Curien P.-L., Mauny M. The categorical abstract machine.In: Science of Computer Programming No.8, Vol.2, 1987, pp. 173-202.
3. Roussopulos N.D. A semantic network model of data bases, Toronto Univ., 1976.
4. Scott D.S. Domains for denotational semantics. In ICALP, 1982, pp. 577-613.
5. Wolfengagen V.E. Event-driven objects. In: 1st International Workshop on Computer Science and information Technologies (CSIT 1999). MEPhI Publishers, Moscow, Russia, 1999, Vol.1, pp. 88-96.
6. Lenat D., Guha R.V. Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project. Addison-Wesley, 1990.
7. Witbrock M., Panton K., Reed S.L., et al. Automated OWL nnotation Assisted by a Large Knowledge Base. In: Workshop on Knowledge Markup and Semantic Annotation at the 3rd International Semantic Web Conference (ISWC 2004), Hiroshima, Japan, 2004, pp. 71-80.
8. Masters J., Güngördü Z. Structured Knowledge Source Integration: A Progress Report. In: Integration of Knowledge Intensive Multiagent Systems, Cambridge, MA, USA, 2003.
9. Reed S., Lenat D. Mapping Ontologies into Cyc. In: AAAI 2002 Conference Workshop on Ontologies For The Semantic Web, Edmonton, Canada, 2002.
10. Forbus K., Birnbaum L., Wagner E. et al. Combining analogy, intelligent information retrieval, and knowledge integration for analysis: A preliminary report. In: 2005 International Conference on Intelligence Analysis, McLean, Virginia, USA, 2005.
11. Gurevich Yu., Veanes M., Wallace Ch. Can abstract state machines be useful in language theory? In: Theoretical Computer Science Vol. 376, (1-2), 2007, pp.17-29.
12. Zykov S.V. Integrated Methodology for Internet-based Enterprise Information Systems Development. In: 1st International Conference on Web Information Systems and Technologies (WEBIST 2005), Miami, FL, USA, INSTICC Press, 2005, pp. 168-175.
13. Zykov S.V. Enterprise Portal Content Management: from Model to Application. In: 2nd International Conference on Web Information Systems and Technologies (WEBIST 2006), Setubal, Portugal, INSTICC Press, 2006, pp. 465-468.
14. Zykov S.V. Enterprise Content Management: the Integrated Methodology. In: Information Systems and Web Technologies (EISWT'07), Orlando, FL, USA, 2007, p.p. 226-233.
15. Zykov S.V. An Integral Approach to Enterprise Content Management. In: Proc. of International World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2007), Orlando, FL, USA, 2007, Vol. I, p.p. 212-216.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

i·Society 2007

Some design considerations in context aware and ubiquitous computing

Daniel MacCormac
Fred Mtenzi
Mark Deegan
Brendan O'Shea
School of Computing
Dublin Institute of Technology
Kevin Street, Dublin 8, Ireland
{dan.maccormac, fred.mtenzi, mark.deegan, brendan.oshea}@comp.dit.ie

Charles Shoniregun
School of Computing & Technology
University of East London
Docklands Campus, University Way, London, UK
C.Shoniregun@uel.ac.uk

Abstract Enabling ubiquitous computing through the development of context-aware and smart environments poses opportunities as well as challenges. Due to the youthful nature of the ubiquitous computing era, design considerations, guidelines, and models are not well established. Existing approaches vary from one implementation to the next, and to date, it appears that there is no common, recurring, standard approach to building applications in the ubiquitous computing domain, as opposed to other fields of computing, in which design approaches, models etc, are well established. Consequently there is a strong case for the clarification of such information to aid developers and researchers in building ubiquitous computing systems. In this paper, we present a set of design considerations and techniques to aid in the development of ubiquitous computing applications. By analysing existing implementations and related work, we can derive key design considerations, as well as avoiding pitfalls experienced in past projects.

Keywords Ubiquitous computing, pervasive computing, context-aware, design guidelines.

1 INTRODUCTION

Ubiquitous computing promises an age of calm technology, an age in which hundreds of invisible computers will recede into the background of our everyday life in a seamless manner, serving our every need in a subtle yet graceful manner. Weiser noted that ubiquitous computing will become a reality when three key requirements are satisfied [1]. Firstly, technology must become widespread and be attainable at a low cost. Secondly, a network to tie these widespread low cost devices together is necessary. Finally, applications that can facilitate ubiquitous computing must be developed and deployed. To date, the first requirement has been well met. In addition to the proliferation of mobile computing devices, people now interact with dozens or even hundreds of microprocessors on a daily basis. With the advent of emerging wireless networking technologies such as Bluetooth, 802.11, and increasing bandwidth being offered by 3G providers, the fulfilment of the second requirement is becoming evident. However, there is a strong need for development of ubiquitous computing applications if the third requirement is to be met. Anhalt states that to minimize distractions, a pervasivecomputing environment must be context-aware [2]. Inferring circumstance is a common approach, which can aid in delivery of seemingly "smart" services to end-users; therefore we also discuss context-aware systems in this paper.

Since the origins of ubiquitous computing in the early 90's, there has been steadily growing interest in research in this field, supported by various projects at institutes such as Berkeley, MIT and Stanford as well as major companies within the industry such as IBM, Microsoft and HP [3-7]. This effort has led to the development of many prototypes, proof of concept frameworks, middleware, and applications helping to validate this vision. Despite the implementation of many such systems, there is still no commonly recurring approach to designing ubiquitous computing systems, and approaches vary from one implementation to the next, as noted by Kranz [8]. This is not the only challenge facing designers. Satyanarayanan discusses the challenges involved

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

in designing and implementing ubiquitous computing systems, noting that it is more difficult to design and implement a ubiquitous computing system than a simple distributed system of comparable robustness and maturity [9]. Similarly, Grimm et al. note in their work that despite the proliferation of mobile computing devices and technologies, very few applications run in the ubiquitous computing infrastructure, which the authors believe stems largely from the fact that it is currently too hard to design, build, and deploy applications in the pervasive computing space [10]. Pervasive computing presents many design challenges. Understanding user needs and clarifying design considerations early on in the development cycle is crucial, yet this can be difficult due to the diversity of past examples and the lack of clearly defined design guidelines. From analysis of existing frameworks, past approaches, and discussions presented in a variety of research papers, we outline and discuss some key design considerations for ubiquitous computing and context aware applications. These include understanding the nature of ubiquitous computing, requirements and components for context aware infrastructures, high-level design patterns in ubiquitous computing, as well privacy considerations. We make no claim of completeness or exclusiveness; the topics outlined are merely a selection of key considerations from a variety of work. By identifying these issues early on in the development cycle, we hope to learn from past examples, and avoid pitfalls experienced in previous implementations.

2 RELATED WORK

Many tools and approaches to aid in ubiquitous computing development have emerged in recent years, yet examples show that approaches are still varied, or specific to a particular class of application [11-14]. Schmidt and Terrenghi have previously discussed methods and guidelines for the development of ubiquitous computing applications in a domestic environments [15]. Gemperle et al. explore the concept of wearable computing, presenting and discussing guidelines for wearable systems [16]. Hall and Bannon present the results of their design process which involved exploration of techniques and the feasibility of using ubiquitous computing to stimulate participation by children visiting museums [17]. Björk et al. have presented their experiences designing ubiquitous computing games. The literature includes reports on the exchange of methods, techniques and technologies usable for future research in ubiquitous computing games [18]. Our work does not focus on a specific subset of applications or deployment environment, but rather on ubiquitous computing in the broader context. In this section we briefly discuss related work from which we draw both examples and techniques. We aim to create a general list of considerations, resulting from analysis of a range of work.

Landay et al. have focused considerable energy on identifying design patterns within ubiquitous computing [19-21]. Additionally, it has been noted by the authors that due to the youthful nature of ubiquitous computing, there are a lack of clearly defined design guidelines and patterns, and reusing knowledge attained by past developers has proven to be successful [21]. Chung et al. have continued Landay's work of identifying design patterns within ubiquitous computing. The authors have invited the public to submit design patterns, as well as discussing how such design patterns can be identified [22-24]. We also draw conclusions from design patterns and principles outlined in various other literature [25, 26]. Weiser and Brown outline the key attributes of *calm technology* in their early article on ubiquitous computing, giving several examples of calm technology in today's world [27]. From this work, we can gain a deeper understanding of how ubiquitous computing applications should behave, and derive key design considerations to aid in the development of calm technology. Dey and Abowd have previously discussed their approach to development of context aware applications [28]. We incorporate design requirements and components from their work into our own design considerations.

Ubiquitous computing redefines traditional interaction between users and computing devices. This new type of interaction creates a strong case for the development of new approaches when building applications, as noted by Beale [29]. In response to this, Beale has proposed an intuitive form of design known as *slanty* design [29]. We discuss slanty design, and it's advantages and disadvantages in Section 5. Ubiquitous computing systems, and location based systems in general, raise many social fears and privacy issues among users, and consequently the topic of privacy within ubiquitous computing has been widely discussed [26, 30-33]. We consider privacy to be a crucial design consideration. We have studied real world examples as well as research papers discussing the topic of security and privacy within ubiquitous computing [34-39]. Additionally, we also derive design considerations from several existing frameworks and applications [36, 40-43]. As an example, Long et al. have previously described their experiences in relation to the development of a mobile tour guide application. In their design reflections, the authors note the need for rapid prototyping and reiterative design [40]. We can incorporate such observations into our own design considerations and consequently, developers can avoid reinventing the wheel.

3 HOW SHOULD UBIQUITOUS APPLICATIONS BEHAVE?

Understanding the role of ubiquitous computing and how applications should behave is important in relation to developing supporting applications. Moving beyond the traditional desktop computer and its associated applications, ubiquitous computing applications require a new approach to application development, and define an entirely new kind of relationship with users. For a system to be truly unobtrusive users must interact with it in a sub-conscious manner. It must fade into the background of daily life, while being readily available to reveal itself in a flexible manner when a user desires so. From this vision, the concept of the *centre and periphery* emerge [27, 44]. Understanding this concept is crucial to building ubiquitous computing applications that act in a subtle yet informative manner.

We use "periphery" to name what we are attuned to without attending to explicitly [45]. Take driving a car as an example. Normally, when driving, our attention is focused on the

road, operating the controls of the car, and perhaps the radio. Such objects are in the centre of our attention. Our attention is not tuned to the noise of the engine, even though we are aware of it. It is in the periphery of our attention. However, if there is a sudden change in the engine noise, we can quickly attend to it, moving it from the periphery to the centre of our attention. Well designed ubiquitous computing applications can easily move objects from the periphery of our attention to the centre and back to the periphery [27].

There are two key advantages to this approach. Firstly, we can handle many more objects in the periphery than we can in the centre of our attention. Objects in the periphery are attended to by the sensory portion of our brain, and consequently are informative without being overburdening. Secondly, by moving an object from the periphery to the centre, we take control of the object, allowing us to attend to and fix discomforts. We may be aware that an object in the periphery is not behaving as desired, so we must move this object into the centre to attend to the problem.

Designing for the periphery allows us to take advantage of ubiquitous computing services, without being dominated by them. While this is not appropriate for all applications, it can certainly be appropriate for many ubiquitous computing applications with which we aim not to overburden the user.

4 APPLYING EXISTING KNOWLEDGE THROUGH THE USE OF DESIGN PATTERNS

A pattern is the abstraction from a concrete form which keeps recurring in specific non-arbitrary contexts [46]. Design patterns are a general reusable solution to a commonly occurring problem. They are written in a flexible manner, and thus can be reused in many situations, supporting the reuse of existing knowledge. It is not an empirical solution, but rather a description of how to solve a problem that can be used in many different situations.

An example of a design pattern in ubiquitous computing applications is shown below [21].

Problem: *Ubiquitous computing devices will be used in a variety of locations and situations, but the device interfaces must not interrupt or distract the user from performing a primary task or annoy a nearby group of people.*

Solution: *Input and output modalities should adapt to the user's current context.*

Due to the young nature of ubiquitous computing research, design patterns are still emerging, and to date, a considerable amount of work can be attributed to Landay et al. [21], Chung et al. [22-24] as well as others [8, 25].

Some further interesting design patterns identified to date include:

- *Ad-hoc Association.* When nomadic users wish to collaborate in some fashion, it is important that they should not have to spend time configuring their devices to work together. When users are within each other's proximity, connecting and associating devices should be made as simple as possible, without the need for manual configuration. The resulting association should enable them to them share information over the life of a session.

- *Service handoff.* Users need to be truly mobile in ubiquitous computing environments, and should not be limited geographically. To support this, we need to provide seamless and transparent handoff of services across multiple services stations.

- *Proxies for devices.* The sheer variety of heterogeneous devices and applications in existence today complicates the task of creating ubiquitous computing applications. To overcome this, transformation and interpretation can be performed at a proxy.

Chung et al. have identified over 60 design patterns [22-24]. In addition to listing design patterns, their work also discusses how one can identify design patterns from existing work in the field. By leveraging this existing work, we can build betterquality applications in a shorter development time.

5 MOVING BEYOND TRADITIONAL DESIGN

Despite the fact that computers have progressed considerably over the last few decades, they are still quite difficult and frustrating to use [47]. Ubiquitous computing promises an end to this. Consequently, to design for the ubiquitous computing era, we must apply more creative approaches.

In the earlier days of software development, developers employed the waterfall model. However, often it did not produce very good results, placing too much emphasis on requirements documents, and resulting in a final product that did not match the changing requirements of the clients. Today, a more modern approach to designing effective software involves the concept of user-centred design [48, 49]. User-cantered design gives greater weight to user experience, which involves consulting users at each stage of the development process, validating or questioning the decisions of designers and programmers. This approach has proven itself effective in recent years [50]. In response to today's ever-changing technological landscape, Beale outlines a new approach to design, which he calls *slanty design* [29].

Central to the concept of slanty design, is the notion that software developers must move beyond user centric design to develop systems with better support for modern day context. The approach involves designing systems so that it is intentionally easy to perform certain desirable tasks, while being difficult or impossible to perform undesirable tasks. Slanty design states that more functionality increases the potential for user error. Take multitasking for example. The user now has access to multiple windows simultaneously. However, if the user is performing data entry, this increases the chance

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

of information being entered into the wrong window. The Apple iPod on the other hand, takes a more slanty approach to design. Performing tasks such as track playback or shuffle is easily achievable, while we are shielded from performing undesirable tasks such as deleting tracks. So, using this approach, we can guide users towards performing tasks, which are desirable, while guiding them away from performing undesirable tasks. In addition to this notion, slanty design also involves incorporates clean usability. By this, we mean delivering usability on the important issues, but without the unforeseen consequences that allow users to create a new set of problems for themselves and possibly others [29].

Slanty design can be achieved using five key design steps [29].

- Identify user goals
- Identify user non-goals -- things users don't want to be able to do easily (such as deleting all their files)
- Identify wider goals being pursued by other stakeholders, including where they conflict with individual goals
- Follow a user-centred design process to create a system with high usability for user goals and high anti-usability for user non-goals
- Resolve the conflicts between wider issues and individual goals, and where the wider issues without ensure that the designmeets these needs.

These five design steps aid in slanty design. However, we also want to develop clean usability, ensuring that the system is beneficial to a wide audience. This may require many iterations of the development cycle before the system becomes cleaner.

6 AN APPROACH TO BUILDING CONTEXT-AWARE FRAMEWORKS

Inferring circumstance is a common approach, which can aid in delivery of seemingly "smart" services to end-users. As a result, we discuss design considerations for context-aware systems in this section.

Context is defined as:

"Any information that can be used to characterize the situation of entities (i.e. whether a person, place or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves. Context is typically the location, identity and state of people, groups and computational and physical objects." [51]

6.1 Design requirements

Dey and Abowd have successfully applied their chosen approach to building context aware in past example [28]. Employing the requirements and components outlined in this work, we can build highly flexible and robust context aware frameworks to aid in ubiquitous computing environments. Below we outline these requirements and corresponding components.

1. *Separation of concerns*. It is important to ensure that the sensors used to acquire context information and the application logic which handles this context are clearly defined as two distinct entities. Some past examples have hardwired the sensor drivers into the application itself, leading to an inflexible solution [52, 53]. Such an approach does not support good software engineering practices, and makes reusability and modification burdensome. Ideally, context information should be handled in a similar fashion to the handling of user input in standard software engineering, i.e. software developers can use input without considering how the input was attained (keyboard, mouse, speech, pen device etc.)
2. *Context Interpretation*. Context information may need to pass through several layers before reaching the application. For example an interpretation layer could convert an ID number to a corresponding name. Such layers should be completely transparent to the programmer. To support this transparency, context must be interpreted before it reaches the application layer. Since different applications may be interested in different levels of information, the approach of separating context interpretation promotes reusability and flexibility.
3. *Transparent, distributed communications*. When developing sensor driven context-aware applications, it is important to bear in mind devices used to sense context are not likely to be attached to the computer running the application, but instead may be distributed across a large physical environment. Additionally, multiple such applications running across a variety of servers may employ these sensors (many-to-many relationship). It is desirable that the distribution of hardware and software is transparent to the programmer. Such an approach helps to simplify the deployment of both sensors and applications.
4. *Constant Availability of Context Acquisition.* In traditional software development, components such user interface items, are instantiated, controlled and used by a *single* application. In the case of context aware applications, applications should be able to query existing sensor components when necessary, as opposed to instantiating their own instances of objects. Multiple applications must be able to access the same piece of context. Thus, the components that attain context information (sensors etc) must execute autonomously from the applications that use them. These components must be continuously available, allowing applications to gather information on demand.
5. *Context storage and history.* Components that acquire context should record a detailed log of all information attained. This context history can consequently be used to view past trends and attempt to predict future trends. Components will collect context information during periods when no particular application may be interested in this information. Since no application will be available to store this information, it is the duty of component itself to store this information. At later stage, applications can then query sensor components for information about previous context events.

It is desirable that the context history is stored at a fine granularity in order to support detailed queries from applications.

6. *Resource discovery*. In order for applications to take advantage of sensors within a particular environment, the application must be aware of the existence of these sensors. Furthermore, it must be aware of sensor details such as what information the sensor can provide, its geographical location, what communication methods it supports, hostname, port etc. To hide these details from the application, the architecture must supports some form of resource discovery [54].

6.2 Framework components

Any context-aware ubiquitous computing application built using this approach can be separated into several components as noted by Dey and Abowd [28]. We now present a brief overview of these components.

1. *Context widgets*. Drawing on the abstraction of traditional widgets used in GUI applications, *context widgets* hide the complexity of the actual sensors being used by the applications, much in the same manner as GUI widgets hide the complex details of how the user input is collected. The properties of the underlying sensor are completely concealed from the application by the widget. As an example, a location monitor widget will notify an application when a user moves from one location to next while concealing the low level data used to make this assumption.
2. *Interpreters.* As discussed in section 6.1, context needs to be interpreted, passing through multiple layers before reaching the application. The role of *Interpreters* is to carry out such actions. Interpretation involves raising the level of abstraction associated with a piece of context. For example, at a low level, context may be represented as a serial number of a tracking device, while at a higher level, this information can be represented as a username. Similarly, location information could be raised from the ID of sensor node, to a corresponding room name or number. This supports separation of concerns, allowing multiple applications to take advantage of the same interpreter.
3. *Aggregators.* Aggregators are responsible for collecting and combining multiple pieces of context information that are related in some manner. The distributed physical nature of sensors, in addition to the complex requirements of applications creates the need for such aggregation. As an example, a ubiquitous computing application in an academic environment may wish to behave accordingly when; a student in course A enters the laboratory, who is taking module B, and has yet to submit assignment C. Aggregators simplify the task of gathering various context-based information, and as with *Interpreters*, multiple applications can make use of output from a single aggregator.
4. *Services*. Services are components within the context-aware framework that execute actions on behalf of applications. Separating services from applications removes the need for each application to implement the service, and furthermore, removes the need for an application to understand the complexity of service operation. Therefore, services are building blocks that are available to multiple applications. The service is responsible for managing the surrounding environment via an *actuator* (output of the service).
5. *Discoverers*. As discussed in section 6.1, there is a need for resource dynamic discovery within a context-aware framework. The role of discovers is to maintain information about what capabilities exist in the framework. This could include information on components such as widgets, interpreters, aggregators, and services available. There are two methods that an application can use to query a discoverer in relation to available components. Firstly, an application can search by name or identity using the *white pages* lookup facility. Secondly, applications can lookup services based on a particular category, which is referred to as the *yellow pages* lookup facility.

7 A NOTE ON PRIVACY IN UBIQUITOUS COMPUITNG

7.1 Overview

Privacy in ubiquitous computing per se is an extremely broad topic, which has prompted a plethora of articles, papers, and discussions; both negative and positive. As a result, we were initially hesitant to discuss privacy is this paper. However, we decided to include a very brief note on privacy due to its pivotal role in ubiquitous computing. Privacy is possibly the most criticised aspect of ubiquitous computing. Criticisms include interviews [55-57], books [58, 59], articles [60, 61] and media coverage [62, 63]. In this section we will highlight some of the issues in relation to privacy within ubiquitous computing, and suggest an architecture and set of techniques that can aid developers in building privacy sensitive applications.

The concept of ubiquitous computing often raises many social fears among users. Because such systems can pinpoint the location of an individual at any given time, users often feel that management may misuse such systems, resulting in the "big brother" effect. However, studies have shown that these fears are generally unfounded, and are quickly dispelled following the initial trial period the system [34-36]. Moreover, the ability to track user location is already present. We are already living in an age of ubiquitous computing, in which we interact with dozens of microprocessors everyday. Networking these devices together in addition to making them smarter is the bigger challenge that poses itself. Cell phones, ATM and credit cards, internet access, GPS navigation, domain logon at work, to name but a few, are all tools which can be employed to paint a picture of our daily routine. Many critics and individuals alike fail to note or realise this. We are not advocating that users relinquish their right to privacy, but merely noting that there are trade-offs to be considered when using such technologies. In this section we outline some approaches to designing ubiquitous

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

computing applications with privacy in mind, a key concern of many users.

Confab provides an architecture and a suite of privacy mechanisms that allow application developers and end-users to support a variety of trust levels and privacy needs within context-aware ubiquitous computing [64]. Confab provides mechanisms and user interfaces that facilitate the creation of three basic interaction patterns for privacy-sensitive applications: optimistic, where an application shares personal information and detects abuses by default; pessimistic, where it is more important for an application to prevent abuses; and mixed-initiative, where decisions to share information are made interactively by end-users.

7.2 Approaches to improving privacy

Control and Feedback The main concern among users is the collection and logging of information about their past and present location. In certain environments, users may fear that management is secretly monitoring their activities. We suggest that users should have the ability to choose whether or not they wish avail of such a service, and if a user desires so, they can simply choose not to carry their tracking device. Additionally, there are some situations in which users may require a finer granularity of control. For example, users may wish to avail of some services, but not all. They may wish to be viewable by some people, but not everyone. To address this, we can employ the principles of *control and feedback* [37], as widely suggested in many articles [30, 33, 38, 65] on privacy in ubiquitous computing.

Control: Empowering people to stipulate what information they project and who can get hold of it.

Feedback: Informing people when and what information about them is being captured and to whom the information is being made available.

Using control and feedback, users can control their level of interaction with the system, as well as their visibility to applications and people. When implementing this approach, it is important that we implement these control mechanisms in the periphery. Users do wish to spend time configuring control and feedback information through traditional GUIs, but rather the system should keep the user informed in a subtle manner, as well performing dynamic and automated control and feedback to some degree. An example of an application which implements the feedback and control approach is the calendar mirror [66]. This application combines a display of a users calendar with information about how the calendar information has been accessed by others.

Decentralisation Another method of increasing privacy is the concept of decentralisation. Decentralisation can reduce the likelihood of large amounts of personal data falling into the wrong hands. Using this approach, instead of storing large amounts of data on central servers, information is disseminated across all devices in the environment. Personal information about each user stored on their workstation, laptop or other computing device for example. An application that needs to query the location of a user must query the computing device of the user in question to attain this information, passing any security and trust checks necessary. EuroPARC have employed this approach in their UbiComp systems [39]. This approach is also desirable as it promotes scalability and robustness [67].

User Education Perhaps the most crucial approach to improving privacy lies in the education of end users. Dissemination of information relating to the potential dangers of ubiquitous computing environments is possibly one of the most powerful tools we can employ in terms of security. This approach can help to decrease the chance users private information being comprised, as well as helping to raise the level of acceptance of the system.

As noted by Weiser, *"the problem, while often couched in terms of privacy, is really one of control. If the computational system is invisible as well as extensive, it becomes hard to know what is controlling what, what is connected to what, where information is flowing, how it is being used, what is broken (vs. what is working correctly, but not helpfully), and what are the consequences of any given action (including simply walking into a room). Maintaining simplicity and control simultaneously is still one of the major open questions facing ubiquitous computing research."* [68].

8 SUMMARY AND CONCLUSION

In this paper we have touched on some crucial design issues in relation to ubiquitous computing applications, as well as applications which leverage context information as means of delivering services which are somewhat "smart". We outlined the concept of the centre and the periphery of our attention in section 3. We believe that this approach can be highly effective in building ubiquitous computing applications that do not overburden the user, and hence have a high level of usability and acceptance. In section 4, we touched on high level design patterns, outlining some problems and lessons learned to date which can help in building better applications in the future. We also discussed the concept of *slanty* design in this paper, an emerging design approach which we believe is applicable to ubiquitous computing applications due to its intuitive nature. Considering the importance of contextual information in ubiquitous computing, we drew on Dey and Abowd's work in the field on context aware computing, and in section 6, we outline an approach to building scalable, reusable and robust context-aware frameworks [28]. Finally, we briefly touched on some privacy issues in ubiquitous computing in section 7, outlining the some approaches to building privacy sensitive applications in the ubiquitous computing domain, as well as nominating a framework and set of techniques to aid in developing such applications [64]. We have attempted to gather and highlight some key design considerations for ubiquitous computing and context-aware frameworks. Due to the infancy of ubiquitous computing, common practices and patterns for implementing such systems are still varied and not well established [8, 22, 23]. Additionally, there are few independent design documents, and design approaches and reflections

are often nested within papers outlining implementations of various projects, which can be difficult to identify. Thus, we have attempted to compile a selection of design considerations from a variety of sources. These sources include design guidelines, reports, discussions, and reflections gathered in post implementation phases. The intuitive nature of ubiquitous computing creates a strong need new approaches to developing frameworks and applications; *Slanty* design being one such example. This approach provides applications with the inherent ability to guide users towards desirable actions in a sub conscious manner, employing the periphery of our attention, thus reducing the level of burdensome interaction. By analysing past examples and existing work, we hope to gain a deeper understanding of the design issues within the ubiquitous computing domain as well as contributing to existing work in the area.

ACKNOWLEDGEMENTS

Dan MacCormac gratefully acknowledges the contribution of the Irish Research Council for Science, Engineering and Technology: funded by the National Development Plan. The authors also wish to thank Dr. Kudakwashe Dube for his discussion on this topic.

REFERENCES

1. Weiser, M. (1991), The Computer for the 21st Century, Scientific American, 265 / 3 pp. 94-104
2. Anhalt, J., et al. (2001), Toward context-aware computing: experiences and lessons, Intelligent Systems, IEEE [see also IEEE Intelligent Systems and Their Applications], 3 / 16 pp. 38-46
3. Wang, H., et al. (2000), ICEBERG: An Internet-core Network Architecture for Integrated Communications, IEEE Personal Communications Magazine.
4. Massachusetts Institute of Technology. (2007), Project Oxygen, http://oxygen.lcs.mit.edu/ (Accessed August 2007)
5. Stanford University. (2007), Interactive Workspaces Project, http://iwork.stanford.edu/ (Accessed August 2007)
6. Singer, M., Ubiquitous Computing Research Spreads. 2004. http://www.internetnews.com/ent-news/article.php/3448171
7. Microsoft Corporation. (2001), Easy Living, http://research.microsoft.com/easyliving/ (Accessed August 2007)
8. Kranz, M., Design Guidelines and Design Patterns for Ubiquitous Computing. 2006.
9. Satyanarayanan, M. Pervasive Computing: Vision and Challenges. 2001.
10. Grimm, R., et al. Systems Directions for Pervasive Computing. in 8th Workshop on Hot Topics in Operating Systems. 2001.
11. Dey, A., et al., a CAPpella: Programming by demonstration of context-aware applications. 2004.
12. Hartmann, B., et al. d.tools: Visually Prototyping Physical UIs through Statecharts. in Ext. Abstracts of ACM Symposium on User Interface Software and Technology (UIST 2005). 2005.
13. Lee, J.C., et al. The calder toolkit: wired and wireless components for rapidly prototyping interactive devices. in DIS '04: Proceedings of the 2004 conference on Designing interactive systems. 2004. New York, NY, USA: ACM Press.
14. Li, Y., J.I. Hong, and J.A. Landay. Topiary: a tool for prototyping location-enhanced applications. in UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology. 2004. New York, NY, USA: ACM Press.
15. Schmidt, A. and L. Terrenghi. Methods and guidelines for the design and development of domestic ubiquitous computing applications. 2006. New York, NY, USA: ACM Press.
16. Gemperle, F., et al., Design for Wearability, in Proceedings of the 2nd IEEE International Symposium on Wearable Computers. 1998, IEEE Computer Society.
17. Hall, T. and L. Bannon, Designing ubiquitous computing to enhance children's interaction in museums, in Proceeding of the 2005 conference on Interaction design and children. 2005, ACM Press: Boulder, Colorado.
18. Björk, S., et al. (2002), Designing Ubiquitous Computing Games – A Report from a Workshop Exploring Ubiquitous Computing Entertainment, Personal Ubiquitous Comput., 5 / 6 pp. 443-458
19. Landay, J. and G. Borriello. (2003), Design patterns for ubiquitous computing, IEEE Computer, 36(8) pp. 93-95
20. Landay, J.A. Informal Tools for Designing Anywhere, Anytime, Anydevice User Interfaces. in Diagrams. 2002.
21. Landay, J. and G. Borriello. (2003), Design patterns for ubiquitous computing, IEEE Computer, 36(8) pp. 93-95
22. Chung, E., J. Hong, and J. Lin, Patterns in Ubiquitous and Context Computing. 2007. http://kettle.cs.berkeley.edu/ubicomp/
23. Chung, E., et al., Design Patterns in Ubiquitous Computing. 2003. http://www.ece.cmu.edu/~echung/ubicomp-patterns-talk.ppt
24. Chung, E.S., et al. Development and evaluation of emerging design patterns for ubiquitous computing. in DIS '04: Proceedings of the 2004 conference on Designing interactive systems. 2004. New York, NY, USA: ACM Press.
25. Riva, O., et al. Unearthing Design Patterns to Support Context-Awareness. in PERCOMW '06: Proceedings of the 4th annual IEEE international conference on Pervasive Computing and Communications Workshops. 2006. Washington, DC, USA: IEEE Computer Society.
26. Salvador, T., S. Barile, and J. Sherry. Ubiquitous computing design principles: supporting human-human and human-computer transactions. in CHI '04: CHI '04 extended abstracts on Human factors in computing systems. 2004. New York, NY, USA: ACM Press.
27. Weiser, M. and J.S. Brown. (1996), The coming age of calm technology, Beyond Calculation: The next fifty years of computing,
28. Dey, A.K. and G.D. Abowd, A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. 2001, Human-Computer Interaction.
29. Beale, R. (2007), Slanty design, Commun. ACM, 50 / 1 pp. 21-24
30. Cadiz, J. and A. Gupta, Privacy Interfaces for Collaboration. 2001, Microsoft Research, Redmond.
31. Kaasinen, E. User needs for location-aware mobile services. 2003. London, UK: Springer-Verlag.
32. Lederer, S., J. Mankoff, and A.K. Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. in CHI '03: CHI '03 extended abstracts on Human factors in computing systems. 2003. New York, NY, USA: ACM Press.
33. Hong, J.I., et al. Privacy and Security in the Location-enhanced World Wide Web. In Workshop on UbiComp Communitites: Privacy as Boundary Negotiation (UbiComp 2003). 2003.
34. Want, R., et al., The Active Badge Location System. 1992: ORL, 24a Trumpington Street, Cambridge CB2 1QA. p. .
35. Barkhuus, L. and A.K. Dey. Location-based services for mobile telephony: a study of users' privacy concerns. in INTERACT 2003, 9th IFIP TC13 International Conference on Human-Computer Interaction. 2003.
36. Colbert, M. A diary study of rendezvousing: implications for position-aware computing and communications for the general public. in GROUP '01: Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work. 2001. New York, NY, USA: ACM Press.
37. Bellotti, V. and A. Sellen. Design for Privacy in Ubiquitous Computing Environments. 1993: Kluwer.

38. Hong, J.I. and J.A. Landay. An Architecture for Privacy-Sensitive Ubiquitous Computing. in Second International Conference on Mobile Systems Applications and Services (Mobisys). 2004.
39. Weiser, M. (1993), Some Computer Science Issues in Ubiquitous Computing, Commun. ACM, 7 36 pp. 74-84
40. Long, S., et al. Rapid Prototyping of Mobile Context-Aware Applications: The Cyberguide Case Study. in Mobile Computing and Networking. 1996.
41. Want, R., et al., The Active Badge Location System. 1992: ORL, 24a Trumpington Street, Cambridge CB2 1QA.
42. Bahl, P., A. Balachandran, and V. Padmanabhan, Enhancements to the RADAR User Location and Tracking System. 2000.
43. Bahl, P. and V.N. Padmanabhan. RADAR: An In-Building RF-Based User Location and Tracking System. Proceedings of INFOCOM (2). 2000.
44. Weiser, M. and J.S. Brown, Designing Calm Technology. 1995.
45. Brown, J.S. and P. Duguid, Keeping It Simple: Investigating Resources in the Periphery. 1993.
46. Appleton, B., Patterns and Software: Essential Concepts and Terminology. 2000. http://www.cmcrossroads.com/bradapp/docs/patternsintro.html
47. Lazara, J., et al. (2006), Severity and impact of computer user frustration: A comparison of student and workplace users, Interacting with Computers, 18 pp. 187-207
48. Norman, D. and S. Draper. (1986), User-Centered System Design: New Perspectives on Human-Computer Interaction, Lawrence Erlbaum Associates Inc., Hillsdale, NJ.,
49. Ibm, IBM Ease of Use: User Engineering. 1996. www.03.ibm.com/easy/page/
50. Vredenburg, K., et al. A survey of user-centered design practice. in SIGCHI Conference on Human Factors in Computing Systems. 2002: ACM Press, New York.
51. Dey, A.K. and G.D. Abowd. Towards a better understanding of context and context-awareness. Proceedings of the Workshop on the What, Who, Where, When and How of Context-Awareness. in Proceedings of the Workshop on the What, Who, Where, When and How of Context-Awareness (affiliated with the CHI 2000 Conference on Human Factors in Computer Systems). 2000. New York, NY: ACM Press.
52. Harrison, B.L., et al. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. in CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems. 1998. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co.
53. Rekimoto, J. Tilting Operations for Small Screen Interfaces. in {ACM} Symposium on User Interface Software and Technology. 1996.
54. Schwartz, M.F., et al. (1992), A Comparison of Internet Resource Discovery Approaches, Computing Systems, 4 / 5 pp. 461-493
55. Barkhuus, L. and A.K. Dey. Location-based services for mobile telephony: a study of users' privacy concerns. 2003.
56. Harper, R., Why do people wear active badges? 1993, Rank Xerox Research Centre, Cambridge Laboratory.
57. Kaasinen, E. (2003), User Needs for Location-aware Mobile Services, Personal and Ubiquitous Computing, 7 pp. 70-79
58. Brin, D. (1998), The Transparent Society, Perseus Books,
59. Garfinkel, S. (2001), Database Nation: The Death of Privacy in the 21st Century, O'Reilly & Associates,
60. Doheny-Farina, S. (1994), The Last Link: Default = Offline, Or Why Ubicomp Scares Me, Computer-mediated Communication, 6 / 1 pp. 18-20
61. Talbott, S. The Trouble with Ubiquitous Technology Pushers or Why We'd Be Better Off without the MIT Media Lab. 2000.
62. Sloane, L., Orwellian Dream Come True: A Badge That Pinpoints You. 1992, New York Times.
63. Whalen, J., You're Not Paranoid: They Really Are Watching You. 1995, Wired Magazine.
64. Hong, J.I. and J.A. Landay. An Architecture for Privacy-Sensitive Ubiquitous Computing. 2004.
65. Harper, R., Why do people wear active badges? 1993, Rank Xerox Research Centre, Cambridge Laboratory.
66. Mynatt, E.D. and D. Nguyen. Making ubiquitous computing visible. 2001: ACM Press.
67. Jiang, X. and J.A. Landay. (2002), Modelling privacy control in context-aware systems, IEEE Pervasive Computing, 3 / 1 pp. 59-63
68. M. Weiser, R.G. and J.S. Brown. (1999), The Origins of Ubiquitous Computing Research at PARC in the Late 1980s, IBM Systems Journal, 38 / 4 pp. 693-696

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e•Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

An information support service for moderators of SME company networks

Heiko Thimm
Institute for Business Information Systems, University of Applied Sciences Kiel, Germany,
heiko.thimm@fh-kiel.de

Kathrin Thimm
Institute for Business Information Systems, University of Applied Sciences Kiel, Germany,
kathrin.thimm@fh-kiel.de

Karsten Boye Rasmussen
Department of Marketing and Management, University of Southern Denmark,
kbr@sam.sdu.dk

Abstract Inquiries received by a business network of collaborating companies often demand that a specific combination of network members needs to be selected for the request handling. For companies being chosen and, in turn, allocated to an inquiry this may result into revenue, ultimately. Therefore, a careful and transparent allocation process is required to avoid that network members feel discriminated. On the other hand, for the purpose of responsiveness of the network this orchestration process needs to be completed rapidly. Proposed solutions include tendering and special group decision processes. Because of the administrative and communication overhead of such approaches, however, company networks of small and medium size enterprises (SME) often deploy simpler solutions where the orchestration is performed as a single person task by the network moderator. We present an approach for an SME-suitable service to support this task of moderators. Our service is based on a heuristic multi-criteria optimization approach. Through the optimization criteria the specific properties of each inquiry may be reflected such as the requestor's reputation and the economic conditions of the network like a balanced business volume across all network members.

Keywords Company Networks, Virtual Organizations, Collaborative Business, Cross-Company Business Process Orchestration, Collaboration Platforms.

1 INTRODUCTION

During the last decades for companies it has become substantially harder to survive on the market due to the globalization of the markets, strong competition among suppliers, and tight constraints imposed by the world economy, to name only a few reasons. It has been proposed from different research disciplines that companies should participate in new forms of alliances with other companies to deal with these conditions [14, 18]. These special forms of alliances include company networks in which companies come together to act on the market collaboratively in a well coordinated form. This means that the members of the company network are required to comply with rules and regulations defined for the network. In the context of this presentation we are not treating the question of the establishment and population of the network. The network is distinguished from an electronic market by having a population and by solving requests that comprise uncertainty by including elements of non-tested products and procedures. For so-called transactional company networks where operative-level collaborative business processes occur within the network and with external companies, an authority has been proposed to coordinate the collaboration [8, 9, 13]. We refer to this authority as moderator. Alternative names found in the literature are coordinator, network manager, and broker.

Internet-based collaboration platforms such as BSCW [1], VICOPLAN [9], and Teamspace [6] are proposed as an approach to offer extensive IT-support to company networks. It appears that only a few of these platforms consider the concept of a moderator at all. From our experience with company networks, we learnt that these platforms do not match very well the specific requirements of transactional company networks of SME companies. For example, with respect to functionality only in research prototypes special information support and decision support services for moderators can be found. The rationale for SMEs to act in coopetition within a network is mainly to gain in potential and thus being able to attract larger tasks while remaining small, flexible, and agile. There is also a geographical aspect involved as these networks

typically consist of companies within a limited geographical area. Theories of such clusters also mention that "Most cluster participants do not compete directly, but serve different industy segments" [15, p. 205]. Our objectives are to develop and evaluate new, especially SME-suitable services of collaboration platforms that are specialized to the moderators' needs. Our initiative is part of the research program that is carried out in the EU-funded international project eBusCo.net which stands for "Electronic Business in Company Networks".

In this paper we present early results of our work on a service that is intended to enable a network moderator to deal efficiently with a specific allocation problem that we regard as "network actor allocation problem". This problem occurs within the inquiry management process where for an inquiry received by the network a proper combination of actors needs to be dynamically orchestrated from the set of network members. In our solution we take an approach that goes beyond a simple matching of services requested with services offered by the network members. We have designed a solution that employs a heuristic multi-criteria optimization scheme. The moderator may choose optimization criteria from a list of predefined choices. For example, it may be defined that the set of actors should be generated such that highest preference is given to product quality. This will adapt the orchestration process to give highest preference to companies which are known for high quality products or production steps, respectively. Also, our scheme can deal with special collaboration oriented conditions such as the need to mutually exclude two companies from being part of the same set of actors. We regard our service as "SME-suitable" because it enables moderators to deal with the allocation problem pragmatically, efficiently, and rapidly. Through the use of our service, the network members are freed from complex coordination and decision processes that are otherwise necessary for collaborative inquiry and order management in company networks. Furthermore, our service will make allocation decisions less dependent on human factors and as preferences and outcomes are documented the service will lead to a higher degree of transparency of moderator actions. This is expected to result in a higher level of trust in the company network. The service also especially helps inexperienced and not-well trained moderators to manage inquiries efficiently.

We present a system architecture for our service and implementation details about our optimization scheme. A prototypical implementation of this architecture is on its way. We will use this prototype to evaluate our service by simulation experiments that will also involve real moderators.

The remainder of this article is organized as follows. Section two contains further background information about company networks and collaborative business processes in such networks. Also, more details about the eBusCo.net project can be found there. In Section three, we analyse the moderator task of allocating proper combinations of network members to incoming inquiries. A system architecture and implementation details for a first prototype are presented in Section four. Related work is discussed in Section five and concluding remarks are given in Section six.

2 COMPANY NETWORKS AND THE EBUSCONET PROJECT

Company Networks. A company network is an organizational form that consists of individual companies which have joined their forces together in order to strengthen their competitive position. Ideally, a company network acts like one big company on the market and thus, e.g., may acquire business opportunities (especially with big companies) for each member that none of the individual companies would be able to gain alone. To join forces may also mean that the members of the network perform joined product development, joined purchasing, maintain a shared inventory, or that knowledge is shared among the members. In order to be successful company networks need clearly defined organizational regulations. In particular, it is proposed that the network operation needs to be moderated and supervised by a neutral authority often referred to as "the network moderator" [8, 9, 13, 18]. The role of such a moderator within the inquiry management process is analyzed in Section 3.

Collaborative E-Business Processes in Company Networks. Often, company networks are developed in order to increase the business volume of the networking members. It is expected that this goal may be achieved through networking benefits such as a large market penetration, cross-selling effects, and better agility in pursuing business opportunities with other companies together. Often the notion of transactional networks is used to refer to this type of network where operative-level business transactions are performed collaboratively by several network members. Examples of such collaborative business transactions include collaborative processing of received inquiries such as Requests for Information (RFI) and Requests for Quotation (RFQ), and collaborative order fulfillment.

Production Networks. Existing transactional networks where collaborative operative-level business transactions are in the foreground often consist of SME companies of the manufacturing sector such as the Production Network Neumünster [16]. In the literature such networks are referred to as production networks. It appears as a general property of such production networks that the products, production and refinement services, respectively, of the companies may be flexibly combined together. This combination or, in other words, value chain will lead to combined semi finished products such as spindles for water pumps or finished products such as chainsaws.

The EU Funded eBusCo.net Project. eBusCo.net which stands for "E-Business in Company Networks" presents an international research project funded by the European Union within the Interreg III A program. The research partners are the University of Southern Denmark in Odense and the University of Applied Sciences in Kiel. In the project we work closely together with public business development agencies, existing production networks, and companies of two particular regions in Northern Germany and Southern

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Denmark. Apart from a thorough empirical study of the current networking status of companies in these two regions, the project aims on investigating the use of Internet-based Collaboration Platforms for company networks. In order to deliver project results that are of relevance for the practice the Production Network Neumünster, a production network that has been existing for more than six years now, is heavily involved in our activities and serves as "test environment" for new concepts that will be developed in our project.

3 THE MODERATOR TASK OF ALLOCATING NETWORK ACTORS TO INQUIRIES

We observed from existing examples of company networks that moderators play an important role in collaborative business processes. They often act as dispatcher for incoming inquiries and orders addressed to the network. It seems to be a best business practice to use a central customer interaction center to channel incoming inquiries to the moderator.

In the work presented in this article we look especially at incoming inquiries that impose a network actor allocation problem to the moderator. That is, the moderator is called to choose certain combination of members among and also on behalf of the network. The members that are chosen present the particular "staff" or in other words the network actors that will be allocated to the inquiry. Hence, in the remainder we use the notion of Network-Actor-Set1 (NAS) to refer to a subset of network members selected within this context. A NAS may be viewed as a coarse-grained work plan for a composite product without scheduling and temporal information.

It appears to be common practice of SME production networks that not only the dispatching of inquiries but also the handling of the network actor allocation problem is performed by the moderator. In other types of networks a different approach might be found. For example, it has been reported that tendering and special group decision processes, respectively, may be used to solve the network actor allocation problem [9]. Such approaches, however, usually impose a lot of extra administrative efforts and communication overhead to the network. This explains why these alternatives seem to be less adequate for SME production networks where the individual companies cannot spend these extra efforts.

Within the task to obtain a concrete NAS for a given inquiry that we regard as "target NAS" in the remainder, one needs to be aware of economically oriented conditions such as price limits and fulfillment deadlines. Also, special conditions related to collaboration issues need to be reflected as follows. The INCLUDE condition requires certain companies of the network to be part of the target NAS by default (this could typically include the company or companies that brought the inquiry to the network), the EXCLUDE condition requires certain companies of the network to be not part of the target NAS by default, the MUTUAL-EXCLUDE condition requires that two specific companies must not be part of the same target NAS together.

To summarize our practical observations about the moderators' typical actions to deal with dispatched inquiries and in turn the implied allocation problem, two steps can be given that will be detailed later:

1. The inquiry is analyzed and complemented by further data until all relevant information about the concrete network actor allocation problem are available. In the remainder, we refer to this information by the notion of Collaboration-Request-Profile (CRP).
2. Given a concrete CRP an extensive data analysis and assessment of the network members is performed with the intention to find a proper target NAS. By "proper target NAS" we mean that particular NAS among possibly many alternatives that fits best to given preferences and constraints. In the remainder, we refer to such a concrete target NAS by the notion of "most-promising NAS" denoted by NASmp.

Obtaining a CRP. In principle, a CRP specifies a combination of products and productions services, respectively, for which corresponding suppliers – that is a corresponding concrete NAS -have to be found in the network. We subsume such products and production services related to an inquiry under the notion of Request Element in the following. Any possible set of network members that will offer the specified combination of Request Elements is regarded as a "valid NAS" in our terminology.

For example, assume that in an inquiry from a manufacturer of water pumps it is asked for an offer for a certain number of hardened steel spindles with a particular length, width and weight of a special steel material quality. This inquiry might be decomposed into three Request Elements, one element requesting proper blank steel pieces, another element referring to the production step where the blank pieces are milled into spindles, and a further element that relates to the hardening process of the spindles. The successful decomposition of elements or design of modularity lies behind sourcing and the possibilities in new division of labor [2]. As the moderator cannot be expected to have expertise at all areas and all levels we foresee that the system in further development will apply some hierarchical structure in posting elements for decomposition and receiving the decomposed descriptions.

In principle, obtaining the CRP can be described as sequence of analytical task as follows:

1. Explore the characteristics and special conditions related to the inquiry.
2. Decompose the inquiry into corresponding Request Elements.
3. Explore the characteristics and special conditions of each Request Element.

As the practice shows, inquiries addressed to a company network usually contain not by themselves all the CRP data. Replacement for the missing data can be derived from special knowledge sources (e.g., construction plans, CAD drawings, bills of materials) using the moderator's own experience, and also by interacting with the requestor. However, the system will have to behave reasonable stable and consistent even

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 1. Sample optimization criteria for the allocation of Network Actor Sets

Optimization Criterion	Explanation
Economically-Oriented Criteria	
Distance	Preference is given to companies that are closest to a given location.
Price	Preference is given to companies that offer the lowest price for the product and service, respectively.
Experience	Preference is given to companies with largest amount of experience in supplying the specified Request Elements.
Product Quality	Preference is given to companies that are assessed as high-quality product suppliers.
Service Quality	Preference is given to companies that are assessed as high-quality service suppliers.
Resource Availability	Preference is given to companies with largest amounts of unused production resources.
Economic Power	Preference is given to companies with strongest economic power.
Collaboration-Oriented Criteria	
Collaboration Experience	Preference is given to companies with largest amounts of collaboration experience.
Collaboration Profit Shares	Preference is given to companies to which the network delivered the smallest total amounts of profit shares so far.

with some amount of missing data as a fully specified CRP is unlikely to be obtained.

Allocating a NASmp. The optimization criteria for orchestrating a concrete target NASmp are dependent on the specific properties of the individual request, the request initiator, and other factors. We divide these criteria into economically oriented criteria and collaboration oriented criteria. Table 1 contains some examples. The given explanations describe the influence of each criterion on the orchestration process.

It is the moderator's tasks to decide what are the right optimization criteria for a given CRP. The factors that drive this decision include: size of business involved in terms of revenue for the entire network, reputation of the request initiator, economic conditions of the network, the market and of the individual network members.

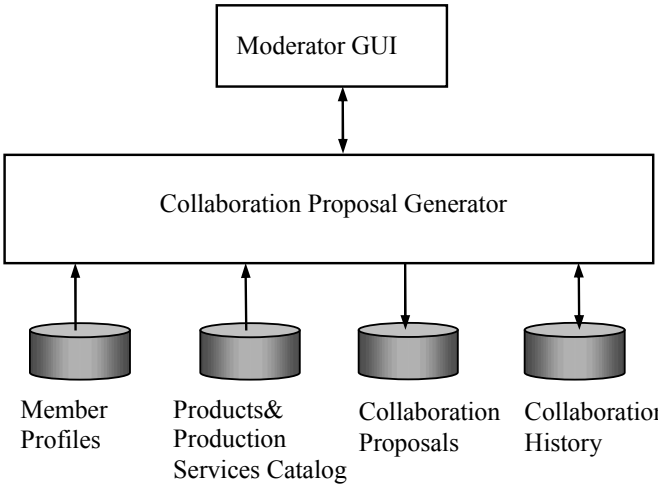
4 COLLABORATION PROPOSAL GENERATOR FOR SME COMPANY NETWORKS

Motivation. Company networks need to react rapidly and carefully to external inquiries. If such an inquiry imposes a network actor allocation problem as described in Section 3 adequate support from the underlying collaboration platform is required. Limitations in this support may lead into an error-prone, too expensive, and too slow orchestration process. As a result the company network and each individual company, too, may loose money. For example, consider the situation where an order is lost due to a bad collaboration of the companies that were assigned to the order. Also, the image of the network and the participating companies may be damaged through such deficiencies.

It appears that only a few available collaboration platforms consider the concept of a moderator at all. We furthermore observed that these concepts do not match very well with the specific requirements of SME networks described in Section 3. This observation has motivated our objective to develop and evaluate a novel, especially, SME-suitable service by which moderators may solve the network actor allocation problem efficiently and also conveniently. Furthermore, the use of our proposal service will lead to more consistent and well-founded decisions about who of the company network is assigned to what inquiries with what potential business volume being associated. Moreover, the tracking and documentation will lead to more transparency and auditability of the moderator's allocation decisions. This benefit is especially useful in circumstances where the placing of orders to companies of the network is being questioned by some members, for example, if some companies feel discriminated.

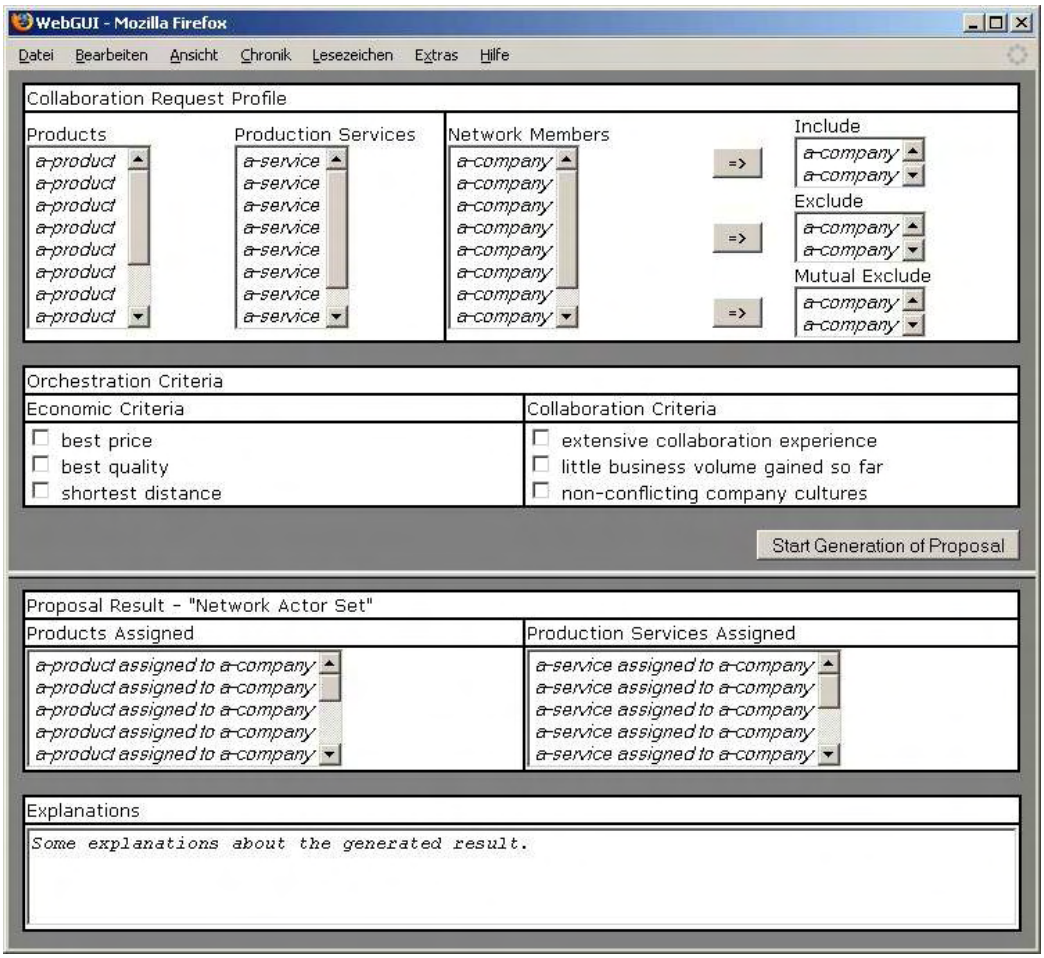
Architectural Overview. Figure 1 shows the "big picture" of our support service for network moderators that we are currently implementing as an isolated stand-alone system. For the long run, we strive to integrate this service into an existing collaboration platform.

Figure 1. Principle System Architecture



The so-called Collaboration Proposal Generator serves as central component where, among others, all those process-

Figure 2. Moderator GUI



ing steps are performed that are described later in this section.

The data considered by our service are organized in corresponding repositories. In the Member Profiles Repository general information about each company can be found such as number of employees, size of production facilities, areas of specialization, compliance to quality standards (and other attributes that are used by the optimization as exemplified in Table 1). The Products and Production Services Catalogue Repository contains detailed information about the products and production services together with corresponding characteristics offered by the companies. Furthermore, all the data that are processed during interactive sessions of a moderator are recorded in the Collaboration Proposals Repository. This includes especially the CRP information and the generated proposals that is NASmp. Data about performed collaborative processes and business transactions that occurred in the network are administered in the Collaboration History. Note that in our current implementation this repository is not automatically populated and maintained. However, the learning aspects of the service arise from the elaborate information on the generation of the decisions.

For the generation of solution proposals for the network actor allocation problem our system offers a specialized Graphical User Interface (GUI) to the moderator. Figure 2 gives an impression of the principle structure of this GUI. The upper part contains GUI elements to describe a network actor allocation problem in the form of a CRP. The Request Elements may be specified by selecting corresponding products and production services from the given product list and production services list. Special conditions such as INCLUDE, EXCLUDE, and MUTUAL EXCLUDE may be edited through usual GUI elements for condition editing know from other software packages.

The choices offered in the selection boxes such as the products, production services, and the companies of the network are dynamically queried from the corresponding repositories. In the middle part of the user interface the optimization criteria may be selected from a given set of check boxes.

A solution proposal (that is a NASmp) generated for an allocation problem is presented together with some explanations in the lower part of the main window. Such a solution is presented in the form of two lists that contain the proposed component products and production services with correspondingly assigned companies. The list box with title "Products Assigned" contains the component products with corresponding suppliers. The list box with title "Production Services Assigned" contains the proposed production services with associated supplier names.

Optimization Scheme for the Allocation of a NASmp. Our scheme follows a heuristic approach which is structured into four steps to compute solution proposals (i.e. a NASmp) for network allocation problems. The resulting NASmp may not include for each Request Element of the CRP a corresponding company of the network. It is up to the moderator to

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

deal with such gaps for example by considering further companies outside the network. The four steps are as follows:

1. For each Request Element find in the Products and Production Services Catalogue companies that qualify. Exclude from the set of found companies those companies that are defined in EXCLUDE conditions.
2. Perform a scoring for all alternative companies that were found in step 1 for a given Request Element. Reflect in this scoring the given optimization criteria by an evaluation of the corresponding data repositories. Manipulate the scoring result in order to comply accordingly with INCLUDE conditions.
3. Obtain a preliminary NASmp' by choosing for each Request Element either the only company found in step 1 or, if several alternatives have been found, the highest scoring company computed in step 2. Obtain the final NASmp by a further manipulation action to comply with MUTUAL EXCLUDE conditions. In this action replace companies that conflict with such conditions with next lower-scoring companies.
4. If the NASmp solution is incomplete and contains gaps for one or some Request Elements the new assigned companies will be treated as INCLUDE companies when a recalculation is performed.

5 RELATED WORK

Using the Internet for inter-organizational collaboration has been an active area of research for several years already. Many of the more recent initiatives such as [4, 10, 11] are targeted on the exploitation of semantic web technologies for more efficient and powerful web-based collaboration as it is possible today.

IT-Support of company networks has been addressed before in other research projects taking different phases of the lifecycle of such networks as focal point. The NETTO tool [5] supports ten different lifecycle phases starting from market analysis over strategic partner selection, operation, and further phases, up to a controlled end of life of the network. VICOPLAN [9] aims especially on the management relevant functions and, therefore, puts emphasis on support of network coordinators. In contrast to our approach, VICOPLAN is focused on tendering and group decision processes to deal with inquiries which, according to our experience, makes the platform less suitable for SME companies. A general discussion of success and failures of collaboration platforms can be found in [7].

Communalities to our research also can be found in the work on the creation of dynamic Virtual Organizations (VO). The alternative regarded as top-down planning approach for VO proposed in [3], from its core idea, is similar to our service. However, a fully automated proposal generation for VO structures, due to lesser time constraints, is not as important as for our work on transactional networks.

The idea to take a flexible optimization approach in order to obtain the best collaboration partners can also be found in [17]. While we have implemented this idea through a heuristic multi-criteria optimization, in this project mathematical equations are employed. Furthermore, this work aims on collaborative product development processes in company networks, whereas our focus is on operative-level collaborative business processes in transactional company networks.

6 CONCLUSIONS AND FUTURE WORK

Moderators of SME company networks in which operative-level business transactions are collaboratively processed impose special requirements to collaboration platforms. In our work we address the moderators' special needs for allocating a proper subset of actors from the entire set of network members that, in turn, are allocated to a given inquiry. We propose a dedicated service for collaboration platforms that will enable moderators to complete this allocation task based on an extensive analysis of various related data in a fast, reliable, and flexible fashion.

The underlying collaboration model for this service reflects the fact that often in SME company networks it is preferred that the network actors are directly determined by the moderator and not through a complicated group decision process or other approaches such as tendering. Among other advantages, it is assumed that through such a moderated SME-suitable model responsiveness of the company network may be improved.

Our proposed new platform service aims on providing a high degree of flexibility so that moderators may deal with the allocation task according to their preferences. We address this requirement through an adaptable heuristic multi-criteria optimization approach that employs scoring. The scoring criteria reflect the optimization criteria chosen by the moderator. Note that these optimization criteria may not only relate to specific findings for the received inquiry such as reputation of the requestor. The solution also considers optimization criteria that relate to the network as a whole such as a well balanced distribution of business volume across the members. The optimization criteria given in this article only present an initial proposal. Many more useful criteria can be found and integrated in our solution which will be part of our future work. Integrating additional optimization criteria may require to extent our system architecture by further data repositories. For example, an optimization that takes the availability of production resources within the network into account will require a further data repository. In this repository the utilization profiles of the companies' production resources and production scheduling information, respectively, need to be available.

Before we will extend the set of available optimization criteria, we will verify our service by simulation experiments and through further tests with real moderators of production networks.

In the context of technical revisions, we intend to evaluate if available and emerging standards based on XML may be exploited for our service. We will in particular evaluate standards for electronic catalogues (e.g. UN/CEFACT) and

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

intercompany workflows (e.g. XPDL) that may be used for our product and production services data repository, and standards for the specification of business transactions (e.g. ebXML, cXML) that may relate to our concept for CRP. We will especially look at the standardized modeling method UN/CEFACT UMM for the definition of business collaboration models and consider recent work in this area such as [10], too.

REFERENCES

1. Appelt, W.: WWW Based Collaboration with the BSCW System, Proc. SOFSEM'99: Theory and Practice of Informatics: 26th Conf. on Current Trends in Theory and Practice of Informatics, Milovy, Czech Republic, November/December 1999, Springer LNCS Vol. 1725/1999, pp. 66-78
2. Baldwin, C.Y., Clark, K.B.: Design Rules. The Power of Modularity, (Vol. 1), MIT Press, Cambridge, MA, 2000
3. Camarinha-Matos, L., Silveri, I., et al.: Towards a Framework for Creation of Dynamic Virtual Organizations, Proc. VE '05, Valencia, Spain, 26-28 Sept. 2005, Springer
4. Dell 'Erba, M., Fodor, O., Höpken, W., Werthner, H.: Exploiting Semantic Web Technologies for Harmonizing E-Markets, Information Technology & Toursim, Vol. 7, 2005, pp. 201-219
5. Fraunhofer IPT, Netto –Network Tool: Tool for the Configuration and Operation of Enterprise Networks, Website:

www.ipt.fraunhofer.de/fhg/ipt/EN/businessactivities/MetrologyandQualityManagement/Qualitymanagement/CurrentResearchProjects/Nettonetwor ktool.jsp, visited 13th Feb. 07

6. Fuchs, L., Poltrock, S., Wetzel, I.: Teamspace – An Environment for Team Articulation Work and Virtual Meetings, Proc. WBC '01, Web Based Collaboration (DEXA 2001), Munich, Germany, 2001, pp. 527
7. Gogolin, M.: Success and Failure of Collaboration Platforms, in: Lechner, Ulrike (Hrsg.), Proc. Tenth Research Symposium on Emerging Electronic Markets 2003, S. 169-183, Bremen, Germany, 2003.
8. Harbilas, C., Dragios, N., Kartesos, G.: A Framework for Broker Assisted Virtual Enterprises, Proc. Collaborative Business Ecosystem and Virtual Enterprises: IFIP TC5/WG5.5 2002, Kluwer Academic Publishers, pp. 73-80
9. Hess, T.: Planning and Control of Virtual Corporations in the Service Industry – The Prototype VICOPLAN, Proc. 35th Annual Hawaii International Conference on System Sciences, 2002, pp. 24-33
10. Hofreiter, B., Huemer, C., Winiwarter, W.: OCL-Constraints for UMM Business Collaborations, Proc. of the 5th Int. Conf. on Electronic Commerce and Web Technologies (EC-Web 2004), Zaragoza, Spain, September 2004
11. Kramler, G., Kapsammmer, E., Retschitzegger, W., Kappel, G.: Towards Using UML 2 for Modelling Web Service Collaboration Protocols, Proc. Int, Conf. on Interoperability of Enterprise Software and Applications (INTEROP-ESA'05), Feb. 2005, Geneva, Switzerland
12. Latour, B.: Reassembling the Social. An Introduction to Actor-Network-Theory. Oxford University Press, 2005
13. Pereira-Klen, A., Klen, E.: Human Supervised Virtual Organization Management, Proc. Collaborative Networks and their breeding environments: IFIP TC5 WG 5.5, Working Conference on Virtual Enterprises, 2005 Valencia, Springer, pp. 229-238
14. Pereira-Klen, A., Rabelo, R., Ferreira, A., Spinosa, L.: Managing Distributed Business Processes in Virtual Enterprises, in Journal of Intelligent Manufacturing, 2001, Vol. 12, No. 2, pp. 185-197
15. Porter, M.: Clusters and Competition. New Agendas for Companies, Governments, and Institutions in " On Competition", Harward Business Review Book, 1998, p. 197-287
16. Production Network Neumünster, Web Site: www.pnw-neumuenster.de, visited 13th Feb. 07
17. Schweinberger, D.: Eine Methodik zur Unterstützung der Suche und Auswahl von Partnern für kooperative Produktinnovationsprojekte, PhD Thesis University of Karlsruhe, 2002, ISSN-1615-8113
18. Tolle, M., Bernus, P., Vesterager, J.: Reference Models for Virtual Enterprises, Proc. PRO-VE 2002, Portugal, Kluwer Academic Publishers, pp. 3

There is (yet) no relation between the Network-Actor-Set and the Actor-Network-Theory (ANT) that is a social theory with semiotics and materialist views mostly focused on the creation of knowledge as described by Bruno Latour [12].

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

i·Society 2007

An analysis framework for the economic potential of process interoperation

Thomas Keller

Reinhard Riedl

Abstract Traditionally, business integration has a medium to long term horizon. However, with the introduction of the concepts of the Virtual Organisation and the Extended Enterprise this horizon is coming closer with the implication that business integration does not pay off anymore. In order to overcome this new challenge semantics are applied to existing integration technologies leading to a new flexibility and hence to new opportunities for the trade-off of higher complexity. This new flexibility based on semantics and realized by ontologies has a much greater impact on the enterprise and its socio-economic context than with existing approaches. Hence, it is becoming more demanding to manage the decision process and to estimate the economic value added. New means to evaluate the economic business potential are required. This paper introduces the interoperation evaluation framework (IEF) as an analysis tool for this purpose.

Keywords Business Process Management, Automated Process Interoperation, Semantics

1 INTRODUCTION

In an effort to gain both economic benefits and competitive advantages, companies and public institutions are increasingly involved in ever more complex networks of cooperation. The level of cooperation varies from simple exchange of information to complex interactions of processes. Process integration across functional borders and across enterprises is increasingly supported by business information technology. Alongside this development, proprietary processes are being replaced by standardized processes or are standardized themselves, leading to a situation where processes become commodities [1]. This opens new possibilities with regard to simplified process composition and orchestration. Hence, in the future, each link in the value chain will potentially be a standardized and interchangeable process.

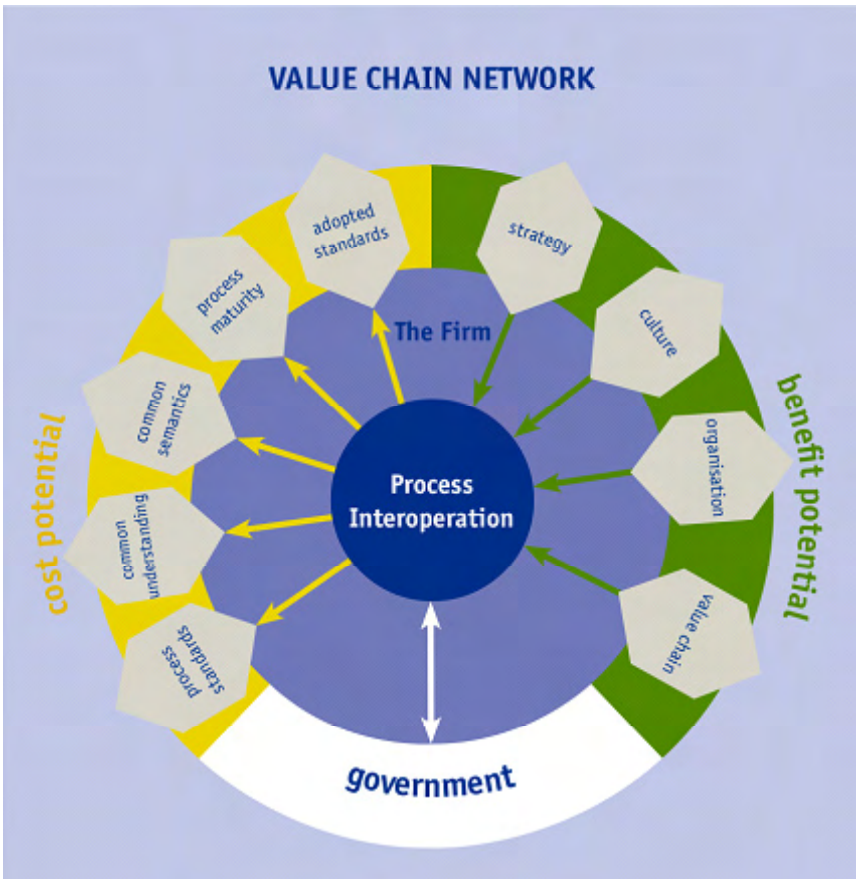
Traditionally, networks of enterprises had a mid to long term duration. Today, short term networks are becoming more and more important as well; they can be adapted to the needs of the individual project, they are more flexible and efficient. This is especially true for virtual enterprises since cooperation may be setup on a project by project basis, striving after the optimal allocation of resources. This new focus on short term cooperation has a significant impact on the business information technology infrastructure [2]. The investment in process automation merely for short term cooperation does not pay off very well. Even for value chain networks where application heterogeneity is low, a considerable amount of resources must be invested in automating processes. With a typical pay-off period of longer than three years, short term cooperation is a real challenge.

In order to overcome this dilemma a recent approach in business information technology is to extend current technologies by semantics. If the cooperation-partners' application interfaces are made self-describing enough to enable automated process composition, the cost of automation will decrease to a level where it pays off even for short term networking. Probably the best-known area used for this approach is semantic web services [9] and automated process interoperation based on a service-oriented architecture [3]. The ultimate goal is process composition based on semantic information which can theoretically be fully automated. Research activities are intense in this field and are mostly concentrating on a purely technical view. However, to reach this goal the complex business aspects of the entrepreneurial (strategic-structural) level need to be considered for the support of short term collaboration networks. A common understanding in an Industry is mostly reflected by the quantity of relevant work on reference processes (or meta processes) describing solutions. Examples can be found in, e.g., the supply chain or in the insurance industry. This work is typically tedious and needs a lot of time and resources [29]. Most important, the responsibility is on business and not on IT.

2 PROCESS INTEROPERATION AND THE ENTERPRISE

Cooperation has a lot to do with compatibility. The compatibility aspect is even further stressed with the introduction of process composition over functional borders and across enterprises, linking together whole value chain networks. It forces enterprises to open and align their processes and define interfaces. With the definition of interfaces, common

Figure 1. Overview of the Interoperation Evaluation Framework (IEF)



semantics are needed to increase the chances of an understanding on a machine interpretable level. A bad example is WSDL, where in general no semantics are present that make it machine interpretable (except for syntactical purposes).

Furthermore heterogeneity in terms of application systems and infrastructure is high in typical industrial value chain networks, which put pressure on harmonization and use of a common set of standards. Standards are meant to provide certain protection for financial investments. In this field, however, standards have lost their status. Many new standards have been released by various organizations, reflecting the economic calculus of the firms behind the respective standardization body [12]. On an entrepreneurial level a common understanding of the problem must be established prior to any technical solution.

The alignment of processes within a network of enterprises along value chains (in addition to the automation of this alignment) has various influences on the enterprise itself, but also on the network as whole. The first factor to consider is definitely the strategy that the enterprise has defined or is willing to follow. But the structure and culture of the enterprise will also have an influence. Other factors that may have an impact are for instance legal issues and how the costs and benefits are shared among the partners [29]. The structure of the value chain network must also be considered. Two extreme cases can be distinguished. The first case consists of a monopoly or oligopoly which features one or a few strong market players who control their respective network. The other case is a truly open market, a polypoly, with equal market players. Whatever the market structure is, it will have an influence on a possible business case.

The bottom line of the above statements is that it is not obvious what strategic-structural and technical circumstances may be necessary to justify automated process composition. A business case must promise a real advantage over a traditional business integration approach where business systems are integrated manually. The following questions need to be addressed:

- What possible economic value added can be achieved within the given entrepreneurial constraints?
- What are the determinants of a successful implementation?
- What criteria have to be monitored?

If the entrepreneurial constraints are not given and open for change, the first question from above can be reversed and a second question added:

- Is my enterprise fit for automated process composition?
- What measures have to be taken to enable added economic value?

Only a well structured and comprehensible business case will attract investment in such a high risk development. Without serious investment automated process composition cannot readily be boosted to an industrial scale.

A conceptual framework that attempts to answer the above questions is presented below.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

3 THE INTEROPERATION EVALUATION FRAMEWORK

The "Interoperation Evaluation Framework (IEF)" has its roots in well established management frameworks [13], [14], [15], [16] which consider strategy, culture, structure and their relationships as views of an enterprise. These views are complemented by additional process and information oriented views as shown in Figure 1.

Each view is composed of a set of aspects or criteria by which we can analyze the particular context where the framework is applied. The aspects have been theoretically derived from existing research and logically deducted based on the nature of automated interoperation. Due to the complexity and the multi-layered nature of automated interoperation, it is not easy to understand the topic in its entirety. This is why it is important to create a conceptual frame by means of which basic interconnections as well as some principles can be shown systematically [30]. Thus, a framework clarifies the research object and can serve as a basis for scientific objectives and research perspectives. Existing theories, or at least hypotheses derived from such theories, are used as a foundation. This fact must be reflected in the design elements and indicators inherent in the framework.

Due to the fact that the context of this framework is subject to intensive research, the framework is also going to be used to position and classify new findings, trends and methods. The framework is therefore also a structuring and classification instrument. In the context of corporate management, the framework is going to provide a coordinating and integrating function, which makes it easier to communicate complex issues and assess the implementation of leadership decisions. It acts as a conceptual screen with which to evaluate both existing and newly developed problem-solving methods.

In the modern philosophy of science, frameworks constitute the basis for both empirical-inductive as well as analytical-deductive research [13]. In the inductive method, a great number of observations can be deduced from generalizable findings. They are then integrated into a single conceptual frame. The quality of the framework that is thus generated very much depends on the objectivity of the observations. These can, in turn, be affected by the value-generating network(s), the industry, the current economic situation or the spirit of the age. The deductive method starts with existing, superordinate phenomena and circumstances which have been accepted by virtue of their logic and plausibility. The quality of a framework that is deduced in this manner is directly dependant on how closely the superordinate circumstances and the logic of the deduction approach reality. The evaluation framework described below is based on the latter, the deductive method.

The innovative feature of the interoperation evaluation framework is the combination of different views, strategic-structural and enabling/technical, and the distinction between the firm and the firm's value chain network. It is technology independent.

3.1 Overview

The Interoperation Evaluation Framework as shown in Figure 1 consists of the part on the left (the enabling perspective), representing the cost potential of the automated interoperation by means of five design elements, and a part on the right (the strategic-structural perspective), describing the benefit potential by means of four design elements. The illustration also takes into consideration the general requirements of the regulations stipulated by the state.

The design elements always concern both the value chain network as well as the enterprise itself, and their indicators must therefore be evaluated both for the enterprise and for the respective context, accordingly. The benefit potential is generated from the strategic-structural perspective, the cost potential from the enabling perspective. Both potentials finally determine the applicability of automated interoperation in the enterprise and thus also within the value chain network, or the industry, respectively.

The distinctions of cost and benefit potentials and the attribution of the enabling design elements to costs and the strategic-structural design elements to benefits originates in the opinion that the benefit potential is primarily determined by the strategic-structural design elements of the enterprise, and is thus prescinded from the enabling design elements. The enabling design elements provide information about the cost side of an introduction of automated interoperation by pointing out the necessary changes in the technical environment. They do not make any contribution to increasing the benefit potential because as enablers they constitute a condition. The strategic-structural design elements determine the business administration context in which automated interoperation is to be deployed, and thus also the possible economic added value. The Evaluation Framework makes it possible to evaluate the four strategic-structural design elements and is thus to be regarded as a general situational analysis that shows the current benefit potential. The Framework can however also be used to define the changes to the strategic-structural design elements that would be necessary to increase the benefit potential.

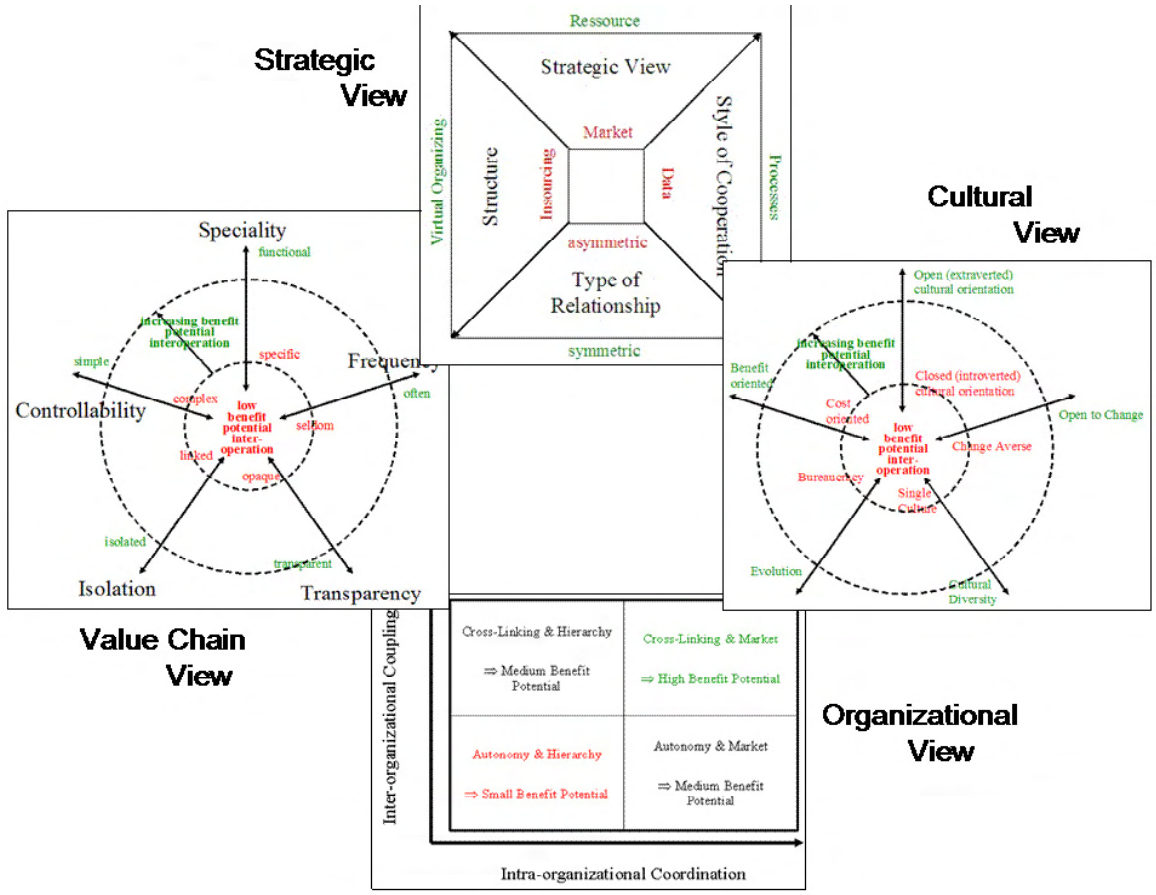
3.2 The Benefit Potential: The Strategic-Structural Views

Figure 2 provides you with a summary of the four design elements: strategy, culture, organization and value chain, together with their indicators. They constitute a strategic-structural view and provide an indication of the benefit potential that an organization might expect if it applies automated interoperation. It is out of scope of this paper to discuss all the indicators in detail.

Strategy is measured with four indicators. The strategy view distinguishes between a market, a relational and a resource based view whereas the resource based view is considered the best suited strategy approach regarding automated interoperation. The reason for this lies in the fact that in the case of the concentration on core competences the other business functions necessary for an extended value chain have to be

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Overview over the strategic-structural views of the interoperation evaluation framework



acquired from the market or from within a network. In the case of the relational view, the focus is on the stakeholder relationships which can be depicted on the process level by means of interoperation. Another strategic indicator is cooperation style. We distinguish between cooperation based on a pure exchange of data and cooperation based on harmonized processes. In the former case, interoperation generates less of a benefit than in the latter case. The reason for this is that a mere data exchange, such as in the case of an enquiry and an offer, the possibilities, or rather the full potential of automated interoperation cannot really be exploited. Cooperation based on the process level works differently, since it is here where automated interoperation, as one of several possible technical alternatives, has a higher benefit potential. Yet another indicator consists of the type of relationship. The two extreme values are a symmetrical relationship in which all the involved enterprises apply interoperation, and an asymmetrical relationship which is the contrary. The former case is considered in favor of automated interoperation because a symmetrical relationship implies a win-win-situation among the participating enterprises.

As opposed to strategy, the influence of culture can easily be defined by means of a classification of types of culture. The relevant literature offers a number of different classification systems which differ in the degree of dimension, i.e. the number of descriptive characteristics. A summary of one-, two- and multi-dimensional approaches can be found in [31]. The chosen classification has been derived from the multi-dimensional approach of [32] and [33] and is shown in Figure 2 on the right side. The cultural indicators are difficult to evaluate and have a strong qualitative character.

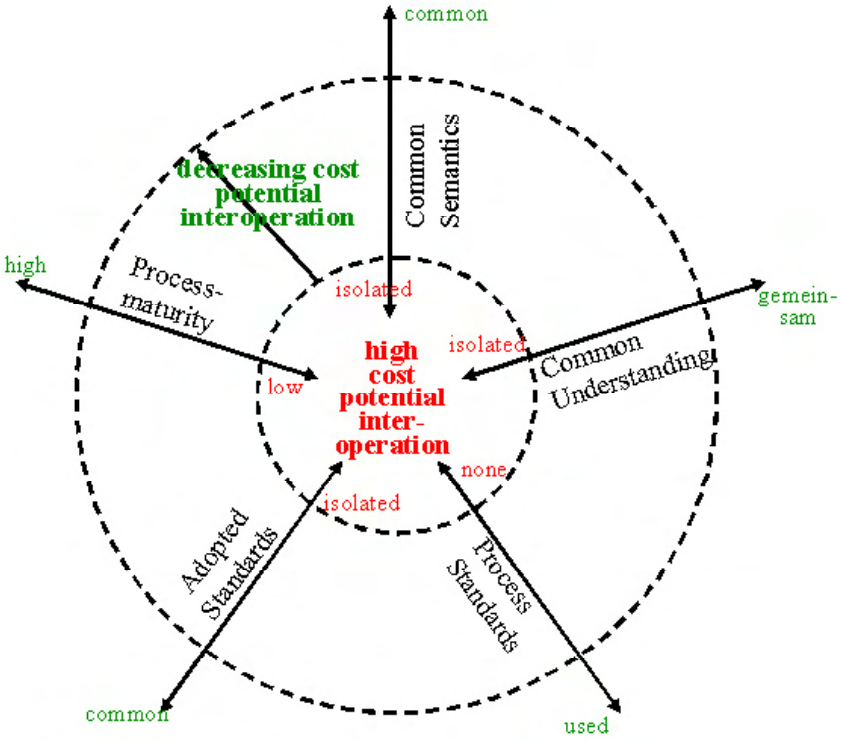
The influence of the organization is illustrated by interorganizational coupling and intra organizational coordination. The interorganizational coupling can be described in an obvious way by means of two bipolar values, autonomy and interconnection. In case of the former, an organization is characterized by few or no coupling relationships to other organizations. If there are relationships to other organizations, these are characterized by low intensity, i.e. loose coupling. Another trait of such features is the fact that the relationships tend to have formed as a result of another organization's initiative [34]. The interorganizational coordination is meant to provide an indication of the manner in which the organization operates. Here, the two bipolar features are the hierarchy and the market. In the first instance, central regulating mechanisms play an important part as opposed to the second case where decentralized regulating mechanisms are applied.

The applicability of automated interoperation is influenced by the type of products and services and their engineering and manufacturing processes, on the one hand, and by the position the enterprise enjoys within the value chain network as a whole. Following five indicators (as shown in Figure 2 on the left side) have been identified:

- Frequency as the number of process invocations per time unit.
- Specificity or degree of innovation signifies the type of product, i.e. whether a product is functional or innovative.
- Transparency constitutes the knowledge which function is carried out using what means, under what pre- and post-conditions and generating what side effects.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. The five Enabling Views



- Isolability means the reusability of individual process elements in other contexts.
- Controllability specifies the degree of influence that has to be exerted on the partial process entity (e.g., user interaction).

3.3 The Cost Potential: The Enabling Views

As opposed to the strategic-structural views introduced above, the five enabling elements as shown in Figure 3 are seen as a prerequisite for the applicability of automated interoperation. Following a short introduction:

Process standards make processes comparable. They define measures regarding process quality and process management. An example of a process quality management standard is the capability maturity model. Reference processes like SCOR [19] define key performance indexes which can also be used for comparison purposes. In the context of process interoperation the comparison of processes is a basic feature, without which it is not possible to choose from alternatives. Therefore, it must be mandatory to have process quality standards and process management standards applied to processes. Various applicable standards exist for that purpose [26], [27], and [28]. If process standards are already well established, the cost potential will generally decrease.

Common problems understanding is essential for process interoperation because it implies a common approach and common solutions. Its manifestation can be found in reference processes (e.g., SCOR [19], VCOR [20]). Most important is that these reference processes consist of the value chain as a whole and not just a small part of it. With regard to process interoperation, reference processes can serve as a source for semantic information and can even be realized as ontology. The more reference processes are available and the more details they model, the lower is the cost potential.

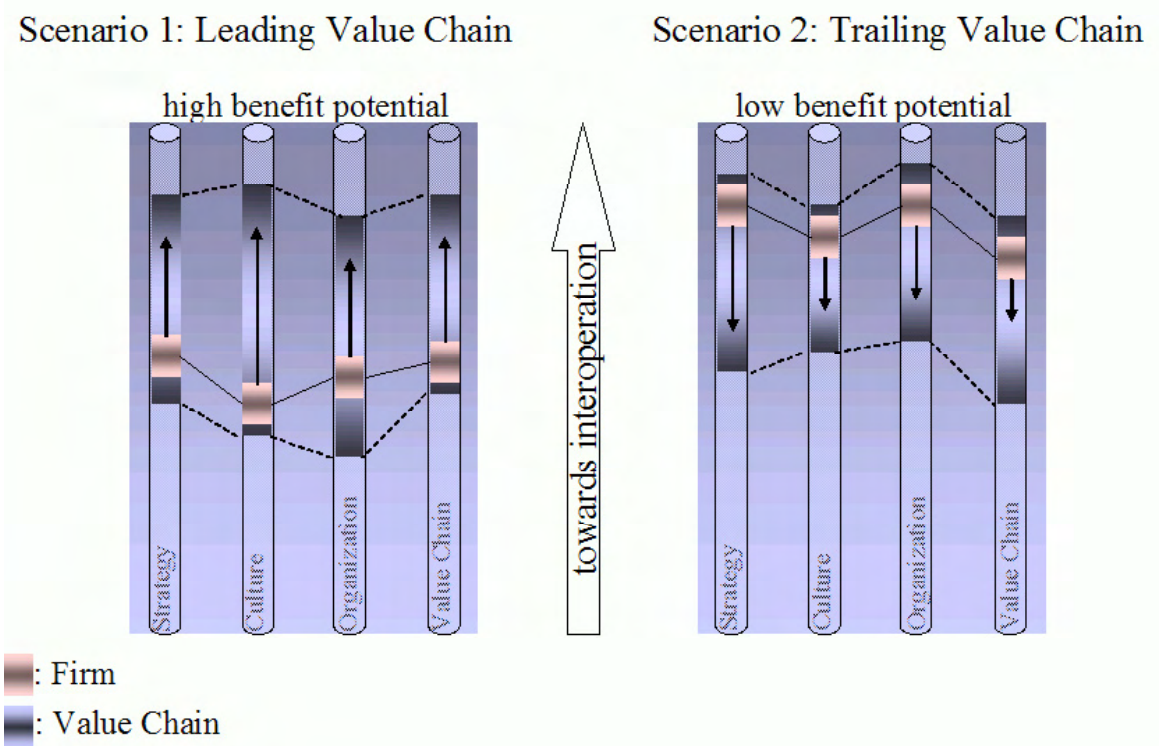
Common semantics refer in general to a shared ontology. It is rather a difficult and complex task to establish a new ontology even in a well defined application domain. Therefore, it is wise to reuse any semantic information that is already in use within the value chain network. Often, enterprises are not even aware that they are using such information. Classification schemes and taxonomies can be useful here as well. An example is the United Nations Standard for Products and Services Code (UNSPSC) [21]. There are dozens of industry specific classifications, taxonomies and even ontologies available [22], [23], [24], [25]. It is considered best practice to integrate them whenever possible, decreasing the cost potential. A complete lack of semantics will generally rule out a project or at the very least lead to a substantial increase in its duration and risk.

Process maturity must allow for executable formalized processes. However, reality shows that many processes are still tacit, some are just on paper and with some luck some are automated. However, merely automated processes are not enough for process composition. Process composition relies on modular processes with the possibility of exchangeability. Any process that should be part of process composition must be interoperable, implying a modular structure and a semantically rich description.

A good technical solution uses well established and adopted standards. This is definitely true from a business perspective because the use of standards implies some protection for the investment. From a technical perspective, however, it does not hold. In any case, standards enable and simplify compatible solutions. They are absolutely necessary for process

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. Alignment Scenarios



composition across enterprises and along the value chain. Missing standards will generally increase the cost potential.

3.4 The Significance of the Value Chain Network

The nature of interoperation implies that the influence of the industry, or rather of the value chain network of the enterprise to be examined, is of particular significance in the framework, since it is one of the prerequisites of interoperation that there are partners that want to support it and are capable of doing so. This is why the design elements are applied both on the added value framework as well as on the enterprise in question and why the position of the enterprise in the context of its added value framework is being determined.

The alignment of the organization with the industry, or rather with the respective value chain network, is a further essential component for the assessment of the benefit potential. It mainly deals with a relative perspective of the organization relative to its network. Basically, two scenarios can be distinguished: The first scenario is based on a progressive value chain network and examines an organization limping somewhat behind its network median. The second scenario takes a look at the reverse case, by studying an organization that is ahead of the network median. Both scenarios are shown in Figure 4.

In the case of a value chain network with a high benefit potential, measured by the indicators of the four design elements, i.e. strategy, culture, organization and added value chain, and an organization that itself has a low benefit potential, the resulting benefit potential for the organization is high because it only has to copy its industry's "best practices". The relative advantage of the partner organization vis-à-vis the organization to be evaluated has a positive effect on its benefit potential. It must be critically noted that copying "best practices" might not be an easy task.

In the case of an industry with a lower benefit potential than the organization's benefit potential, the organization has a problem: The whole industry would first have to be heaved up to a sufficiently high benefit potential. In order to do so, the organization would have to be in a position to sufficiently affect the indicators of the industry's four elements, strategy, culture, organization and added value chain. However, this will only be possible in some rare cases which is why this second scenario represents a negative effect on the benefit potential of the organization to be evaluated.

4 BUSINESS CASE EVALUATIONS

The above presented framework has been applied to two different industries. In the healthcare industry several hospitals in Switzerland have been investigated and characterized. In the medical industry a conglomerate of enterprises in the South of Germany has been investigated. The research question was whether automated process composition would be applicable and under what conditions it would lead to further improvements in the cross-linking of the various processes. Following some findings based on the latter case:

- The investigated conglomerate features a shared quality management. This has led to shared and common processes which are partly automated. The application landscape in the different enterprises is rather homogeneous due to the central management. The products and services available and offered by the various enterprises are catalogued and classified by a shared classification scheme according to UNSPSC. The govern-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

mental influence is limited. In short, decision-making is relatively straightforward.

- Strategic analysis shows a medium to high benefit potential because within the conglomerate a resource based view is applied combined with outsourcing of non core competencies. In certain cases the level of cooperation approaches virtual organization. Within the conglomerate the type of relationship is symmetric, with external partners rather asymmetric, showing the market power of the conglomerate.
- The culture and the organization are both suitable for automated process composition. This is definitely the result of the internal structure of the conglomerate. The value chain consists mostly of processes that are frequently used. They are transparent, modular, partly isolated and range from specific to functional; all features which lend themselves to the application of automated process composition. The overall benefit potential is medium to high.
- Meanwhile, the cost potential is low to medium. Due to the centralized quality management, common semantics and common problem-understandings exist already. In addition, the governmental influence forces a common problem understanding by enforcing certain regulations. Process standards are defined and applied due to the quality management. Process maturity has reached the level of automation. In certain cases processes are modular and ready for interoperation. The applied standards vary depending on the enterprise. There are some standards that are used on conglomerate level which are forced upon all enterprises. Only this last point increases the cost potential; all the other technical aspects suggest low cost potential.
- Both the benefit potential as well as the cost potential are in favor of automated process composition. The conglomerate is a good business case for automated process composition.

As for the healthcare industry, the framework has already proved to be a valuable tool in analysing the situation. The results can be used for argumentation and communication with the various stakeholders. Attracting and convincing business sponsors is getting easier.

5 SUMMARY

The evaluation framework proposed here represents a tool with which the economic added value of automated interoperation can be qualitatively determined. It lists the different influences and groups them by separating them into strategic-structural design elements, on the one hand, and enabling design elements, on the other. The strategic-structural influences are divided into the design elements strategy, culture, organization and added value chain; together, they form the benefit potential and represent the firm's current situation regarding the benefit of automated process composition. The enabling design elements form the cost potential. They constitute a condition.

The Evaluation Framework and its strategic-structural design elements, strategy, culture, organization and added value, is geared to established approaches, such as the management approach of the University of St. Gallen. It defines strategy, structures and culture as a structuring means and distinguishes between management, business and support processes. The interactions of the organization in its added value network are divided into several fields: resources, standards & values and concerns & interests. In the Evaluation Framework, these fields again make an appearance as the enabling design elements. What is new compared with these established approaches, are the indicators which are specifically tailored to the context of automated interoperation. Here again, existing findings have in some cases been made use of. This includes findings from organization and administration research, from empirical studies on factors of influence concerning the success or failure of organizations and from research data concerning collaborative corporate networks. These indicators have been theoretically deduced and tested for plausibility, and are in the process to be confirmed by means of an on-going empirical investigation.

The enabling side uses such tried and tested approaches to a much lesser degree. Yet in the context of automated interoperation, certain design elements appear plausible. The starting point of these reflections is always the semantics and the standards which are necessary for the introduction of automated systems. Standard, or reference processes take centre stage, since they to a large extent reflect today's orientation where organizations are concerned.

The Evaluation Framework has an interdisciplinary structure, being divided into a strategic-structural and an enabling part. This is why its operation and application by just one person is estimated to be rather improbable in practice. It will in all probability require a team of professionals to handle the evaluation in practice. What makes this more difficult is in particular the fact that it is not only one's own organization but the whole added value network that must be evaluated.

REFERENCES

1. Davenport: The Coming Commoditization of Processes in Harvard Business Review, June 2005
2. Ader, W: Technologies for the Virtual Enterprise, Work-flow & Groupware Strategies, France
3. Milanovic: Current Solutions for Web Service Composition, in: IEEE Internet Computing, November/December 2004, IEEE Computer Society
4. Business Process Execution Language for Webservices, BEA, IBM, Microsoft, SAP, Siebel, 2003
5. Mandell, McIlraith. Adapting BPEL4WS for the Semantic Web: The Bottom-Up Approach to Web Service Interoperability, International Semantic Web Conference 2003
6. Lassila, Swick, Resource Description Framework (RDF) Model and Syntax Specification, W3C recommendation, 1999
7. Fikes, McGuinness, An axiomatic Semantics for RDF, RDF-S and DAML+OIL, Manuscript, 2001
8. http://www.daml.org/services/owl-s/1.0/, last access 2007/09/01

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

9. Sivashanmugam, Adding Semantics to Web Services Standards, Proceedings of the International Conference on Web Services, pages 395-401, 2003
10. Semantic Web Services Framework (SWSF), W3C, 2005
11. Web Service Modeling Ontology (WSMO), W3C, 2005
12. Lahiri: Web Service Standards: Do we need them?, in: Workshop on Emerging Web Services Technology (WEWST06), Zurich, 2006
13. Rühli: Unternehmungsführung und Unternehmungspolitik Band 1, Haupt 1996
14. Rüegg-Stürm: Das neue St. Galler Management Modell, Haupt, 2003
15. Bach: Zukunftsfähige Organisation - Stand und Entwick-lungstrends der Organisation deutscher Unternehmungen und Verwaltungen, in: soFid Organiations- und Verwal-tungsforschung 2002/2
16. Joyce/Nohria/Roberson: What really works – the 4+2 Formula for Sustained Business Success
17. Camarinho/Afsarmanesh: Processes and Foundations for Virtual Organizations, Springer, 2004
18. Camarinho/Afsarmanesh: Collaborative Networks and Their Breeding Environments, Springer, 2005
19. Supply Chain Operations Reference Model (SCOR), www.supply-chain.org, last access 2007/09/01
20. Value Chain Operations Reference Model (VCOR), www.value-chain.org, last acess 2007/09/01
21. United Nations Standard Products and Services Code, www.unspsc.org
22. Health Level 7, www.hl7.org, last access 2007-09-01
23. Open Application Group, www.oagi.com, last access 2007-09-01
24. United Nations Economic Commission for Europe, www.unece.org/cefact, last access 2007-09-01
25. SWIFT, http://www.swift.com, last access 2007-09-01
26. Web Services Agreement Specification (WS-Agreement), Global Grid Forum, 2005
27. Web Service Level Agreement (WSLA) Language Specification, V1.0, IBM, 2003
28. Web Services Policy Framework (WS-Policy), V1.2, 2006
29. Hepp: Possible Ontologies: How Reality Constrains the Development of Relevant Ontologies, in: IEEE Internet Computing, vol. 11, no. 1, 2007, pp 90-96
30. Anthony: The Management Control Function, Boston (Mass.), 1988
31. Wagner et al: Typologie von Lernkulturen in Unternehmen, QUEM-report, Schriften zur beruflichen Weiterbildung, Heft 73, Berlin, 2001
32. Bleicher: Das Konzept integriertes Management: Visionen – Missionen – Programme, Frankfurt/Main 1999
33. Bleicher: Organisation: Strategien – Strukturen – Kulturen, Wiesbaden 1991
34. Bach: Zukunftsfähige Organisation - Stand und Entwicklungstrends der Organisation deutscher Unternehmungen und Verwaltungen, in: soFid Organiations- und Verwaltungsforschung 2002/2

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e•Centre for Infonomics

i·Society 2007

A model of evolution and ontological development for trust transferring in e-business

Omer Mahmood, John D Haynes

School of Information Technology
Charles Darwin University
Darwin, NT, 0909, Australia
Omer.Mahmood@cdu.edu.au
John.Haynes@cdu.edu.au

Abstract Trust plays a pivotal in role in the adoption and success of a business particularly in the electronic environment. It has a much stronger impact on consumer behavior in an e-environment in relation to the physical world, primarily due to the gap in the exchange of funds and the delivery of goods and services. With the growing share of electronic commerce in the global economy, distance trust building has become imperative; hence better models to represent and transfer online trust are required for wider adoption of e-commerce. To enhance online trust, businesses employ a wide range of trust transferring techniques that are categorized as individual and collective trust transferring techniques. The commonly used individual trust transfer techniques include, web interface design, use of privacy and disclosure policies and familiarity of merchant sold brands. The individual techniques are subjective and personal to both online merchants and potential consumers. Moreover the users' cultural background, physical and social context and the users' personal experiences largely influence these techniques. This paper introduces the idea of an evolution of trust as an ontological development similar to the evolvement of a culture and particularly through customer loyalty development and concentrates on 'Collective Trust Transferring Techniques' as they rely on the combined effort of several users and service providers. Collective trust transferring techniques broadly rely on exchange and presentation of selected information in the electronic environment and impact a wide potential customer base. This paper firstly identifies and evaluates the online trust transferring techniques that are primarily based on the exchange and the presentation of information in the electronic environment. Followed by this, it proposes a conceptual model based on the evolvement of a culture that relies on the, in this case, mutual, trust of the participants. The proposed model aims to assist online businesses to effectively develop, build, employ and evolve online trust transferring capacity for sustained business growth.

Keywords Online Trust, Electronic Business, Electronic Information and Trust, Ontology, Evolution, Culture.

1 INTRODUCTION

Online trust must be allowed to evolve (in the special sense of "evolve" used throughout the paper) and this can be a goal which is developed from the very beginning of the E Business cycle. In this paper four trust transferring techniques are considered, namely, trusted referrals, online reputation, unified merchants and third party trust seals. In terms of the evolution of trust, online reputation is the pivotal technique for which customer loyalty through the emotional and transactional interaction between the firm and client are crucial, which is very much like the ontological evolution of a culture, insofar as mutual trust is the key element in the evolvement of a culture. This paper then considers recommended trust transferring phases (akin to cultural phase shifts in cultural trust).

2 ONLINE TRUST TRANSFERRING TECHNIQUES

Recent trends in electronic commerce are the use of trusted referrals, the display of aggregate online reputation, the operation under known unified merchants and the use of trusted third parties seals, as collective trust transferring techniques. Out of four identified techniques, trusted referrals and online reputation entirely depend on the existence, development, evolution, feedback and presentation of online information. The use of trusted third party seals and executing business operations under the trust brand only deal with presentation of information to enhance online trust. The recognized techniques are discussed below.

2.1 Trusted Referrals

Information regarding a product, physical or online business acquired from either the user's physical, or online trusted social network impacts the user's initial and subsequent levels

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007

of trust in an online business. The impact is directly associated with the user's level of trust on the source of information in terms of its credibility, honesty and ability. Trusted referrals [ ] "are the primary means of disseminating market information when the services are particularly complex and difficult to evaluate. This implies that if one gets positive word-of-mouth referrals on e-commerce from a person with strong personal ties, the consumer may establish higher levels of initial trust in e-commerce", (pp. 538). Fullam et al. [ ] defined that the user's belief in relation to information accuracy and certainty conveyed in the information, form the level of trust in the information source. Thus one can conclude that "a referral from the user's personal trusted source actually transfers trust from the referring source to the referred entity".

2.2 Online Reputation

In the absence of trusted referral or past experience, the online reputation can be one of the crucial factors for the user to establish relations with online service providers. Zacharia [ ] states "reputation is usually defined as the amount of trust inspired by a particular person in a specific setting or domain of interest" (pp. 163). Online reputation regarding an e-business is built by collating the past experiences of the users who have previously interacted with the same service provider. This technique in the form of reviews, feedback and point ratings, is also used by several online auction sites like eBay.com and some web retailers like Amazon.com to enhance the user's level of trust on web merchants. In an empirical study by Sarah et al. [ ] it was identified that most users give high value to the previous customer endorsements to judge the ability of the web merchant, even more than third party affiliation. In the study 80% of the respondents reacted positively to establishing trust on online merchant, due to the positive feedback from the previous customers. However under such circumstances the users' level of trust in the information source plays a decisive role.

Another consideration for the development of the online reputation, especially in terms of its evolution, is that customer loyalties must be developed and this should be set in place right from the very beginning. Customer loyalties are developed from a process of providing not only transactional support but also emotional support to existing and continuing customers (which can be elicited from Figure 2).

2.3 Unified Merchants

Unified platforms like Amazon and eBay are changing the way we learnt to buy goods and services online. Such companies usually provide access to monetary insurance to their customers and have dispute resolution policies. Moreover, they facilitate the end users to search for a product or service, provide feedback on the sellers, standard user interface and optional access to enhanced monetary insurance services like PayPal. Such unified online platforms like eBay and Amazon are making it tougher for the large online merchants to compete with the small specialized merchants who do business under the umbrella of established brands like eBay and Amazon.

2.4 Third Party Trust Seals

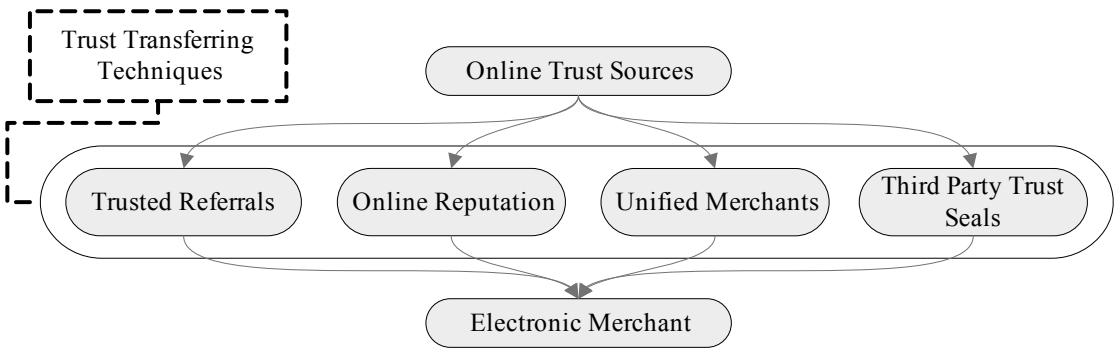
Third party web seals are used to provide consumers with a trusted view of an e-merchant. Such seals are mostly used by new or small businesses. The dominant trust seals used on the Internet include BBBOnLine Privacy [ ], BBBOnline Reliability [ ], TRUSTe [ ], and WebTrust. While evaluating the impact of WebTrust [ ] on user's perceived trustworthiness Portz [ ] identified that 94% of the participants noticed the presence of WebTrust seal. A recent study by Egger [ ] identified that the web-based trust seals contribute to trustworthiness in case of US respondents. The same study recognized that Americans consider the presence of familiar brands and credit card companies' logos, like MasterCard or Visa, as less of an indication of trustworthiness than web-only trusted third parties like VeriSign or TRUSTe. Cheskin Research [ ] investigated the international validity of online seals. Cheskin research concluded that the VISA brand is most trusted in Latin America while TRUSTe is most trusted in the US. However contrary to other studies, Princeton University [ ] survey estimated that only 19% of respondents identified seals as important in their trust formation.

The four recognized collective techniques used by electronic merchants to transfer online trust are presented in the figure 1.

3 CRITICAL EVALUATION

This section critically examines the evolution and ontological development of trust and then evaluates the online trust transferring techniques by identifying their advantages, limitations and dependencies.

Figure 1. Online Trust Transferring Techniques



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

3.1 Evolution and Ontological Development of Trust

Few academics would disagree that over the past two decades, organizations, and E-Businesses in particular, have become intensively technologically based [ ]. But what is not entirely obvious is that technological change alone does not evolve by itself into an ontology for trust in relation to an E Business. We can see that initially as a metaphor, an ontological development is like a culture. Haynes [13] makes that point that "technology alone cannot succeed in dominating a learning culture; the culture would simply disintegrate. In varying degrees then, any attempted technological domination would fragment and inhibit the development of a learning culture. It is rather the case that technology needs to be integrated into the organizational learning culture." An evolution of the development of trust in an E Business is as fragile as a learning culture and that fragility can be captured as a formalized model (or ontology). It is the understanding and use of the ontological development of these models over time (as evolving) that are "essential in an increasingly dynamic and uncertain business environment" [ ].

In relation to an evolution of trust it is the online reputation which is the most fragile. In terms then of this evolution we can see from the following diagram that the on-line reputation is based on the three intersections of trusted referrals, unified merchants and third party trust seals.

Further considerations for the development of the online reputation are customer relationships and follow ups to customers which then develop customer loyalties. If the customer is loyal then there has been an evolution of trust in that customer. That evolution of trust is based on the development of positive relationships between the firm and the customer. As Osterwalder et al [14] note "it is often forgotten in most cases that it is much cheaper to incite existing customers to do repeat business than to acquire new customers". Of course this assumes that there has been a sufficient build up in the client base, but nevertheless provision for customer loyalty should be considered from the start of the E Business cycle. It should be noted that the term "customer loyalty" has been chosen rather than 'user loyalty', because it is the term "customer loyalty" that will appear as part of the on-line trust experience display to on-line users.

3.1 Trusted Referrals

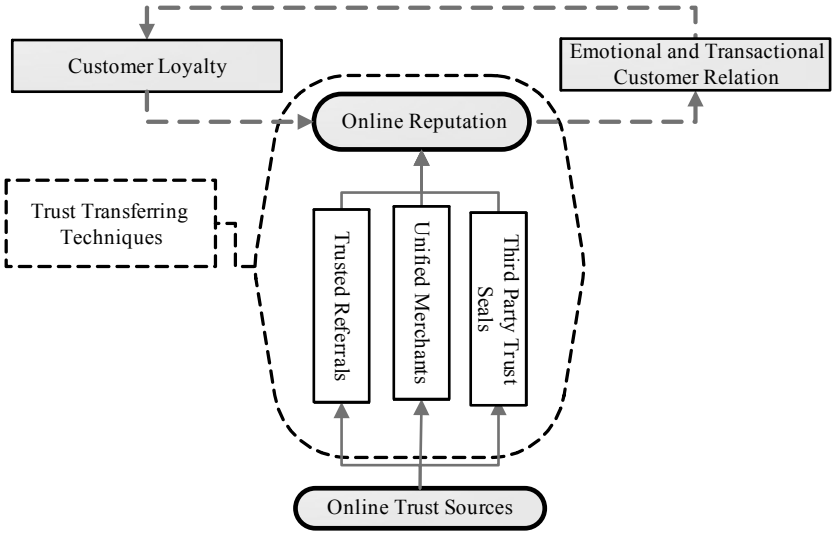
While analyzing the impact of trusted referrals Fullam et al. [2] identified that the users' prior confidence in the information source that is, users' perceived trustworthiness of the information source, contents of the reported information, number of sources reporting similar information, certainty conveyed in the information and age of the information affect the users' belief on acquired information.

Besides the above factors, the existence of trusted referrals in small isolated segments, if any, in the case of new and emerging merchants further act as a limitation to fully utilize their business potential. Any new and emerging online merchant will always have relatively lower brand value and small existing customer base in relation to established existing e-merchants. Therefore they have to employ other individual and collective techniques to further increase users' trust by transferring it from other sources. It can be concluded that trusted referrals should be established as a parallel process which the e-merchants should execute as they establish their brand and online business.

3.2 Online Reputation

Resnick et al. [ ] identified that from the potential customer's perspective, the absence of a process to validate and assure the honesty of feedbacks, for instance a group of people may collaborate to inflate their reputation, acts as a hurdle in wide acceptance of the online reputation system. In the same study, the absence of a mechanism to exchange/share online feedback information between reputation systems decreases the overall effectiveness of such systems. This lack of inter-integration makes the collection of information from all trusted sources expensive and a tiring activity for the potential customers. The above limitations work as a hindrance to fully transfer online trust from the previous customers to potential customers. Resnick et al. also recognized that

Figure 2. Evolution of Trust in E Business



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

mostly previous customers do not bother to provide any feedback and if they do, they usually only provide feedback if they go through very bad experience.

Moreover since new and emerging online merchants have a relatively small existing customer base, therefore they cannot fully rely on online reputation systems to build trust. Thus online reputation building should be a parallel process which the e-merchants should execute as they establish their brand and online business.

3.4 Unified Merchants

Unified e-merchant model works well for unestablished small business, as the new merchant inherits trust from the containing merchant e.g. eBay, Yahoo or Amazon. Extra services like, template site building, insurance services, searchable catalogues and transaction tools greatly reduce the burden of the new sellers. However unified merchants have associated high running costs and different charging models. For example, eBay charges US$15.95 to US$499.95 per month for each eBay store. eBay also charges US$0.05 to US$0.10 for inserting each item and an average of 10% of the final selling price of the product. Besides above the sellers are required to pay extra charges for product list upgrades and additional picture services and product promotions.

In relation to eBay, Yahoo Shopping essentially charges US$0.10 to US$0.75 every time someone clicks on the sellers' product, US$149 to US$999 per month for the use of marketing tools and US$249 to US$1599 per month for using search optimization service.

Amazon has adopted a different model from eBay and Yahoo. It only collects fees when the product is sold. At that time, Amazon, charges commission from 6% to 15% of the product sale price, a fee of US$ 0.90 for each transaction and a variable closing fee. The variable closing fee is charged as fixed price for certain items or a combination of fixed price and some other fixed charges depending on the weight of the product.

The above discussed pricing models clearly indicate that all three major unified merchants charge approximately 25% of the final product price. The unified merchant model is good for very small merchants who want to save on setup costs and plan to have small number of transactions. But this model clearly fails to work for merchants who want to establish their own independent brands, as the merchants will always be operating under the brand of unified merchant and may fail to establish their independent identity.

A unified merchant model is excellent for initial trust transfer, but due to high associated operation costs and branding issues, the unified merchant model is ideal for a new e-merchant to analyse the market and gain experience. Any e-merchant targeting for high transaction volume should setup an individual independent brand. This model should be used as a step to learn the process of selling products online and to develop online reputation and trusted referrals network.

3.5 Third Party Seals

The third party seals are mostly used by new or small businesses, as it only makes sense to use them when the trust value generated by the seals used on the website exceeds the trust value generated by the merchant's brand. This explains why major e-brands are not part of online seal programs. Such seal programs are usually not cheap and small e-merchants may not be in a position to afford associated costs for example, the TRUSTe Email Privacy Seal [ ] application fee range from US$450 to US$1,875 and the annual license fee for each brand range from US$1,000 to US$15,200 per year. McKnight et al. while evaluating the effectiveness of trusted third part seals and icons in promoting consumer trust in electronic commerce states that "web-related consumer decisions take place within the context of an individual's personal tendencies and his/her perceptions about the web environment. Thus, user trust in web vendors and the effect of third party icons may depend, to an extent, on individual characteristics … and an individual's prior experience with the web." [ ]

A recent study by Egger [10] identified that web-only privacy and security seals were perceived as trustworthy only by the people who know them. Burke, et al. [ ] study identified that consumers who attended to the seal, were those who were exposed to the advertisement about the seal.

Although the third party seal programs are generally quite effective in transferring and enhancing user's trust on e-merchant. But as discussed above, they are expensive, require a lot of effort to get and their impact is correlated with cultural background and awareness of the users. Thus the seal programs should be used for the development and establishment of online reputation and trusted referrals.

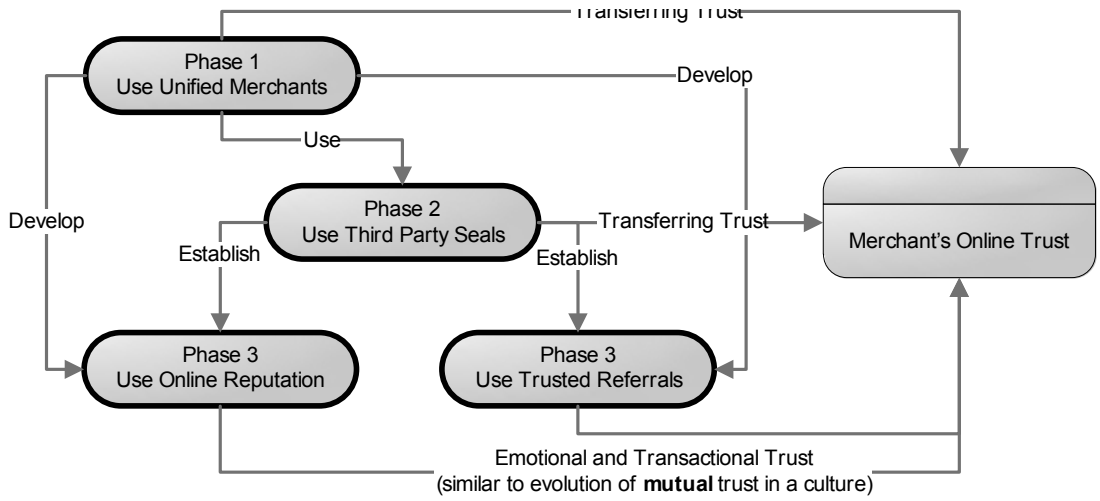
4 RECOMMENDED TRUST TRANSFERRING PHASES

After analyzing the advantages and shortcomings of each collective trust transferring technique, a model is proposed in this section. The proposed model can be used by e-merchants as a guide to employ a unique combination of collective trust transferring technique at each stage of business development.

The proposed model suggests the use of unified merchants as a starting point to gain experience and minimize setup costs. At this stage the e-merchant, it is suggested, should use every available opportunity to develop an online reputation and a trusted referral network. After developing trusted referrals and an online reputation the e-merchant moves to Phase 2. In Phase 2, the e-merchant should display the company logo, personal contact details and provide external links to its external independent website – from the unified merchant's store location. At this stage the e-merchant is recommended to display the third party seals to further transfer trust from other sources. Just like in Phase 1, the e-merchant should concentrate on establishing an online reputation and trusted referrals network and developing customer loyalties.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. Proposed Application of Trust Transferring Techniques



In Phase 3, the e-merchant should start to operate independently and should fully utilize established online reputation and trusted referrals network. At this stage the e-merchant is recommended to keep on displaying the third part trust seals along with online reputation and customer loyalty information to further attract new customers. By the end of this stage, the e-merchant has fully established brand value, online reputation and trusted referrals. In phase 4, followed by phase 3, the e-merchant may remove the third party trust seals, as its brand value is higher than the value of online seals. Figure 3 visually presents the proposed model.

4.1 Strengths of the Model

We have emphasised in this paper an evolution of the ontological development of trust in an E-Business and further noted that this development is both similar to, and as fragile as, a learning culture. When a culture is finally formed it is continuously achieved by passing certain milestones. Moreover cultures are not transactional alone but essentially evolve through transactional and emotional elements. Completed formative milestones in the evolution of a culture are, when encapsulated into a formal model, best depicted as distinct phases of completion.

The particular strength of this proposed model is that it exemplifies the passing of certain phases and because it recognises the importance of the emotions of human beings it thereby aligns itself to cultural evolution. The successful completion of each phase strengthens trust, and thereby allows, during each incremental trust phase completion, an opportunity for the display of that strengthened trust. Since the recognition of trust is not purely transactional but engenders, as well, emotional connections in human beings, a step by step phase completion is highly appropriate.

The key strength to the recognition of trust embedded in this model is the premise that the effective transfer of trust is not merely transactional but also emotional and because it is also emotional that emotional building of confidence is best achieved by satisfying developed and established (and displayed) successfully completed incremental phases much like the development of a culture through successive mutual trusting phases.

5 CONCLUSION

Trust is a significant stumbling block in the development and wide adoption of electronic commerce. It is often forgotten that customer loyalties and the effort in terms of a continuation of trust to maintain those loyalties are a significant element in the evolution of trust. When physically present established merchants move into the electronic environment they face none or little restrictions in capturing online consumer market, primarily because of the trust value associated with their brand. In the case of established merchants the total trust value is the composite of individual and collective trust transfers, such as; trust transferred through existing online trust and trusted referrals. In comparison to established e-merchants, new and establishing e-merchants have to rely on different individual and collective trust transferring techniques.

This paper has provided an overview of the limitations and advantages of recognized trust transferring techniques for the purpose of formulating a proposed model to both retain advantages and minimise limitations (limitations which arose because there was not a clear recognition of how trust really evolves). We have argued that trust really evolves like a culture evolves. Therefore our proposed model is inspired by the development of a culture. A culture successfully evolves because each of the participants has trust in that evolution. That trust, which also relies on the emotions of the participants, arises over time and over the completion of certain cultural milestones. Therefore we have provided a model that emulates cultural evolution insofar as the model exemplifies the passing of certain phases, where, importantly, each successfully completed phase connects at the emotional level of a human being who is participating in the capacity of a user seeking proof of trust.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

REFERENCES

1. Kim, K. and Prabhakar, B. 2000. Initial Trust, Perceived Risk, and the Adoption of Internet Banking. In Proceedings of the Twenty First International Conference on Information Systems, December 2000.537 – 543
2. Fullam K. and Barber K.S. 2004. A Belief Revision Algorithm Based on information valuation. Laboratory for Intelligent Processes and Systems. Technical Report TR2003-UT-LIPS-021
3. Zacharia, G. 1999. Trust management through reputation mechanisms. In Proceedings of the Second Workshop on Deception, Fraud and Trust in Agent Societies, Seattle. http://www.istc.cnr.it/T3/download/aamas1999/Zacharia.pdf. (pp: 163 - 167)(August, 2006)
4. Sarah P. W. S, Choon-Ling S, Kai H. L. 2002. A Preliminary Assessment of Different Trust Formation Models: The Effect of Third Party Endorsements on Online Shopping, In Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)
5. BBBOnLine (2006). BBBOnLine Privacy Seal Program. Retrieved from http://www.bbbonline.org/privacy. (12 October 2006)
6. BBBOnLine (2006). BBBOnLine Reliability Program. Retrieved from http://www.bbbonline.org/reliability. (12 October 2006)
7. TRUSTe (2006). TRUSTe.org, Retrieved from http://www.truste.org (12 October 2006)
8. WebTrust (2006) WebTrust Seals, Retrieved from http://www.webtrust.org/ (12 October 2006)
9. Portz, K (2000) The effect of WebTrust on the perceived trustworthiness of a website and the utilization of electronic commerce. University of Nebraska, Lincoln, NB
10. Egger, F.N. (2003). From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce. PhD Thesis, Eindhoven University of Technology (The Netherlands). ISBN 90-386-1778-X.
11. Cheskin Research (2000). Trust in the Wired Americas. Cheskin Research, California, USA. Retrieved from http://www.cheskin.com/assets/report-CheskinTrustIIrpt2000.pdf (10 October 2006)
12. Princeton Survey Research Associates (2002) A Matter of Trust: What Users Want from Web Sites. Consumer WebWatch, April 2002. Retrieved from http://www.consumerwebwatch.org/dynamic/web-credibility-reports-a-matter-of-trust-abstract.cfm (11 October 2006)
13. Haynes John D (1999) Phenomenological Aspects of Churchman's Hegelian Inquiring System, Philosophical Foundations of Information Systems, Proceedings of the Americas Conference on Information Systems, August. Association for Information Systems, USA, pp 633-636.
14. Osterwalder, A, Lagha, S B, and Pigneur Y (2002) An Ontology for Developing e-Business Models. Website; http://inforge.unil.ch/yp/Pub/02-DsiAge.pdf
15. Resnick, P., Zeckhauser, R., Friedman E. and Kuwabara, K. (2000). Reputation Systems: Facilitating Trust in Internet Interactions. Communications of the ACM, 43 (12) pp: 45-48
16. TRUSTe Email Privacy Seal Program (2006), TRUSTe Email Privacy Seal Program Fees, TRUSTe, Retrieved from http://www.truste.org/businesses/invoice_generator_email_privacy_seal.php (12 October 2006)
17. McKnight, D.; Choudhury, V. & Kacmar, C. (2000) Trust In E-Commerce Vendors: A Two-Stage Model, International Conference on Information Systems, Proceedings of the twenty first international conference on Information systems, pp 532 – 536 (pp: 533).
18. Burke, K.G., Kovar S.E. and Kovar, B.R. (2000). "Selling WebTrust: An Exploratory Examination of Factors Influencing Consumers" Decisions to Use Online Distribution Channels ," The Review of Accounting Information Systems, Vol.4, No. 2: 39-52, Spring 2000.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e•Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Clustering e-satisfaction factors in tourism industry

Masoomeh Moharrer

Department of Management, Economics and Industrial Engineering,
Politecnico di Milano, Milan, Italy
masoomeh.moharrer@polimi.it

Hooman Tahayori

Universita Degli Studi di Milano,
Dipartimento di Scienze dell'Informazione, Milan, Italy
hooman.tahayori@unimi.it

Abstract The emergence of the Internet has led to the rapid growth of e-commerce which has had an effect on the nature of different businesses. Tourism industry is not exempted from that. Despite the growing body of literature dealing with tourism industry and online consumer behavior, few researches have been done on satisfaction in e-tourism. This research is going to fill this gap by clustering items that affect customer satisfaction in online tourism which results in investigating main factors which are the determinants of e-satisfaction. In the next step it has been shown that customers' satisfaction will make customers recommend tourism product to the others. This research has collected the data among a group of European people who has experienced e-tourism at least once before. The result helps managers of tourism firms by introducing them the factors namely Convenience, Product offered, Offered information, Web site design which can affect the satisfaction of their customers.

Keywords e-tourism, e-marketing, customer satisfaction, cluster analysis

1 INTRODUCTION

Today Internet, with an estimated 813 million users worldwide, provides at modest cost, an unprecedented level of connectivity and the ability to communicate efficiently, effectively and directly with customers. Its potential to generate more revenue is no longer a matter of debate [7].

E-tourism, takes into account when, traditional travel agencies, tour operators, national tourist offices, airlines, car hire firms, hotels and other accommodation providers offer their services online which enable the tourists to schedule their trip online, and hence describes a new way of doing business.Tourism industry is an information-intensive industry in which e-commerce is already playing a significant role by allowing information to flow through the Internet on a worldwide basis with virtually no entry barriers. E-commerce is not only about transactional activities but satisfaction and retention of customers as well. A Research has shown that a 5% increase in customer retention results in a 25–95% increase in profits [12]. This relationship between customer retention and profits underlines the importance of customer and customer satisfaction. Customer satisfaction will increase loyalty hence if e-tourism firms can manage to increase customer satisfaction they can guarantee an increase in their revenue from tourism. E-tourism has had a dramatic growth in recent years. Marcussen in [6] has mentioned that online travel sales has increased for 31% from 2005 to 2006 and has reached EURO 38.3 billion in the European market in 2006. He also suggested that at the end of 2007 a further increase of about 22% may be expected.

E-tourism is enhanced with different factors like strong information search capabilities offered by the Internet, and in general with factors which can increase customer satisfaction.

The main aim of this survey was to characterize the main factors, which determine e-satisfaction in tourism industry or in other words satisfaction in e-tourism.

2 LITERATURE REVIEW

Internet has stimulated dramatic changes in tourists' information search and purchasing behavior. Due to the nature of the tourism products and services (e.g., intangibility, complexity, diversity, and interdependence), customers cannot test them at the point of the sale so they are more eager than ever for product-related information in order to minimize their purchase risk and close the gap between their expectations and the actual travel experience. A research about satisfaction factors in websites of the hotels has been conducted whose results indicated that potential online customers were only moderately satisfied with hotel websites. They found

that website design, sufficient information, and customers' perceptions of security for online transactions were crucial to increase the number of Internet sales Satisfaction can be defined as the perception of pleasurable fulfillment of a service [9].

E-tourism firms use internet to better serve their customers. Customer relationship management is a crucial subject which firms take special attention to, that directly or indirectly results in Customer satisfaction, Customer loyalty and finally Customer retention. Among these concepts customer satisfaction can relatively influence customer loyalty and retention which in turn increases firms profit and efficiency. CRM as a research topic has attracted much attention since the beginning of the 1990s. However, in a new e-commerce context the concept of CRM and its core subject, customer satisfaction, yet has not been studied sufficiently.

E-tourism firms use internet to better serve their customers. But making tourism service information accessible to customers is not enough for effective distribution. Well-designed mechanism must allow customers to process their purchase [5]. If a tourism organization can better represent its destination (services) on the internet than another then it may win tourist who is uncertain about where to travel [16]. Since travelers cannot pre-test the product or easily get their money back if the trip does not meet up to their expectations, access to accurate, reliable, timely and relevant information is essential to help them make an appropriate choice [8].

There have been several researches related to satisfaction in traditional ways of businesses, e.g. [2], [11], [10]. Moreover there have been also enough researches related to tourists' satisfaction in off-line environments, e.g. [1], [3]. The authors have introduced several models and determinants of tourists' satisfaction. But unfortunately there have been fewer researches related to satisfaction in on-line industries especially on-line tourism, some of existing researches will be stated in the following paragraphs.

Jeong and Gregoire in [4] investigated consumer perceptions of hotel websites. The results indicated that potential online customers were only moderately satisfied with hotel websites. They found that website design, sufficient information and customers' perceptions of security for online transactions were crucial to increase the number of Internet sales. In this research it will be demonstrated that the satisfaction of a tourism website positively affect online customers' intention of buying from that website.

Szymansky and Hise in [15] conceptualized e-satisfaction as the consumers' judgment of their Internet retail experience as compared to their experiences with traditional retail stores. They have created a model which is the base of our work, but here it is consumers' judgment of their e-tourism experience comparing with their experience of traditional travel agencies.

3 METHODOLOGY

The first stage of this work was qualitative, which was to design the questionnaire and the next part was quantitative and focused on gathering survey data to assess determinants of the e-satisfaction model. Overall satisfaction with e-tourism was measured by the degree to which the consumer was satisfied/dissatisfied. The initial questionnaire for the current study was designed in different sections. In a focus group interview, we used items of Servequal Model [1] and e-satisfaction model [15]; the final questionnaire was made after conducting pilot test.

The first section of the questionnaire was designed to obtain the respondents' demographic data and behavioral characteristics: gender, age, occupation, purpose of their trip, use of travel agency and e-tourism websites per year.

Sections 2 and 3 of the questionnaire consist of main questions related to the model in this research. Items were gathered from the previous studies, Servequal Model [1] and e-satisfaction model [15]. In section 2, the items of these factors investigated perceptions of tourists toward on-line tourism organizations versus traditional travel agencies. This comparison between e-shops and their counterparts in traditional market is exactly the way, which was used by Szymansky and Hise [15] in e-satisfaction model. The traditional travel agencies are the reference because almost everybody has an experience in using them so it would be a suitable comparison while respondents make judgment for satisfaction. In this section respondents were asked to compare each item in on-line tourism with traditional travel agency, relating to their previous experiences. Their e-satisfaction level measure was in a 5-point scale, (1) Much worse than (2) Worse than (3) The same (4) Better than and (5) Much better than. In the next section, section 3, items related to tourism websites' design are investigated. In this section respondents are asked to mention their satisfaction level on a 5-point scale namely (1) very dissatisfied, (2) dissatisfied, (3) Fair, (4) satisfied and (5) very satisfied.

After designing the questionnaire a pre-test was conducted. In this part the questionnaire was given to the tourism experts and those who were expert in designing the questionnaire. Then a pilot test was conducted too, so the questionnaire was given to 10 people from our sample. The aim was to find if all items and questions are easy to understand and if it requires any change. Some modifications were implemented after finishing these two phases.

The suitable population for data collection of this research were the people who have experienced e-tourism at least once before. Beauvais airport in Paris accepted to cooperate with this research. This airport was used mainly by Ryanair (www.Ryanair.com) which is an airline company that has cheap flights for most of the countries in the Europe. All of the tickets of this company sold directly to the passengers on-line.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Totally 150 questionnaires were distributed, 115 questionnaires was given back and after eliminating unusable and incomplete responses and a further process of eliminating outliers the data was pared down to 99 cases. Totally 53% of respondents are female and more than 50% of the respondents are between 18-34 years old. The findings of this research will be representative of European countries, since all the respondents are European.

A Cluster analysis was performed on the questions of second and third sections of the questionnaire to ascertain factors which reflect consumers' underling mental model. Usually cluster analysis is used to cluster observations, but in this research it is used to cluster variables. To this end we have transposed the matrix of our dataset in order to change the place of our observations and our items and then tried to cluster our items to achieve main factors which explain satisfaction in e-tourism. Using cluster analysis for clustering items is a novel technique which is rarely used in the literature. Rencher in his book [13] has mentioned that it would be interesting to cluster variables rather than observations. The items which were used in cluster analysis and the result of cluster analysis is illustrated in Table1.

4 RESULTS

A Cluster analysis with Average Linkage method using STATA software was performed with independent variables to determine the factors which measure the customer satisfaction in online tourism. The result showed a four-factor structure for predictors of e-satisfaction in tourism industry. In fact it has clustered items into four groups. Factors are logically and easily interpreted (Table 1). They are Convenience, Product offered, Offered information and Website design.

Table 1. Result of Cluster analysis using Average Linkage Method.

Factors	Items
Convenience	Time efficiency, Purchase anywhere, Convenience at any time, Direct access, Price of tourism services,
Product offered	Number of tourism services, Variety of tourism services,
Offered information	Quantity of information, Quality of information, Safe feeling in transactions, Formal privacy,
Website design	Friendliness of web sites, Attractiveness of web sites, Interactiveness of web sites, Uncluttered screens, Download time

The next finding of this research is that whether customers' satisfaction will make customers recommend tourism product to the others. The result is shown in Table 2. It illustrates that all the respondents whose overall satisfaction is rated 4 and 5 (satisfied and very satisfied), except two of them, will recommend others to use e-tourism.

Table 2. Results of recommending tourism products

Overall Satisfaction	Recommendation to Others	
	No	Yes
1	-	-
2	-	2
3	4	10
4	1	57
5	1	25

5 CONCLUSION AND IMPLICATIONS

As information is the life-blood of the travel industry [14], utilizing and managing Information Technology efficiently is essential for tourism organizations to satisfy their customers. Some factors play key roles as the predictors of tourists' satisfaction in e-tourism, but yet there has not been enough research on the topic. In this research different determinants of tourists' satisfaction in on-line tourism were investigated.

The result showed that Convenience, Product offered, Offered information and Website design are the factors which are the determinant of customer satisfaction in online tourism. Convenience includes increasing time efficiency, enables customers to shop anytime and from anywhere in the globe which is the most important factor for tourists. E-tourism service providers can increase the satisfaction of their customers by providing adequate quantity of information, high quality information, large number of tourism services and variety of tourism services. Good site design includes fast, friendly and uncluttered sites. This finding illustrates to managers of e-tourism firms the factors which can affect customers' satisfaction with the tourism websites. The study also clearly demonstrates that satisfied customers will recommend others to use e-tourism. This result shows the importance of customer satisfaction in e-tourism. Since tourist can not see or check the product or service of tourism before the selling point, recommendation to other or in other words, word of mouth can be a great marketing channel for tourism service providers. This can be achieved by satisfying current customers.

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciations to Professor Rocco Mosconi from Politecnico di Milano for all his guidance and kindness.

REFERENCES

[1] Akama, John.S, Mukethe Kieti, Damiannah,(2002), "Measuring tourist satisfaction with Kenya's safari", Tourism management, 24, pp73-81.

[2] Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). "Customer satisfaction, market share and profitability: Findings from Sweden". Journal of Marketing, 58(2), 112–122.

[3] Haber, S., Lerner, M. (1999). "Correlates of tourist satisfaction", Annals of Tourism Research, 26, 197–201.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

[4] Jeong, M., Oh, H. , Gregoire, M. (2001), "An internet marketing strategy study for the lodging industry", American Hotel and Lodging foundation.

[5] Kim, G.W, Xiaojing Ma, Dong Jin Kim (2005), "Determinants of Chinese hotel customers' e-satisfaction and purchase intention", Tourism management.

[6] Marcussen C.H., (2007), Trends in European Internet distribution of travel and tourism services. Trends in European Internet Distribution - of Travel and Tourism Services, Centre for regional and tourism research, Denmark, http://www.crt.dk/uk/staff/chm/trends.htm.

[7] Maswera, T., Dawson, R., Edwards, J., (2006),E-commerce adoption of travel and tourism organizations in South Africa, Kenya, Zimbabwe and Uganda, Telematics and Informatics.

[8] O'Conner, Peter (2000), "Electronic information distribution in tourism and hospitality. Oxon, England: Cabi Publishing.

[9] O'Connor, P., Frew, A.J. (2002), "The future of hotel electronic distribution: Expert and industry perspective", The Cornell hotel and restaurant administration quarterly, 43(3),33-45.

[10] Oliver Richard L. (1997), "Satisfaction: A behavioral perspective on the consumer", New York, McGraw-Hill.

[11] Parasuraman, A., Zeithaml, V. A., Berry, L. L. (1988). SERVQUAL: a multiple-item scale for measuring consumer perceptions for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), pp.12–40.

[12] Prewitt,2002, Build customer loyalty in an internet world, www.cio.com- Scottish Parliament, Tourism e-business, Aug.2002.

[13] Rencher Alvin C. (2002), Methods f Multivariate Analysis, 2nd Ed., John Wiley & Sons, USA.

[14] Sheldon P., Information technologies for tourism, CAB International. Oxford, 1997.

[15] Szymansky, David M., Hise, Richard T. (2000), "E-satisfaction: an initial examination, Journal of retailing, 76(3), 309-322.

[16] World tourism organization business council (1999), Marketing tourism destination on-line-Strategies for information age. Madrid: WTO Publication.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Motivation in organisations: trends in modern ICT companies

Olatubosun Olubusuyi Ojo

1 INTRODUCTION

The purpose of this paper is to research top performing and effective organisations operating globally in the information technology industry and to understand the techniques that have been used to motivate its workforce.

Technology transfer from one part of the world to another has demonstrated that there in no need to reinvent the wheel but it is essential to understand how the wheel was made in order to be able to make improvements and to increase its efficiency.

This same idea will be used in this paper after researching into those good practices in motivation that has made Cisco, Google and Microsoft household names all over the world.

This paper will seek to find a general model that will be effective anywhere in the world and to project into the future the right motivation techniques that managers will need to adopt as technology improves our lives.

1.1 OBJECTIVES OF THE RESEARCH

Many papers have been written on the subject of motivation in organisations but very few on the comparative analysis of the effect of motivation in firms that operate globally. All organisations have differences in their goals and objectives; some were set up to offer services to people while others were created as a result of a scientific invention or innovation and provide unique products to its customers.

The common link between all organisations is that they have employees who put in time and effort to ensure that the goals and objectives of the organisations they work for are met.

A well motivated workforce is the building block of a successful organisation therefore it is important to study the science of motivation in order to deploy proven techniques and to ensure that the goals and objectives of the organisation are achieved.

This research will study the techniques used in motivating employees working in information technology organisations that have been able to achieve a global influence within the past few decades with a view of establishing the efficacy of those techniques.

It will also analyse those methods that have been found effective and how improvements can be made as ICT becomes part of our daily lives.

This paper is intended to serve as a reference point for future entrepreneurs to enable them learn from the experience of successful organisations and to seek ways in which improvements and innovation can be achieved.

2 LITERATURE REVIEW

"Motivation can be defined as the psychological forces that determine the direction of a person's level of effort, as well as a person's persistence in the face of obstacles" (Daniel, T.A & Metcalf, G.S, 2005).

An individual may be motivated due to intrinsic or extrinsic sources.

Intrinsic motivation is evident when an individual acts independently or without prompting for its own sake. Intrinsic motivation comes from within an individual and can be closely associated with those habits that give a sense of satisfaction to the individual in relation to work.

Extrinsic motivation on the other hand leads to actions that are performed to acquire material or social rewards or to avoid punishment. It derives its source from the consequences of action or inaction.

What differentiates intrinsic motivation from extrinsic motivation is the reward expected or outcome (Lepper, M.R. 1988). The outcome of intrinsic motivation for an individual is a sense of responsibility, autonomy, a feeling of accomplishment and the pleasure of doing enjoyable work while extrinsic motivation leads an individual to tangible benefits such as pay, stakeholder shares, public recognition and prestige.

Organisations employ people to perform certain tasks; these tasks are vital for the organisation to achieve its goals therefore it is good management practise to be able to motivate

employees to make valuable contributions to the organisation.

There are many management theories that have been put forward to explain what creates a motivated workforce but for this paper we will consider the following theories: expectancy, need, equity, goal-setting and learning theories.

2.1 Expectancy Theory:

Postulated by Victor H.Vroom in the 1960s states that motivation is high when employees believe that high levels of effort lead to high performance which in turn leads to the attainment of desired outcomes (Daniel, T.A & Metcalf, G.S, 2005).Managers need to make sure that those under them must believe that if they do try hard, they can actually succeed. It is a popular theory because it focuses on all three parts of the motivation equation: inputs, performance and outcomes.

2.2 Need Theories:

These are theories which provide managers with knowledge about what outcomes or rewards motivate employees to perform at optimum levels and suggest inputs to assists organisations achieve its aims in order to survive (Smith, G.P. 1994).

Under the Needs theories we have Maslow's Hierarchy of Needs, Herzberg's Motivator-Hygiene Theory, McClelland's Needs for Achievement, Affiliation and Power.

2.2.1 Maslow's Hierarchy of Needs

This theory is named after the psychologist Abraham Maslow who proposed that all people seek to satisfy five basic kinds of needs which are psysiological, safety, social, ego and self-actualization needs (Maslow, A 1947).

These needs are arranged in order of importance, from the basic to the complex and an individual moves to the next level only after the lower level is at least minimally satisfied.

2.2.2 Herzberg's Motivator-Hygiene

Fredrick Herzberg proposed that two factors in the workplace result in job satisfaction, while others do not, but can lead to dissatisfaction if absent (Herzberg, F.1959).

These are Motivators; for example challenging work, recognition, responsibility which gives positive satisfaction.

While Hygiene factors for example salary, fringe benefits, status and job security which do not motivate if present but will lead to demotivation if absent.

2.2.3 McClelland's Needs for Achievement, Affiliation and Power

David McClelland also a psychologist like A. Maslow instead proposed that individuals have three needs which are the need for achievement, the need for affiliation (good social skills) and the for power but that people differ in the degree in which the various needs influence their behaviour (Daniel, T.A & Metcalf, G.S, 2005).

2.3 Equity Theory:

This was postulated by J.Stacy Adams in the 1960s and it focuses on people's perceptions of the fairness of their work outcomes in relation to, or in proportion to, their work inputs. It stresses that motivation is influenced by a comparison of one's own outcome/input ratio with the outcome/input ratio of a referent. If an individual feels that the ratio is unfair then the person's performance will fall; however, when the ratio is seen to be fair and equity perceived to exist then people are motivated to continue contributing their current levels of inputs in order to sustain their current levels of outcomes (Adam, J.S. 1963).

2.4 Goal-Setting Theory:

Ed Locke and Gary Latham are the leading researchers of this theory and they suggest that the goals employees strive to obtain are the major determinants of their motivation and performance (Daniel, T.A & Metcalf, G.S, 2005).It also suggests that to encourage a high level of motivation and performance, goals must be both specific and challenging.

2.5 Learning Theories:

The learning theory suggests that people learn to perform behaviours that lead to the desired outcomes and learn not to perform behaviours that lead to undesired consequences (Skinner, B.F.1950) having this knowledge means that managers can increase the motivation of employees through the ways they link the rewards that employees receive to the performance of the desired behaviours.

By linking the performance of specific behaviours to the attainment of specific outcomes, managers can motivate employees to perform in ways that help the organisation meets its goals.

This paper will now consider the growth and performance of selected modern information technology organisations and the extent to which some of the theories of motivation has assisted in the success and also perform a comparative analysis afterwards.

3 CASE STUDIES

A global study in the Harvard business review showed that more than 50% of male executives and more than 80% of women executives working 60 hours a week or more said they would not be able to keep it up for more than a year.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Globalization, instant communication to keep in touch with offices in different time zones of the world and also increased travel commitments are now a reality in today's world therefore organizations at the cutting edge of technology will need to create alternative work models to the ones used in previous decades.

Some organisations have already adopted radical changes to work in order to motivate employees to meet the challenges of today. The organisations under study in this research were all started by university colleagues and provide a good model on which future entrepreneurs can improve upon.

3.1 Cisco:

Cisco systems Inc with headquarters at San Jose, California USA is a global company that designs and sells networking and communication technology services under three brands: Cisco, Linksys and Scientific Atlanta.

The company was formed in December 1984 by two members of Stanford's Computer science support staff, Len Bosack and Sandy Lerner.

The Xerox Palo Alto Research Centre gave Stanford some of its Alto Workstations and Ethernet networking boards and then through innovation and enterprise Len Bosack and his wife Lerner found a way to make the first multiprotocol router marketable and commercially viable working from their living room.

Initially there was some controversy between Stanford University and Cisco in relation to the technology behind the routers, the important point is that it takes entrepreneurship to be able to convert invention into commercial success. The dispute was settled eventually as Stanford licensed the router software and two computer boards to Cisco.

Gradually the company grew from two founders into a global organisation with fifty one thousand four hundred and eighty employees as at November 2006.

Also with a current revenue of close to $30 billion it is undisputedly one of the major information technology organisations on the globe.

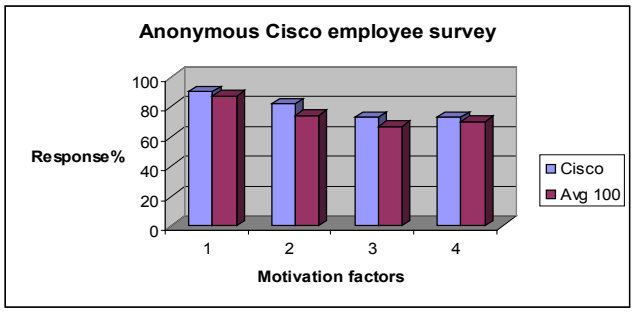
Cisco has consistently been voted as one of the "100 Best Companies to Work For" by fortune magazine and several other reputable ratings organisations within the past few years which makes it clear that it is adopting modern motivation techniques which will be studied in detail in this research paper.

According to data obtained from the Great Place to Work Institute based in Canada after an anonymous survey evaluating staff attitudes to company values, trust in management and pride in work it can be deduced that Cisco performs better than the average 100 best company to work for.

In a recent employee survey conducted as part of the Best Companies to Work For evaluation process, 95% of the employees who responded said "Taking everything into account I'd say this is a great place to work" which is quite remarkable for a young, fast paced and stressful but exciting work environment.

Cisco compared with the "Average 100 Best Company to work for in America"

Figure 1. Motivation factors



Source: Great Place to Work Institute

Key to employee responses:

1: Management is honest and ethical in its business practices.
2: People are paid fairly for the work they do.
3: I feel I receive a fair share of the profits made by this organisation.
4: This is a physically safe place to work.

3.1.1 Analysis:

Analysing the responses of Cisco employees it is clear that equity theory which focuses on people's perception of fairness either in relation to company ethics or work inputs is a factor that has motivated the staff to perform at optimum level and seems to be slightly more important to them than the needs theory.

Table 1.

No	Motivation factor	Theory
1	Management is honest and ethical in its business practices.	Equity
2	People are paid fairly for the work they do.	Equity
3	I feel I receive a fair share of the profits made by this organisation.	McClelland's Needs for Achievement, Affiliation and power
4	This is a physically safe place to work.	Herzberg's Motivator-Hygiene

3.2 Google

Google Inc was first incorporated in California, USA in 1998 and then reincorporated in Delaware, USA in 2003. The main business of Google is to maintain the index of web sites and other contents and to make this information available to anyone with an internet connection.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

It was founded by Larry Page and Sergey Brin who were graduates students of computer science at Stanford University at the time, when the first office was opened there were only there staff, a year later there were eight employees. As at December 2006 the company had grown to a global organisation with ten thousand six hundred and seventy four employees.

The revenue for Google in 2006 was almost $11 billion and it came tops on the list of "100 Best Companies to work for 2007" a ranking for fortune magazine by the Great Place to Work Institute. In 2006 it was the only information technology company to achieve the very rare distinction of zero percent employee turnover.

Google Management attributes its success within a short time to its policy of employing creative, principled and hard-working stars and then rewarding them generously.

Table 2. Google strategy

No	Motivation Technique	Theory Applicable
1	Treating employees with respect	Maslow's Hierarchy of needs
2	Supporting individual creating endeavours	Maslow's Hierarchy of needs
3	Maintaining company motto of "Don't be evil"	Equity

3.2.1 Analysis:

Google has been a pacesetter in the area of employee motivation in this information technology age; it has adapted some of the most radical ideas of workplace attitudes in the twenty-first century.

According to the Great Place to Work Institute article titled "Why is Google so great?" Google seeks out intelligent, creative and entrepreneurial people to join its ranks and also provides strong support for people's professional growth and development. They also provide opportunities to have fun at work as well.

Having fun at work as against just taking a break from work is a new idea which has the ability to reduce stress, build camaraderie and enable creative thinking. These cover the first two points in the table above while the third point is embedded in Google philosophy and motto which is to fair and to be seen that way by its employees.

There are many schemes within Google which enable employees to express themselves for example through internal e-mail lists dedicated to the discussion of new ideas, issues and complaints.

The Management team takes the forum seriously and provides adequate feedback to encourage and motivate staff.

The theories most applicable to Google are the Equity theory and Maslow's Hierarchy of needs with radical innovative ideas to make it a truly twenty-first century global organisation.

3.3 Microsoft

Microsoft Corporation has played a major role in the information technology age. Its main line of business is to develop, manufacture, license and support a wide range of software products for computing devices.

Microsoft Corporation was founded by Bill Gates and Paul Allen who met at an exclusive prep school called Lakeside in Seattle, United States of America.

From this friendship and a love of computing devices grew the idea of starting a company which took the world of programming by storm.

With seventy six thousand employees worldwide and revenue of $44.3 billion dollars in the year 2006, Microsoft is one of the biggest global Information technology companies.

It was voted the No 1 Best Company to Work For in 2003 by fortune magazine and its current ranking in 2007 among the 100 Best Companies to work for is 50.

Motivation of employees using Maslow's Hierarchy of Needs theory is a long established practice at Microsoft for example between 1986 and 1996 Microsoft stock increased more than a hundredfold due to the popularity of its windows operating system and office applications in the PC industry.

The rewards of this success was evident among the employees as four Microsoft employees became billionaires while a further twelve thousand employees became millionaires, this was an unprecedented application of Maslow's theory.

Morale is very high at Microsoft among the employees due to the introduction of "morale budgets" for socialising which is a way of applying McClelland's theory of Needs for Achievement, Affiliation and Power.

Table 3. Microsoft's approach

No	Motivation Technique	Theory Applicable
1	Allowing employees to contribute ideas into projects	Maslow's Hierarchy of needs
2	Reward of employees through the option of buying stocks	Maslow's Hierarchy of needs
3	Introduction of morale budgets for socialising	McClelland's Needs for Achievement, Affiliation and Power.

3.3.1 Analysis

Microsoft has played a central role in the information technology age; it has inspired several start-up companies all over the world to follow its model.

The company has continued to encourage employees to remain within the organisation by offering stocks to them es-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

pecially those employed after the boom of the nineties that made several older employees millionaires.

Although it is no longer the best company to work for according the fortune ratings it still retains a prominent place among information technology organisations.

The theories most applied in Microsoft are the needs theories of Maslow and McClelland, its pioneering role in the software industry has opened the doors of ideas to many companies today.

3.4 Critical Analysis of Case Studies

All three organisations have been rated among the "Best Companies to Work For" at various times by fortune magazine within the past decade.

This consistency demonstrates that employee satisfaction and motivation is needed to achieve goals set by the managements of these global organisations and that is why priority has been given to them.

The highest motivating factor in Cisco based on the anonymous Cisco employee survey was the perception of a honest management and ethical business practice by the company followed closely assurance of been fairly paid for work done.

For Google the main motivating factor is that the company treats employees with tremendous respect while also supporting individual creativity.

Amongst the three companies it is the Google strategy that appears to keep the most employees satisfied as it has the lowest turnover of workers, and those that leave do so with utmost reluctance.

The exceptional observation about Google is the innovative use of Information Technology Communication (ICT) tools in motivating employees for example the use of internal e-mail lists to discuss new ideas, issues and to resolve complaints. All these prove that money alone is not the biggest motivator in organisations.

Microsoft continues its policy of empowering employees through the offers of stocks and allowing creativity in the workplace.

Over the past few years a lack of career opportunities and fear of uneven advantage of other software developers among employees at Microsoft has lowered the appeal of working with the organisation.

Although Microsoft remains a major force in the global software business it is the newer companies providing services to the ever expanding users of the internet that are coming up with innovative ways to motivate employees and to sustain interest in the workplace.

The introduction of the personal computer in the 1970's has revolutionised the way we live and work. Information has been travelling almost at the speed of light throughout the world which has empowered those of us living in this time and age in ways that previous generations will find astonishing.

The companies studied in this research paper have all contributed greatly to the revolution of empowerment by information which individuals have today.

In almost equal measure these companies have also been pioneers in introducing new methods of motivating employees.

Microsoft was the first organisation in the world to empower its employees financially through generous stocks which made millionaires in thousands within a decade. This alone was not sufficient to motivate its employees to remain in the organisation as some left to invest the money made in other ventures while others just quit in pursuit of other interests.

The organisations that were formed after Microsoft realised that in order to motivate employees it will be necessary to be more creative than just fulfilling the basic needs of staff through generous financial rewards.

Philip Rosenzweig argues in his book the halo effect…and the eight other business delusions that deceive managers that most of the evidence used in research such as this one to measure motivation and consequently the success of companies cannot be relied upon due to the "delusion of single explanations".

The delusion of single explanations often times leads to wrong conclusions about what drives organisations to succeed. Some studies have shown that a single factor for example company policy towards its social responsibility can lead to high performance. Whereas in actual fact many factors are interrelated and any given factor could have had a much smaller effect on motivation than a research study indicates.

There is a need to be careful about jumping to conclusions on the greatest motivating factor in an organisation because if conditions suddenly change in the stock market and there is a slide in the market capitalisation of an organisation what will happen to the morale of its employees?

Nothing is certain and motivation cannot be tied to one or two factors alone, this research has focused heavily on employee satisfaction and company performance of three of the best dotcom organisations in the world.

Motivation strategies in organisations will remain different as long as there is competition and as long as what works in one company may not work as well or have the same effect at another company.

From this research we can infer that the following strategies have been adopted mainly for each organisation:

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Cisco: Equity theory mainly combined with the needs theory (McClelland's & Herzberg's Motivator-Hygiene).

Google: Equity theory combined with the needs theory (Maslow's Hierarchy of needs).

Microsoft: Needs theory mainly (McClelland's Needs for Achievement, Affiliation and Power & Maslow's Hierarchy of needs).

Using the results of this individual research and based on the current performance of Google as an organisation with a highly motivated workforce one can propose that companies should motivate its employees by applying the equity and needs theories.

For effective management we need to be careful not to make decisions based on the halo effect which are delusionary and can be misleading but be prepared to take risks based on sound strategy.

4 IMPACT OF INFORMATION COMMUNICATION TECHNOLOGY (ICT) ON MOTIVATION:

Motivation in global organisations remains a big challenge; there are no quick answers to the secrets of a well motivated workforce.

The findings from a survey conducted by the Creative group on two hundred and fifty high-level executives revealed that 30% of the respondents found "motivating existing employees" to be their biggest challenge.

Recruiting qualified staff came second as their most difficult challenge with 28% of the respondents.

Table 4. The result of the survey

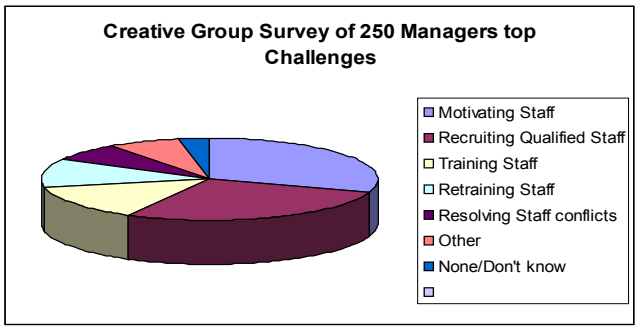
Challenge	Responses (250)	Percentage(%)
Motivating staff	75	30
Recruiting qualified staff	70	28
Training staff	35	14
Retraining staff	28	11
Resolving staff conflicts	17	7
Other	17	7
None/Don't know	8	3

From the above it is clear that modern technology will have to play a central role in enhancing motivation: in today's world ICT already plays an important role in peoples' lives and culture.

It is the means by technology has allowed information and communication to empower people and to break down previous barriers.

Several studies have shown that investing in ICT has a positive effect on employee motivation.

Figure 2.



Work has become more flexible due to the fact that more people can work from home or away from the office while in transit, thus flexibility has been noted as a key motivating factor in today's high tech world.

This flexibility also provides the opportunity to work with colleagues in different places and time zones thus enhancing teamwork and boosting productivity.

For the individual, flexibility mainly involves changes in the relationship between employer and employee and amongst employees, associated with the so-called 'death of distance' (Bradley & Bradley, 2002). This death of distance is directly linked to Maslow's hierarchy of needs in particular esteem and self respect leading to self actualization (ability to complete projects without undue pressure and supervision) and also Herzberg's Motivator-Hygiene Theory which include pleasant and comfortable working conditions.

ICT has brought crucial changes in our perception of time and space which in turn has lead to a necessary increase in trust between employers and employees.

This necessary trust in order to achieve company goals can best explain why recruiting of the best qualified staff available comes second in the survey of challenges faced by managers.

Garfinkel (1967) pointed out that trust is in itself 'unreflective and based on expectations of persistence, regularity, order and stability in everyday and moral world'. We have to take note that today's workplace structure is different from the workplace structure of the industrial age which was based on hierarchy and physical control; therefore the less physical space becomes, the greater the need for trustworthy relationships.

Managers will need to adapt using all the opportunities provided by ICT to remain firmly in control of projects and tasks.

Managers today can monitor the progress of tasks and assignments given to staff without the need for physical contact or closeness, thanks to the benefits of ICT.

The extensive use of emails, intranet, group work software, videoconferencing in organisations to share information rather than carrying hardcopies of data around has been

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

proven to increase employees motivation and efficiency but more research is needed in order to the balance the ratio of physical contact versus ICT use between employers and employees.

5 SUMMARY

This research has shown that Google has the highest motivated workforce among the three organisations studied but will not rush to conclude that it has a perfect formula for motivating employees.

Clearly it has benefited from one of the strong pillars of success which is being in the right place at the right time, in an age where the World Wide Web has enabled individuals around the world to have access to enormous amounts of information it was just a matter of time before an organisation came up with a user-friendly means of searching for information and it is exactly this gap that Google has filled.

Microsoft which came some years before Google also succeeded because it was in the right place at the right time.

Currently Cisco motivates its employees in a way that can be considered in-between what obtains in Microsoft and Google.

Money alone is not a good motivator but can demotivate if absent according to one of the need theories (Herzberg's Motivator-Hygiene).

Information Communication technology (ICT) has enhanced the art and science of motivation by providing effective alternatives to the old and traditional methods of motivating employees in organisations. In particular it has made it possible to motivate a very large number of employees in a global organisation without the burden of physical contact through e-mails and electronic-recognition schemes.

When an organisation is focused and is a specialist in a particular technology one of the strategies for success is to recruit employees with skills in the specialist area and to keep them focused by sharing the company vision with them consistently and also to allow individual creativity to flourish.

Outward performance of such organisations for example employee satisfaction, turnover and whether the company stock is rising or falling can lead researchers to make far reaching conclusions which can be misleading.

What is important is that organisations remain focused on its area of expertise with a dynamic management that is ready to take risks based on sound strategies.

It is also important to keep lines of communication open whether upwards or downwards in any organisation that hopes to be successful; this can be facilitated by the use of ICT tools.

Another area where the use of ICT has been found to be effective is the issue of trust, trust between organisations and employees and also between employees themselves is increasingly becoming a crucial factor in what makes an organisation successful or not.

6 CONCLUSION

In concluding this research study it is important to take note that there is no one magic formula that can cure all the motivation challenges that organisations face in bringing out the best in employees.

It is important for managers to provide a flexible work environment which can empower individuals to be creative and also to contribute positively to the overall goals of the organisation.

Finally as technology keeps advancing and improving our lives it is absolutely important to be dynamic, have a good strategy in place and be prepared to take risks.

REFERENCES

Birk, J. (2005), "The Microsoft Millionaires come of Age", The New York times May 29, 2005. http://www.nytimes.com (Last accessed 01/04/07)

Carey, P. (2001), "A Start-up's true tale (The Story of Cisco)", Mercury News Dec.1, 2001. http://www.mercurynews.com (Last accessed 29/03/07)

Cisco Corporate Citizenship report (2006), Cisco Systems, Inc. 2006

http://www.cisco.com (Last accessed 10-04-07)

Damme, M.V., Haan, J.D., Iedema, J. (2005), "Modelling a Multidimensional Concept: ICT-access at work", European Conference on the Knowledge Society and Changes in work, the Hague, June 2005.

Daniel, T.A., Metcalf, G.S. (2005), 'The Science of Motivation'

Donaldson, C. (2007), "Employee motivation a key driver in CSR", Human Resources magazine. April 14, 2007.

Harvard Business School Publishing Corporation (2005) Motivating People for Improved Performance

Lindner, J.R. (1998), "Understanding Employee Motivation", Journal of Extension Volume 36 Number 3 June 1998.

Lepper, M.R. (1998), "Motivational considerations in the study of instruction. Cognition and Instruction" 5(4).289.309

Rosensweig, P (2007), 'The halo effect … and eight other business delusions that deceive managers' New York Free Press.

Smith, G.O. (1994), Motivation in W. Tracey (ed.), Human resources management and development handbook (2nd ed.).

Using ICT to make work flexible, http://www.beepwork.com (Last accessed 23-03-07)

Who says Cash is King?, The importance of non-cash recognition in building motivation across a global organisation. http://www.globoforce.com (Last accessed 15-04-07)

Why is Google so great?, http://www.greatplacetowork.com (Last accessed 16-04-07)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

The impact of Internet marketing banks in Tanzania

Happiness Joseph Mbuna, Ali Alao Babatunde

School of Computing & Technology,
University of East London
Docklands Campus, University Way
London E16 2RD United Kingdom
u0615040@uel.ac.uk u0110353@uel.ac.uk

Abstract The present rivalry in the international market place, an internet presence to identify, and satisfy customers needs is extremely important for a bank to make profit because customers' have several alternatives in their finger tips, switching from one bank to another just require the customers' name and the rest would be done by the competing bank. Identifying and providing services to completely satisfy these needs should be the focus of the bank.

The rapid growth of the internet have revolutionized the way businesses are been conducted and similarly enhanced the viewpoint of customers in term of the service to be received thereby forcing banks to improve their services because supremacy has been shifted from service provider to the service receiver which happens to be the customer themselves.

This research will be focusing on banks in Tanzania and will be discussing ways by which internet marketing could help these banks in identifying and pleasing their customers' needs and wants so as to be lucrative. The benefits of internet marketing to the banks in the country would be explained, growth strategies that could be used by these banks with internet marketing to expand using matrix, and the hindrances facing these banks from using internet marketing in the country.

It is a well known fact that identifying and satisfying customers wants and need is important for success, and with the accuracy and timeliness of the internet, the banks in the country are enabled to get hold of information required to identify the needs of their customers and as well channel their resources towards providing and satisfying those needs for their customers because acquiring potentially new customer is known to be more expensive than maintaining an existing customer who is already pleased with the services he or she is receiving from the provider of the service.

Beyond all, this study identifies the contribution of internet marketing in the provision of enhanced and better services by banks for their customers.

1 INTRODUCTION

The internet is an international information path comprising numerous interconnected computer networks, reaching out to very many users all over the world while marketing is the managing process that is responsible for identifying, analysing, and rewarding customers wants and needs satisfactorily which involves creating new customers and building link with the existing / new customers by completely satisfying their needs and wants through the production of goods customers services that meet their needs. According Preston D.ICT Handbook (2007) define Marketing of art as a creating value for customers and encourage purchases of its product or services such as product, price, promotion and distributions. Also involve communicating to customers what would be gained from these products and services and ensuring that these goods and services are accessible to the customers and available at reasonable price.

Internet marketing on the other hand is the use of internet and related digital technologies to attain marketing objectives and sustain the modern marketing concept. These technologies include internet media and other digital media such as wireless mobile, cable and satellite media. (Chaffey D, 2003)

Internet marketing first began in the 1990s as a simple, text-based websites that offered product information. It then evolved into advertisements complete with graphics. The most recent step in this evolution was the creation of complete online businesses that use the internet to promote and sell goods and services. (Wikipedia, 2007)

In this new era, where banks usually have an internet presence, i.e. having a website, the internet marketing activities performed would be referred to this website through the link from online promotion strategies such as e-mail marketing, search engine marketing, traffic analysis, link strategies, affiliate marketing programs, competitor analysis, cost per click,

banner advertising to acquire completely new customers and making available services to already acquired customers of the organisation.

2 SCOPE AND DEFINITION OF INTERNET MARKETING

As mentioned previously that internet marketing is the use of interconnected computer networks to communicate to customers online of the availability of products and services for sale. This involves making use of online communication methods such as e-mail to inform customers or users of services and product availability, sorting customers' queries, and taking customer orders electronically. Online marketing can also be used of internet to pass information into the global market, i.e. richness of marketing communications with the global market, promote products and services, market and distribute to the global market.

Internet marketing is part of electronic commerce which can include information management, public relation, customer service and sales. Internet marketing has become popular as internet access is becoming more widely available and used. With the fact that are said to have been done using the internet (See Table 1).

Table 1. Impact of unique features of e-commerce technology on marketing

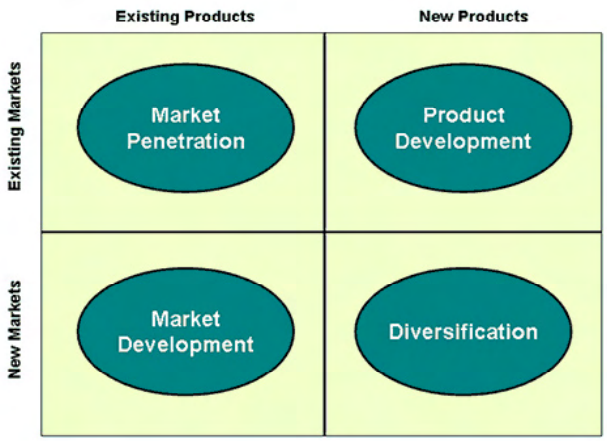
E-commerce technology	Significance for marketing
Ubiquity	Marketing communications have been extended to the home, work and mobile platforms. Geographic limits on marketing have been reduced. Customer convenience has been enhanced and cost reduced.
Global reach	Worldwide customer service such as marketing communication and messages.
Universal standards	Cost for delivering marketing messages and feedback as been reduced because of shared, global standards of the internet.
Information density	Highly detailed information on consumer, "data mining" internet tech permits the analysis of terabytes of consumer data everyday for marketing purposes.
Customization	Feature enables product and services differentiation down to the level of the individual, ability and strengthening.

Limitations of internet marketing create problems for both companies and consumers.

1. Slow internet connections can cause difficulties.
2. If companies build overly large or complicated web, internet users may struggle to download the information.
3. Internet marketing does not allow shoppers to touch, smell, taste or try-on tangible goods before marking an online purchase.
4. Another problems is internet market bank doesn't protect customer thought safety because some customer using there internet on the public net which is not safe the may put a software that recording customers information.

The marketing opportunities available to these banks for using the internet marketing can be appreciated by applying the marketing strategic grid. (Chaffey D, 2003). The internet can be used to achieve the four strategic directions using Ansoff product / market Matrix below.

Figure 1. Ansoff product / market Matrix



It involves increasing sales of these banks' existing services in their existing market(s). E.g. Encouraging customers to pay their credit card bills online (a credit card bill online payment might be encouraged by the bank by offering lower rates of longer period for their repayment); or persuading a new customer to apply online (by offering free gift with new credit card account open online). (CIM Study Text, 2003)

Market penetration can be achieved by advertising the banks services over the internet, increasing the service awareness and the profile of these banks among potential customer in an existing market. (Chaffey D, 2003)

It involves the expansion of these banks services into a new market using their existing services. New market may be different geographically, or new uses of existing product or service. These banks could encourage foreign customers to open a business account with them so as to be able to transact businesses from their home country without needing to visit the branch of these banks in the foreign country. Most of the transactions made by the customers are on the internet e.g. transferring funds from their account to their clients. (CIM Study Text, 2003)The internet is used to sell to new market, taking advantage of the opportunity provided by the internet without the necessity of visiting the bank branch.

It entails redesigning or repositioning of the banks existing services or the introduction of completely new service so as to appeal to existing market. Like the introduction of switch cards where a customer would upgrade his account from a current to a switch account in which the account holder or customer will be opportune to transfer fund from his account to any of his clients or partners or the development in the mortgage market, e.g. illustrate product development

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

as the traditional standardized mortgage account is rapidly being supplemented by variant which offers low rates, special terms for particular. (CIM Study Text, 2003)New products and services are developed which are delivered by the internet. These can be information products such as reports which can be purchased using electronic commerce. (Chaffey D, 2003). This involves moving into areas where these banks has little or no experience. New products are developed and are sold into new markets. This involves a high risk and would only be pursued by these banks if profitability is guaranteed. It is innovative. (Chaffey D, 2003)

3 INTERNET MARKETING STRATEGIES

Internet marketing is the use of techniques, strategies or tactics to promote a business, product, service online; and the most important part of the promotion is the website while others like e-mail marketing and affiliate marketing etc, may be considered as part of the entire online marketing strategies. Well structured online marketing strategy could bring more visitors to a banks' website which would be targeted and the target is likely to be interested in the banks services. When marketing and advertising online, the bank has to choose a sound internet marketing strategy. Choosing the right internet marketing strategy is quite important for the bank to achieve the very best result for the online marketing.

There are many online marketing strategies available for the banks to choose from and these include cost per click, banner advertising, and pay on page placement, search engine marketing and e mail marketing. An internet marketing strategy may be one or more of the followings which are examples of internet marketing strategies and techniques:

- E-mail marketing
- Search engine marketing
- Traffic analysis
- Link strategies
- Affiliate marketing programs
- Competitor Analysis
- Cost Per Click
- Banner advertising

This will involve the promotion of the banks site thereby getting accurate information on the site visitors' activities. Search engine optimisation is a process of getting the banks' website high as possible within the search engine ranking. This online marketing strategy is cost effective and perhaps the first in internet marketing strategy. Search engine ranking is assessing the performance of the banks on the search engines and directories. A report could be made to help measure the effectiveness of existing campaigns and identify which internet marketing activity is most effective for the bank.

3.2 Paid promotion

The bank would have to set up and manage a pay par visitor campaign so as to create and maintain their presence on the major search engines and directories, which is particularly effective for searches on niche market terms or for short term promotions.

3.3 Traffic analysis

This involves analysing the banks' website visitor activity to improve conversion rates and to improve website usability. Traffic analysis is essential to measure the effectiveness of internet marketing.

3.4 Link building

This involves sourcing and acquiring high quality incoming links to the banks' website to attract more targeted visitors by the bank and also to improve link popularity, which can enhance search engine prominence.

3.5 Affiliate marketing

Management of affiliate marketing campaigns is to implement and increase sales of the banks' services, to prominence of the banks' own website and others things.

3.6 Costs per Click

With cost per click technique, these banks will only pays for each click their website listing receives from search engines such as Google.com. One of the big benefits of these forms of internet marketing is that these banks would effectively target their online audience.

This is because these banks choose which search-terms or keywords they want their website listing to appear under. The only downside to this online marketing strategy is that these banks have to pay for each click their website receives which does not guarantee that any click will turn into a sell. Therefore the conversion rate of click to sales is pretty poor

Banner advertising is another online marketing strategy applicable to these banks. With this online marketing strategy the banks will only pay search engines or web directories to place a banner at the top of their search pages once a particular search term is typed

Table 2. The Consumer Decision Process and Supporting Communication

Market Communication	Awareness need recognition	search	Evaluation Alternatives	purchase	Post-purchase Behaviour Loyalty
Online communication	Targeted banner, interstitials and promotions	Search engines, online catalogs, site visit and e-mail	Search engines, product information and user evaluations	Promotions and discounts	Communication, Customer e-mail and online updates

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

into the search. One of the main advantages of banner advertising is that the banner will always appear for a particular search term and stay at the top of the page of the banks' website.

If a banner is produced with detail and professional image, it can also be attractive for people to click through to the banks site. It is not always effective to choose a creative banner (moving banner) when carrying internet marketing promotions, it all depends on the banks' service and what the bank want from the end user. Another good benefit of this online marketing strategy is that these banks can sometimes take out a cost per click system on the banner so that they only pay for the clicks on the banner as opposed to a set monthly or annual fee.

This is the use of appropriate technique to gather information from the banks site visitors and then managing the mailing lists collected, using e-mail as an effective internet marketing tool. A major strength of email campaigns is that delivery is prompt, with outcomes measurable within days. Targeted email marketing could be extremely effective if based on permission with a direct link with the banks site. In the creation of such campaign, it is absolutely necessary to put conversion into consideration i.e. the percentage of customers performing a desired marketing activity, like completing a transaction or filling out a form.

4 PURPOSE OF INTERNET MARKETING

With the internet, customers' needs could be acknowledged through the conduct of a research over the internet to determine how the bank would utilise its' resource to satisfy this need. i.e. the banks are able to identify what customers out there need through customer information obtained over the internet and could make provisions for such needs. The banks have to build an online brand equivalent with there offline reputation and there products and services availability to the customers.

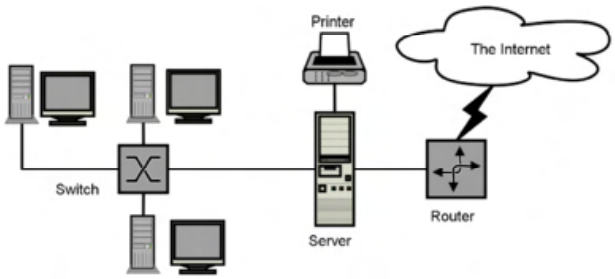
Likewise, the provision of additional channel through which customers would be enabled to assess information of their banks' services thereby making use or purchasing such services. Furthermore, internet marketing can also be used to recognize customer's reactions by getting feedback from online research asking customers for their comment about the way they receive the banks services over the internet with questions like:

- Is the site easy to use?
- Does the site perform adequately?
- Are the products and services properly dispatched?
- Do you feel secure use the internet banks?
- Does information satisfy customers?

So understanding the opportunities that could be gained for using internet marketing, the bank can then apply online marketing activities to implement their internet marketing activities. The followings are the opportunities that the banks can take advantage of using online marketing to implement their marketing activities.

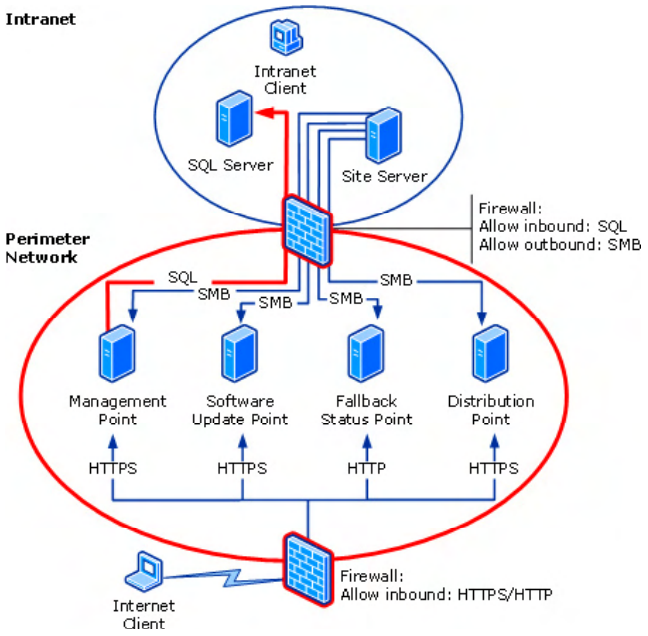
1 Sell: - In this situation, the banks are able to get customers over the internet as well as retain those customers thereby extending their customer base.
2 Serve: - Through the creation of online services with added value by the banks, there customers have access to service that is available to them at any time of the day.
3 Learn: - The banks are able to identify their customers buying behaviour and habit as well as determine what customers buy, learn about their references through tracking i.e. what are the content the customers are most interested in.
4 Save Cost: - Serving customer online is extremely cheaper than customer visiting the bank branch. An extremely reduced rate is required to serve banks customers online without charges like rent, lighting, or staffs' physical presence
5 Build: - Customers would be aware of the banks services online. The bank is able to extend the brand online, build its' image, and maintain relationship with its customers.

Figure 2. Customers diagram for using the internet.



Source from: http://en.wikipedia.org/wiki/Network_diagram

This network diagram shows server placement firewall configuration to supported internet-based client management.



Source from: http://www.microsoft.com/technet/prodtechnol/sms/smsv4/smsv4_help/

Banks in the Tanzania are examining themselves in term of their organisation, processes and structures in the rapidly

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

changing environment locally and globally, so as to be more competitive, engaging adequately in activities requiring the use of the internet. The following are the competitive advantages of banks and benefiting for using internet marketing:

The banks are using e-mails, teleconferencing internet calls to communicate with their staffs and partners to reorganize their business, i.e. improving the processes involve in the delivery of their services. It is a well known fact that the internet is the fastest communication network and all the banks have taken advantage of it in their business process reengineering using internet marketing.

In addition, the banks are able to get recent and accurate information about potential market and the industry and as well been able to have a competitive edge within the industry. Learning what other banks are doing and introduce different services form other banks, are able to benchmark or set a standard higher than any other bank or looking for means of innovating to provide better services for their customer. With this kind of information available to them, they are able to identify potentially new markets and as well create and maintain a competitive advantage that is be enhance through their connection to the Internet.

Likewise, the banks use the Internet in the search for best ways. To be profitable in the banking industry, these banks have to be creative and innovative to create and maintain a competitive advantage. I.e. modifying existing practices or finding new means of servicing customers that could help to improve their activities. And with the interactive nature of the internet, the banks are able to share and understand their customer's need which is critical for the identification and satisfaction of customers' needs and wants.

Moreover, these banks use communications abilities of the Internet to engage in internal marketing plan using intranet. These banks use the Internet to maintain process control across all their business locations which might include those in foreign countries. These banks use the Internet to search and develop a global practice for their corporate and service improvement. An understanding of the banking industry's current state gave them a competitive advantage because they are able to determine their position in the industry and as well set a standard above others in the industry. In addition to that, these banks gained a competitive advantage because they have access to accurate information on services, processes or new ideas.

There are questions do banks ask themselves such as:

1. What are other banks doing?
2. What are the kinds of information these banks are having?
3. Who is the main competitor and possible competitors within the industry?

Through internet marketing which is a very good source of information for the banks, they are able to know where and how to compete within the industry. Most of these banks use the internet to watch out for emerging and new technologies, and the market response to these technologies.

Moreover, the information from public discussions over the internet provides an idea of what's going on in the industry to these banks and the conclusions of such discussion is more like a result of a research to the banks which is quite difficult to obtain in through other means. In addition, people from different industry researches and exchange information on marketing issues relating to services provided by banks as well as technological developments i.e. emerging technology, internal processes like customer relationship management or external activities such as public relations.

5 BENEFITS OF INTERNET MARKETING TO THESE BANKS

Internet marketing has had a large impact on several industries including banking, clothing, music, and airline. It has affected the banking industry more because more and more banks now need to offer the ability to perform banking services online.Online banking is believed to appeal to customers because it is more convenient than visiting these banks' branches.Online banking is now the fastest growing internet activity.

The increasing speed of the internet connection is the main reason for the fast growth. Of those individual who use the internet, 45% now perform banking activity over the internet. (Wikipedia, 2007)Some of the benefits associated with the internet marketing for these banks include the followings:

5.1 Availability of Service Information

The customers of these banks can log onto the banks website and learn about their services are, as well as buy them at any time. Banks using internet marketing also find it useful as they are able to save money due to reduction in their physical branch services. Overall, Internet Marketing can help expand from a local market to both national and international market place. (Wikipedia, 2007)

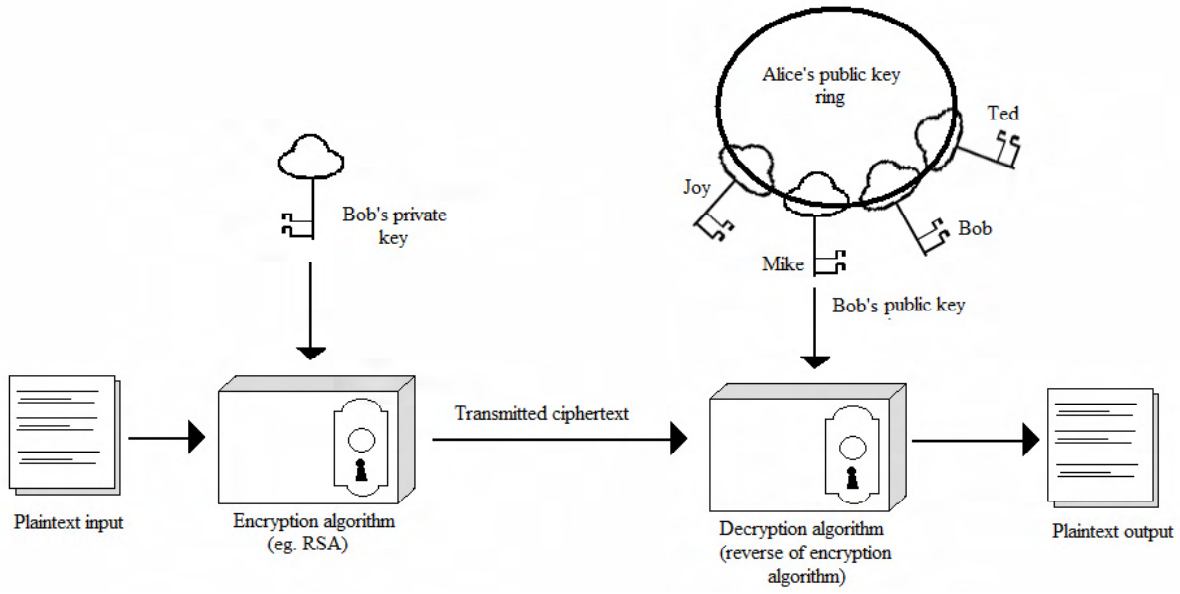
5.2 Research and customer service

The Internet is able to provide several needs of these banks in situations like marketing what is been revealed through research, customer service and the exchange of information. With the aid of the Internet, these banks can develop new services, take orders from customers, and receive electronic publications and documents.

5.3 Convenience

Another thing to consider about the internet is how easy and convenient it is to use by customers. If it is difficult to use by the customers, it wouldn't have been an ideal channel to communicate with customers but because of its ease of use and convenience, these banks are opportune to make use of internet marketing to reach their customers.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics



5.4 Cost Saving

It is less expensive to send and receive information from millions of people online and these banks would definitely benefits from this because they could conduct an online research to find out about their customers service satisfaction and this is quite easy and fast to conduct unlike the traditional survey with paper questionnaire but in this situation, their customers could send their comments about the services they are receiving or respond to questionnaire at a much more reduced cost for these bank.

5.5 Global Coverage

For these banks, there is the possibility of reaching out to the global market place by developing a virtual stall and having their presence on the Internet, it is absolutely possible for these banks to reach customers in tens of millions of people at an extremely reduced cost. They would be able to with foreign big banks on the internet and even customers won't know how big or small these banks are, but would still be servicing their customers profitably.

6 HINDERANCES FROM THE USE OF INTERNET MARKETING BY THESE BANKS.

Knowing fully well that these banks are located in a developing country, there are some problems that faces the country which will definitely have an effect on these banks as well. The following are problems face banks in Tanzania. Due to the instability of electricity in the country, most of these banks local customers found it difficult to benefit from the opportunities provided by internet marketing for these banks. Their customer could not have internet access due to power shortage that happens frequently in the country.

Moreover, the literacy level of the local customers does not really encourage the use of internet marketing practices or activities. These banks might perhaps be able to reach many of there customers through internet marketing but be unable to obtain information about there needs or wants because giving personal information or responding to research questions is not widely practiced especially in the northern part of the country where a few happens to be literate.

The cost and skill required for the maintenance of internet marketing might not available by these banks. Internet marketing requires a technical skill for it's maintenance because the information on the internet need to be updated regularly and if these requires hiring an expert all the time, these banks will have to compare the cost as well as revenue generated from internet marketing (most banks do business to survive in the country and put little consideration on benefits) to determine its' effectiveness and compare it with their entire marketing activities. The cost of living in the country is high that quite a few of these banks' customer could afford to have a personal computer at home to access his account information at any time. It is believed to be unnecessary to use internet marketing by these banks as far as their customers don't have a round the clock access to the internet. Unlike in UK where a large percentage of the population is believed to have a round the clock access to the internet.

7 EFFECT ON THE INTERNET MARKETING INDUSTRY

Internet marketing has had a large impact on the industry necessitating it for these banks to perform banking services online. Although, internet marketing hindrances in the country might perhaps reduce this threat but for the growth of these banks, they need to meet up with the global standard to expand thereby leading to the followings.

Form above pic show that information can be transfare from one side to another,when the infromation transfare it's will be easy if the isn't security for a third part to track those information and misuse its.

7.1 Employment

These banks are forced to employ people with IT skills to create and maintain websites. A constant maintenance of the website is required because there customers expect to see

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

updated information on the banks site all the time likewise a high quality performance siteand also to keep the customer safe using internet.

7.2 Competition

Moreover, the introduction of internet marketing had led to an intense competition among these banks because these banks will have to prove there superiority on other banks to their customers that they meet up with the global standard by having a web presence. This help builds reputation and courage that their customers have in them.

7.3 Reputation

Having a web presence is a means of building trust and confidence in customers mind by these banks. Although this is more a psychological thing because most of these banks site has nothing to offer there customers online, customers still have to go to there bank branches for services.

7.4 Corruption

Likewise due to poverty that led to corruption in the country, using internet marketing will require a high security to be place to secure customers information because most of their customer engage in advance fee fraud normally referred to as "419" in the country and they are ready to make use of another persons detail to make purchases on the internet and this implies that these banks would need a security measure that would be able to confirm authenticity of ownership and transaction.

Table 3. Security Breaches and Requrement to Prevent in the Internet Banks.

Security invasions/ breaches	Security requrements
User/Password	Confidentiality
Trojar horses	Integrity
Password breaking	Authenticity
Denial of service attacks	Non-repudiation.
Packet sniffing on net	
Spooning websites	
Dumpster diving	

Above all, internet marketing had led to many corrupt practices in the industry whereby information about customers' details are stolen or bought from internal source and sold to outsiders to make purchases. These outsiders are often called 'yahoo yahoo' because buy card or account information of the wealthy people to make purchases believing that making purchases with them will not be noticed by the owner.

8 CONCLUSIONS

The Internet marketing has provided new opportunities for these banks to change their path and excel. Currently, these banks are trying to explore the value of the internet marketing and still trying to comprehend its benefits so as to integrate it their marketing activities for their success. There might be several hindrances in the use of internet marketing for these banks but due to the fact that the opportunities created by it cannot be underestimated .i.e. considering there presence in the global market place, they will find means of reducing these hindrances so as to make use of the opportunity available to them for using the internet.

It is the most marketing channel of 21st century and almost all banks if not all have internet presence. Integration of online marketing activities with offline marketing activities is widely used by banks to determine the effectiveness of their marketing performance. So being a strong marketing tool, if fully implemented, these banks would be able to communicate their marketing messages with the use of expressions that is globally accepted and understood by customers, hence promoting their brand and as well as marketing themselves and their services which would be supported by offline promotions and advertising locally and globally.

Even thought banks a highly on the successfully in performance but still need internet strategy that will use the short-term and long -term payoff. This exploration on surface of internet bank which give service though internet technology and support on conveniences, cost savings and the fact that the banking industry as a whole has emerged as a leader when it comes to internet security-many consumers, remain doubtful about the security risk and other perceived complexities of putting their financial lives online.

REFERENCES

1. Preston D., (2007) global sustainable information and communication technology management. University of East London.
2. Chaffey D., Mayer R., Johnston K., Ellis-Chadwick F.: Internet Marketing: Strategy, Implementation and Practice: (2003): Pearson Education Ltd Essex England.
3. Wikipedia (2007) Internet Marketing http://en.wikipedia.org/wiki/Internet_marketing(Accessed 22/03/07
4. Chaffey D., Mayer R., Johnston K., Ellis-Chadwick F.: Internet Marketing: Strategy, Implementation and Practice: (2003): Pearson Education Ltd Essex England.
5. http://www.tutor2u.net/business/strategy/ansoff matrix.htm(Accessed 21/03/07)
6. CIM Study Text (2003) Advanced Certificate (Stage 2) Paper 6: Marketing Planning London BPP Professional Education.
7. Digital Explosion-Glossary (2003) the future of marketing communication. http://www.denow.com/6gloss/ (Accessed 28/03/06)
8. http://www.microsoft.com/technet/prodtechnol/sms/smsv4/smsv4_help.
9. Intense Development (2006) Intensify your business http://www.intensedevelopment.net/website-design-O.html (Accessed 24/03/06)
10. Pearson Education (2004) Glossary http://wps.pearsoned.co.uk/wps/media/objects/1452/1487687/glossary/glossary.html#I (Accessed 22/03/07)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

11. Sylvia Kantor (1998) Internet Marketing for Farmers http://www.metrokc.gov/wsu-ce/agriculture/PDFs/internetMrkt.pdf (Accessed 29/03/07)
12. Wikipedia (2007) Internet Marketing http://en.wikipedia.org/wiki/Internet_marketing(Accessed 22/07/07)
13. Wikipedia (2006) Online Marketing http://en.wikipedia.org/wiki/Online_marketing (Accessed 21/03/07) Bureau of National Affairs, Inc., The, 1992, Daily Report for Executives, Special Supplement, April 3.
14. Carlton, D., 1986, "The rigidity of prices," American Economic Review. Vol. 76, No. 4, pp. 637-658.
15. Carlton, D., and J. Perloff, 1989, Modern Industrial Organization, Glenview, IL: Scott Foresman/Little, Brown.
16. Cornett, M., and H. Tehranian, 1992, "Changes in corporate performance associated with bank acquisitions," Journal of Financial Economics, Vol. 31, April, pp. 211-234.
17. Cowling, K., and M. Waterson, 1976, "Price-cost margins and market structure," Economical, Vol. 43. August, pp. 267-274.
18. DeYoung, R., 1997, "Bank mergers, X-efficiency and the market for corporate control," Managerial Finance, Vol. 23, No. 1, pp. 32-47.
19. Di Salvo, J., 1997. "Federal Reserve market definitions methodology by District," Casework Conference, Federal Reserve Bank of Philadelphia, October 23-24.
20. Gelfand, M., and Spiller, P., 1987, "Entry barriers and multi product oligopolies," International Journal of Industrial Organization, Vol. 5, March, pp. 101-113.
21. Hannan, T., 1997, "Market share inequality, the number of competitors, and the HHI: An examination of bank pricing," Review of Industrial Organization, Vol. 12, February, pp. 23-35.
22. Hannan, T., and A. Berger, 1991, "The rigidity of prices: Evidence from the banking industry," American Economic Review, Vol. 81. No. 4, pp. 938-945.
23. Hannan, T., and J. Liang, 1993, "Inferring market power from time-series data. The case of the banking firm," International Journal of Industrial Organization, Vol. 11, June, pp. 205-218.
24. Holder, C., 1993a, "The use of mitigating factors in bank mergers and acquisitions: A decade of antitrust at the Fed," Economic Review, Federal Reserve Bank of Atlanta, March-April, pp. 32-44.
25. Hughes, J., W. Lang, L. Mester, and C. Moon, 1996, "Efficient banking under interstate branching," Journal of Money. Credit, and Banking, Vol. 28, No. 4, Part 2, pp. 1045-1071.
26. Iwata, G., 1974, "Measurement of conjectural variations in oligopoly," Econometrica, Vol. 42, September, pp. 947-966.
27. Jackson, W., III, 1997, "Market structure and the speed of adjustment: Evidence of non-monotonicity," Review of Industrial Organization, Vol. 12, February, pp. 37-57.
28. Kravitz, P., 1997, "Antitrust policy in banking: Comments," in Proceedings of a Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, pp. 180-183.
29. Kwast, M., M. Starr-McCluer, and J. Wolken, 1997, "Market definition and the analysis of antitrust in banking," Antitrust Bulletin. Winter, pp. 973-995.
30. Litan, R., 1994. "The ABC of Justice's antitrust assessment of bank acquisitions," Banking Policy Report, Vol. 13, No. 10, pp. 16-20.
31. Neumark, D., and S. Sharpe, 1992, "Market structure and the nature of price rigidity: Evidence from the market for consumer deposits," Quarterly Journal of Economics, Vol. 107, May, pp, 656-680.
32. Peristiani, S., 1995, "Do mergers improve the X-efficiency and scale efficiency of U.S. banks? Evidence from the 1980s," Federal Reserve Bank of New York. working paper.
33. Pilloff, S., 1996, "Performance changes and shareholder wealth creation associated with mergers of publicly traded banking institutions," Journal of Money, Credit, and Banking, Vol. 28, No. 3, pp. 294-310.
34. Prager, R., and T. Hannah, 1998, "Do substantial horizontal mergers generate significant price effects? Evidence from the banking industry," Journal of Industrial Economics, December.
35. Radecki, L., 1997, "The expanding geographical reach of retail banking markets," Federal Reserve Bank of New York, mimeo.
36. Rhoades, S. A., 1995a, "Market share inequality, the HHI, and other measures of the firm composition of a market," Review of Industrial Organization, Vol. 10, No. 6, pp. 657-474.
37. Roberts, M., 1984, "Testing oligopolistic behavior," International Journal of Industrial Organization, Vol. 2, December, pp. 367-383.
38. Shaffer, S., 1993a, "A test of competition in Canadian banking," Journal of Money, Credit, and Banking, Vol. 25, No. 1, pp. 49-61.
39. Shaffer, S., and J. Di Salvo, 1994, "Conduct in a banking duopoly," Journal of Banking and Finance, Vol. 18, December, pp. 1063-1082.
40. Simmons, K., and J. Stavins, 1998, "Has antitrust policy in banking become obsolete?," New England Economic Review, March/April, pp. 13-26.
41. Spiller, P., and E. Favaro, 1984, "The effects of entry regulation on oligopolistic interactions: The Uruguayan banking sector," The Rand Journal of Economics, Vol. 9, No. 2, pp. 305-327.
42. Suominen, M., 1994, "Measuring competition in banking: A two product model," Scandinavian Journal of Economics, Vol. 96, No. 1, pp. 95-110.
43. Whalen, G., 1995, "Non-local concentration, multitasked linkages and interstate banking," paper presented at the Conference on Antitrust and Banking, Office of the Comptroller of the Currency, November 16.
44. Wolken, J., 1984. "Geographic market delineation: A review of the literature," Board of Governors of the Federal Reserve System, staff study, No. 140, pp. 1-38.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

i·Society 2007

Issues and challenges related to online shopping in Saudi Arabia

Inam Abousaber
Anastasia Papazafeiropoulou

Brunel University,
Department of Information Systems and Computing,
Uxbridge, UB8 3PH, United Kingdom

Ziad Hunaiti

Anglia Ruskin University,
Department of Computing/Design and Technology,
Chelmsford, CM1 1SQ, United Kingdom

Abstract This paper presents the outcome of a study carried out to identify the challenges and barriers hindering the development and progress of e-commerce in the Kingdom of Saudi Arabia (KSA). Questionnaires have been distributed to Internet users, which have been designed to address all aspects that might affect e-commerce and to reflect the real situation in the KSA. The outcome from analysing the survey results has been used to identify the challenges and barriers facing the development and progress of e-commerce in the KSA. The study showed that the vast majority of the KSA citizens are keen to use online shopping. In addition, the results reflected the real problem holding the e-commerce growth in the kingdom, which includes: Internet pricing is still high, Internet security is still seen as not secure enough, legal system does not govern e-commerce, having Arabic websites could encourage online shopping and internet filtration is a major obstacle.

Keywords E-commerce, Online Shopping and Saudi Arabia

1 INTRODUCTION

In the mid-1990s the world saw a substantial growth of the Internet. Such a growth has never been witnessed in the history of information and communications technology. The Internet is an interconnection of computer networks linked by a communication medium such as cable and satellites. The Internet idea started with connecting four supercomputers, nowadays internet links over 300 million machines spread in 170 countries around the world [1].

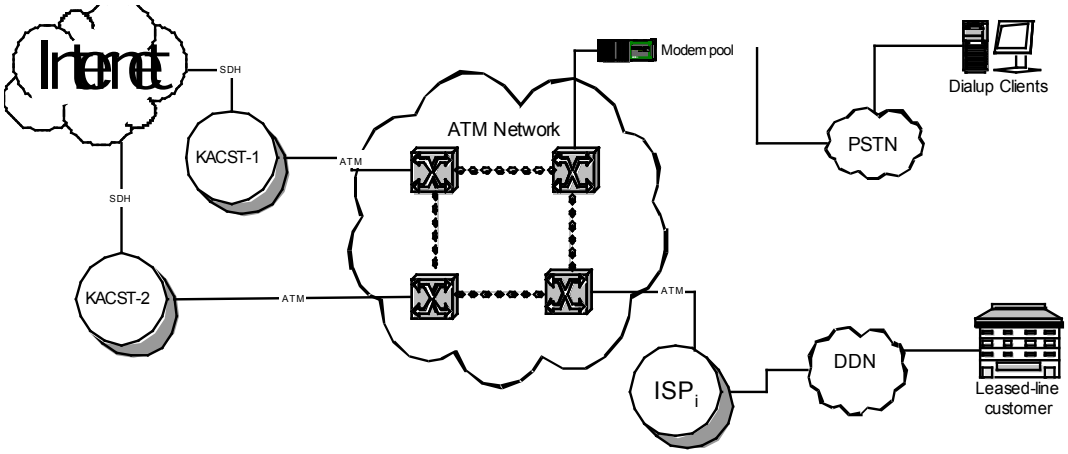
Internet's contribution has transformed people's lives forever and its significance is clearly seen by its impact on industries. One of the main sectors, which has been directly influenced by the introduction of the Internet, is the business industry. The Internet has changed the face of businesses and the way people are doing business. A new form of trading over World Wide Web (www) has become a main competitor to the traditional way of trading. As a result of that, the birth of what is called e-commerce (Electronic commerce) emerged, resulting in rapid expansion in the applications supporting it [2].

The KSA is a wealthy country, located at the heart of Arab World with over 27 million in population and more than one-quarter of the world's oil reserves i.e. largest oil producer. It is also amongst the twenty strongest economies in the world [3-4].

The KSA boasts a conservative society, where things like nudity, drugs and any substances that might contaminate ones mind is prohibited according to the shariah (Islamic law). This is being one of the key reasons why the initiation of the Internet service in the KSA was delayed until 1998 [5]. The ban was lifted after a decision made by the Council of Ministers' in March 1997. This allowed the use of Internet services under supervision of an elected body. The council allowed the use of Internet with limitations to its users. The elected body observes the limitation by restricting access and filtering any material/content that goes against the Islamic belief and the values of Saudi society.

The birth of e-commerce has a greater impact in shaping the global economy and contributed in increasing of innovations in business practices, and transforms their impact from local to become global [2]. According to the IT finance newsletter Computer Economics, it was expected that the worldwide e-commerce activities volume to grow from $5,520 billion in 2001 to $11,999 billion in 2004 [2] For the Arab world the overall e-commerce activities have been estimated to be $3 billions for year 2000 and 2001, and by 2005 it was ex-

Figure 1. Internet infrastructure in the KSA [10]



pected to grow by $500 million. E-commerce activities in the Gulf region1 have been estimated to be $1.9 billion in 2001. The KSA being the big player here generated the most e-commerce activities i.e. $1 billion of the total e-commerce activities in the Gulf region in 2001[4].

Internet service in The KSA is operated by three bodies: Internet services unit at KACST2, Saudi Telecom, and ISPs. The Internet unit at KACST manages and controls the outer world connections, where the national Internet network is linked to the global network; it monitors the Internet gateway and restricts sites with un-accepted contents. Managing and maintaining the telecommunication infrastructure in the country is the responsibility of Saudi Telecom. It connects subscribers to their ISPs, ISPs and the KACST core network, as well as KACST and the global network. Figure 1 illustrates a simplified diagram of the Internet infrastructure in The Kingdom of Saudi Arabia [6].

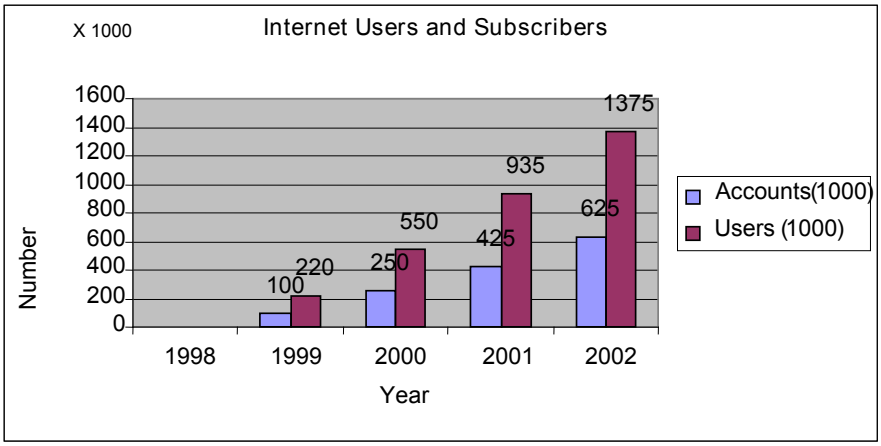
According to the KSA's Ministry of Communications and Information Technology, April 2003 report, the number of the internet users has shown significant increase between years 1999 and 2002, as shown in figure 2. The estimated number of Internet subscribers was 100 000 at the end of 1999. Subscribers considerably increased to 425 000 in 2001, and by the end of 2002 the numbers reached 625 000. In 2002, the number of users rose to 1, 375 000; this represented 6.41% of the kingdom's population. In addition, there were around 3500 digital subscriber lines and 2500 leased lines [6]. According to the growth experienced, it was expected that the number of the users will increase to 2 million by 2005 and over 5 million by 2010 [7].

Despite the significant growth in the number of Internet users in the KSA and the high demand in the IT products i.e. to become the world largest market with over 33% of its PC sales in 2001, there is no indication for potential growth of e-commerce [4]. In comparison to the worldwide size, e-commerce in the KSA is significantly low around $150 million [8]. There were limited studies in the field of e-commerce in the Kingdom of Saudi Arabia. Moreover, many issues have not been motioned in the pervious studies. Therefore, this study has been carried out to assess consumers' behaviour towards e-commerce in the KSA and to identify the issues facing the ecommerce adoption in the KSA from their point views.

2 METHODOLOGY

This section presents the research methods which have been used in the evaluation study conducted on consumers in the KSA. The main objective of this study is to assess consumers' behaviour towards e-commerce. It also identifies the main barriers facing the e-commerce industry growth in the KSA. Data can be collected using two main approaches namely quantitative and qualitative methods [9]. During this study two kinds of research methods have been used:

Figure 2. The Number of Internet accounts and users in the KSA [9]



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 1. Consumers' Attitude towards e-commerce issues in the KSA

Rating (Percentage %)					Questions
5 Strongly Agree	4	3	2	1 Strongly Disagree	
12	17	43	10	18	Q6: I do trust buying good over the internet
17	15	31	15	20	Q7: It is easy to do online shopping with current internet speed you use
25	25	18	17	15	Q8: Internet access cost in KSA is too high to be used for online shopping
12	18	26	22	22	Q9: Internet is secure enough to perform financial transaction
12	12	22	29	25	Q10: Internet is safe enough to give my personal information.
8	12	37	15	28	Q11: I think the existing legal system is strong enough to do online shopping
20	10	22	22	26	Q12: Postal services in KSA are suitable for the delivery of online-bought
43	28	15	7	7	Q13: It is important to physically examine goods before purchasing them (not digital items)
53	25	7	7	8	Q14: I prefer to go out for shopping for socialising and meeting people
30	27	20	13	10	Q15: Internet filtration in KSA discourage online shopping
32	30	20	13	5	Q16: Suitability of banking system in KSA for online shopping (does your bank provide you with Debit card, Credit card, etc)
42	18	17	15	8	Q17 Having websites in Arabic eases online shopping
37	27	23	3	10	Q18: I'm willing to do online shopping in the future

the first method was based on gathering secondary literature resources while the second was a survey conducted through collecting primary data through questionnaire. Jeddah city was the scope here, as it acts as a central market for most retailers in the KSA. The questionnaire has been distributed on sixty consumers. It has been structured based on the aspects adopted from previous studies [10-12], as well as the gaps identified in the literature review have been used to design the survey. The following factors have been included in this study:

- Computer literacy: this is to assess the knowledge of the participants using computer and Internet simply by asking consumers whether they have email or not.
- Access/availability, consumers will be asked if they have internet access and what type of connection they might have.
- On-line shopping experience, in order to evaluate how many consumers might have experience with e-commerce. They will be asked if they ever did on-line shopping.
- E-commerce trust, consumers will be asked if they trust trading over the internet.
- Internet speed, this question will be asked to consumers to reflect their point of view if the internet speeds impact e-commerce.
- Internet pricing, similarly, consumers will be asked if the internet access price is preventing them from practicing e-commerce.
- Internet security, if the internet is secure enough to perform online trading between both businesses and consumers, basically if the consumers can trust Internet security.
- Culture, this question will be asked to consumers if they like seeing and touching goods before they buy them. Another question will be asked if consumers prefer to do shopping for social reasons, these two questions can reflect if culture has impact on e-commerce.
- Legal regulations consumers will be asked whether the current legal regulations allow them to practice e-commerce.
- Internet filtration, in order to evaluate the impact of controlling the content of the internet by Saudi authority on e-commerce, consumers will be asked to give their opinions.
- Willingness for e-commerce, to assess whether consumers are willing to use e-commerce in the future or not is a question to be asked.
- Banking systems, a question will be asked to consumers if the current banking systems support the e-commerce.
- Postal services, this question is to obtain consumers' opinion about the reliability of the mail services for e-commerce applications.
- Arabic Web content, language of the might have an impact on e-commerce growth therefore, this question

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

will be directed to the consumers, in order to evaluate if it is a significant issue or not.

3 FINDINGS

This section presents the findings that were obtained from the questionnaire which were distributed and collected from sixty consumers (12 female and 48 male) based on Jeddah city. 50% of the participants held undergraduate qualifications, while 25% have postgraduate qualifications and the remaining 25% have high school education level. The findings from the questionnaire, where the issues related to the development of e-commerce in the KSA have been evaluated from the consumers' point of view are listed as followed:

Q1: Computer literacy

The study shows that 88% of the consumers have e-mail account.

Q2: Internet access availability

Also the survey shows that 85% of the consumers have access to the internet.

Q3: Type of the Internet connection

49% of the consumers have broadband DSL connection; 37% use dial up connection and the remaining 14% use leased line connection.

Q4: Place of Internet access

The result shows that 88% of the consumers access the Internet at homes, while 8% access internet from their offices and 4% depend on the internet cafés to have access.

Q5: Experience with online shopping

When the consumers have been asked if they ever used the internet for online shopping, only 23% answered Yes and 77% answered No.

The following table shows the results obtained from the remaining questions using Likert scale questions type allows the evaluation of the subject feeling towards particular topics by indicating how strongly they agreed or disagreed on the given statements.

4 DISCUSSION

The results showed that the vast majority of the participants are highly educated (75%), and over half of those with high school education have e-mail addresses increasing to a percentage of 88%. In addition, 85% of the samples have internet access. This will lead us to the conclusion that our sample has sufficient computer literacy which enables them to do online shopping.

As seen from the results, the majority of those having access are connected via broadband internet; 27 out of 51 are using DSL and 7 out of 51 using leased line; the remaining 19 participants still using dial-up connection. The results showed that there is a rapid increase in the number of DSL subscribers. Although the results showed that there is an increase in the number of home internet connections as 88% are connected from home. Therefore, availability and speed of the internet will not be a major issue for the future of e-commerce in the KSA. Despite, the study showed only 23% of the sample has experienced online shopping. The lack of experience with online shopping has made the participants reluctant, due to not having trust on buying goods over the internet. Lack of trust is one of the main barriers facing the internet in the Arab world, which has been also shown in the previous study [10].

Due to the fact that the participants are using different internet connection namely DSL, leased line and dial-up, that has affected their opinion about the suitability of the current internet speed in the KSA for online shopping, it can be seen that those using DSL or leased line do see it suitable for online shopping and majority of their votes were 4 or 5. On the contrary, those with dial-up connections see it is hard to do online shopping with their connection, therefore, the majority voted 1 or 2. This outcome does agree with similar study that has been conducted within the UK [11] which showed when people switch to DSL connection, the time that people spent on the internet has significantly increased, because DSL does allow access to larger data content with better speeds which it is faster than the dial-up connection.

Even with all the wealth of Saudi Arabia, internet access cost is still too high and beyond the purchasing power of average citizens [10] therefore, the vast majority see the current internet costs is too high to be used for online shopping; the current monthly cost for internet DSL connection is about £50 [15] while the average cost with the gulf region is £20 per month.

Consumers showed that Internet security and information privacy are still major issues facing the e-commerce in the KSA, which is similarly facing the Arab world according to the previous study [10]. Consumers' response to these two issues was high, mainly when it comes to financial transactions.

The absence of a legal frame work to govern e-commerce is one of the major obstacles facing the development of e-commerce in any country that does not have it and particularly in the KSA [16]. Legal framework has been rated as the second important issue in another study [10], similarly the vast majority of the participants in this study agreed with that.

In spite of the introduction of the new postal system in the KSA [17] and the efforts that have been done to make it suitable for delivering parcels for online shoppers, the majority of the consumers still not satisfied.

The study showed that there is an impact of the Saudi culture on e-commerce growth in the country that has been

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

confirmed through two points; the first one is that Saudis are concerned about buying goods over the internet without trying or touching them physically. Shopping is one of the activities in the KSA, where people can meet and socialise with each other. This is one of the main reasons stopping people from doing online shopping; it is also a unique issue within the Saudi society, due to the nature of the society as conservative society [18].

The control of the Internet contents by the KACST [6] has made itself a barrier for the advancement of the e-commerce industry within the KSA. This has been seen by the vast majority of the consumers as discouraging factor for online shoppers, while the filtration will not allow them to view some web pages they might be interested in.

The recent development in the banking system across the KSA, and the services provided to the consumers, seem to have a good impact on the e-commerce, consequently, most of the consumers are satisfied with their banks and see them as suitable for performing online shopping.

As it was reported by the previous study [10], about the impact of the web contents language on e-commerce growth in the Arab World; it has been also confirmed in this study by most of the consumers that having websites in Arabic eases online shopping.

The vast majority of the consumers who participated in this study showed a big interest, and willingness to do online shopping in the future.

5 RESEARCH CONCLUSIONS

The study has been conducted by evaluating all the aspects that might have impact on the e-commerce industry in the KSA, as a conclusion, the following issues have influenced the advancement of e-commerce in the KSA

- Internet pricing: As internet access cost in the KSA is too high and beyond the purchasing power of average citizens, it can be seen as a major obstacle for online shoppers.
- Internet security: Saudi society sees lack of Internet security and information privacy when it comes to perform financial transactions. Such fear has been emerged as a barrier facing e-commerce development in the KSA.
- Legal regulations: the delay in establishing a legal framework to govern e-commerce is one of the major obstacles facing the development of e-commerce in the KSA.
- Postal services: postal system services in the KSA is not suitable for delivering parcels for online shoppers; therefore, this becomes another issue facing e-commerce growth in the KSA.
- Culture: Shopping is one of the activities in the KSA, where people can meet and socialise with each other. This is one of the main reasons stopping people not to do online shopping. This issue is more obvious within the Saudi society than other societies in the same region. In addition, Saudis are concerned about buying goods over the internet without trying or touching them physically.
- Internet filtration: The control of the Internet contents by the KACST has contributed in discouraging both businesses and online shoppers. They impose restrictions on the contents of the web pages making it a barrier for the advancement of the e-commerce industry in the KSA. Another difficulty is not allowing the consumers to view some web pages they might be interested in.
- Arabic Web content: the study has shown that the language of the web contents could be seen as an obstacle for some shoppers. Therefore, having web contents in Arabic language will contribute to the advancement of e-commerce in the KSA.
- E-commerce trust: it has been noticed from this study that consumers do have doubts about buying goods through online shopping, which can be another barrier decelerating the development wheel of e-commerce in the KSA.

6 IMPLICATIONS FOR RESEARCH AND PRACTICE

There have been continuous efforts from the Saudi government to facilitate e-commerce industry in the KSA, However, it has been concluded from this study that the e-commerce growth in KSA is still very slow; that is mainly due to the obstacles mentioned earlier discouraging internet users. These barriers are mainly associated with the regulations governing the e-commerce industry in the KSA. Therefore, any solution to overcome these issues should be initiated by the KSA authorities. That could be achieved through two phases:

First Phase, Saudi authorities may need to establish a national framework which focuses on a well designed plan for addressing issues reported by this study and other related studies. In particular by addressing the major obstacle including:

- Internet pricing: through reducing the internet access cost to a level which makes it affordable by the vast majority of consumers that can be similar to the cost of the internet access in the region.
- Security: this can be achieved through adopting high security system similar to the one used by other countries with much advanced experience in the field.
- Legal system: legal framework should be improved to promote e-commerce in the KSA.
- Arabic websites: Businesses should launch websites in Arabic in order to make the online shopping easier for Saudi consumers.
- Internet filtration: Saudi authorities should look at the negative impact of the internet filtration on e-commerce and consider a plan to overcome this negative impact.

Second Phase: Saudi authority should be engaged with citizens to facilitate and encourage online shopping culture in the KSA. One step forward, Saudi government agencies

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

could be encouraged to shift a large part of their purchases online which can be very useful to enhance the e-commerce in the country.

Due to the limited time allocated for this research, the author tried to conduct it within Jeddah involving a sample of sixty consumers, in order to obtain more adequate results. The study could be conducted to include a large number of consumers.

In order to ease the e-commerce growth in the KSA, a national framework can be established to achieve this target. Such a framework can be successful if the previously mentioned issues such as, legal regulations, internet filtration, and Internet pricing are properly addressed. Solving these three main issues will contribute to relieving other issues.

REFERENCES

1. Vulkan, N. (2003), The economics of e-commerce: a strategic guide to understanding and designing the online marketplace, Princeton University Press.
2. Kaynak, E., Tatoglu, E., Kula, V. (2005), 'An analysis of the factors affecting the adoption of electronic commerce by SMEs: Evidence from an emerging market', International Marketing Review, 22(6), pp. 623-640.
3. The Kingdom of Saudi Arabia, (2007), Available from World Wide Web: http://www.infoplease.com/ipa/A0107947.html (Accessed 30 March 2007).
4. Moores, S. (2002), Information technology opportunities for British business in The Kingdom of Saudi Arabia an executive summary, Available from World Wide Web: http://www.egovmonitor.com/reports/SA_ExecSumm.pdf (Accessed 2 May 2006).
5. Al-Furaih I. (2002), 'Internet Regulations: The Saudi Arabian Experience', Proceeding of the Internet Society's 12th Annual INET Conference, 18-2 June, Arlington, Virginia, pp. 1-16.
6. ITU, (2003), 'Information and Telecommunication Technology in The Kingdom of Saudi Arabia', Available from World Wide Web: http://www.itu.int/dms_pub/itu-s/md/03/wsispc3/c/S03-WSISPC3-C-0025!!MSW- E.doc (2 May 2006).
7. ESCWA, (2003), 'National Profile of the Information Society in the Kingdom of The Kingdom of Saudi Arabia', Available from World Wide Web: <http://www.escwa.org.lb/wsis/reports/docs/SaudiArabia_2005-E.pdf (06 April 2007).
8. U.S. Commercial Service, (2006), Available from World Wide Web: < http://www.buyusa.gov/saudiarabia/en/130.html (01 March 2007).
9. Saunders, M., Lewis, P., and Thornhill, A. (2003), Research Methods for Business Students, 3rd ed., N.J.: Prentice Hall.
10. Aladwani, A. (2003), 'Key Internet characteristics and e-commerce issues in the Arab countries', Information Technology and People, 16(1), pp. 9-20.
11. Ahmed, A., Zairi, M., and Alwabel, S. (2006), 'Global benchmarking for internet and e-commerce applications', Benchmarking: An International Journal, 13(1-2), pp. 68-80.
12. Sait, S., Al-Tawil, K., and Hussain, S. (2004), 'E-commerce in Saudi Arabia: adoption and perspective', Australian Journal of Information (AJIS) systems, 12 (1), pp.54-74.
13. Gunina S. and Haj M. (2004), The Arab Book: Telecommunication Policies in the Arab Region, 3rd ed., Beirut: ITU, Available from World Wide Web: http://www.ituarabic.org/arabbook/ArabBook04.htm (15 March 2007).
14. Parfett, M. And Szarun, J. (2005), Broadband for business, the institute of chartered accountants of England and Wales, Available from World Wide Web: http://www.ecommerce.ac.uk/Default.aspx?page=1959 (26 March 2007).
15. STC DSL prices site, (2006), Available from World Wide Web: http://www.saudinet.com.sa/html/prices.php (13 Feb 2007).
16. Arab News, (2007), Available from World Wide Web: www.arabnews.com (10 Feb 2007).
17. Saudi Post, (2007), Available from World Wide Web: http://www.sp.com.sa (26 Feb 2007).
18. Cordesman, A. (2003), Saudi Arabia enters the twenty-first century: the political, foreign policy, economic, and energy dimensions, Praeger, London.

(Endnotes)

Gulf region countries include Bahrain, Kuwait, Oman, Qatar, The Kingdom of Saudi Arabia (KSA) and the United Arab Emirates (UAE).

1 KACST: King Abdulaziz City for Science and Technology- a government research centre in The Kingdom of Saudi Arabia (www.kacst.edu.sa).

i·Society 2007

Improving outsourcing operations by integrating outsourcing determinant index & outsourcing cycle effectiveness

A. Adnan, S. Arunachalam, A. Cazan

School of Computing & Technology
University of East London, UK
(A2adnan@uel.ac.uk)

Abstract Most of the developed countries particularly United Kingdom, Europe and United States have witnessed a sharp increase in outsourcing operations. These operations, range from the service sector to manufacturing of various components such as automotive components. Research has shown that organisations are outsourcing their operations not only to reduce cost but also for competitive advantage (Fan, 2000). As a result, different organisations are adopting various outsourcing models and frameworks to acquire competitiveness (Brannemo, 2006). In fact, many organisations practicing outsourcing are failing to achieve their objectives for a number of reasons. In depth literature search revealed that weakness in outsourcing operations is due to a large number of defects ranging from 'order request' to 'invoice payment confirmation'. It is concluded that lack of quality in the outsourcing system is the predominant reason for weaknesses in the outsourcing operations. The current frameworks do not have integrated assessment methodology in order to assess the quality of outsourcing operations and use that feedback to improve the outsourcing operations. The determinants of the outsourcing system and operations are defined and indexed to assess the performance of the outsourcing operations. The ODI assessment process is relatively static, whereas, the outsourcing operations are dynamic. So it is required to take extreme care in integrating ODI assessment and outsourcing framework in order to achieve continuously improving outsourcing operations. The quality of outsourcing systems and operation will be assessed by using outsourcing determinant index (ODI). The framework is developed by incorporating continuous feedback from ODI assessment to improve outsourcing operations in order to meet the dynamic market demand.

This paper will present the assessment methodology for outsourcing operations, formulation and implementation of frameworks to achieve continuous improvement. Development of framework to improve Outsourcing operations by incorporating ODI for assessment has not been attempted before. The integration of ODI for assessment into Outsourcing framework could contribute further improvements to outsourcing operations already practiced by companies.

In conclusion the ODI evaluation highlights the components of outsourcing system requiring improvement for improving the quality of outsourcing operations. The continuous improvement is achieved by assessing outsourcing operations using ODI, incorporating feedback for continuous improvement in a closed loop cyclic process. At the end Outsourcing Cycle Effectiveness (OCE) is also used to assess the outsourcing operations in terms of time flow.

Keywords outsourcing, outsourcing determinant index, continuous improvement, framework, Outsourcing Cycle Effectiveness

1 INTRODUCTION

Outsourcing trend is rising globally to acquire competitiveness ranging from service industry to the manufacturing sector. The concept of outsourcing is transferring in-house activities to an outside organisation without compromising its functional integrity. Decision to outsource is not a solution to acquire advantage for organisations but it generates a shift of services, resources, technology and manufacturing all over the Globe. The shift has contributed significant transitional changes on various aspects such as economic growth, international trade, and composition of workforce and trend of education (Koong et al. 2007). The outsourcing has its advantages and disadvantages. The organisations can perform better in globalisation process by adopting outsourcing philosophy effectively. The determinants of outsourcing need to be identified and analysed to assess the outsourcing performance. The research is planned to define set of determinants and taxonomy that can be employed assessing performance of outsourcing operations and use that as a feedback improving operations continuously. The taxonomy will help the or-

ganisations practicing outsourcing to assess the performance of the outsourcing operations easily and precisely.

A large number of scholars have contributed in carrying out research on outsourcing. Each research is unique in its own way but has limited scope as it was focused on one or several aspects of outsourcing. By integrating into a systematic framework or taxonomy, outsourcing develops a meaningful view. Combining outsourcing system components (Communication system, Delivery system, Outsourcer, Outsourcee) and the outsourcing system activities (Order preparation, Message communication, Order processing, Delivery, Delivery matching, Invoice matching), a multi-dimensional framework is developed. A large number of variables those are characteristics attributes of the components may impact the improvement. In addition to the ODI, outsourcing cycle effectiveness (OCE) that can be explained as the ratio of the manufacturing / processing time to total outsourcing cycle time will be used to assess the performance.

2 OUTSOURCING SYSTEM TAXONOMY

Practicing outsourcing is based on a rich array of theories, determinants and variables. Each of these attributes is important because they contribute in improving outsourcing operations that has multi-dimensional cause. The outsourcing system taxonomy is built on the basis of these theories. It is an integrated model that is formulated from the series of variables, in relation to outsourcing system characteristics attributes, characteristics of the organisation, market characteristics and characteristics attributes of external elements. The process of practicing outsourcing and the determinant variables affecting the performance of outsourcing operations are also integrated into the framework. Each of the characteristics variables and their relationship to the model is also described. Outsourcing determinant index is utilised to evaluate outsourcing performance.

2.1 Improving Outsourcing System

Improving outsourcing system involves the evaluation of current state and determines whether improvement can be achieved in outsourcing operations. In case of improving outsourcing system, the improvement in outsourcing operations can be understood in the context of quality. The characteristics attributes of outsourcing system (Outsourcer, Outsourcee, Communication system, Delivery system) are evaluated in terms of quality. The quality of the outsourcing system involves a number of variables that influence the decision to in improvement. The framework is to assess the performance before and after the improvement is carried out.

2.2 Determinants of Outsourcing System

There are four factors that influence the improvement of outsourcing system. These factors are Outsourcer characteristics, Outsourcee characteristics, Communication system characteristics and Delivery system characteristics. These characteristics consist of one or more dominant contributing attributes and few are explained as follows:

2.2.1 Resource

The capacity is determined by its resource attributes (value, rareness, imitability and substutability) and resource allocation. As a resource, asset specificity and functional complexity are considered important. In order to measure the performance of outsourcing operations asset specificity and functional complexity are considered as determinants.

2.2.2 Strategy

Strategy plays an important role in resource acquisition. Outsourcing is a function of the strategy-wise need and dimension of resources. By adopting 'cost leadership strategy', the capability reducing the peripheral costs is sacrificed and by adopting 'differentiation strategy', the capability improving quality is reduced.

2.2.3 Technology

Outsourcing system quality is also influenced by technology. Improvement in outsourcing is driven by the technology core of the innovation process. Outsourcing is considered effective medium speed technology change compared to extremely fast or extremely slow technology change.

2.2.4 Environment

Change in environment plays a vital role in improving outsourcing operations. The improvement in outsourcing operation performance is directly influenced by the environmental dynamism (Degree of the activity, Uncertainty and complexity of the market). The uncertainty resulting from fast changing and unpredicted market environment is an important variable that influence the quality of the outsourcing system. The organisational volume uncertainty and technological uncertainty affects the performance of outsourcing operations. It is important including quality, cost, delivery, flexibility, capacity utilization, procedure, personnel, software and transport media as integral part of the determinant.

2.3 Outsourcing Determinant Index for assessing improvement

On the basis of the taxonomy of outsourcing system, an ODI is formulated to evaluate the improvement periodically. The ODI is utilised to assess whether improvement has taken place in present compared to the past period. The ODI is utilized in three steps.

Step1: The relative importance of each of the 37 determinant variables is rated by experts based on a total weight of one hundred points. Then the expert assigned score is used to indicate the weights of the outsourcing determinants.

Step 2: The total weight for the period is evaluated respectively in terms of each determinant variable. Outsourcing

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 1. Outsourcing Determinant Index for Outsourcing System

Determinants (Weight)	PS1 = Period 1 Scale (1-5)	PW1 = Period 1 WeightXPS1	PS2 = Period 2 Scale (1-5)	PW2 = Period 2 WeightXPS2
Outsourcer Characteristics (30)	1 to 5	X1-1		X2-1
Resources (4)	.....			
Strategy (4)	.....			
Technology (4)	.....			
Environment (3)	.....			
Quality (3)	.....			
Cost (3)	.....			
Delivery (3)	.....			
Flexibility (3)	.....			
Capacity Utilization (3)	.....			
Outsourcer Characteristics (30)	.....			
Resources (4)	.....			
Strategy (4)	.....	.....		
Technology (4)	.....	.....		
Environment (3)	.....	.....		
Quality (3)	.....	.		
Cost (3)	.....	-		
Delivery (3)	.	-		
Flexibility (3)	.	-		
Capacity Utilization (3)	.	-		
Communication System (20)	.	-		
Procedure (2)	.	-		
Personnel (2)	.	-		
Data (2)	.	-		
Software (2)	.	-		
Hardware (2)	.	-		
Quality (2)	.	-		
Cost (2)	.	-		
Delivery (2)	.	-		
Flexibility (2)	.	-		
Capacity Utilization	.	-		
Delivery System (20)	.	-		
Procedure	.	-		
Personnel	.	-		
Software (2)	.	-		
Transportation (2)	.	-		
Quality (2)	.	-		
Cost (2)	.	-		
Delivery (2)	.	-		
Flexibility (2)	.	....		
Capacity Utilization (2)	.	X1-n		X2-n
Total Weight Score		$\sum_{i=1}^{n} x_{1-i}$		$\sum_{i=1}^{n} x_{2-i}$

system performance is rated using a five point scale with 5= full preference and 1= least preference.

Step 3: The weight obtained from the determinants in the step 1 and the score obtained in the step 2 are multiplied to obtain the weighted score of the outsourcing system for a

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

particular period in terms of each determinant. The sum of the weighted score is the total weighted score and an increase in the total weight score shows that the improvement has taken place.

The model provided in Table 1 is used to evaluate the outsourcing system for the specific period and is a continuous process. If the total weighted score of the period 2 is greater than the period 1, shows that there has been an improvement in outsourcing system. The evaluation on the outsourcing is holistic from the system point of view and is easy to use practically in order to assess the improvement in outsourcing operations.

2.4 Outsourcing Time Measurement Model

(Olve et al. 2004) listed various short-term and long-term performance indicators employed by organisations. (Kaplan et al. 1996) expressed manufacturing cycle effectiveness as the ratio of the processing time to the throughput time. The same expression can be translated for outsourcing operations as outsourcing cycle effectiveness (OCE) and that can be explained as the ratio of the manufacturing / processing time to total outsourcing cycle time.

Like the ODI, Outsourcing Cycle Effectiveness (OCE) can also be to assess the outsourcing operation. The step by step equations are given as follows:

$t_{\underrightarrow{rpo}}$ = Time required preparing order in time without error

t_{e-rpo} = Time required preparing order in time with error

$$\frac{t_{e-rpo}}{t_{rpo}} \geq 1$$

$$t_{e-rpo} \cong t_{rpo1} \pm t_{rpo\,2} \pm \cong \sum_{i=1}^{i=n} t_{rpoi}$$

t_{rc} = Time required for error free communication

t_{e-rc} = Time required for communication with error

$$\frac{t_{e-rc}}{t_{rc}} \geq 1$$

$$t_{e-rc} \cong t_{rc1} \pm t_{rc2} \pm \cong \sum_{i=1}^{i=n} t_{rci}$$

There are 11 communication steps between company and the supplier from order request preparation to invoice payment. In order refine the model the probability theorem will be applied based on decision theory.

t_{cop} = Time required for correct order processing without error

t_{e-cop} = Time required for correct order processing with error

$$\frac{t_{e-cop}}{t_{cop}} \geq 1$$

$$t_{e-cop} \cong t_{cop1} \pm t_{cop2} \pm \cong \sum_{i=1}^{i=n} t_{copi}$$

$t_{e-cop} \cong t_{cop}$ + Time to correct error

t_{cd} = Time required for correct delivery without any error

t_{e-cd} = Time required for correct delivery with error

$$\frac{t_{e-cd}}{t_{cd}} \geq 1$$

$$t_{e-cd} \cong t_{cd1} \pm t_{cd\,2} \pm \cong \sum_{i=1}^{i=n} t_{cdi}$$

t_{cd_note} = Time required for correct delivery note preparation without any error

$t_{e\_cd-note}$ = Time required for correct delivery note preparation with error

$$\frac{t_{e\_cd-note}}{t_{cd\_note}} \geq 1$$

$$t_{e\_cd-note} \cong t_{cd1-note} + t_{cd\,2-note} \cong \sum_{i=1}^{n} t_{cdi-note}$$

t_{in} = Time required for correct invoice preparation without any error

t_{e-in} = Time required for correct invoice preparation with error

$$\frac{t_{e-in}}{t_{in}} \geq 1$$

$$t_{e-in} \cong t_{in1} + t_{in2} \cong \sum_{i=1}^{n} t_{ini}$$

t_{dnm} = Time required for delivery note matching

$$t_{dnm} \cong t_{dnm1} + t_{dnm2} \cong \sum_{i=1}^{n} t_{dnmi}$$

t_{inm} = Time required for invoice matching

$$t_{inm} \cong t_{inm1} + t_{inm2} \cong \sum_{i=1}^{n} t_{inmi}$$

t_{tm} = Time required for invoice matching

$$t_{tm} \cong t_{tm1} + t_{tm2} \cong \sum_{i=1}^{n} t_{tmi}$$

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Total time for the complete cycle from 'Order request preparation' to the 'Invoice payment' = t_{cyc} = Time for order request preparation + 10* Time for communication + Time for order processing + Time for preparing delivery note + Time required for delivery + Time to do delivery matching + Time required for invoice preparation + Time required for invoice matching +Time for money transfer

$$\therefore t_{cyc} = \sum_{i=1}^{i=n} t_{rpoi} + 10^* \sum_{i=1}^{i=n} t_{rci} + \sum_{i=1}^{i=n} t_{copi} + \sum_{i=1}^{n} t_{cdi-note} + \sum_{i=1}^{i=n} t_{cdi} + \sum_{i=1}^{n} t_{dnmi} + \sum_{i=1}^{n} t_{ini} + \sum_{i=1}^{n} t_{inmi} + \sum_{i=1}^{n} t_{tmi} \pm error$$

The 'Outsourcing Cycle Effectiveness' (OCE) can be expressed as follows:

$$OCE = \frac{Manufacturing/\Pr oces\sin g\_Time}{Total\_Outsurcing\_Cycle\_Time}$$

$$OCE = \frac{\sum_{i=1}^{i=n} t_{copi}}{t_{cyc}}$$

In outsourcing, manufacturing or the processing time (value added time) is less than 5% of the total cycle time. The total outsourcing cycle time may be in multiples of weeks, whereas the manufacturing or the processing time in days. In an ideal outsourcing operation, the difference between the outsourcing cycle time and the processing time is reduced to minimum.

3 CONCLUSION

Based on the literature review, this paper provides an integrated taxonomy to use in evaluating improvement in outsourcing system. The model is incorporated with major determinants and their variables that influence the performance of the outsourcing system. It is a systematic model and practical framework that is used quantitatively to evaluate the improvement. The determinants and their variables can be used to monitor and predict changes. In addition Outsourcing Cycle Effectiveness (OCE) is also included as a new concept in evaluating the improvement in outsourcing operations.

REFERENCES

Adnan, A., Arunachalam, S. (2007), 'Improving outsourcing framework by integrating with lean', Advances in Computing and Technology 2nd Annual Conference, London.UK.pp.137-144.

Brannemo, A. (2006), 'How does the industry work with sourcing decisions? Case study at two Swedish companies', Journal of Manufacturing Technology Management, Vol.17.No.5 2006. pp.547-560.

Fan, Y. (2000), 'Strategic outsourcing: evidence from British companies', Marketing Intelligence & Planning, ISSN 0263-4503. pp.213-219.

Kaplan, R.S., Norton, D.P. (1996), Translating Strategy into Action The Balanced Scorecard, Harvard Business School Press, Boston, MA.USA.

Koong, K.S, Liu, L.C., Wang, Y.J. (2007), 'Taxonomy development and assessment of global information technology outsourcing decisions', Industrial Management & Data Systems, Vol.107.No.3 1007.pp.397-414.

Olve, N., Roy, J., Wetter, M. (2004), A Practical Guide to Using the Balanced Scorecard, performance drivers, John Wiley & Sons Ltd, Chichester, England.

Weir, J.W., Moore, J.T., Stoecker, M.G. (2001), 'An improved solution methodology for the Arsenal Exchange Model AEM)', The Journal of Operational Research Society, Vol.52.No.1.pp.48-54.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Taxonomy and frameworks for improving outsourcing operations

A. Adnan, S. Arunachalam, A.Cazan

School of Computing and Technology
University of East London
(A2Adnan@uel.ac.uk)

Abstract The literature search has shown that organisations are outsourcing their operations not only to reduce cost but to perform better. Most of the developed countries particularly United Kingdom and United States have witnessed a sharp increase in outsourcing operations. These operations, range from the service sector to manufacturing of various components. Despite differences organisations are adopting various outsourcing models to acquire competitiveness advantage (Brannemo, 2006). In depth literature search revealed that weakness in outsourcing operations is due to a large number of defects ranging from 'order request' to 'invoice payment confirmation'. By reconciling the defect taxonomy and its frequency, it can be concluded that lack of quality in the outsourcing system is the predominant reason for weaknesses in the outsourcing. The organisations practicing outsourcing are failing to achieve their objectives for a number of reasons. A survey by (Fan, 2000) shows that organisations are outsourcing rather than focusing on price and other issues of the business. Koong et al. 2007 developed outsourcing determinant index to practically evaluate outsourcing. The model and the assessment methodology is quite comprehensive and practical but relatively static, whereas, outsourcing is a dynamic phenomenon. The defects and weaknesses in the outsourcing operations are identified and classified according to their taxonomy. The major determinants of outsourcing operations are formulated into framework and outsourcing determinant index will be used to evaluate the outsourcing operations. The framework will be validated on a company as a case study.

The purpose of this paper is to present the taxonomy and assessment methodologies for outsourcing operations, formulation of frameworks and implementation to achieve continuous improvement. Furthermore, the aim is to analyse, how effectively this can be done to achieve an improved outsourcing model.

REFERENCES

Brannemo, A. (2006), 'How does the industry work with sourcing decisions? Case study at two Swedish companies', Journal of Manufacturing Technology Management, Vol.17.No.5 2006. pp.547-560.

Chapman, R.L., Soosay, C., Kandampully, J. (2003), 'Innovation in logistics services and the new business model', International Journal of Physical Distribution & Logistics Management, Vol.33.No.7 2003. pp.630-650.

El-Ansary, A.I. (2006), 'Marketing strategy: taxonomy and frameworks', European Business Review, Vol.18.No.4 2006.pp.266-293.

Fan, Y. (2000), 'Strategic outsourcing: evidence from British companies', Marketing Intelligence & Planning, ISSN 0263-4503. pp.213-219.

Koong, K.S., Liu, L.C., Wang, Y.J. (2007), 'Taxonomy development and assessment of global information technology outsourcing decisions', Industrial Management & Data Systems, Vol.107.No.3 2007.pp.397-414.

La, V.Q., Patterson,P.G., Styles,C.W. (2005), 'Determinants of export performance across service types: a conceptual model', Journal of Services Marketing, Vol.19.No.6 2005.pp.379-391.

Leachman, C., Pegels, C.C., Shin, S.K. (2005), 'Manufacturing performance: evaluation and determinants', International Journal of Operations & Production Management, Vol.25.No.9 2005.pp.851-874.

Mesquita, L.F., Lazzarini, S.G., Cronin, P. (2007), 'Determinants of firm competitiveness in Latin American emerging economies', International Journal of Operations & Production Management, Vol.27.No.5 2007.pp.501-523.

Power, D. (2005), 'Determinants of business-to-business e-commerce implementation and performance: a structural model', Supply Chain Management: An International Journal, Vol.10.No.2 2005.pp.96-113

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Trust and e-procurement transaction management

Joy Okah, Sonny Nwankwo, Charles Shoniregun

School of Business, School of Computing
University of East London. United Kingdom
joyokah@uel.ac.uk, s.nwankwo@uel.ac.uk c.shoniregun@uel.ac.uk

Abstract In common with all other forms of trading, e-Procurement is predicated on trust. E-Procurement refers to the use of electronic methods in every stage of the purchasing process from identification of requirement through to payment, and potentially to contract management. E-Procurement is in theory, the easiest and safest of e-business transaction and companies seek to establish and project the trustworthiness of their e-Procurement services to the market. This is an important step in the process of building sound business relationships with their suppliers. Trust is a central theme on many articles on business to business relationship, yet it remains an obscure target. The process of sustaining and projecting trust is on the possibility that trust must be managed in order for the company to operate effectively and efficiently.

This paper is focused on trust in e- Procurement and proposes trust and e-Procurement transaction model. The model ensures that trust is being sustained in a business to business relationship.

Keywords e-Procurement, trust, transaction, supplier, buying organisation

1 INTRODUCTION

Trust is an essential issue in e-Procurement. There has been enormous increase in transactions and cooperative computing services on the internet. This is a technical and a social phenomenon. The goals in trading are to convince customers to buy products from a particular supplier and persuade them to buy it again. The key factors influencing customers' decision to buy goods via Internet are the wide assortment, the reliability of the retailer, the safety of the transaction with the security of personal data and clear information about products (Gregor and Stawiszyński, 2002). Except for the wide variety of available goods, all indicated causes of buying are connected with trust. The Internet vendors perceive a lack of trust in electronic buyers as one of the main barriers in the development of the electronic market. Among the traditional marketing tools in selling for electronic retailers, reputation and building long lasting relationship with customers seem to be the most important. Both of them are directly based on trust. Erkki Liikanen – member of the European Commission, responsible for issues of Information Society, in one of his speeches about perspectives of the e-Procurement development said: "No trust, no transactions" (Liikanen 2000).

In transaction management, trust is an instrument for carrying forward trade. Organisations write trade contract, which helps them as a legal document if some dispute among them arises in future. It thus becomes an increasingly important factor for the electronic community to have means and methods for tackling trust related issues for e-transactions. We feel that trust in e-Procurement can be enhanced by reliability, familiarity, high performance standards and continuous interaction from both the buying organisation and the supplier. Research has shown that the e-Procurement concept includes the business to business (B2B) transaction, business to customer (B2C) transaction, customer to business (C2B) transaction, business to administration (B2A) transaction and customer to customer (C2C) transaction. In this paper, we focused on the business to business transactions and we looked into the benefits and vulnerabilities of e-Procurement, hence proposing a model that will sustain the trust of the end users involved in e-Procurement transaction.

2 E-PROCUREMENT

E-Procurement is an electronic method of conducting business transactions. According to Uninova (2000) , e-procurement is the catalyzing tool, which enables companies to integrate its supply chains from the beginning to the end, sharing information on prices, availabilities and performance, allowing buying organisations and suppliers to work with mutually beneficial prices and planning. Furthermore Davila (2002) proposed another definition of e-Procurement from technological perspective: a technology designed to facilitate goods acquisition by one organization through Internet. The Scottish Enterprise e-business magazine (2005) defines e- as a term used to describe the electronic methods used in every stage of the purchasing process from the identification of requirements through to payment, and potentially contract management. Chaffey's (2002) definition is one we find particularly interesting and it defines e-Procurement as "the electronic integration and management of all

procurement activities including purchase request, authorisation, ordering, delivery and payment between a purchaser & supplier".

A good e-procurement system enhances customer's interactions and provides built-in monitoring tools to help control costs and assure maximum supplier performance. It provides an organized way to keep an open line of communication with potential suppliers during a business process. The system allows managers to confirm pricing, and leverage previous agreements to assure each new price quote is more competitive than the last.

3 PROJECTING TRUST IN E-PROCUREMENT

Trust is one of the most prominent issues in e-Procurement today. Collins English Dictionary defined trust as a "reliance on and confidence in the truth, worth and reliability of a person or thing". Trust has been variously defined in the extant literature; however the Moorman, Zaltman & Deshpande (1992) definition is frequently quoted in the terms of relationship as "a willingness to rely on an exchange partner in whom one has confidence." Kaplan & Sawhney (2000), states that Trust is the extent, to which one party is willing to depend on somebody, or something, in a given situation with a feeling of relative security, even though negative consequences are possible. It is a pervasive notion and, as such, has been studied thoroughly in a variety of different fields, including the social sciences, economics and philosophy. In terms of transaction management, trust is an instrument for carrying forward trade. From the approach of business to business e-Procurement transaction relationship, finding suppliers simply was not feasible a decade ago, today it has allowed both businesses and suppliers to reap some significant benefits, based on a trusting relationship. In a detail auction, the buying organisation creates a listing of the company's specific project needs. For example, if the company needs to purchase spools of wire from a supplier, then the company would detail the exact specifications and quantity of the wire they required. The buying organisation then posts this listing of requirements and invites potential suppliers to give quotes on the cost of fulfilling those needs. Essentially, the suppliers are competing against one another so lower quotes have an advantage. At the end of the auction, the buying organisation then chooses a winner based on several important factors, including cost, delivery time, and supplier reputation.

The suppler transaction relationships are different from simple purchasing transactions in several ways. Firstly, there can be a sense of commitment to the organisation because of the ease in information flow in e-procurement. For example, if a supplying organisation sells light bulbs, he can feel confident that the buying organisation will come to him the next time it requires a new shipment of light bulbs. Another element of these supplier transaction relationships is advanced planning. A buying organisation don't just communicate with the supplying organisation when a procurement need arises; they also contact them in order to discuss their future needs and to determine how best to satisfy those needs by working together. Secondly the company's attitude and view of its suppliers matters a lot for business success. Companies that build supplier relationships think of these suppliers as partners and not just simple commodity providers, hence manage their suppliers trust in their e-Procurement system. This difference in orientation can have a profound effect on the way an organization communicates and works with its suppliers.

Overall, suppliers and buying organisations are both better served when they come together to form strong, mutually beneficial, and secure business relationships for e-procurement of goods and services. When these relationships exist, they can drive the growth and profitability of both organization and prevent any problem that arises in the course of execution of e-procurement. We would hence look into the development of e-Procurement in some industries as the case study.

4 CASE STUDY OF E-PROCUREMENT TRANSACTIONS IN SOME INDUSTRIES.

Case study has multiple meanings. Shoniregun (2005) states that case study could be used to describe a unit of analysis or a method. Saunders et al (2003) defines a case study as a research strategy that involves the contemporary phenomenon within its real life context using multiple sources of evidence. A case study is considered a good method for research. For this study we would investigate four types of industry namely, the automotive industry, oil and gas industry the hospitality and the telecommunications industry. Each industry has a unique combination of occupations, production techniques, inputs and outputs, and business characteristics. The occupations found in each industry depend on the types of services provided or goods produced but the all have commonly developed and implemented the e-Procurement transaction system.

4.1 E-Procurement in the Automotive Industry

E-Procurement has a significant development in the automotive market since 1999. Although Ford has implemented their e-Procurement programme called e-Steel since 1995 because they were interested in gaining maximum cost savings and control over supply partners. In 1999, Ford and General Motors simultaneously and independently announced that they were to launch business to business electronic trade exchanges for supply and procurement throughout their supply chains. The exchanges intention was to expand them globally and encourage other auto manufacturers to join. Ford then commissioned e-Steel in May 2000 to design and implement a supply network to provide a secure, real-time environment for its steel procurement, estimated at over $1bn pa. By April 2000 other auto manufacturers like DaimlerChrysler, Renault and Nissan joined the e-Procurement trend. BMW also experimented with e-business since 1999, and after completion of a route-finder with suppliers, launched its own private electronic procurement platform in March 2000 (Howard et al, 2002). The automotive industry

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

is one of the largest and most complex in the world and it has many activities concerning electronic markets in which e-procurement is the centre. The purpose of e-procurement in the automotive industry was to reduce cost and facilitate information flow but was noted that trust was a barrier for its efficiency (Howard et al, 2004).

4.2 E-Procurement in the Oil and Gas Industry

The oil and gas industry is now harnessing effective purchasing and procurement as a strategic function to increase significant profit margins. The industry is realising the fact that costs savings is essential and hence have taken to e-Procurement. Oil and Gas is one of the world's largest electronic exchanges. E-procurement in the oil and gas industry came from a business led initiative, not from Information Technology. According to Chris Miller, Shell International's vice president of strategic sourcing in 2000, business leaders have driven the move towards e-Procurement regardless of resistance from IT units. The e-Procurement scheme was founded in the oil and gas industry in January 2000 by Shell, BP-Amoco, Statoil, Mitsubishi and Total Finaele collectively (Riley, 2000). It was expected to facilitate $ 125 worth of exchanges which makes it the largest trade exchange of year 2000. Apart from cost savings, other reasons why the oil and gas industry took to e-Procurement were for facilitating financial, logistic and auctioning services.

Shell International which is rated the highest as the most sustainable oil company for four consecutive years (United Business Media, 2007) projected its savings relatively when using e-Procurement with standardisation for 10%, Compliance 37%, Leverage, consolidation and supplier cost reduction 23%, process efficiencies 7% and Engineering man hours 3%. Crabtree et al (2000) noted that trust would be an issue for the usage of e-Procurement in the oil and gas industry in the U.K.

4.3 E-Procurement in the Hospitality Industry

The hospitality industry includes the hotels, motels and restaurant business. Large hotel companies are highly complex organizations with locations spread across the globe. Each hotel requires tight controls and automated processes for procuring food and beverage and managing requisitions and inventory. The hospitality industry started to embrace electronic procurement in 2002, resulting in enormous supply chain cost savings, according to a report by a national research agency, Aberdeen Group. The study shows the hospitality industry spends an estimated $50-60 billion per year on supplies and services, and that number is expected to continue a steep upward trend in the coming years. The report indicates the sector stands to benefit enormously by moving its purchasing operations to the Internet. The resulting increased efficiency is expected to save companies millions of dollars each year, as purchasing managers gain better control over their procurement business and also cut down on off-contract purchases, a costly problem for the sector. According to Mark Withington, an Aberdeen analyst who coordinated the report, "the hospitality sector has traditionally lagged behind other industries, such as manufacturing, when it comes to electronic procurement". But now hotel companies have taken advantage of the Web-based applications that are available to easily upgrade their purchasing systems, resulting in immediate dividends." A research conducted by Roger et al (2001) conducted under the Hospitality banner 2000, indicated that trust would be a serious challenge in the use of e-procurement in the hospitality industry.

4.4 E-Procurement in the Telecommunications Industry

The telecommunications industry is at the forefront of the information age delivering voice, data, graphics and video at ever increasing speeds and in an increasing number of ways. In the late 1990s, the telecommunications industry experienced very rapid growth and massive investment in transmission capacity. Eventually this caused supply to significantly exceed demand, resulting in much lower prices for transmission capacity, hence this brought about the development of e-Procurement. The industry realised that cost can be cut with the new technology, leading to substantial savings. BT, for example, is the second largest procurer in the UK after the Government, with £1.3m procurement transactions annually - amounting to costs of £7bn. BT decided to procure goods and services electronically in June 1999, and aims to migrate 95% of its total procurement to internet-based systems by the middle of the year 2000. BT claimed that, the average cost of a purchasing transaction has decreased from £70 to £50 and will fall by a further £5 as a result of increasing use of BT MarketSite thus e-Procurement has dramatically saved cost. (E-consultancy, 2000). Ericson Services white paper in March 2007, states that trust should be effectively managed for the effectiveness and efficiency of e-Procurement.

5 BENEFITS AND VULNERABILITIES OF E-PROCUREMENT

5.1 Benefits

Of all the elements of the e-business revolution, e-Procurement is the strand that has claimed the greatest benefits. Apart from cost savings which has been extensively discussed in the case study industries, e-Procurement offers other benefits that includes contributing to the formalisation of information-flow through several companies, and that is a more important issue for large companies. Another benefit is that sourcing is much easier now than previously, when people had to make do with trade directories and overwhelming manual catalogue searches. Now, most applications allow for easy uploading of keyword-searchable catalogue information, so it is no longer necessary for purchasing managers to flip through piles of catalogues.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

5.2 Vulnerabilities

When it comes to electronic interactions trust is always an issue. Sometimes data are extensively misused and that can cause great concern for companies involved in e-Procurement. Miller the Shell International vice president of sourcing (Shell Bulletin, 2002) issued a warning when making a presentation that companies need to be careful when issuing e-Procurement exchanges because they can give legitimacy to bad deals as well as good deal. E-Procurement like any e-business application has its risks. The vulnerabilities in e-Procurement are not just with money losses, the reputation also counts. Another vulnerability identified is the major changes e-Procurement entails, compliance is an issue when it comes to changes in corporate culture.

6 TRUST AND E-PROCUREMENT TRANSACTION (T & E-PT) MODEL.

In this research, we proposed a model to sustain trust in business to business transaction relationship. The main drivers for e-Procurement are - Transactional benefits, Compliance being Improved, Management of Information, Improved Price and Payment and Trust.

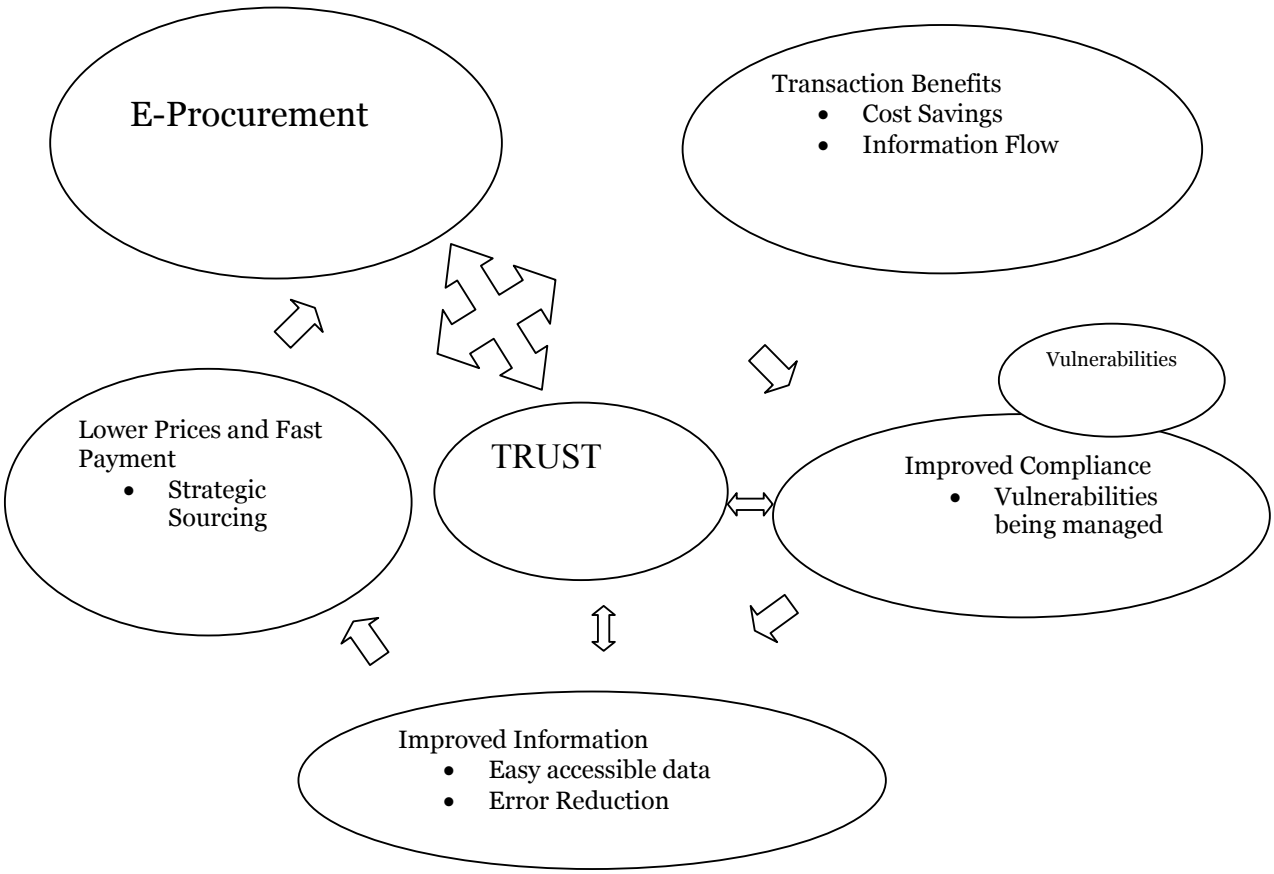
Transaction Benefits: Until the conception of e-Procurement in many organisations, only the higher value suppliers have been actively managed (Aberdeen Group, 2001). E-Procurement enables the purchase to pay process online. Automating the purchase to pay process leads to greater time savings and efficiency due to:

- global, automated, processes incorporating best practice and eliminating unnecessary activities
- e-enabled relationship with suppliers speeds procurement cycle times and facilitates supplier performance improvements
- greater data accuracy minimises ordering inaccuracies and provides the essential foundation for better management through measurement and analysis.

Improved Compliance: Compliance is always a challenge within an organisation, not just because employees deliberately purchase outside of preferred arrangements, but rather through lack of awareness that a preferred arrangement is in place. E-Procurement addresses this issue through tools such as catalogues and standard order processes. Electronic catalogues of the goods and services provided by a supplier are at the heart of all e-Procurement systems. An electronic catalogue will typically contain the name of the products, the product hierarchy, a description, its price and all relevant supplier codes and internal codes. These catalogues may contain several hundred or several thousand items per supplier and have to be created, approved and updated so the end users can have access the goods and services they require. End users would always comply when they perceive the benefits. Compliance is also improved when vulnerabilities like change and perceived risks are being managed effectively.

Improved Information: The fact that key information (cost centre, commodity codes etc) is hard coded, the user dramatically reduces coding errors and provides highly detailed and easily accessible data. This is essential to maximise the

Figure 1. Model to sustain Trust in e-Procurement business to business Transaction.



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

financial benefits of strategic sourcing. A successful e-Procurement implementation will provide high quality detailed management information and will minimise the need for data warehousing or resource-heavy data mining.

Reports from the management system will enable the improvement of service to end-users and allow effective monitoring as to where the system is not being used, so that the necessary action can be taken to improve the service, communicate this to end-users and convince them of the value of using the e-Procurement system.

Lowered Prices and Fast Payment: Implementing an e-Procurement transaction system will not itself reduce the individual price of goods and services provided by a supplier. E-Procurement is, however, a powerful way to ensure benefits captured during a strategic sourcing effort effectively translate into savings and are not lost through poor contract compliance. In turn, e-Procurement can become a source of data for strategic sourcing activities and lead to:

- identification of cost saving opportunities through supplier spend consolidation, which might lead to placing improved national or global contracts
- the ability to treat low value, high volume spend strategically.

This will enhance ability to negotiate down prices. Manpower will also be reduced and this can guarantee suppliers a prompt payment.

Trust: In the e-Procurement system, trust is effectively sustained when compliance is continually improved. Compliance is improved by managing vulnerabilities. Trust is also sustained when information is being improved. Information is improved when errors in the transaction process is highly reduced. In any business system, perceive risk which turns up as vulnerabilities in the system and when this is effectively managed through the improved compliance and improved information, all the end-users of e-Procurement system tends to keep on trusting the system.

7 CONCLUSION

Trust is a higher-order construct that is difficult to directly measure. The changing nature of trust during transaction relationships requires the understanding of the transaction in e-Procurement process and projecting trust on it. We investigated the development of e-Procurement in four types of industries. In the course of investigation, we realised that the main reason for e-Procurement strategy was cost reduction and information flow in contract management for business transaction. The e-Procurement process involves transaction benefits, some vulnerabilities, improved compliance, improved information, lowered price and fast payment. Our model is a graphical representation of the sustainace of trust in the business to business relationships in e-Procurement transaction.

REFERENCES

1. Aberdeen Group Report, More Hospitality E-Procurement'(January 17, 2001)
2. Chaffey (2002), E-Business& E-Com. Mgt. Prentice Hall Hallow.
3. Crabtree, E., Bower, D. and Keogh W (2000) Manufacturing strategies of Small technology-based firms in the UK Oil Industry, International Journal of Manufacturing Technology and Management. 4-5 (1), 455-463
4. Davila, A. (2002) Moving Procurement Systems to the Internet: The adoption and use of E-procurement technology models, Research Paper Series, Graduate School of Business Stanford University, research papers no 1742.
5. E-business Fact sheet E-Procurement (2005), Scottish Enterprise business magazine. July.
6. E-consultancy: e-Procurement, who is using it and why? (April 18, 2000)
7. Ericsson Services white paper (2007), Managed Services Impact on Telecom Industry
8. Gregor B., Stawiszynski M. (2002), e-Commerce, Bydgoszcz – Lódz: Oficyna Wydawnicza Branta
9. Howard, M. Vidgen, R, Powell, P. and Graves, A. (2002), Are hubs the centre of things? E-Procurement in the automotive Industry, University of Bath United Kingdom.
10. Howard, M. Powell, P. and Vidgen R. (2004), Inter-organisational Collaboration and Value Creation in the automotive Industry.
11. Kaplan, S. and Sawhney, M., 2000, E-hubs: The New B2B Marketplaces", "Harvard Business Review, May- June, 97-103.
12. Lacity, M. and Hirschheim, R. (1993), Information Systems Outsourcing: Myths, Metaphors and Realities, John Wiley and Sons, Chichester, UK.
13. Liikanen E. (2000): Trust and security in electronic communications: The European contribution, Speech/00/344, Information Security Solution European Conference "ISSE 2000".
14. Moorman, Christine, Zaltman, Gerald and Deshpande, Rohit (1992), "Relationship Between Providers and Users of Market Research: The Dynamics of Trust Within and Between Organisations", Journal of Marketing Research, 29 (3), 314-328
15. Sauders M, Lewis P and Adrian T (2002) Research Method for Business Student, third edition. Pretence Hall Financial Times.
16. Shoniregun C, (2005) " Impacts and Risk Assessment of Technology for Internet Security: Enabled Information Small- Medium Enterprises', ISBN: 0-387-24341-0
17. Shell International Bulletin June 2001.
18. The Scottish Enterprise e-business magazine, Service to business. July 2005
19. Riley J (2000) "Oil Giant of e-hub" Computer Weekly.com, Reed Business Information Limited (June 22, 2002).
20. Roger S. Cline and Dr. Mark Warner (2001), The Future, HITEC 2001 - Hospitality E-Business. Hospitality.net http://www.hospitalitynet.org/news/4008386.search?query=trust+and+e (June 8th, 2007)
21. Uninova S, (2000) The impact of e-Procurement on the role of corporative purchasers http://www.iamot.org/conference/viewpaper.php?id=1963&print=1&cf=10 (June 2, 2007).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

i·Society 2007

Mitigating effect of number of bidders on perceived uncertainty

Ossama Elhadary

Felician College, Lodi, NJ

Abstract This is a paper in progress in which the author is presenting the hypothesis, the theoretical background and the research model and will present the final conclusions and the results of the data analysis in a future paper. Imagine you are an American tourist in Bangkok, walking aimlessly and then you feel hungry. You do not recognize any of the restaurants around you and you can not even read the names on the signs. Which restaurant would you choose? Probably the most crowded one. In this restaurant-choosing dilemma, it seems that the more customers there are in a certain restaurant, the more confidence we have in this restaurant. Somewhere in our minds we are probably arguing that everybody inside that crowded restaurant must think it is good one otherwise they would not have chosen it themselves. The author in this paper is hypothesizing that in electronic markets like eBay, this restaurant choosing dilemma will still be manifested. In an auction with tens of similar products, a buyer will probably bid on products that other buyers are bidding on. In a sense the number of bidders in an auction can be seen by a potential bidder as a manifestation of the Trust other bidders have in the auction and in the seller. It would make sense in this case to assume that this already established trust between the seller and the existing bidders will help lower the potential bidder's level of perceived uncertainty and thus help build trust. In this paper, the author is presenting a research model that shows the relationship between the number of bidders, trust, the perceived uncertainty and the intention to buy. The author is currently conducting a number of experiments to collect data that will then be analyzed and used to validate the model.

Keywords Trust, online auction, eBay, bidder, electronic markets, Uncertainty

1 Introduction

In order to deal with an uncertain world and to decide to act and pursue our goal without perfect knowledge, we have to take the risk by trusting our information, beliefs, our action, and other agents we are relying upon in order to fulfil our needs (Yuan & Sung, 2004). Pavlou e al. (2007) define perceived uncertainty in a buyer-seller relationship as the degree to which the outcome of a transaction can not be accurately predicted by the buyer due to seller and product related factors. They then explain that uncertainty consists of seller quality uncertainty (seller making false promises, shirking or defrauding, and hiding its true characteristics), and product quality uncertainty (product condition not being as promise, or product quality being compromised). They then identified four antecedents of perceived uncertainty in online buyer-seller relationships: perceived information asymmetry, fears of seller opportunism, information privacy concerns, and information security concerns and then proposed four uncertainty mitigating factors: trust, website informativeness, product diagnosticity, and social presence.

Trust is defined as the buyer's intention to accept vulnerability based on her beliefs that the transaction will meet her confident expectations (Pavlou e al, 2007). As explained by Ba and Pavlou (2002) there are two distinct types of trust: benevolence (the belief that one partner is genuinely interested in the other partner's welfare and has intentions and motives beneficial to the other party even under adverse conditions) and credibility (the belief that the other party is honest, competent and reliable) with the latter being the most prevalent in electronic markets like eBay.

Bolton et al. (2004) investigated trust among Internet traders in computer-mediated online markets such as eBay and explained some of the challenges in establishing Trust in such markets: transactions on these platforms are characterized by asynchronous actions of anonymous traders, operating at spatially disperse locations. They then explained that in a medium of communication such as Computer-mediated communication it is more difficult to signal trustworthiness and to promote cooperation that richer communication media such as face-to-face communication (Frank, 1988; Brosig et al., 2003). According to Bolton et al. (2004) other challenges include the fact that is easier for a buyer or a seller to choose a trader identity other than one's true identity, as well as the fact that lasting personal relationships on Internet market platforms are infrequent. They then concluded that cyberspace makes it particularly difficult to develop social and economic bonding that supports the emergence of trust and trustworthiness in more traditional markets.

Nikitov (2006) believes that a number of factors can assist in building trust between a buyer and a seller in an online auc-

tion: 1) seller's reputation rating, 2) quality and quantity of visual disclosure of the product sold, and 3) the amount of textual disclosure of the product sold. Ba and Pavlou (2002) on the other hand believe that buyers develop trust in the sellers' credibility partly as a result of feedback mechanisms and that trust has a substantial effect on the transaction by generating price premiums. Their study also identified product price as a moderating factor and that transaction specific risks are highly intertwined with trust. Resnick, and Zeckhauser, (2001) believe that trust has emerged due to the feedback or reputation system employed by eBay and other auction sites. A reputation system according to them must meet three challenges: (1) must provide information that allows buyers to distinguish between trustworthy and non-trustworthy sellers (2) must encourage sellers to be trustworthy, and (3) must discourage participation from those who aren't.

Resnick, and Zeckhauser, (2001) explained that the presumptive challenge to Internet-based feedback systems is to get buyers to provide feedback with reasonably high frequency, and to provide it honestly. In their study of the reputation system in eBay they concluded that the frequency is not a Problem as more than half of transactions receive feedback. However, the 0.3% negative feedback rate on transactions (.6% of those that provided feedback) and 0.3% neutral feedback numbers from eBay are highly suspicious. They then concluded that although this reputation system may not work well in the statistical tabulation sense it does seem to work in terms of facilitating transactions and they attributed this to two reasons: 1) even if the system is unreliable or unsound, it might still work if its participants think it is working, and 2) it may function successfully if it swiftly turns against undesirable sellers, a process they called stoning; and if it imposes costs for a seller to get established which they called initiation dues.

2 RESEARCH DESIGN

What seems to be missing from all the researches that dealt with uncertainty and trust is an investigation of the relationship between the number of bidders in an auction and the perceived uncertainty associated with buying from that auction. It is worth noting here that in a sense the number of bidders is an auction can be seen by a potential bidder as a manifestation of the Trust other bidders have in the auction and in the seller. It would make sense in this case to assume that this already established trust between the seller and the existing bidders will help lower the potential bidder's level of perceived uncertainty and thus help in building trust. The relationships between the number of bidders, trust and perceived uncertainty are depicted in the research model (Figure 1).

The hypotheses in this research are:

H1: There is a significant negative relationship between the number of bidders in an online auction and the uncertainty perceived by a potential bidder.

H2: There is a significant positive relationship between the number of bidders in an online auction and the trust a potential bidder has in the auction.

H3: There is a significant negative relationship between the trust a potential bidder has in the auction and this bidder's perceived uncertainty in the auction.

H4: Inexperienced users are more likely than experienced users to rely on the number of bidders in an online auction to mitigate uncertainly

The author will create an experiment and will then ask a number of subjects to participate in this experiment. All the subjects will have to be eBay users to ensure that this research captures the behaviour of the actual users on this online auction site.

The experiment will consist of three sets of auctions. In the first auction, the buyers will only be shown an indicator of the seller's reputation, in the second auction the buyers will only be shown the number of bidders on the item they want to buy, and in the third auction, the buyers will be shown both the number of bidders as well as an indicator of the seller's reputation. In each case, the audience will be asked to choose which item they will bid on as well as asked to answer a few questions related to the user's trust in the auction they have chosen. The Chi square test will then be used to compare the data collected from the three cases and prove that buyers do consider the number of bidders when choosing an auction. Regression analysis will also be used to measure the relationships between the number of bidders in the auction, the potential bidder's trust, and the uncertainty perceived by that bidder.

The author will also attempt to collect some information about the users' experience in online auctions and then use regression analysis to prove that inexperienced users are more likely than experienced users to rely on the number of bidders construct to mitigate uncertainty.

The Trust construct in this paper was borrowed from Pavlou e al. (2007) and Gefen (2002) and then slightly modified to reflect trust in an online auction setting. The Uncertainty construct on the other hand is borrowed from Pavlou e al. (2007) and Torkzadeh & Dhillon (2002).

Figure 1. Research Model

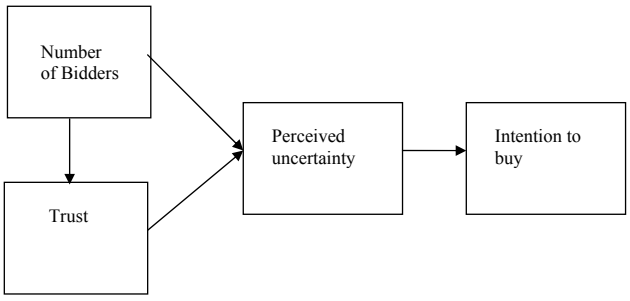


Table 1 shows all the measurement items for the Trust construct and table 2 shows the measurement items for the perceived uncertainty construct.

Table 1. The Trust Construct

Trust
The seller understands the market he/she works in (competence)
The seller knows a lot about the product he/she sells (competence)
Promises made by the seller are likely to be reliable (integrity)
I do not doubt the honesty of the seller (integrity)
I expect the seller to keep any promises he/she makes (integrity)
I expect the seller to have good intentions towards me (benevolence)
I expect the seller's intentions to be benevolent (benevolence)
I expect the seller to be well meaning (benevolence)

Table 2. The Uncertainty Construct

Perceived Uncertainty
I feel that buying from this eBay seller involved a high degree of uncertainty
I feel the uncertainty associated with buying from this eBay seller is high
I am exposed to many transaction uncertainties if I purchase from this eBay seller
There is a high degree of product uncertainty (ie. The product you receive may not be exactly what you want) when purchasing from this eBay seller

3 THEORETICAL IMPLICATIONS

This research aims at adding a new dimension to our understanding of uncertainty and trust, a dimension that had been overlooked by many researchers. After the experiments are conducted and the data is analysed and the number of bidders is shown to have a moderating effect on uncertainty and trust, then this will be seen as validating Ba's and Pavlou's (2002) research in which they indicated that transaction specific risks are highly intertwined with trust. This will add another factor to those identified by Nikitov (2006) as assisting in building trust which are: 1) seller's reputation rating, 2) quality and quantity of visual disclosure of the product sold, and 3) the amount of textual disclosure of the product sold.

This research also aims to prove that trust building in online auctions is not merely a process occurring between two entities, but instead is done within a certain social context with all the bidders participating in this process. As the restaurant choosing dilemma presented at the very beginning shows, the effect of this social setting is not restricted to online auctions only but can also be extended to daily non-online transactions between any two entities trying to build trust. But since our focus is mainly on online auctions, it is important to note that by proving that the number of bidders influences perceived uncertainty and trust, we thus add the social context dimension to trust building and we improve our understanding of how users perceive uncertainty and use different tools to build the trust needed to conduct transactions.

The research will also try to discover if there are differences between the various categories of users (example: experienced and inexperienced users) of online auctions in their dependence on the number of bidders to mitigate the perceived uncertainties associated with the transaction or not. Do new inexperienced users for example use the number of existing bidders as a tool to mitigate uncertainty versus more experienced users who rely on their own evaluation of the seller and of the transaction itself to decide if they should bid or not? Or is it actually the opposite, where the experienced users learn from experience that the number of bidders can be used as a tool to improve their success and profitability in online auctions.

4 IMPLICATIONS FOR PRACTICE

In this research, the author is suggesting a dimension that has been overlooked by other researches as they studied uncertainty and trust, which is the influence the existing number of bidders has on the uncertainty perceived by a potential bidder. After the experiments are conducted and the results are analysed, this research will have the potential to change the way we look at how electronic market users engage in auctions and how they perceive uncertainty and use tools to mitigate it. If the hypothesis are proven, then new tools might have to be devised to help build trust taking into consideration that users not only evaluate the sellers and the products being sold but also rely on other users bidding on the product to have an additional level of comfort and trust that can help mitigate the uncertainties associated with participating in the auction.

By differentiating between various categories of users (ex: experienced and inexperienced) in their reliance on the number of bidders construct, we can have a better understanding of how different groups of users address the issue of uncertainty in online auctions. This can aid in developing specific tools that can help different categories of users build trust needed in online transactions. This differentiation can also lead us to gain a better understanding of how users' actions evolve as they become more experienced in online auctions.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

5 CONCLUSIONS

In this paper, the author presented a research model that suggests that users of an online auction use the number of bidders bidding on a certain item to mitigate the uncertainty perceived in participating in this auction. The author argues that the number of bidders in an auction can be seen by a potential bidder as a manifestation of the Trust these bidders have in the auction and in the seller and so it seems that this already established trust will help lower the potential bidder's level of perceived uncertainty and thus help build trust. The author is currently collecting the data needed to validate the hypotheses presented in this paper and the final conclusions will be presented in a future paper. The author was able though to show that if the hypotheses are proven, this research will have a number of implications both on the theoretical level as well as on the level of practice.

REFERENCES

1. Ba S., and Pavlou P. (2002), "Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behaviour", MIS Quarterly, Vol. 26, No. 3, pp. 234-268.
2. Bolton G., Katok E., and Ockenfels A. (2004). "Trust among Internet traders: A behavioral economics approach," Analyse und Kritik, No. 26, pp. 185-202.
3. Brosig J., Ockenfels A., Weimann J. (2002), "The effect of communication media on cooperation", German Economic Review, Vol. 4, No. 2, pp. 217-341.
4. Frank R. (1998), Passions within reason: the strategic role of the emotions, Norton & Company, New York.
5. Gefen, D. (2002), "Customer loyalty in Ecommerce", Journal of AIS, No. 3, pp. 27-51.
6. Nikitov, A. (2006), "Information assurance seals: How they impact consumer purchasing behaviour", Journal of Information Systems, Vol. 20, No. 1, pp. 1-17.
7. Pavlou A., Liang H., and Xue Y. (2007), "Understanding and mitigating uncertainty in online exchange relationships: a principal agent perspective", MIS Quarterly, Vol. 31, No. 1, pp. 105-136.
8. Resnick P., and Zeckhauser R. (2001) "Trust among strangers in Internet transactions: empirical analysis of eBay's reputation system", Working Paper, University of Michigan.
9. Torkzadeh G., and Dhillon G. (2002) "Measuring factors that influence the success of Internet commerce," Information Systems Research, Vol. 13, No. 2, pp. 187-2004.
10. Yan, S., and Sung H. (2004), "A learning-enabled integrative trust model for e-markets", Applied Artificial Intelligence, No. 18, pp. 69-95.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

An adaptive routing protocol for censorship-resistant communication

Michael Rogers
University College London, UK
m.rogers@cs.ucl.ac.uk

Saleem Bhatti
University of St Andrews, UK
saleem@cs.st-andrews.ac.uk

Abstract In open-membership networks such as peer-to-peer overlays and mobile *ad hoc* networks, messages must be routed across an unknown and changing topology where it may not be possible to establish the identities or trustworthiness of all the nodes involved in routing. This paper describes a decentralised, adaptive routing protocol in which nodes use feedback in the form of unforgeable acknowledgements (U-ACKs) to discover dependable routes without knowing the identities of the endpoints or the structure of the network beyond their immediate neighbours. Our protocol is designed to survive faulty or misbehaving nodes and reveal minimal information about the communicating parties, making it suitable for use in censorship-resistant communication.

1 INTRODUCTION

Internet pioneer John Gilmore famously claimed that "The Net interprets censorship as damage and routes around it" [13]. Unfortunately, at a technical level this statement is becoming less accurate every year: censorship of the internet is becoming more widespread and sophisticated, and dozens of governments now filter the information available to their citizens [21]. However, at a social level Gilmore's statement remains a maxim for activists and researchers working on censorship-resistant communication.

In this paper we examine the problem of routing messages across an untrusted, open-membership network, where a node cannot establish the identities or trustworthiness of any nodes other than its immediate neighbours. Each communication exchange involves a series of messages sent from an *originator* to a *destination* by relying on the forwarding behaviour of intermediate *relays*. We assume that any node in the network can function as an originator, destination or relay. Examples of such networks include peer-to-peer overlays and mobile *ad hoc* networks.

In an untrusted network, messagesmay be lost, reordered, or modified for any number of reasons, and it may not be possible to determine whether such events are accidental or deliberate in nature. Rather than trying to identify the node or link responsible for each failure, we take the pragmatic approach of measuring dependability without attempting to distinguish between deliberate and accidental failures. We show that by observing end-to-end *unforgeable acknowledgements* (U-ACKs), relays can adaptively discover dependable routes without knowing the origins or destinations of the messages and acknowledgements they forward. Lightweight *flow identifiers* can be used to improve the efficiency of adaptive routing.

The next section discusses previouswork in this area. Section 3 gives an overview of our adaptive routing protocol. In Section 4 we describe simulations to evaluate the protocol's efficiency and scalability. Section 5 discusses the results of the simulations and considers possible applications of our protocol. We conclude the paper with some ideas for future work.

2 BACKGROUND AND RELATED WORK

Many routing protocols use feedback from the destination to guide forwarding decisions at relay nodes. This adaptive approach has the advantage of propagating information about the state of the network only to those nodes to which the information is relevant; however, the absence of information about unused routes can make the cost of initial route discovery relatively high.

Q-routing [4] uses reinforcement learning to find the quickest route to a destination. Each node updates its estimate of the time it will take a message to reach the destination, including time spent in the node's own queue, based on immediate feedback in the form of the next node's estimate of the delivery time. Information is thus passed back from the

destination towards the source, taking into account congestion and any other causes of delay.

AntNet [10] and AntHocNet [11] are routing protocols inspired by the collective foraging behaviour of ants, which use chemical markers to discover short paths between food sources and their nest. AntNet uses routing messages known as forward and backward ants. Each node periodically dispatches a forward ant to a destination chosen probabilistically from its routing table. The ant records the time at which it enters and leaves each node. When it reaches its destination it returns to the source as a backward ant; information gathered on the forward journey about link states and latencies is left at each node along the path, where it is incorporated probabilistically into the routing table.

In both AntNet and Q-routing, nodes base their routing decisions on information about network paths that is supplied by other nodes. This approach is unsuitable for untrusted networks, where path information could be corrupted by malicious or faulty nodes.

A different kind of ant-inspired routing is used in the MUTE file sharing network [20]. MUTE is a peer-to-peer overlay in which each node adopts a random overlay address for sending and receiving messages. Messages are routed across the overlay using a probabilistic reverse-path forwarding protocol: every message carries the overlay addresses of its originator and destination, allowing relays to learn the reverse path to the originator before forwarding the message towards the destination. If no path to the destination is known, the message is broadcast; if several paths are known, a path is chosen at random with probability proportional to the number of messages received from the destination along that path. Through this simple process of local adaptation, it is possible to discover routes across the network without knowing its membership or structure. However, this form of reverse-path forwarding is vulnerable to address spoofing: an attacker could divert messages addressed to a victim by sending messages using the victim's overlay address.

Address spoofing raises an issue that is central to routing protocols: identity. Most routing protocols use a globally unique name or address to identify each node; it is usually assumed that accidental address collisions can be resolved (for example by choosing a new address when a collision is detected [6]) and deliberate collisions can be prevented (for example through the use of signed route advertisements [23, 18], or by using certified identities for all nodes [1, 2]). However, these assumptions may not hold in an untrusted network without a centralised public key infrastructure. Routing protocols for such networks must therefore be designed to cope with attackers who may deliberately use the identities of other nodes [7, 8], use more than one identity at once [12], or change identities to escape the consequences of past behaviour [15]. Open-membership networks must also cope with the more mundane aspects of identity management, such as scalability and churn: it may not be practical for every node to be aware of the structure and membership of a constantly changing network, even if no attack is taking place.

Open networks also present new challenges to privacy: anyone who can participate in a network can gather information about the activities of other users. Even if all traffic is encrypted, traffic analysis can reveal a great deal of information about who communicates with whom, when, and for how long [9, 22, 17]. We believe that protocols for open networks should aim to protect the privacy of users by minimising the information revealed to eavesdroppers, so our adaptive routing protocol is designed to preserve *unlinkability* between originators and destinations [24].

3 ADAPTIVE ROUTING IN THE DARK

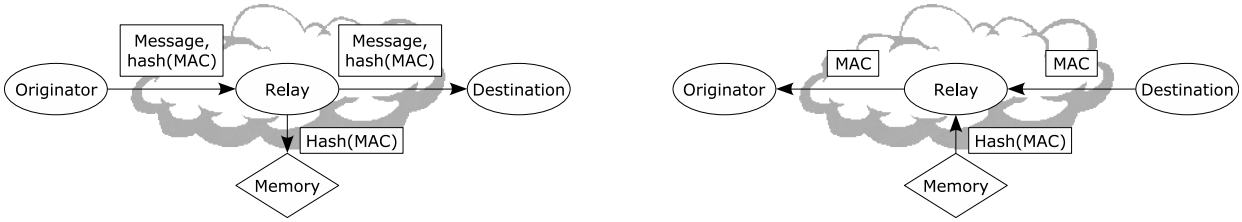
Because of the difficulties surrounding identity and privacy in open-membership networks, we are interested in the question of whether routing can operate succesfully when nodes have only minimal knowledge of the network. In this section we describe an adaptive routing protocol that uses end-to-end feedback to guide routing decisions without identifying the endpoints to the relays.

3.1 Unforgeable acknowledgements

Our adaptive routing protocol uses end-to-end (originator to destination) *unforgeable acknowledgements* (U-ACKs) that can be verified by relays without establishing a security association with either of the endpoints. Unlike a digital signature scheme, relays do not need to share any keys with the originator or destination, or to know their identities.

A full description of the U-ACK mechanism can be found in [14]. Unforgeable acknowledgements make use of two standard cryptographic primitives: *message authentication codes* (MACs) and *collision-resistant hashing*. Before transmitting a message, the originator computes a MAC over the message using a secret key shared with the destination. (Any standard key agreement mechanism appropriate to the application can be used to establish the shared key.)

Figure 1. The U-ACK protocol: relays use the hash attached to the message to verify the acknowledgement.



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Instead of attaching the MAC to the message, the originator attaches the *hash of the MAC* to the message. Relays store a copy of the hash when they forward the message. If the message reaches its destination, the destination computes a MAC over the received message using the secret key shared with the originator. If the hash of this MAC matches the hash received with the message, then the destination has validated the message, and it sends the *MAC as an acknowledgement*. The acknowledgement is forwarded back along the path taken by the message. Relays can verify that the acknowledgement hashes to the same value that was attached to the message, but they cannot forge acknowledgements for undelivered messages – they lack the secret key to compute the correct MAC, and the hash function is collision resistant. Thus a U-ACK proves to the originator and relays that the message was delivered unmodified to its intended destination, without revealing the destination's identity to the relays.

In the above discussion we have assumed that all links between neighbours are bidirectional – if a message can be sent in one direction, an acknowledgement can be sent in the opposite direction. For the purposes of our protocol, nodes that are only connected by unidirectional links are not considered to be neighbours, and the protocol cannot discover routes that contain unidirectional links. This issue is discussed in more detail in [14].

3.2 Local adaptation

Nodes in our protocol require only minimal knowledge of the network; in fact we assume that each node knows nothing about the network beyond its immediate neighbours. Each node must be able to identify its neighbours for the purposes of forwarding, but these identities need not be cryptographically verifiable, and a node is free to use a different identity when dealing with each neighbour. *Our protocol does not use end-to-end addresses*. U-ACKs allow relays to discover which messages have reached their destinations without identifying those destinations; using this information, nodes can attempt to learn which messages should be forwarded to which neighbours in order to maximise the number of messages delivered. As with Q-routing, AntNet and MUTE, this process of local adaptation can lead to globally efficient routing.

We use the following general approach for discovering dependable routes:

- Each node keeps a small pool of messages that are waiting to be sent
- For each message in the pool and each potential next hop, the node estimates the probability that an acknowledgement will be received if the message is sent to the next hop
- The message and next hop with the highest probability are chosen
- The next hop is removed from the message's list of potential next hops, and the message is sent to the next hop
- Information about the message and the next hop is recorded in the message table (see Section 3.9)
- When the message is acknowledged or times out (see Section 3.7), the information in the message table is used to update the node's dependability estimators

The size of the message pool is limited; in the simulations described in Section 4, the pool can hold five messages. Messages that have been sent to all potential next hops are removed from the pool. When a new message is added to the pool and the pool is already full, the message with the lowest remaining probability of being acknowledged (which may be the new message) is discarded.

3.3 Dependability estimators

Within the general framework described above there are many possible ways to evaluate a message's dependability, but to achieve the best results, any information that may indicate the message's relationship to earlier messages should be taken into consideration. Even without end-to-end addresses, the identities of the previous and next hops provide some information that can be used to distinguish between messages. Timing is also significant: changes in the network topology and traffic levels, even if they are not directly visible to the node, make new information more relevant than old information when estimating dependability. Thus our first attempt at a dependability estimator is a simple exponentially weighted moving average for each pair of neighbours (previous hop and next hop). The moving average is adjusted upwards whenever a message is acknowledged (equation 1), and downwards whenever a message times out (equation 2):

$$x_{i+1} = \alpha x_i + (1 - \alpha) \quad (1)$$

$$x_{i+1} = \alpha x_i \quad (2)$$

where x_i is the estimate before updating the moving average and x_{i+1} is the estimate afterwards. The parameter a determines the sensitivity of the estimator; in our simulations the value of a was 0.9. We will see in Section 5 that even this simple per-pair estimator provides a considerable improvement in efficiency when compared with flooding; further improvements may be possible by designing more sophisticated estimators.

3.4 Flow identifiers

In addition to discovering implicit relationships between messages, the efficiency of adaptive routing can be improved if related messages are explicitly grouped together. We define a *flow* as any sequence of messages that have the same origin and destination and that are semantically related in some way, such as the sequence of messages that make up a single file transfer. To indicate the existence of a flow, the originator marks all messages in the flow with an arbitrary *flow identifier*. The contents of the flow identifier are not significant – it is just a label, and it is not covered by the message authentication code. All messages in a flow are marked with the same flow identifier.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Flow identifiers have local scope: as a flow travels across the network, it may be assigned a different identifier on each link it traverses. However, messages belonging to the same flow should have matching identifiers on any given link. Each flow traversing a link must be assigned an identifier that distinguishes it from any other flows currently traversing the same link; in particular, flows arriving at a node from different upstream neighbours must be assigned distinct identifiers on any downstream link, even if they happen to have matching identifiers on their respective upstream links.

The use of flow identifiers with local scope is similar to the use of label-swapping in virtual circuits or multiprotocol label switching, but there is no requirement to establish state in the relays before data transfer begins – identifiers can be assigned to new flows on the fly.

Although they do not identify the endpoints, flow identifiers enable fine-grained dependability measurement: messages arriving from the same previous hop with the same flow identifier are likely to have the same (unknown) origin and destination, so the dependability of earlier messages in the flow can be used to estimate the dependability of later messages.

Nodes can make use of this information by keeping a separate dependability estimator for each active flow. As with the simple per-pair estimators described above, new information is more likely to be relevant than old information, so an exponentially weighted moving average is again appropriate.

To estimate the dependability of new flows, nodes also keep per-pair estimators that are only updated by the first message in each flow. This provides an estimate of the dependability of a message *given that it is the first message in a new flow* – a per-pair estimator updated by every message would tend to overestimate the dependability of new flows. The per-pair estimators are used to initialise per-flow estimators, which are thereafter updated independently.

3.5 Locally generated messages

In the preceding discussion we assumed that every message has a previous hop, but in fact any node may originate messages as well as forwarding them.

It would be inefficient to add every local message to the pool before the first copy is sent, so nodes keep a queue of messages for each locally generated flow (using a separate queue for each flow prevents head-of-line blocking). When choosing a message and a next hop, the node considers the first message in each local queue as well as the messages in the pool. If a local message is chosen, it is removed from the queue and added to the pool, since additional copies may later be sent to other neighbours.

The dependability estimators for locally generated messages are similar to those described above for forwarded messages.

3.6 Duplicate detection

Duplicate messages should be detected and discarded to prevent routing loops and reduce redundancy. However, before discarding a duplicate message, the previous hop is added to the corresponding record in the message table (see Section 3.9). If a U-ACK for the message is received, a copy of the U-ACK is returned to every previous hop listed in the record. This proves to all the previous hops that the message was delivered, providing a relatively lightweight way for nodes to maintain information about alternative routes in case the existing route fails.

3.7 Timeouts

Information about outstanding messages cannot be stored indefinitely, and it is important to update dependability estimators in a timely fashion. Therefore a node must at some point conclude that an outstanding message is not going to be acknowledged, decrease the dependability estimator, and remove the corresponding record from the message table.

A relay that receives an acknowledgement after discarding the corresponding record cannot verify or forward the acknowledgement, so there is no reason for a relay to keep records for longer than its upstream or downstream neighbours. Fixed timeouts are a simple way to ensure that adjacent relays discard their records at approximately the same time, minimising wasted storage; the choice of an appropriate timeout is discussed in [14].

3.8 Aging and discounting

From the description given in Section 3.2 it might appear that adaptive routing is likely to produce a large number of redundant messages. Aging and discounting are two techniques designed to improve the efficiency of routing by reducing the likelihood of sending redundant messages.

The technique of *aging* is based on the observation that an acknowledgement arriving after the timeout will not be recognised, since the corresponding record will have been discarded. Similarly, an acknowledgement arriving near the timeout is likely to miss the timeout at the next node. Thus the probability of an acknowledgement reaching the originator decreases for as long as the message is held in the pool, reaching zero at the timeout. The dependability of each message in the pool should ideally be aged using the expected arrival time of the acknowledgement, but that would require relays to keep an estimate of the round-trip time for each flow. A simpler alternative is to use the current time, for example by decreasing a message's dependability linearly from the time it enters the pool to the time it expires.

The second technique, *discounting*, is based on the observation that each additional copy of a message that is sent is increasingly likely to be redundant. The higher the dependability of the copies sent so far, the more likely it is that an additional copy will be redundant, so an additional copy should be sent if and only if the dependability of the copies sent so far is low. (This is also desirable for the network

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

as a whole, because it will lead to more exploration on new or damaged routes, and less exploration on well-established routes.)

Ideally we would like to calculate the conditional probability of an additional copy of the message being acknowledged, given that none of the previous copies is acknowledged first. However, it would be impractical to store all the information needed to estimate the conditional probability for each neighbour given any possible combination of previous neighbours, so in practice it is necessary to treat the probabilities as independent.

Let x_i denote the probability of the *ith* copy of the message being acknowledged, and let y_i denote the conditional probability of the *ith* copy being acknowledged, given that none of the previous copies is acknowledged first. Then, under the simplifying assumption of independence:

$$y_1 = x_1$$

$$y_i = (1 - \sum_{j=1}^{i-1} y_j)x_i \quad (3)$$

where x_i is the estimate before discounting and y_i is the discounted estimate. As the sum of the estimates for previous copies approaches one, the estimate for an additional copy will approach zero. The calculation can be made more efficient by keeping a running total.

3.9 Storage overhead

Our protocol has modest bandwidth and computation overheads: each acknowledgement is the size of a message authentication code (typically around 20 bytes), and only one hash computation is required to verify an acknowledgement. However, the storage overhead may be more significant. We offer some rough calculations below; the exact figures will depend on implementation decisions such as the choice of hash function, and application characteristics such as the link speed and message size.

Nodes store information about outstanding messages in the *message table*. Each record includes the hash of the expected acknowledgement, the previous hops of all received copies of the message, the next hops of all sent copies, and pointers to any dependability estimators that will need to be updated when the message is acknowledged or times out.

The size of a node's message table depends primarily on its outgoing bandwidth. If we assume a timeout of 60 seconds and a typical message size of 1000 bytes, a node may have up to 60 messages outstanding for every kB/s of outgoing bandwidth. If each record occupies 100 bytes, a typical peer-to-peer node with 32 kB/s of outgoing bandwidth would need to allocate 192 kB of storage for its message table.

Flow identifiers introduce a second source of storage overhead: the *flow table*. For each flow a node is currently forwarding, the node must record the mapping between the flow identifier on the incoming link and the flow identifier on the outgoing link, together with a per-flow dependability estimator. The number of active mappings is limited by the node's outgoing bandwidth, since each outgoing message reactivates one mapping. If we assume that mappings are discarded after 60 seconds of inactivity and each mapping occupies 100 bytes, the node described above would need a further 192 kB of storage for its flow table.

All the information in the flow table is soft state: it does not need to survive across restarts, and information about inactive flows can be discarded to reclaim space. When information about a flow is discarded, the flow is not cut off, but the route will need to be rediscovered if the flow later becomes active again.

4 SIMULATIONS

In this section we describe simulations designed to test the feasibility of adaptive routing without end-to-end addresses. The experimentswere conducted using a discrete event-based simulator written in Java; the simulation code is available from the authors on request.

Each data point was based on five independent runs of the simulator with different random seeds – the error bars in the figures show the maximum and minimum values obtained in any run. Each run lasted for two hours of simulated time, and the measurements were taken over the course of the second hour.

The network topologies were classical Erd¨os-R´enyi random graphs with an average degree of 10; we obtained similar results for scale-free graphs. The number of nodes was varied from 20 to 1000.

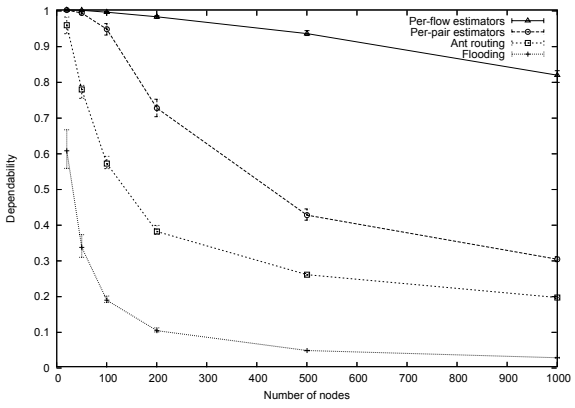
Throughout each run, churn was simulated by removing nodes from the overlay at random and replacing them with new nodes. Node lifetimes were exponentially distributed with a mean of two hours.

Messages were sent from randomly chosen originators to randomly chosen destinations; no node was the destination of its own messages. Messages were organised into flows of exponentially distributed length, with an average of 1000 messages per flow (the effect of varying the flow length is examined in Section 4.3). We obtained similar results for flows with constant and exponentially distributed inter-message delays; the results presented here are for constant delays. All messages were 1000 bytes in size and acknowledgements were 50 bytes. Each node had 32 kB/s of outgoing bandwidth and unlimited incoming bandwidth – these values are meant to represent the approximate capacity of nodes in current peer-to-peer networks. Each node was the originator of one flow at a time on average, for an average offered load of 1 kB/s per node.

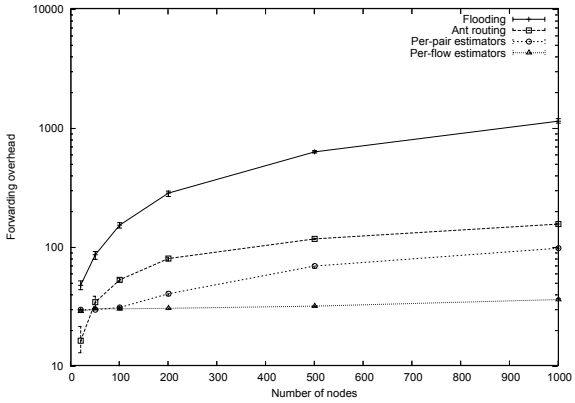
In each run we measured the *dependability*, defined as the fraction of messages that were successfully delivered, and the *forwarding overhead*, defined as the number of messages

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Comparing the scalability of adaptive routing, ant routing, and flooding.



(a) Dependability as a function of the network size.



(b) Forwarding overhead as a function of the network size (note the logarithmic scale on the vertical axis).

forwarded divided by the number of messages successfully delivered.

4.1 Scalability

The first experiment compared the scalability of adaptive routing to two existing protocols: flooding and ant routing. We chose these protocols for comparison because, unlike most other routing protocols but in common with our protocol, they only require local knowledge of the network. The flooding implementation used a drop-tail queue with a capacity of 20 messages. The ant routing implementation was similar to that used by MUTE, as described in Section 2, with a 20-message drop-tail queue. Varying the queue size did not have a significant impact on the results.

Two variants of adaptive routing were tested. The first variant, per-pair estimators, maintained one dependability estimator per pair of neighbours. The second variant, per-flow estimators, used flow identifiers as described in Section 3.4 and maintained one dependability estimator per flow, plus one estimator per pair of neighbours to initialise the estimators for new flows. Both variants selected the message and next hop with the highest probability of receiving an acknowledgement. The pool had a capacity of five messages; varying the pool size did not have a significant impact on the results.

The results of the scalability experiment are show in Figures 3(a) and 3(b). Flooding cannot support dependable communication even in small networks, due to the large number of redundant messages it produces. Ant routing works well in small networks but does not scale. When compared with flooding, per-pair estimators clearly improve dependability and reduce forwarding overhead at all network sizes. Per-flow estimators perform even better, with the result that adaptive routing in a 1000-node network has higher dependability and lower overhead than flooding in a 20-node network. However, even with per-flow estimators, dependability is only 82% in a network of 1000 nodes, suggesting that adaptive routing may not scale to very large networks.

4.2 Resilience to faulty nodes

The second experiment examined the impact of faulty nodes, which do not forward messages for other nodes, but continue to place load on the network by originating and acknowledging messages. Routes passing through faulty nodes are unusable, but the neighbours of a faulty node cannot trivially detect that it is faulty, because it will still acknowledge messages for which it is the destination; a relay does not know whether a node that returns acknowledgements is the destination or just another relay. To achieve our goal of censorship resistance, it is important to know how our protocol is affected by these faulty nodes.

Figure 4(a) shows the impact of faulty nodes on the four routing methods described in the previous section: flooding, ant routing, per-pair estimators and per-flow estimators. The performance of flooding is largely unaffected by the presence of faulty nodes. Interestingly, the performance of ant routing actually increases when up to 40% of the nodes are faulty, and the same is true of per-pair estimators with up to 20% faulty nodes. This is accompanied by a decrease in the forwarding overhead (not shown here due to space restrictions), suggesting that the improvement might be due to a reduction in the number of redundant messages.

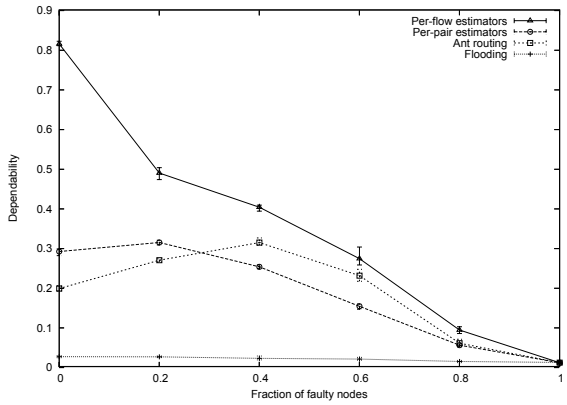
Increasing the number of faulty nodes eventually reduces the performance of ant routing and adaptive routing to the same level as flooding. Dependability does not reach zero even when all nodes are faulty, because communication can still succeed when the originator and destination are neighbours, even though no forwarding is taking place. It would be interesting to compare these results with the effect of simply removing the faulty nodes from the network.

4.3 Flow length

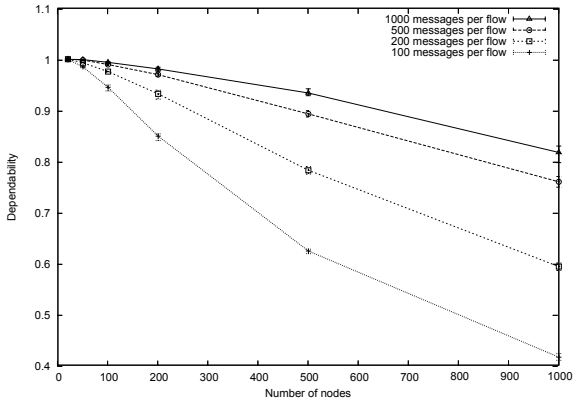
All of the simulations so far have involved an average of 1000 messages per flow. Figure 4(b) shows that as the average flow length decreases, the dependability of adaptive routing with per-flow estimators also decreases. This suggests that our protocol would be most suitable for applications that involve long flows, such as video conferencing, file transfer or instant messaging. (Note that it is the number of messages sent be-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. The effect of varying the number of faulty nodes (left) and the flow length (right).



(a) Dependability as a function of the fraction of faulty nodes, in a network of 1000 nodes.



(b) Dependability of per-flow estimators as a function of the network size, for various flow lengths.

tween the same endpoints, rather than the amount of data transferred, that determines the suitability of our protocol.)

5 DISCUSSION

The simulation results clearly show that local adaptation without knowledge of the endpoints can outperformflooding. This is not much of a boast – flooding is known to perform poorly in large networks – but it shows that knowledge of the network topology is not a precondition for making intelligent routing decisions. Our protocol also outperforms ant routing without relying on end-to-end overlay addresses that might be vulnerable to eavesdropping or spoofing.

Adaptive routing works by identifying the implicit and explicit relationships between messages – these include timing, previous and next hops, and flow identifiers. The performance of the protocol therefore depends on the number and strength of those relationships. Long-lived flows give the network time to identify dependable routes, and the initial cost of route discovery can be amortised over the lifetime of the flow (the same is true of routing protocols with an explicit route discovery phase, such as DSR [16]). We therefore believe the overall performance of our protocol would be improved by a more scalable method of route discovery.

5.1 Applications

Adaptive routing could be useful in private peer-to-peer overlays or *darknets*, where information about the structure and membership of the network is intentionally withheld for reasons of security and privacy. The protocol might also be applicable to mobile *ad hoc* networks, where accurate information about the topology may be unavailable due to node mobility, variable signal conditions, and the previously mentioned problems of identity and trust common to all open-membership networks.

In terms of traffic analysis, flow identifiers reveal less information than end-to-end addresses: an eavesdropping relay can determine how much information is being sent, and when, but not from whom or to whom. The roundtrip time between sending a message and receiving an acknowledgement might reveal the network distance to the destination, but the network distance to the originator would still be unknown – an attacker would need to observe multiple nodes or network links to trace a flow from its origin to its destination. An attacker with knowledge of the network topology would be able to assign higher probability to some originators than others [3]. We do not expect our protocol to provide unlinkability under the stronger attack models typically used to evaluate high-latency mix networks [25, 26].

Our protocol would not be suitable for devices with very little storage capacity, such as mobile phones, due to the overhead described in Section 3.9. The simulation results also suggest that our protocol is unlikely to be suitable for very large networks – the route discovery process does not scale well in its current form. Because of the need to amortise the cost of route discovery over the length of the flow, adaptive routing is more likely to be useful for applications that produce long flows of messages between the same endpoints – such as video conferencing, file transfer and instant messaging – than for applications that produce short flows between a large number of endpoints, such as web browsing and email.

5.2 Future work

The simulation results presented here are only preliminary. Much work remains to be done to evaluate adaptive routing in a wider range of scenarios: issues to consider include churn, mobility, heterogeneous bandwidth, and complex topologies such as social networks.

We are currently exploring ways to improve the scalability of route discovery using a meet-in-the-middle technique based on the U-ACK mechanism. The simple dependability estimators described in this paper can doubtless be improved upon, perhaps by using control theoretic techniques such as Kalman filters.

We would also like to explore our protocol's resilience to a wider range of malicious and strategic behaviour. Classical approaches to Byzantine fault tolerance involve strong authentication [19, 23, 5]; we would like to see whether it is possible to achieve weaker probabilistic guarantees with-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

out authentication, by reducing the information available to faulty nodes in order to restrict the potential complexity of their misbehaviour.

REFERENCES

1. Avramopoulos I., Kobayashi H. and Wang R. (2003), 'A routing protocol with Byzantine robustness', in IEEE Sarnoff Symposium.
2. Awerbuch B., Curtmola R., Holmer D., Nita-Rotaru C. and Rubens H. (2005), 'On the survivability of routing protocols in ad hoc wireless networks', in Proceedings of the 1st International Conference on Security and Privacy for Emerging Areas in Communication Networks (SecureComm 2005), Athens, Greece, IEEE Computer Society Press, pp. 327–338.
3. Borisov N. (2005), Anonymous Routing in Structured Peer-to-Peer Overlays, Ph.D. thesis, UC Berkeley.
4. Boyan J. and Littman M. (1994), 'Packet routing in dynamically changing networks: A reinforcement learning approach', Advances in Neural Processing Systems, 6:671–678.
5. Castro M. and Liskov B. (1999), 'Practical Byzantine fault tolerance', in 3rd Symposium on Operating Systems Design and Implementation, New Orleans, LA, USA.
6. Cheshire S., Aboba B. and Guttman E. (2005), 'RFC 3927: Dynamic configuration of IPv4 link-local addresses', IETF standard.
7. Cla ers/cache/ipspoof1.txt (accessed September 2007).
8. Claerhout B. (1997), 'A short overview of IP spoofing: Part II', Available from http://www.cs.ucl.ac.uk/staff/ mrogers/cache/ipspoof2.txt (accessed September 2007).
9. Danezis G. and Clayton R. (2007), 'Introducing traffic analysis', in A. Acquisti, S. di Vimercati, S. Gritzalis and C. Lambrinoudakis

(eds.), Digital Privacy: Theory, Technologies, and Practices, CRC Press.

10. DiCaro G. and Dorigo M. (1998), 'Ant colonies for adaptive routing in packet-switched communications networks', in

Proceedings of the 5th International Conference on Parallel Problem Solving from Nature, Amsterdam, The Netherlands,

Springer-Verlag, vol. 1498 of Lecture Notes in Computer Science, pp. 673–682.

11. DiCaro G., Ducatelle F. and Gambardella L. (2005), 'AntHocNet: An adaptive nature-inspired algorithm for routing in mobile ad hoc networks', European Transactions on Telecommunications, 16(5).
12. Douceur J. (2002), 'The Sybil attack', in P. Druschel, F. Kaashoek and A. Rowstron (eds.), Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), Cambridge, MA, USA, Springer-Verlag, vol. 2429 of Lecture Notes in Computer Science, pp. 251–260.
13. Elmer-DeWitt P. (1993), 'First nation in cyberspace', Time, 142(24).
14. This reference has been removed during the anonymous review process.
15. Friedman E. and Resnick P. (2001), 'The social cost of cheap pseudonyms', Journal of Economics and Management Strategy, 10(2):173–199.
16. Johnson D., Maltz D. and Broch J. (2001), 'DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks',

in C. Perkins (ed.), Ad Hoc Networking, Addison-Wesley, chap. 5, pp. 139–172.

17. Karagiannis T., Papagiannaki D. and FaloutsosM. (2005), 'BLINC: Multilevel traffic classification in the dark', in SIGCOMM 2005, Philadelphia, PA, USA.
18. Kent S., Lynn C., Mikkelson J. and Seo K. (2000), 'Secure border gateway protocol (S-BGP) - real world performance and deployment issues', in ISOC Symposium on Network and Distributed System Security, San Diego, CA, USA.
19. Lamport L., Shostak R. and Pease M. (1982), 'The Byzantine generals problem', ACM Transactions on Programming Languages and Systems, 4(3):382–401.
20. MUTE website, http://mute-net.sourceforge.net/ (accessed September 2007).
21. Pain J. (ed.) (2006), Internet Annual Report, Reporters Without Borders.
22. Partridge C., Cousins D., Jackson A., Krishnan R., Saxena T. and Strayer W. (2002), 'Using signal processing to analyze wireless data traffic', Technical Memorandum 1321, BBN Technologies, Cambridge, MA, USA.
23. Perlman R. (1988), Network Layer Protocols with Byzantine Robustness, Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
24. Pfitzmann A. and K¨ohntopp M. (2001), 'Anonymity, unobservability, and pseudonymity - a proposal for terminology', in H. Federrath (ed.), Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability, Berkeley, CA, USA, Springer-Verlag, vol. 2009 of Lecture Notes in Computer Science, pp. 1–9.
25. Raymond J. (2001), 'Traffic analysis: Protocols, attacks, design issues and open problems', in H. Federrath (ed.), Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability, Berkeley, CA, USA, Springer-Verlag, vol. 2009 of Lecture Notes in Computer Science, pp. 10–29.
26. Serjantov A., Dingledine R. and Syverson P. (2002), 'From a trickle to a flood: Active attacks on several mix types', in F. Petitcolas (ed.), Proceedings of the 5th International Workshop on Information Hiding (IH 2002), Noordwijkerhout, The Netherlands, Springer-Verlag, vol. 2578 of Lecture Notes in Computer Science, pp. 36–52.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Secure communication with SSL remote access VPN

Olalekan Adeyinka, Charles Shoniregun

School of Computing and Technology,
University of East London
olavalen@hotmail.com and c.shoniregun@uel.ac.uk

Abstract Web enabled applications improve the accessibility of business information using SSL VPNs Remote-access that allows the secure access to corporate resources by establishing an encrypted tunnel across the internet using SSL protocol that became the standard for secure web communications. The SSL does not need to be downloaded onto the device being used to access corporate resources and the use is ideal for the mobile users. Security is the cornerstone of any remote access implementation but by sharing information over the web can lead to security risks that must be carefully addressed, SSL VPN appliances can quickly integrate into the network and providing companies with a rapid deployment solution. The SSL which is available without additional software deployment on all standard Web browsers as a secure transport mechanism for users who wants VPN access from anywhere at any time required the use of SSL VPN which is flexible and offers Platform Independence compare to IPSec VPN because they connect to the network through a web browser. This paper investigates the concept of SSL VPN as a technology for remote access communication.

1 INTRODUCTION

The Secure Sockets Layer Virtual Private Network (SSL VPN) is the goal of extending web usage to actual business activity involving the transmission of confidential information over the internet and the need to eliminate eavesdropping by unauthorized parties on web communications between client computers over the internet.

Web pages are delivered using the Hypertext Transfer Protocol (HTTP), the protocol does not offer encryption or any protection of data transmitted between users and web servers. There are two primary methods for deploying remote-access VPNs, IP Security (IPSec) and SSL. Each method has its advantages based on the access requirements of users and IT organizations. In today's markets, security products such as SSL VPNs have become the solution for remote access connectivity, as it Provides secure communications with access rights to individual users, for instance employees, contractors, or partners and are often sold as appliances which consist of standard computers running SSL VPN software on a standard operating system, that actually reduces the overhead cost of installing, configuring, and maintaining IT systems. SSL VPN products tend to provide more granular tools, because they operate at the session layer, they can filter and make decisions about user or group access to individual applications (ports), selected Uniform Resource Locator (URLs), application commands embedded objects and even content. IPSec VPN has it already addressed the requirements for site-to-site network connectivity, for mobile users, they were often too costly, while for business partners or customers they were impossible to deploy as they require software be installed and configured on each endpoint. It is in this environment that SSL VPNs were introduced, providing remote/mobile users, business partners and customers with the easy, secure access to corporate resources they needed. The SSL-based VPNs provide remote-access connectivity from almost any Internet-enabled location using a Web browser and also it does not require any client software to be pre-installed on the system; this makes SSL VPNs capable of connectivity from company-managed desktops and non-company-managed desktops, such as employee-owned PCs, contractor or business partner desktops, and Internet kiosks. SSL VPNs provide two different types of access: clientless and full network access.

Clientless access requires no specialized VPN software on the user desktop and VPN traffic is transmitted and delivered through a standard Web browser, Since all applications and network resources are accessed through a Web browser, only Web-enabled and some client-server applications such as intranets, applications with Web interfaces, e-mail, calendaring, and file servers can be accessed using a clientless connection.

The network resources used in the office or any client server application that cannot be delivered across a web based clientless, Network access enables access to virtually any application, server, or resource available on the network. Full network access is delivered through a lightweight VPN client that is dynamically downloaded to the user desktop (through a Web browser connection) upon connection to the SSL VPN gateway. The VPN client is dynamically downloaded and updated without any manual software distribution or interaction from the end user requires little or no desktop support by IT organizations. Both IPSec and SSL VPN technologies offer access to any network application, but

SSL VPNs offer additional features such as easy connectivity from non-company-managed desktops, little or no desktop and user-customized Web. SSL VPNs has spurred vendors to expand remote access to include applications that aren't Web-based, which result are three tiers of access, each with separate requirements. These tiers are clientless, Browser-Plus, and Network Access. With clientless SSL connections, users can only run Web-based applications in a Web browser. Browser-Plus SSL connections download a small ActiveX control or Java applet that lets the browser communicate with the application (Conry-Murray, 2004). Both IPSec and SSL technologies are actually similar in terms of transport security at a high level, effectively secure network traffic, and each has associated trade offs, which make them appropriate for different applications. The protocol implementations are differ but the two systems share similarities, including encryption and authentication and session keys. Each of them offers encryption, adapt integrity and authentication technologies; 3-DES, 128 bit RC4, AES, MD5 or SHA

2 SSL PROTOCOL

The Secure socket layer (SSL) is considered a suite of protocols that actually uses many different standards of key exchange, authentication and encryption to get its job done. Server typically provides regular web service http on port 80, and SSL-encrypted web traffic https over port 443.

The SSL 1.0 protocol was designed and released in 1994. Version 2.0 was also released in 1994 to fix bugs presented in version 1.0 and to clarify some aspects of the protocol. SSL version 2.0 was vulnerable to some attacks and SSL version 3.0 was released in 1995 to fix the flaws in version 2.0. The flaw discovered in version 2.0 includes cipher suite rollback attack and version rollback attack. The SSL protocol is owned by Netscape and they approached the Internet Engineering Task Force (IETF) to create an internet standard, an IETF protocol definition RFC. 2246 is the process of becoming an internet standard (Zwicky, 2000). The protocols based very heavily on SSL version 3 and are called transport layer security (TLS). Both TLS and SSL use exactly the same protocol. The most well known application protocol that is run on top of the secure socket layer is the Hypertext Transmission Protocol, commonly known as HTTPS when run over Secure Socket Layer. HTTPS connections are based on HTTP protocol over SSL connections to provide authentication, confidentiality and integrity using symmetric and asymmetric cryptographic algorithms using private or public key, (Vicenc, 2004).

SSL is a standard way to achieve a good level of security between a web browser and a website. SSL is designed to create a secure channel, or tunnel, between a web browser and the web server, the secure tunnel is just for you and the web server, so that any information you exchange is protected within the secure tunnel. SSL are good choices for adding end-to-end protection to applications, it protects against eavesdropping, session hijacking and Trojan servers. SSL can be applied to online security and privacy that provide Authentication, Integrity, Confidentiality and Non-repudiation. SSL provides for authentication of clients to server through the use of certificates, Clients present a certificate to the server to prove their identity. The primary security service that the Secure Socket Layer provide is the "protection of data" while the data is on the wire (Shin, 2006).

Secure Socket Layer (SSL) allows client/server applications to communicate in a way that is designed to prevent eavesdropping, tampering, or message forgery. To obtain these objectives, it uses a combination of public key and private key cryptography algorithms and digital certificates (X.509) which provides the basic four data security requirements, confidentiality of the data while it is on the wire, It also supports data integrity known as tamper-proofing where the recipient of a message can verify that the contents have not been altered since it was generated by a legitimate source and It ensures authentication where both ends of a communication must identify each other. SL connections require two parties, (Vicenc, 2004). On one hand is your SSL-enabled web browser. On the other hand is the SSL-enabled website you are visiting. You are the client and the web site is the server. For SSL to work, both parties must support it. SSL can also be used to secure the authentication and delivery of e-mail using the pop3 and SMTP protocol (Zwicky, 2000). It is important to realize that SSL is not strictly a web protocol but function at the session and transport layer of the OSI model and can establish encrypted communication tunnels for various application-level protocols that may sit above it. SSL enabled website are typically using the server's digital certificate to establish the SSL session. That is the server's certificate is the trusted source for the authentication of the server's identity and public keys that are used for encryption. A client certificate is one that you possess and that your web browser uses in the SSL connection. When both client and server side certificates are used both side have complete trust and both parties know each other's identities.

SSL can provide protection for many types of communications not just web surfing, in fact, SSL can be used to secure email file uploads/ downloads using file transfer protocol (FTP). It is flexible in protecting so many types of digital communications; also it offers mechanisms for authenticating clients.

SSL will work well with NAT; however the end-to-end encryption will prevent the NAT system from intercepting embedded addresses. The SSL protocol encrypts all application layer data with a cipher and short-term session key negotiated by the handshake protocol. Independent keys are used for each direction of a connection as well as for each different instance of a connection. SSL version 3.0 record layer resists powerful active attacks, an example of an attack is the cut and paste attack. This attack cuts an encrypted cipher text from some packet containing sensitive data, and splices it into the cipher text of another packet which is carefully chosen so that the receiving endpoint will be likely to inadvertently leak its plaintext after decryption, (Wagner and Schneier, 1996).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

2.1 Encryption Algorithms and Key exchange

An encryption algorithm is a mathematical algorithm for encryption and decryption of messages (arrays of bytes). There are two types of encryptions used that are electronic key. The key could be private so that both the sender and the receiver could use the same key to encrypt and decrypt the data. Most encryption is now public-key encryption which involves the use of two different keys. The message is encrypted using one of the keys and decrypted with the other key. And, this type of encryption is referred to as the asymmetric encryption. The secured socket layer protocol uses both symmetric and asymmetric types of encryption. In an SSL-enabled web site, the server has a public key that is widely distributed to anybody who wants to encrypt communications with the Web server. The server also has a private key that it uses to decrypt the communications, the server gives its public key to the client, the client then uses the key to encrypt data for the server, but it cannot decrypt the data, but the server decrypts the data with its own private key. Public key cryptography is only used to exchange symmetric encryption keys that will be used for the majority of the SSL session.(Bahadur, 2002) Secure Socket Layer uses the Public key encryption to securely exchange the symmetric keys. Symmetric keys are the same keys used by both the server and browser to encrypt and decrypt the data. The Secure Socket Layer (SSL) client only performs RSA public key encryption rather than the private key operations for signature verification and encryption. (Gupta, V and Gupta, S, 2002)

The RSA encryption is an encryption algorithm developed in 1977 in MIT and was named after its investors Ronald Rwest, Adi Shamar and Leonard Adleman. The algorithm offers both encryption as well as digital signatures (authentication). Each participant has a private key shared with no one else and a public key that is known to everyone.

SSL supports different encryption algorithms; the algorithms that are available for a particular encrypted session vary based on SSL version, company policies, and governmental restrictions. There are two types of cryptography used within each SSL session, Symmetric and Asymmetric. While symmetric encryption is used for encrypting all the communications within an SSL session, an asymmetric algorithm is used to share the symmetric session key securely between the user and the SSL VPN, (Steinberg and Speed, 2005)

Symmetric cryptography: data confidentiality, Symmetric algorithms use the same key for encryption and decryption and, therefore, both parties in a conversation must share a common key. Fig (1)

Asymmetric cryptography: data confidentiality, Addresses the problem of key exchange, one key in a pair is called a public key and the other is called a private key. The public key is shared with the public and is not secret while the private key remains private and only its owner should have access to it, Fig (2).

Data encrypted with one in a key pair can only be decrypted with the corresponding key in the pair. It cannot be decrypted with the same with which it was encrypted. Public key are not secret, asymmetric cryptography does not suffer from an issue of key sharing.

Public keys can easily be transmitted over the internet; however, asymmetric cryptography is extremely processor intensive and not practical for encrypting large amounts of data. It cannot be used to encrypt an entire SSL session. Nevertheless, it is ideal for use as a mechanism to transfer symmetric keys securely across an insecure network, and it is exactly for this purpose that SSL uses it. So SSL uses asymmetric cryptography to share a secret key between the remote user and a server, and then uses that key to perform symmetric

Figure 1. Symmetric Cryptography

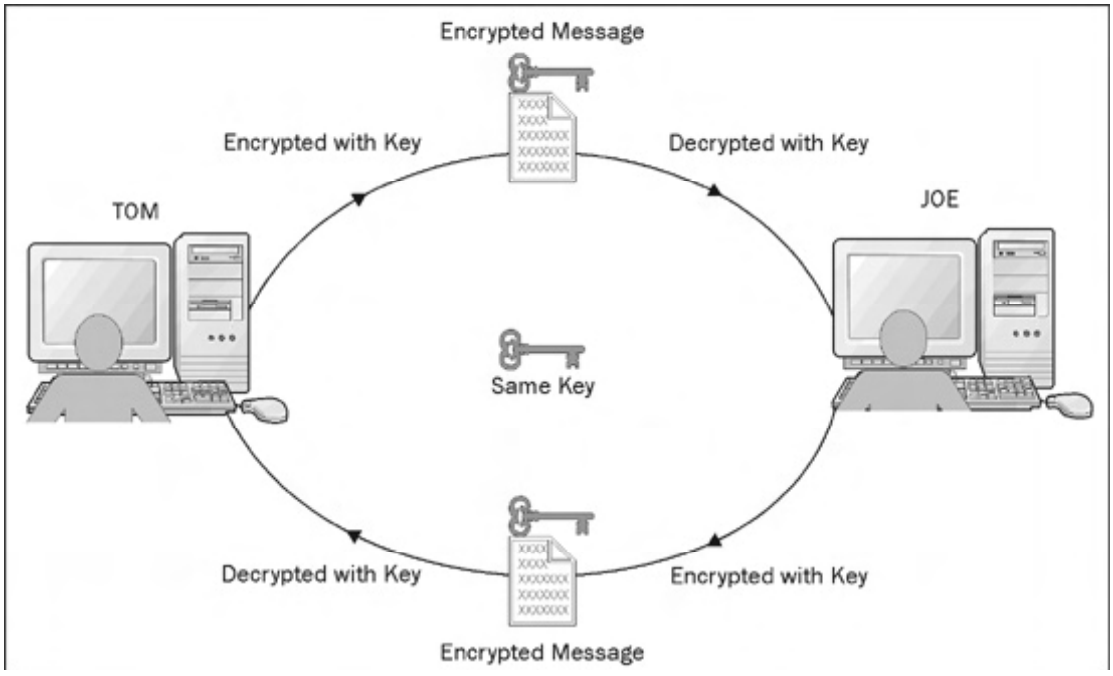
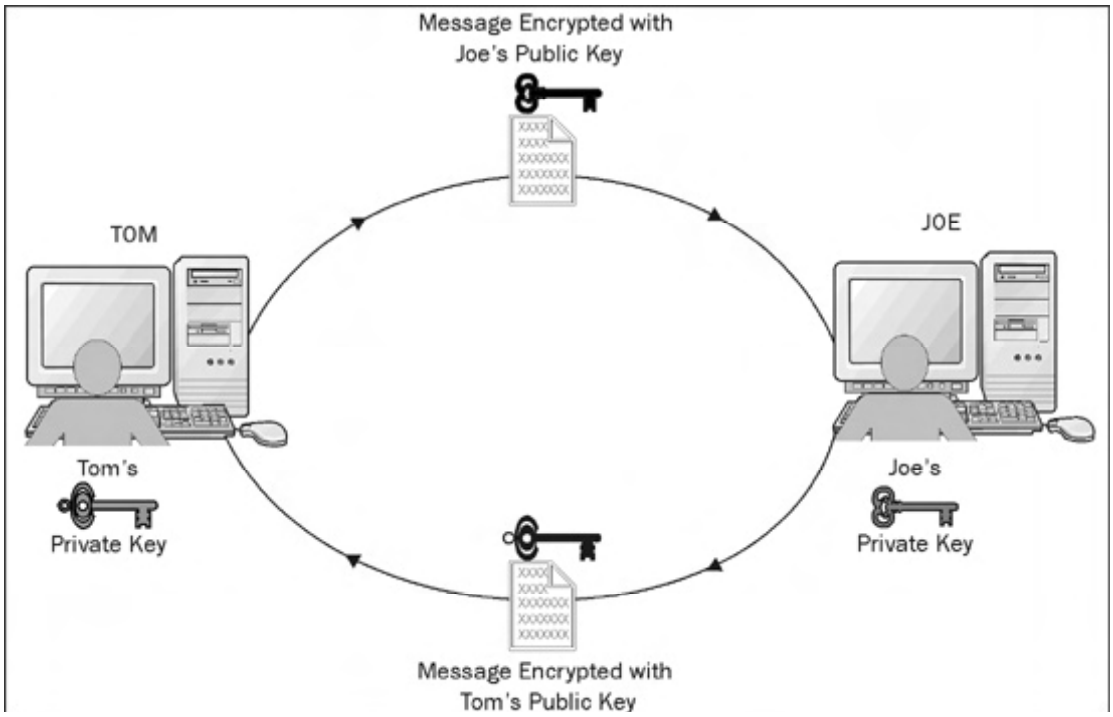


Figure 2. Asymmetric Cryptography



encryption/decryption on the data sent during the SSL session.

Asymmetric cryptography: server authentication, SSL uses asymmetric cryptography to share a secret key between the remote users and the server, then uses that key to perform symmetric encryption / decryption on the data sent during the SSL session, Fig(3).

SSL certificates are a mechanism by which a web server can prove to users that the public key that it offers to them for use with SSL is in fact, the public key of the organisation with which the user intends to communicate. A trusted third party signs the certificate thereby assuring users that the public key contained within the certificate belongs to the organisation whose name appears in the certificate.

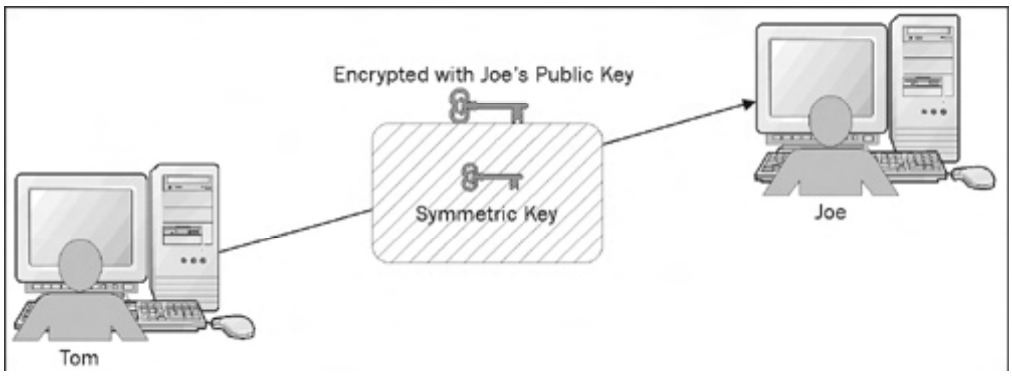
Despite the server-authentication capabilities of SSL, phishing-type fraud has reached epidemic levels. The technical expertise required to appreciate SSL anti-impersonation capabilities properly has severely limited its usefulness in the real world. You then use the server's public key to encrypt data for the server, but you cannot decrypt the data. Only the server can decrypt the data with its private keys. Public key cryptography is only used to exchange symmetric encryption keys that will be used for the majority of the SSL session.

Asymmetric cryptography: Client Authentication. SSL provides authentication of clients to servers through the use of certificate, and the client present a certificate to server to prove their identity. It allows SSL VPN servers to identify client machine of different trust levels.

3 OVERVIEW OF IPSEC VPN

IPSec based VPNs are the deployment-proven remote access technology used by most organizations today to established connection using pre-installed VPN client software to connects hosts to entire private networks by protecting the IP packet exchanged between remote networks or hosts and an IPSec gateway located at the edge of your private network. IPSec VPNs can support all IP-based applications to

Figure 3. Asymmetric Cryptography server authentication



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

an IPSec VPN product, all IP packets are the same, (Phifer, 2003) With the IPSec client software, organizations can control the function of the VPN client fuse in applications such as unattended kiosks, integration with other desktop applications, and other special use cases. IPSec authentication employs Internet Key Exchange (IKE), using digital certificates or pre shared secrets for two-way authentication. They differ significantly on how these extensions are implemented. Many organizations find that IPSec meets the requirements of users already using the technology. But the advantages of dynamic, self-updating desktop software, ease of access for non-company-managed desktops, and highly customizable user access make SSL VPNs a compelling choice for reducing remote-access VPN operations costs and extending network access to hard-to-serve users like contractors and business partners. As such, organizations often deploy a combination of SSL and IPSec approaches. IPSec vendors, for example, offer alternatives such as Extended Authentication (XAUTH) and Layer 2 Tunneling Protocol (L2TP) over IPSec. However, XAUTH, which is frequently deployed using pre shared group secrets and DHCP, is vulnerable to several known attacks, (Phifer, 2003). And while L2TP over IPSec is embedded in Windows 2000/XP, it isn't broadly supported by VPN gateways or used by non-Microsoft shops. IPSec VPN integrate poorly with Firewall and Network Address Translation (NAT) and also have a limited granular access, by operating at the network layer, it's protocol support Authentication Header (AH) and Encapsulating Securing Payload (ESP) as tunnelling and DES, 3DES, 128/192/256 bit AES as encryption.

IT administrators must determine who should have remote access to the network, because IPSec VPNs require a client to be installed on each user machine, which entails deployment, configuration and maintenance, the solution becomes resource intensive and cost prohibitive when deployed across large enterprise (Harding, 2003).

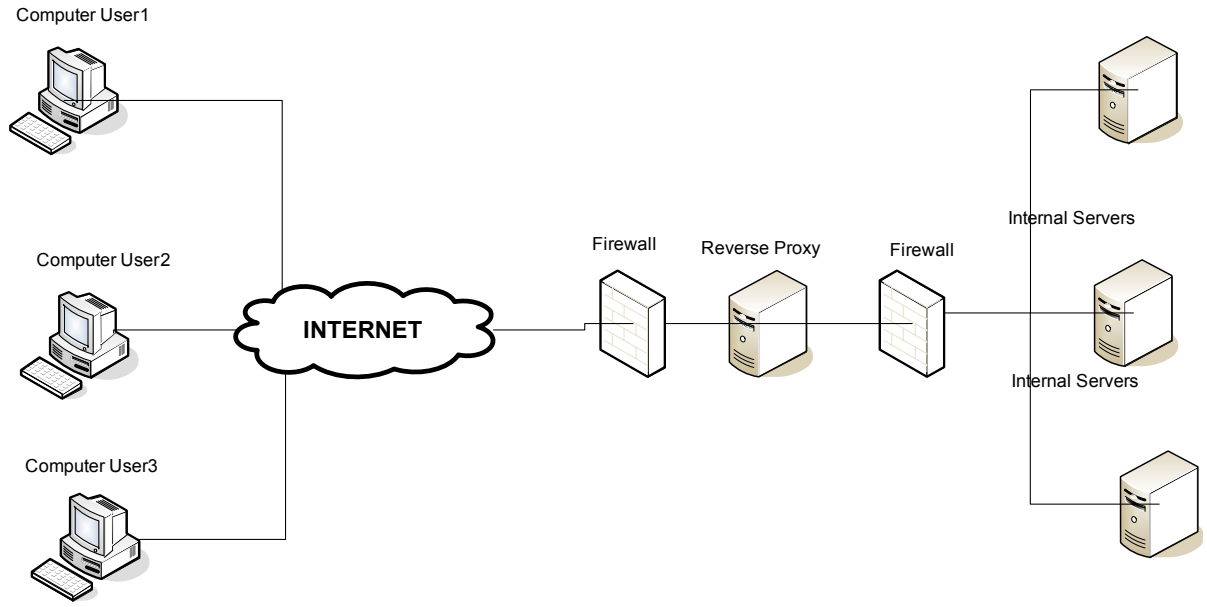
4 REVERSE PROXY TECHNOLOGY

Many SSL VPN servers utilize reverse-proxy-like technology which is also a gateway-type function called reverse proxing that accept web-based user requests and transmit them to internal resources. Today's SSL VPN adds significant additional functionality to those reverse-proxy-type functions. The most basic function of an SSL VPN is its received user's requests and relay them to internal servers. Reverse proxies are often deployed as part of load-balancing schemes, as part of a layered security strategy, or simply to hide real servers from users for security reasons, (Steinberg and Speed, 2005).

A reverse proxy server is a computer that sits between an internal web server and the internet, and appears to external clients as if it were the true web server. Most reverse proxy deployed in the middle of the firewalls, Fig (4).

Web reserve proxies are certainly appropriate technologies for improving the security of internet-accessible web-based systems and for simplifying load-balancing scenarios. However, they lack the ability to deliver many remote-access requirements and also they cannot encapsulate client/server network traffic over SSL, they cannot transform many internal applications to be internet accessible, do not provide secure access to file systems, and do not offer a remote-access-oriented user interface, That is where SSL VPN technology comes in, SSL VPNs were designed to provide granular access from any endpoint by ensuring that each endpoint is in compliance with a minimum corporate security policy is mandatory, this can be done via dynamic endpoint security checks which should be done both before a session is initiated and periodically throughout the session. Some vendor's SSL VPNs can also provide a dual mode network-layer access capability that detects the best method of connection between IPSec and SSL transport to ensure the highest level of connectivity supported by the network environment that enables the high performance required for accessing latency and jitter sensitive applications like VOIP, while providing

Figure 4. Reverse Proxy



the ubiquity and reliability that SSL VPNs are known for with none of the IPSec VPN management overhead.

5 SSL VPN TECHNOLOGY

Secure Sockets Layer (SSL) Virtual Private Networks (VPNs) are quickly gaining popularity as serious contenders in the remote access marketplace. Analysts predict that products based on SSL VPN technology will rival or even replace IP Security Protocol (IPSec) VPNs as remote-access solutions.

There are significant differences between the accesses methods, IPSEC VPNs were designed to provide site-to-site access (branch-to-branch) access.

SSL VPNs were designed to provide remote access for a mobile user to a corporate resource. The SSL VPNs is built on Secure Socket Layer (SSL), a protocol originally developed by Netscape Communications in the mid-90s which as has undergone years of public Scrutiny as the standard for secure electronic commerce (e-commerce) transactions on the Internet. SSL VPNs create secure tunnels by performing authentication from users before allowing access so that only authorized parties can establish tunnels and encrypting all data transmitted to and from the user by implementing the actual tunnel using SSL. The methodology used by SSL VPN to transport data across the public internet is different; SSL VPN technology has evolved to include a variety of different types of access via dynamically downloaded agents; these advances enable the delivery of client/server applications, as well as network-layer connections which are enabled via SSL. Dynamic delivery facilitates the use of agent-based access methods, without the cost or hassle of installing and configuring individual client software. SSL tunnel is established by the exchange of different configuration information between the computers on either end of the connection. The SSL VPN connectivity which functions at the levels 4-5 , also encapsulate information at levels 6-7 and communicate at the highest levels in the OSI model.

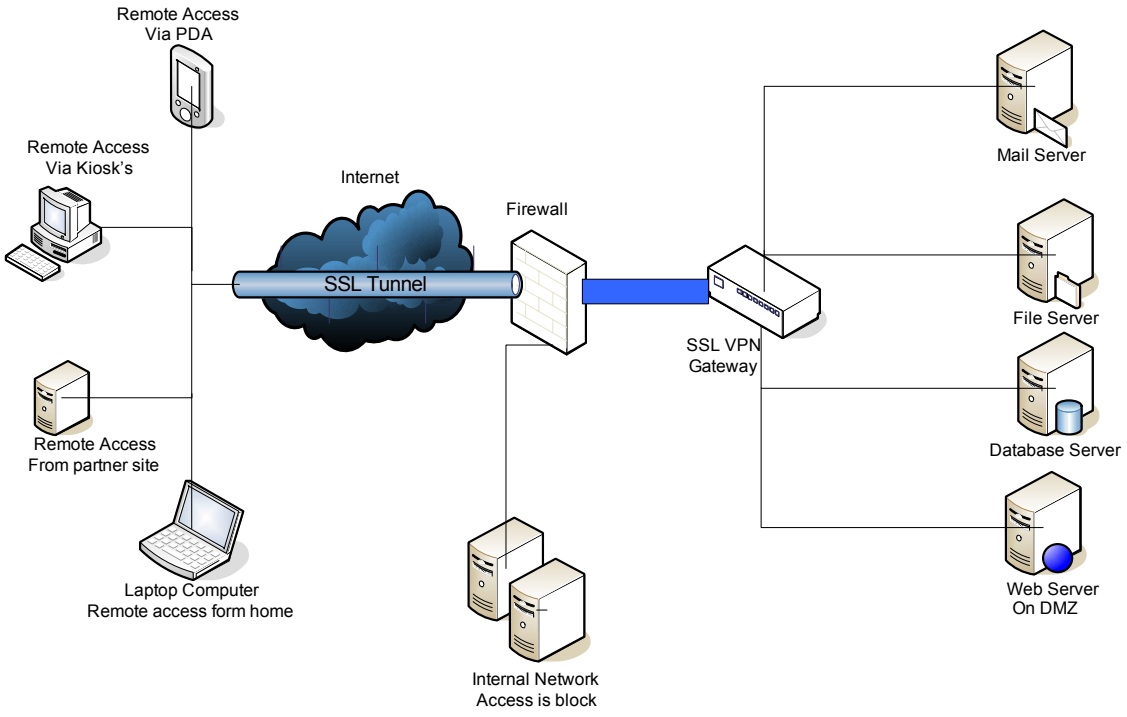
The Web servers often reside in the Demilitarized zone (DMZ) or other security zone for public access, securing this environment is risky and requires high-maintenance.

SSL VPNs have made the technology increasingly popular among IT professionals and empower organizations to avoid deploying application and Web servers in a DMZ or other security zone, where they would have to be exposed to the public Internet, with SSL VPNs, application servers remain safe on the private network, behind the firewall and are never directly exposed to the public network since SSL VPNs combine the security and confidentiality provided by SSL and the mobility of a Virtual Private Network, together they enable remote users to connect to their office networks using standard web browsers.

The SSL VPN gateway that resides in the DMZ behind a corporate firewall, where it intercepts encrypted traffic passing through port 443, by decrypting the traffic depending on the access method provides the user with a portal that includes a menu of accessible applications, or a network connection that the user's in-office experience, (Conry-Murray).

Today, most SSL VPNs provide secure access to Microsoft Outlook Web mail, network file shares and other common business applications, (Phifer, 2003). Some SSL VPNs products protect application streams from remote users to an SSL gateway and also able to tunnel network-level information over SSL, making SSL the most versatile remote access VPN technology available. However SSL VPN application services vary because each product has its own way of presenting client interfaces through browsers relaying application streams through the gateway and integrating with destination servers inside the private network. It is observed that when the web browser establishes a connection with the gateway, the

Figure 5. SSL VPN Gateway



gateway's digital certificate is verified and the session traffic is encrypted providing a secure connection, more importantly it is found that SSL VPNs complicated security pictures since users will often link to applications using untrusted computers, (Conry-Murray, 2004). SSL is already installed on any Internet-enabled device that uses a standard Web browser and no configuration is necessary simply because it's operates at the application-layer that offer extremely granular access controls to applications, making it ideal for mobile workers and those users coming from an insecure end-point, also is independent of any operating system and any changes to the operating systems do not require an update in the SSL implementation, (Rissler and Sorensen, 2005).

SSL VPN operates at the application with existing firewall infrastructure and have a high-level granular access control for applications. It encryption is DES, 3DES, AES 256bit and authentication is local user Database, Microsoft Active Directory, LDAP, NT Domain and Radius.

6 REMOTE-ACCESS SSL VPN SECURITY CONSIDERATIONS

Worms, Viruses, Spyware, hacking, data theft, and application abuse are considered among the greatest security challenges in today's networks. Remote-access and remote-office VPN connectivity are common points of entry for such threats, due to how VPNs are designed and deployed. The existing IPSec and SSL VPN installations are often deployed without proper endpoint and network security which leads to Unprotected or incomplete VPN security that allows the following network threats, (Cisco systems 2006), where remote-user VPN sessions brings Malware into the main office network, causing virus outbreaks that infect other users and network servers, also allowing users to generate unwanted application traffic, such as peer-to-peer file sharing, into the main office network causing slow network traffic conditions and unnecessary consumption of expensive WAN bandwidth, the implication enables theft of sensitive information, such as downloaded customer data, from a VPN user desktop, enables hackers to hijack remote-access VPN sessions, providing the hacker access to the network as if they were a legitimate user. To combat these threats, the user desktop and the VPN gateway that the user connects to must be properly secured as part of the VPN deployment. The user desktops should have endpoint security measures such as data security for data and files generated or downloaded during the VPN session, Anti-Spyware, antivirus and personal firewall. The VPN gateway should offer integrated firewall, antivirus, Anti-Spyware, and intrusion prevention. Alternatively, if the VPN gateway does not provide these security functions, separate security equipment can be deployed adjacent to the VPN gateway to provide appropriate protection. SSL VPNs don't require client address assignment or changes to routing inside your network, because they control access to applications and content (e.g., URLs) rather than network-layer entities, such as subnets and hosts, Typically, SSL VPN gateways are deployed behind a perimeter firewall, which requires punching a hole through that firewall to deliver SSL to the VPN gateway which means delegating trust from the firewall to the VPN gateway, that enforces security policy on SSL-encrypted streams.

Technologies required for mitigating Malware such as Worms, Viruses, and Spyware and for preventing application abuse, data theft, and hacking exist in the security infrastructure of many organizations' networks. In most cases, they are not deployed in such a way that they can protect the remote-access VPN, due to the native encryption of VPN traffic.

Additional security equipment may be purchased and installed to protect the VPN, the most cost-effective and operationally efficient method of securing remote-access VPN traffic is to look for VPN gateways that offer native Malware mitigation and application firewall services as an integrated part of the product, (Cisco systems, 2007). If you're implementing an SSL VPN, try to choose products that support TLS, which is slightly stronger than the older SSLv3. TLS eliminates older key exchange and message integrity options, ensuring strong defence against key cracking and forgery.

7 BENEFIT OF SSL VPN

The growth of SSL VPN allows controlled, secure and managed access to any application. The benefits of SSL VPNs include no client software requirement for accessing web-enabled applications which significantly increases the flexibility of the VPN solution that produced a types of services that can be accessed from anywhere in the world at anytime, Only servers require digital certificates to establish the encrypted session and many more. The SSL is available wherever there is a standard web browser and does not need to be configured by the end user, it has the ability to allow both web and non-web application to utilize the SSL tunnel for communication. Most web enabled applications such as Outlook, Exchange and Lotus Notes already support SSL there is very little configuration required, (Nettilla Networks, 2007). This also reduces the cost of implementing and supporting a VPN. SSL is built into all the leading browsers, which means easy remote access from any pc connected to the internet. The SSL VPN is operating systems independent, which means users can access the VPN regardless if they are using a UNIX or a Mac machine and regardless of if they use Internet Explorer or Mozilla browser. SSL uses port TCP/443, which is normally opened on the firewall to the DMZ, which means SSL has the benefit of not requiring any configuration changes to firewalls, and also NAT tables are not required as all information is passed via the browser and not in IP, (Ferrigi, 2003).

One of the major benefits of SSL VPN is the ability to access resources from any computer on even handheld devices at any location. All data packet exchanged include message integrity check failure which causes a connection to be closed, the identity of the server a client is connecting to is always verified, and this identity check is performed before the optional client user authentication information is sent, the client and server negotiate encryption and integrity protection algorithms, by the use of key exchange algorithms to prevent

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

man-in-middle attacks and at the end of the key exchange is a checksum exchange that will detect any tampering with algorithm negotiation. Since SSL does not allow access to subnets the danger of Trojans, Viruses and Malware being able to access internal resource is significantly reduced, does not completely eliminate the risk of malicious intent but goes a long way to reducing it.

8 SSL VPN ISSUES

Establishing connectivity at higher levels in the OSI model, involves some costs and drawbacks. The disadvantage include optical user authentication which is a major security weakness and requires java or activeX downloads to facilitate access to non-web enabled applications, (Kilpatrick, 2007). Security is another drawback, Such "tunnelling-based" VPN solutions create a direct connection from the third-party client via the SSL VPN to an application server that hosts the target application; In this case there is no intermediation and data translation, Instead application data enters the network without having been analyzed by the SSL VPN appliance, A scenario where authorization and policy occurs at the application server inside the private network rather than in the security zone at the network edge where it might have been processed and controlled, such arrangements do not gain the advantages of an integrated policy, authentication, and authorization frame-work as defined by the appliance, and leave target-application servers vulnerable to attack.

SSL VPN only provides VPN access to web enabled applications systems such as IBM or mainframe, enabling access across an SSL VPN would require many hours of development if it were even feasible.

It is also difficult for administrators to gain low-level access to run commands such as SSH or Telnet, which still requires the download of a thin client of some form and the download consists of Java applet or ActiveX component that is loaded within the browser, (Ferrigni, 2003).

Most of the internet kiosks or cafes block the downloading and running of these types of applets as they are primary means for spreading of Viruses and Malware, SSL technology does offer a sound method for users to verify server identities, but the technical sophistication required in order to do so have rendered this capability impractical for mass acceptance. The problem of counterfeit websites has become an epidemic and deployed ubiquitously and universally that the phishing type of crime in which users are tricked into surrendering confidential information to mischievous parties impersonating valid businesses was virtually unheard of before SSL adoption. Phishing is believed to be costing millions of money in fraud-type damage every month.

At present, this deficiency in SSL poses a far greater risk to online commercial activity than it does to SSL VPN implementations.

Multiple key exchanges may be required during one session; this slows performance of the web server due to the load of performing constant SSL. The fact that SSL VPN provides browser-based access to applications means that internet access is always required, mobile user does not have access to internet which means they cannot work offline.

Users may become the means for bypassing the corporate gateway security infrastructure, giving Malware such as worms and Trojans a free ride onto the corporate network, a situation called split-tunnelling. Network Access packages also fail to answer the new wave of challenges facing SSL VPNs, which includes the ability to prevent Application-layer attacks and filter out Malware. This capability will become crucial as enterprises open themselves to ever-larger numbers of encrypted sessions that pass unmonitored through the firewall and by so doing Network architects may need to invest in additional security products to protect the network from attacks that come through the SSL VPN gateway.

Although SSL VPN tunnels are launched through from the user's browser, often a desktop agent a Java applet or ActiveX control--must be downloaded for access to thin client, client/server or other applications that don't lend themselves to Web page presentation (e.g., Citrix, IBM green screen, Windows Terminal Service). Moreover, applications that require Java applets or ActiveX controls and plug-ins may conflict with a browser security policy that prohibits active content, (Phifer 2003). Most organizations block "unsigned" Java/ActiveX, which can be used to install Trojans, retrieve or delete files, etc. Some organizations block all active content to be on the safe side. As a result, you may have to reconfigure some browser clients to use an SSL VPN.

Passing traffic through an additional security device means adding latency to the connection, as well as introducing another point of failure. On the other hand, piling security checks onto the SSL VPN gateway will likely affect the overall performance of the system.

9 CONCLUSION

Most debates over IPSec and SSL VPN remote access have largely focused on the technical details of the protocols rather than focusing on what should be the most significant deciding factor between these methods and how it can be utilized fully. SSL provides a means of protecting internet users from having people eavesdrop on their communications, tampers with their data in transit, or impersonate the identity of an entity that they trust. However SSL VPNs can access applications from any Web browser and providing more options for users than IPSec remote access VPNs.

The deciding factor between IPSec and SSL VPN lies not in what each protocol can do, but in what each deployment is designed to accomplish, When one considers the cost benefit of each type of deployment as well as what problems each technology was designed to address, the deployment choices become clearer that SSL VPNs are designed to address the needs of diverse audiences that need secure access to administrator specified corporate resources from anywhere and to change both the access methods and the resources allowed as

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

the users' circumstances change that is why administrators that need to allow mobile employees, contractors, offshore employees, business partners or customers access to certain corporate resources will be well served by SSL VPNs.

In certain instances level of access may be unnecessary or unfeasible for example, mobile users that just need to check e-mail or retrieve certain documents from the Intranet don't need a dedicated pipeline to all the resources on the network because this level of access could introduce security risks if the "end-point" that the user is coming from is insecure or easily compromised.

The risk of web server residing in the DMZ which is exposed to the public internet is reduced by SSL VPN gateway residing in the DMZ behind a corporate firewall.

It should be noted that security policy for SSL VPNs is implemented and enforced at the gateway (SSL proxy). Thus, there's no user involvement and no client policy to remotely manage. As SSL VPN products mature, they must deliver on this promise in large successful deployments, grow their turnkey support for common business applications, and demonstrate their ability to withstand Internet threats and enterprise performance demands.

REFERENCES

Bhimani, A., 1996, 'Securing the commercial Internet', Communications of the ACM: Vol. 39, No. 6, Pgs. 29-35

Bahadur, G., Chan, W., and Weber, C., 2002, 'Privacy Defended-Protecting yourself online'

Conry-Murray, A., 2004, 'SSL VPNs: No comprise?' Network Magazine: Vol. 19, No. 11, 2004, Pgs 58-59.

Cisco Systems Inc., 'Remote-Access VPNs: Business Productivity, Deployment and Security Consideration'

http://www.cisco.com/application/pdf/en/us/guest/products/ps6120/c1244/cdccont_0900aecd804fb79a.pdf, (Accessed date: 03/06/2007)

Dierks, T., and Allen, C., 1999, 'The TLS Protocol Version 1.0', RFC 2246, IETF Internet Working Group

Doraswamy, N., and Harkins, D., 2003, 'The New Standard for the Internet, Intranets, and Virtual Private Networks: Prentice-Hall' ISBN: 013046189-X

Enomoto, N., Yoshimi, H., Sai, C., Hidaka, Y., Takagi, K., and Iwata, A., 2005 'A Secure and Easy Remote Access Technology', 6th Asia-Pacific Symposium of Information and Technologies, APSITT -Proceedings

Frankel, S., 2001 'Demystifying the IPSec Puzzle', Arctech house computer series' ISBN: 1-58053-079-6

Gupta, V., and Gupta, S., 2002, 'Experiments in Wireless Internet Security', Wireless

Communications and Networking Conference: Vol. 2, Pgs. 860-864.

Harding, A., 2003, 'SSL Virtual Private Networks', Computer and Security: Vol. 22, No. 5, Pgs. 416-420.

Kilpatrick, I., 2007, 'The rise of SSL VPNs' Software World: Vol. 38, No. 3, Pgs. 15-16.

Netgear Inc., 'SSL VPN Technical Primer',

http://www.watterson.com.au/Downloads/Netgear/WhitePapers/SSL_VPN_WP_23Aug06.pdf, (Accessed date: 18/06/2007).

Netilla Networks Inc., 'Achieving Versatility and Network Protection in an SSL VPN', http://www.aepnetworks.com/products/downloads/wp_achieving_versatility_protection.pdf, (Accessed date: 15/05/2007).

Phifer, L., 2003, 'Tunnel Visions: How do SSL VPNs match up with their older cousins?' Information Security Magazine: Pgs. 31-43.

Rissler, R., and Sorensen, S., 2005 'VPN Decision Guide: IPSec or SSL VPN Decision Criteria' Juniper Networks Inc.,

http://www.juniper.net/solutions/literature/white_papers/350037.pdf, (Accessed date: 4/06/2007).

Steinberg, J., and Speed T., 2005, 'Understanding SSL VPN', Packt Publishing

Shin, S., 2006, "Secure Socket Layer" Sun Microsystems Inc.,

http://www.javapassion.com/j2ee/SSL_speakernoted.pdf, (Accessed date: 28/06/2007)

Vicenc, B., et al., 2004, 'Performance Impact of Using SSL on Dynamic Web Applications' European Centre of Parallelism of Barcelona (CEPBA)

Wagner, D., and Schneier, B., 1996, 'Analysis of the SSL 3.0 Protocol', Proceedings of the Second USENIX Workshop on Electronic Commerce Oakland, November, 1996

Zwicky, E. D., Cooper, S., and Chapman, D.B., 2000, 'Building Internet Firewalls', O'Reilly & Associates Inc., Second Edition, Pgs. 368-373. ISBN: 1-56592-871-7

i·Society 2007

A new taxonomy for analyzing smart card-based authentication processes

Ramaswamy Chandramouli

Computer Security Division, Information Technology Lab
National Institute of Standards and Technology (NIST)
Gaithersburg MD, USA
mouli@nist.gov

Abstract As part of E-Government and physical security initiatives, smart cards are now being increasingly deployed as authentication tokens. The existing classification of authentication factors into – What you Know, What You Have and What You Are- does not provide a good framework for characterizing the strength and robustness of authentication processes involved in smart card-based authentications. The purpose of this paper is to identify the entities involved in this type of authentication processes, study the threats to those processes in terms of these entities involved, and then determine the list of properties associated with these entities that need to be verified to detect exploitation of these threats. A new taxonomy called Smart Card-based Authentication Taxonomy (SCBA) has been developed by classifying the property verification approaches under three authentication classes. The authentication profiles specified in two well-known recent government smart card specifications have been analyzed using the taxonomy to determine the relative strengths and assurances provided by these profiles. The use of the taxonomy in developing authentication specifications for future smart card deployment projects has also been pointed out.

1 INTRODUCTION

As part of E-Government and Security initiatives, Smart Cards or ICCs (integrated chip cards) are now being increasingly deployed as authentication tokens (for identity verification). Typical applications include controlling physical access to secure facilities, logical access to government IT systems and for encrypting and signing documents transferred between government personnel [8]. The systems that implement physical access control are called PACS (Physical Access Control Systems) and those implement logical access control are called LACS (Logical Access Control Systems). In any LACS (whether or not it uses smart cards), the user is allowed entry into the IT application system after verification of a claimed identity through a process called authentication. It is common practice to classify the various authentication mechanisms under the following classes called Authentication Factors:

- What you Know (AF1)
- What you Have (AF2)
- What you Are (AF3)

The above taxonomy for authentication mechanisms seems logical when the entity to be authenticated is a human user. However, when authentication is performed based on a set of electronic credentials resident on a smart card, using the above taxonomy does not facilitate robust characterization of the various authentication processes involving smart cards in terms of their relative strengths and consequent assurance levels. More specifically, using the above taxonomy for characterization of authentication profiles specified in many real-world smart card –based authentication scenarios provides elevated assurance levels (and a false sense of being more secure) than what is truly the case. For example, when a computer application authenticates a user based upon the credential presented through a smart card (a smart token) and the user is required to provide a PIN to activate the card so that it can be read by the authenticating system, the whole process is erroneously characterized as two-factor authentication consisting of what the user knows and what the user has. There is a flaw in this characterization since the application is in fact only authenticating the user purely based on what the user has (electronic credentials present on a token) without any other form of authentication. The PIN that the user provides is in fact not a secret shared between the computer application and the user. The PIN in this context is strictly a secret shared between the smart token and the user and merely serves to establish the binding between the user and the token and does not in any way enhance the authentication assurance from the application point of view.

The reason the interaction dynamics is different when a smart card is used in an authentication mechanism is due to the fact that there are three primary entities involved: electronic credentials, smart card and the card holder. Hence a different taxonomy is required for analyzing this process. This is the prime motivation for this paper. The approach used for arriving at this taxonomy is to study the entities and their interactions involved in this authentication technology

and the possible threats that could subvert the integrity of these interactions. To provide assurance against these threats, the properties of the entities involved in smart card-based authentication that need to be verified are to be determined. This is the focus of section 2. The classification taxonomy for the smart card-based authentication process that follows from threat analysis and property verifications is described in section 3. We have called it the SCBA taxonomy where the acronym SCBA stands for Smart Card-based Authentication. In section 4, we analyze the authentication profiles specified in two U.S government smart card specifications in terms of our taxonomy in order to obtain a clearer understanding of their strengths and assurance levels. In section 5, a brief comparison with related approaches is made. Section 6 provides some benefits of using the SCBA taxonomy for improving the overall security of the authentication process in the usage scenarios where smart cards are deployed for identity verification.

2 THREAT ANALYSIS FOR SMART CARD_BASED AUTHENTICATION PROCESS

In order to understand the threats involved in smart card-based authentication processes, it is first necessary to understand the lifecycle activities involved in issuance of smart cards. Smart cards are generally issued by an issuing enterprise or an issuing authority to legitimate/authorized individuals (after some form of identity proofing) for the purpose of carrying out a specific task (entering a building or accessing an IT system). To enable this business process, a centralized repository called Identity Management System (IDMS) is often used by an enterprise. An IDMS server provides the dual functions of gathering/importing electronic credentials from multiple sources and then distributing (provisioning) these credentials (or appropriate subsets) to various authentication points. Typical sources of credentials are:

- An organization's Human Resource or Personnel Management systems – for supplying basic demographic information about a person ,nature of affiliation of the person to the organization (employee, contractor etc) and possibly a unique number associated with the person (an Employee Number, a large unique number for electronic identification etc).
- Enrollment or Registration systems – for collecting and transmitting information about identity proofing documents (birth certificates, passports etc), biometric information such as fingerprints, facial image etc.

The typical authentication points (also called target systems) to which an IDMS provisions credentials to support smart card-based authentication are:

- Physical Access Control Systems (PACS) – consists of a PACS server (which receives information from IDMS needed for enforcing physical access) and a PACS panel (which contains a cache of the information from PACS server needed for fast authentication –such as the Unique Identification Number of the person to be allowed physical access, the expiration date for this number, the name of the person and in some cases the photograph).
- Logical Access Control Systems (LACS) - this can include any type of IT resource that can support smart card-based authentication – such as an Operating System (Work Station), Single Sign-on (SSO) modules, access control (entitlement) servers etc.

In addition to the authentication points, there is another target to which an IDMS has to provision the credentials in order to support smart-card based authentication. That target is what is known as the Card Management Systems (CMS). The CMS is a software module that can establish secure sessions with a smart card, load programs that: (a) perform the functions of populating credentials (called card personalization) and (b) execute the function calls required for authenticating those credentials. A CMS also interacts with PKI Certificate Authority (CA) servers, to request and obtain digitally signed public key certificates (attesting the credentials) and populate these digital certificates on the smart card to form an integral part of the card-based credential set.

Coming back to our discussion of the three entities involved in smart card-based authentication – electronic credential, the smart card and the card holder, we could easily see that the authentication processes are nothing but a set of transactions between these entities and the authentication points (i.e., PACS and LACS) discussed above. The security of the authentication processes therefore depend upon the integrity of these transactions. Hence, it is necessary to identify the factors that will affect the integrity of these transactions. The factor identification exercise then leads us to examine the threats associated with the entities participating in the transaction. The three primary entities are:

- Electronic Credentials Resident on the Card (E1)
- The Smart Card itself – the physical card stock (E2)
- The Card Holder – the legitimate human being authorized to use the card (E3)

In addition there is the following secondary entity:

- The authentication databases at the authentication points (PACS & LACS) (E4)

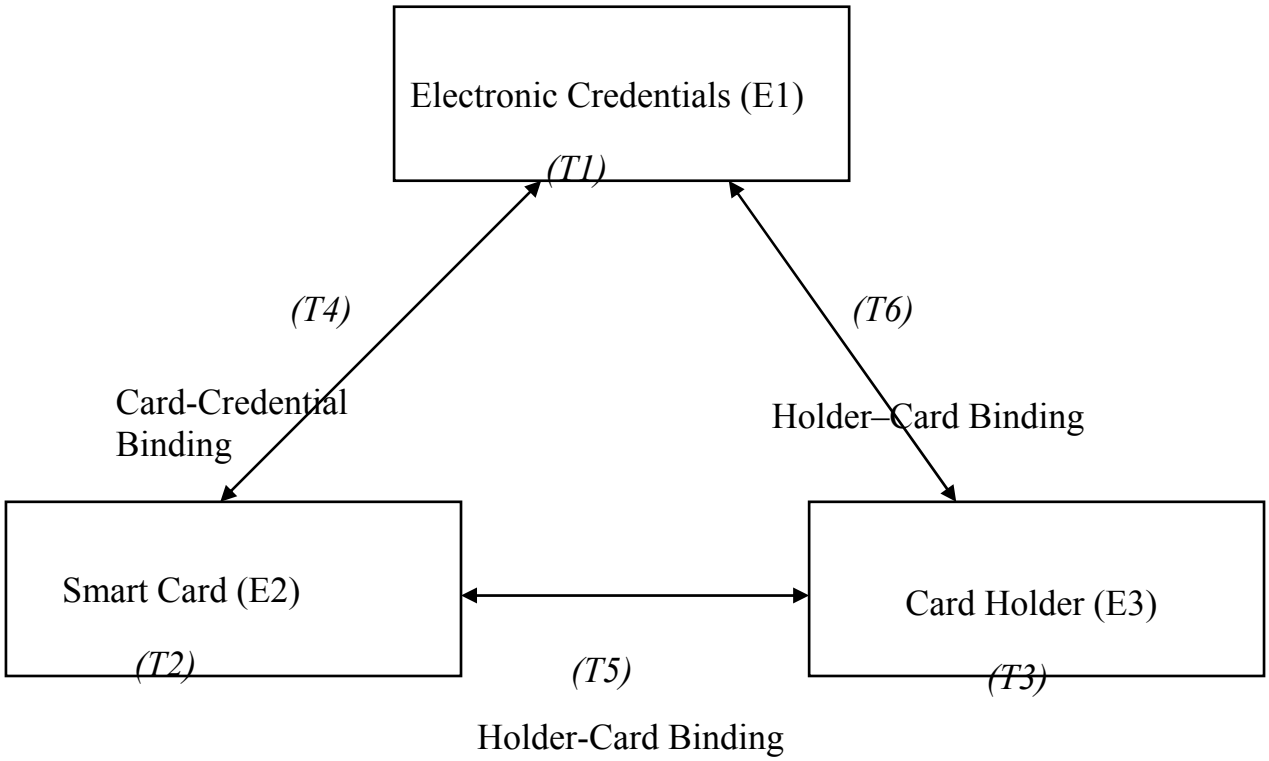
Out of the four entities listed above, managing the threats to the authentication databases (E4) through security countermeasures is under the daily operational control of the organization. However the entities (E1, E2 and E3) go outside the scope of this continuous operational control once the smart card is issued to a human being affiliated with the organization. Hence, in this paper, our focus is on these three entities. Out of these entities, Credentials is an example of an electronic entity, the Smart Card is an example of physical entity while the Card Holder is an example of human entity. The list of threats to these entities that are relevant from the authentication process viewpoint is as follows:

- Threats to Electronic Credential Entity (T1)
- Threats to Smart Card (T2)
- Threats to Card Holder (T3)
- Threats to Smart Card – Credential Binding (due to cloning valid credentials on an unauthorized card) (T4)
- Threats to Card Holder – Smart Card Binding (T5)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 1. Threat Targets (Entities & Bindings) in Smart Card-based Authentication Processes



- Threats to Card Holder – Credential Binding (T6)

Out of the last three threats, the threats to Smart Card-Credential binding are considered under the Smart Card entity while the other two are consider under Card Holder entity. The relationship between the threats outlined above and the entities/bindings is shown in Figure 1. Let us now proceed to analyze in detail the threats to the three primary entities: Credential, Smart Card and Card Holder in the following sections.

2.1 Threats to Credential Entity

In the legitimate use scenario, the authentication system would expect the credential present on the card to have been issued by the right authority and its contents are identical to the one populated by the issuer. Based upon this trust assumption, the authentication system can then proceed to verify the validity or correctness of the credential, its currency (not past its designated expiration date) and its status (revoked, terminated or suspended). However, it is possible that the credential could have been obtained from an illegitimate source or the credential could have been altered or tampered with (done mostly to substitute with credentials that have higher privileges or access rights). The countermeasures to verify whether these threats have been exploited are to perform verification of additional properties such as credential's origin and integrity (in addition to correctness, currency and status). The origin of the credential is verified by examining the associated seal (digital signature) provided by the issuer and in turn verifying with a trusted authority who can attest to the cryptographic key used in generating the seal. The integrity of the credential can be verified by using the same attested key to compare the seal with the content of the credentials. In summary, the properties to be verified are the following:

- Credential Correctness
- Credential Currency
- Credential Status
- Credential Origin
- Credential Integrity

2.2 Threats to Smart Card Entity

The authentication system would expect only the cards issued by the rightful authority to be presented for granting authentication. The threat that breaks this expectation is that the card with legitimate credentials that is issued by the right authority could be cloned or duplicated. Hence the card issued by the issuer and the card presented for authentication is no longer the same. To detect cloning or duplication, the card can be verified to possess a unique tamper-proof identifier known to the issuing system (such as a unique number associated with the integrated circuit chip of the card) or a unique tamper-proof secret such as a cryptographic key whose presence on the card can be verified as a result of an operation that the card is directed to perform and whose creation is known to the issuing system.

However, a moment of reflection reveals that it is not merely enough to verify whether the particular card is one of the valid cards in the batch procured and used by the organization. The integrity of the card issuance and subsequent authentication process (during usage) depends upon the ability to associate a particular smart card (carrying or possessing a unique identifier) with a unique credential set associated with a particular holder. This particular binding or association is needed to exercise control over use of cards reported missing or lost. The properties to be verified for the physical entity (smart card) then become:

- Possession of a tamper-proof Unique Identifier
- Possession of a tamper-proof Valid secret

- Binding between the Card and the Credential set using issuance inventory data
- Binding between the Card and the Credential set using cryptographic Methods

2.3 Threats to Card Holder Entity

The card issuing authority and the authentication system would expect only the user to whom the card was issued uses the card. The events that could nullify this expectation are that the smart card is lost or stolen and before the legitimate holder can inform the authority of the loss/missing state, an unauthorized user who has got possession of the card uses the card to perform functions not authorized for that individual. In this situation, the legitimate binding between the card and its lawful holder is lost. This property can be verified using two different approaches as follows:

However, the binding between the card holder and the card can be faked by tampering with a stolen card. To detect this threat exploitation, the binding between the card holder and the credential needs to be verified. This can be achieved using the following property verification approach: Binding between the Card Holder and Credential through Cryptographic Methods.

- Binding between the Card Holder and the Card through Proof by Knowledge
- Binding between the Card Holder and the Card through Proof by Trait

3 SMART CARD-BASED AUTHENTICATION (SCBA) TAXONOMY

Having analyzed the threats to the entities involved in smart card-based authentication and the properties to be verified to detect the exploitation of those threats, we now proceed to develop the overall authentication taxonomy. We do this through a two-step process as follows:

- Designate an authentication class by grouping together property verifications associated with an entity
- Develop a canonical authentication process description associated with each property verification

The designated authentication classes and the properties verified under each class in the SCBA Taxonomy are shown in Table 1. As discussed in the previous section, some of the property verifications are in response to potential threats, while others stem from the core authentication process logic. Table 1 captures these corresponding threats addressed as well.

The purpose of developing the canonical authentication process description for each of the property verifications is to identify the minimal set of functions (or baseline security mechanisms in case the property verification address the security threats) that each property verification process has to support. These process descriptions under the three authentication classes are given in sections 3.1, 3.2 and 3.3 respectively

Table 1. Authentication Classes and Properties to be Verified in SCBA Taxonomy

Authentication Class	Properties Verified	Threats Addressed
Credential Authentication (CLA)	Credential Correctness (CL-P1)	Use of Tampered Card
	Credential Currency (CL-P2)	Use of Obsolete Card
	Credential Status (CL-P3)	Use of Revoked Card
	Credential Origin (CL-P4)	Unauthorized Source
	Credential Integrity (CL-P5)	Data Tampering
Card Authentication (CDA)	Possession of a tamper-proof Unique Identifier by the card (CD-P1)	Cloning or Duplication of a credential on an unauthorized card stock
	Possession of a tamper-proof valid Secret by the card (CD-P2)	Cloning or Duplication of a credential on an unauthorized card stock
	Binding between the Card and the Credential set using issuance inventory data (CD-P3) (subsumes CD-P1)	Loss of control over cards reported missing or lost as well as Cloning or Duplication of a credential on an unauthorized card stock
	Binding between the Card and the Credential set using Cryptographic Methods (CD-P4) (subsumes CD-P2)	Loss of control over cards reported missing or lost as well as Cloning or Duplication of a credential on an unauthorized card stock
Card Holder Authentication (CHA)	Binding between the Card Holder and the Card through Proof by Knowledge (CH-P1)	Impersonation by stealing the card
	Binding between the Card Holder and the Card through Proof by Trait (CH-P2)	Impersonation by stealing the card
	Binding between the Card Holder and the Credential through Cryptographic Methods (CH-P3)	Impersonation by stealing the card as well as tampering with/retrieving Card Holder Identifier Data

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

3.1 Credential Authentication Class – Canonical Process Descriptions for Property Verifications

From Table 1, one could see that this authentication class involves five property verifications. From a process viewpoint, these five property verifications can be grouped as followed.

- Credential Correctness (CL-P1), Credential Currency (CL-P2) and Credential Status (CL-P3)
- Credential Origin (CL-P4) and Credential Integrity (CL-P5)

CL-P1, CL-P2 and CL-P3: All electronically issued credentials are verified to be correct by checking against entries in the database maintained by the issuing organization or authority (CL-P1). In physical access control situations, the cache of the database containing credentials usually resides in a module called "Panel" of a physical access control system (PACS). An organization with multiple facilities and multiple facility access points may have many panels. Hence the list of credential numbers needed for authentication for each of the panels located at various sites is refreshed periodically from a centralized enterprise database containing organization-wide credential numbers. The logic of the verification processes for correctness, currency and status depends upon the methodology adopted for the database refresh process. If the refresh process involves populating only the active set of credentials (current and not carrying any status flags – revoked, terminated or suspended), then the database comparison for correctness (CL-P1) implicitly performs verifications for currency (CL-P2) and status (CL-P3) as well. On the other hand, if the refresh logic sends in all the credentials appropriate for the panel along with expiration dates and status information, then explicit verification processes have to be performed for correctness, currency and status. The more frequently the panel entries are refreshed; more assurance is obtained for all three verification processes (CL-P1, Cl-P2 and CL-P3). Authentication against panel entries refreshed daily provides better assurance than authentication against panel entries refreshed only once a week.

CL-P4 and CL-P5: A credential found on a presented card, even if it is verified to be correct, current and not carrying revoked/terminated/suspended status cannot be deemed to be valid unless it carries a proof of authenticity with respect to its origin (CL-P4) and its integrity (CL-P5). The most common proof of authenticity in a smart card-based credential is a digital signature. The digital signature string is generated using a private key and the entity that signed the credential demonstrates its own credential by providing a certificate that contains the corresponding public key so that the signature can be verified. Trust in the certificate is established by establishing a trust anchor chain from a known trusted third party to the party that actually signed the certificate (called Certificate Validation). The currency of the certificate is established by verifying the non-presence of the certificate in the list of revoked certificates called Certificate Revocation List (CRL) or obtaining the currency status or through a query directed against a software module called On-line Certificate Status (OCSP) responder. Verification of the digital signature of the credential using the public key present in a trusted, current certificate then establishes the fact that the credential originated from the right authority (CL-P4) and has not been tampered with (CL-P5).

3.2 Card Authentication Class – Canonical Process Descriptions for Property Verifications

Referring again to Table 1, we see that this authentication class contains four property verifications:

- Possession of a tamper-proof Unique Identifier (CD-P1)
- Possession of a tamper-proof valid Secret (CD-P2)
- Binding between the Card and the Credential set using issuance inventory data (CD-P3)
- Binding between the Card and the Credential set using cryptographic Methods (CD-P4)

CD-P1: Any organization that has issued smart cards has to keep an inventory of the list or range of unique identifiers associated with the card stock it has personalized. This is to ensure that valid credentials are not cloned or duplicated on external card stock and presented to the organization's authentication system. This will result in unnecessary proliferation of credentials with attendant security risk. Hence every card presented to the authentication system must be verified for possessing the unique identifier that falls within the range or list of numbers in the organization's card stock inventory. This verification provides the required assurance only if the unique card identifier is tamper-proof.

CD-P2: Another approach for the organization to ensure that the card presented during the authentication process is one of the cards issued by it, is to verify that the card possesses a valid secret. In this process, the card demonstrates possession of a secret by revealing an artifact related to the secret and then participating in a cryptographic protocol. This cryptographic protocol is called the challenge-response protocol using asymmetric keys. The secret the card possesses is therefore the private key and the artifact related to the secret that the card presents is the public key embedded in a PKI certificate. The private key is tamper proof and cannot be revealed without destroying the physical entity (plastic card). The PKI certificate on the card is signed/issued by a trusted authority and carries the name of the asymmetric algorithm. The authenticating system reads the certificate, establishes trust in the certificate through PKI Certificate validation, verifies that the certificate is current using CRL or OCSP mechanisms and then sends a challenge that is consistent with the key size of the asymmetric algorithm through an appropriate command. The card encrypts the challenge using its hidden private key and sends back the encrypted challenge as a response to the command. If the authenticating system, on decrypting this encrypted challenge using the public key gets back the challenge it sent, then it indeed authenticates the smart card (physical entity) by virtue of the following:

- It contains a trusted certificate issued by the valid issuing authority

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

- The card is in possession of the valid secret (private key) associated with the public key string listed in the certificate.

CD-P3: The process for verifying the binding or association (property) between the smart card (physical entity) and credentials (electronic entity) depends upon how the uniqueness property of the smart card is verified. If the uniqueness property is verified through testing the unique identifier (such as the integrated circuit chip ID), then this verification process involves retrieving the unique credential (or credentialing number) that the card carries and then comparing the retrieved combination (Card Identifier – Credential Number) with combinations recorded in the organization's card issuance database.

CD-P4: On the other hand, if the uniqueness property is verified through testing the possession of the valid secret, then the verification is enabled by including the unique credential as one of the fields in the PKI certificate. The binding between the certificate and the credential is established when the digital signature of the certificate is verified. Since the binding between the card and the certificate (its public key) is already established through the card's demonstration of its possession of a valid secret (private key held by the card) (through the challenge-response cryptographic protocol), the binding between the card and the credential is established through the transitive relationship.

3.3 Card Holder Authentication Class – Canonical Process Descriptions for Property Verifications

The credential residing on the card may have been validated for having originated from the right authority (CL-P3) and proved to be not tampered with (CL-P4). Even the binding between the Card and the Credential may have been established using the issuance database (CD-P3) or through cryptographic methods (CD-P4). Still a security problem exists when the card itself is being used by a person to whom it is not rightfully issued. This problem can only be solved if the binding between the card holder and the card can be established at the time of usage. The binding can only be established use one of the following two property verification approaches.

- Binding between the Card Holder and the Card through Proof by Knowledge (CH-P1)
- Binding between the Card Holder and the Card through Proof by Trait (CH-P2)

In addition the faking of this binding through card tampering can be detected by performing the following property verification:

- Binding between the Card Holder and the Credential through Cryptographic Methods (CH-P3)

CH-P1: In this approach, the card authenticates the user of the card based on a shared secret such as PIN. The strength of authentication depends upon the size of the secret. The security of the process comes from the fact that the initial secret is either granted by the authority that issues the card (and made known to the user through a secure communication channel – face to face verbally or postal mail) or chosen by the user of the card in the physical presence of an official of the issuing authority. Subsequently the secret can be changed to a value the user wants (subject to some entropy/strength requirements) but since the change process requires demonstration of knowledge of the existing secret, the security is maintained. Apart from exhaustive search (called password cracking) whose difficulty increases with the size of the secret, the other threats to the security of this process are social engineering (giving out the secret to another human) and negligence of the user (writing down the secret and leaving it at a place where it can be seen or easily accessed). This form of authentication setup is conceptually similar to a system administrator setting up an userid for a user cleared for access with an initial password to access a system that can then be changed by the user.

CH-P2: Another method of authenticating the cardholder is by using a biological characteristic of the person such as fingerprint biometrics or hand geometry. An example of this is the one where the card may store a biometric data such as fingerprint templates. The user authenticates to the card by providing fingerprints which are then extracted, converted to templates and then matched with the ones found on the card (by special devices called scanners and augmented with special software modules called template generators and template matchers). This form of authentication is called biometric authentication. The matching of the live scan biometric (the one provided by the user) with the stored biometric (one on the card) can take place either outside the card or on the card itself if the card contains the matcher program running within itself (such cards are called match-on cards).

CH-P3: The previous two property verifications (CH-P1 & CH-P2) merely establish the binding between the card and the card holder. This binding could easily be faked if the person who has stolen the card guesses the PIN correctly or injects his/her biometric data into the card. Hence an additional property to be verified is the binding between the card holder and the credential. To enable this verification, the unique credentialing number is often combined with the card holder identifier information (e.g., biometric template) and digitally signed. The verification of this signature establishes the binding between the card holder and the credential while simultaneously performing the origin authentication of both the card holder identifier information and the credential.

3.4 Analysis of the set of Property Verifications for overall Authentication Assurance

Practical smart card-based authentication mechanisms will involve subsets of the property verifications in the SCBA taxonomy shown in Table 1. The choice of a given subset determines the assurance level associated with the authentication mechanism. However we find that the following set of property verifications are common to all authentication mechanisms due to the fact that these verifications involve testing the validity of the credentials read from the card and

credential validation forms the core function of any authentication process:

- Credential Correctness (CL-P1)
- Credential Currency (CL-P2)
- Credential Status (CL- P3)

However, it is important to note that any authentication mechanism with a high level of assurance should include verifications relating to all the possible combinations of binding between the three entities involved in smart card-based authentication, i.e., Credential, Card and Card Holder. Hence, a high assurance authentication mechanism should involve the following property verifications:

- Card to Credential Binding (using CD-P3 or CD-P4)
- Card Holder to Card Binding (using CH-P1 or CH-P2)
- Card Holder to Credential Binding (using CH-P3)

Further we find that even within the same type of property verification, one particular property verification approach provides more assurance than another. For example within the Card to Credential Binding, we find the authentication process based on cryptographic method (CD-P4) provides higher assurance than the one based merely on database comparison (CD-P3). Similarly, verification of the binding between the Card Holder and Card through biometric data matching (CH-P2) is certainly more robust than matching of the shared secret such as PIN (CH-P1).

4 ANALYSIS OF GOVERNMENT SMART CARD-BASED SPECIFICATIONS USING THE SCBA TAXONOMY

In this section, we look at the authentication processes specified in two recent U.S government smart card usage profiles and assess their assurance capabilities using the property verification approaches outlined in our SCBA taxonomy and the subsequent analysis outlined in section 3.4.

4.1 PACS 2.3 Specifications

To promote interoperability among smart card based physical access control systems (PACS) across various agencies of the U.S Federal government, the Physical Access Interagency Interoperability Working Group (PAIIWG) within the Government Smart Card Interagency Advisory Board (GSC-IAB) drafted this specification [9]. The two salient features of this specification are:

- Standardized container for Credentialing Elements (called CHUID) containing a series of optional and mandatory tagged objects. One of the mandatory elements is FASC-N (Federal Agency Smart Credential Number). The container includes a tag for storing the asymmetric digital signature of the credential.
- A graded set of assurance profiles – Low, Medium and High – that provide for increased assurance for the authentication of credentials read from the CHUID container.

Let us now analyze the authentication processes specified under the PACS 2.3 assurance profiles in terms of the authentication processes in our SCBA taxonomy to determine the assurance levels that each of them provide.

PACS 2.3 Low Assurance Profile: Under this profile, the card reader first reads the Card Unique Identifier (CUID) and then the contents of the CHUID container. The entire contents or a subset of the CHUID container elements that constitute the credentials are sent to the security panel of the PACS. The smart card holder is allowed entry into the physical facility based on the matching of the credentials sent by the reader with the list of credential numbers present in the panel. Since the list of credential numbers is refreshed periodically (weekly, daily or several times within a day depending upon the type of physical facility), this authentication mechanism verifies the Correctness, Currency and Status of the credential (CL-P1, Cl-P2 & CL-P3). No other property verification approach is used in this process.

PACS 2.3 Medium Assurance Profile: In this process, the PACS security panel stores along with the list of correct, current credential numbers, the HMAC (Hashed Message Authentication Code) of the concatenation of the credential data from CHUID and the Card Unique Identifier (CUID), thus creating a cryptographic binding between the credential and the specific card from where the credential is expected. The HMAC is computed using a site-specific secret key and a site-specific cryptographic algorithm. When a user presents the card, the reader retrieves the CUID and reads the CHUID contents. Using these two, it computes the HMAC using the same site-specific secret key and algorithm. The selected credential elements read from CHUID along with the computed HMAC are sent to the panel. Authentication is done based on matching of the credential as well as the matching of the associated HMACs. Further, the matching of the HMACs implicitly provides assurance that the credential has not been tampered with. This process therefore performs correctness, currency, status and integrity verification of credentials (CL-P1, CL-P2, CL-P3 & CL-P5) and binding of the card to the credential (CD-P3) through HMAC matching. Card Holder to Card binding and Card Holder to Credential binding properties are not verified. However an interesting aspect of this authentication process is that the authentication system expects a HMAC computed using a site-specific algorithm and a site-specific key. Hence, this process provides authentication of an additional entity (i.e., the card reader which is an infrastructure entity).

PACS 2.3 High Assurance Profile:This process verifies whether the card is in possession of a cryptographic key that is derived using the site-specific secret key and a concatenated text string made up of Card Unique Identifier (CUID) and CHUID contents. During authentication, the reader retrieves CUID, reads the contents of CHUID and computes the cryptographic key based upon a site-specific secret key using the algorithm information in a data structure called Authentication Key Map. To verify whether the card is in possession of the same cryptographic key, the reader sends a random challenge and receives the encrypted challenge (encrypted by the card using the cryptographic key

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

injected into it) from the card as a response. The reader then encrypts the challenge using its generated cryptographic key and looks whether the two cryptograms (one computed by it and the other received from the card) match. Finally of course the extracted credentials are sent to the PACS security panel for matching. This process therefore performs correctness, currency, status and integrity verification of credentials (CL-P1, CL-P2, CL-P3 & CL-P5) and binding of the card to the credential (CD-P4). Thus we see that the same property verifications as found in PACS 2.3 Medium Assurance Profile are performed in this profile. However the verification approach used for Card to Credential binding (CD-P4) is based on a cryptographic protocol and is therefore much more robust than the corresponding approach (CD-P3 and HMAC based) used in the Medium Assurance Profile. Since the success of the cryptographic protocol depends upon the reader's ability to generate the right cryptographic key based on the combination of site specific secret key and the card and credential data read from the card, this serves to authenticate the reader as well.

The results of the analysis of the authentication processes used in PACS 2.3 Authentication Profiles in terms of the property verification approaches outlined in SCBA Taxonomy are summarized in Table 2 below.

4.2 FIPS 201 Specifications

In response to a Presidential Directive called HSPD-12, the U.S Government developed a set of specifications for use of smart cards to provide physical access to federal facilities and logical access to government IT systems using a set of uniform, interoperable and tamper-proof credentials. These specifications are embodied in a document called FIPS 201 [5] and its various companion documents [1,2,6]. In terms of the credentialing elements, FIPS 201 uses the same CHUID container defined in PACS 2.3 specifications (discussed in previous section) with some minor variations. FIPS 201 outlines a set of authentication use cases classified into three graded assurance levels – "SOME confidence", "HIGH confidence" and "VERY HIGH confidence". As we did in the discussion of PACS 2.3 specifications, let us now analyze the FIPS 201 authentication processes in terms of the property verification approaches in our SCBA taxonomy. (The summary is shown in Table 3).

SOME Confidence: One of the processes under this assurance level is "Authentication Using the PIV CHUID". Under this process credential elements in the CHUID container are read by the card and their origin and integrity are verified using the associated digital signature. Eventually, the credentials that are read from the CHUID container are sent to the PACS system for matching against a periodically refreshed list. This process therefore performs correctness, currency, status, origin and integrity verification of credentials (CL-P1, CL-P2, CL-P3, CL-P4 & CL-P5) thus covering all property verifications relating to credentials. The Card to Credential binding, Card Holder to Card binding and Card Holder to Credential binding properties are not verified under this process.

HIGH Confidence: FIPS 201 provides a single authentication process called "Authentication using PIV Biometric" under this assurance level and labels it as BIO. This process calls for the user of the smart card to provide his/her fingerprint biometric data through a live scan and also provide a PIN to enable the reader to read the stored biometric data on the card. A key credentialing element FASC-N is embedded in the data structure containing the biometric data and verification of the digital signature associated with biometric data implicitly verifies the origin and integrity of the credential. This process therefore performs correctness, currency, status, origin and integrity verification of credentials (CL-P1, CL-P2, CL-P3, CL-P4 & CL-P5) thus covering all property verifications relating to credentials. The Card Holder to Card binding property is verified through the verification approach CH-P2 as the card holder is authenticated to card using biometric matching. The Card Holder to Credential binding is verified through the verification approach CH-P3 as the identifying credential is embedded with biometric data structure. The only property that this process does not verify is the Card to Credential binding.

Table 2. Characterization of PACS 2.3 Authentication Profiles

Authentication Profile	Property Verification Approaches	Additional Property Verifications
Low	Credential Correctness (CL-P1) Credential Currency (CL-P2) Credential Status (CL-P3)	
Medium	Credential Correctness (CL-P1) Credential Currency (CL-P2) Credential Status (CL-P3) Credential Integrity (CL-P5) – through HMAC Binding of Card to Credential (CD-P3)	Authentication of the Reader (Infrastructure element)
High	Credential Correctness (CL-P1) Credential Currency (CL-P2) Credential Status (CL-P3) Credential Integrity (CL-P5) Binding of Card to Credential (CD-P4) – through a cryptographic protocol	Authentication of the Reader (Infrastructure element) The cryptographic key injected into the card is based on a site-specific key. Hence limits the use of the card to specific designated sites where that key is used.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 3. Characterization of FIPS 201 Authentication Use Cases

Authentication Use Cases	Property Verification Approaches	Additional Property Verifications
SOME Confidence	Credential Correctness (CL-P1) Credential Currency (CL-P2) Credential Status (CL-P3) Credential Origin (CL-P4) Credential Integrity (CL-P5)	
HIGH Confidence	Credential Correctness (CL-P1) Credential Currency (CL-P2) Credential Status (CL-P3) Credential Origin (CL-P4) Credential Integrity (CL-P5) Card Holder to Card binding (CH-P2) Card Holder to Credential binding (CH-P3)	
High	Credential Correctness (CL-P1) Credential Currency (CL-P2) Credential Status (CL-P3) Credential Origin (CL-P4) Credential Integrity (CL-P5) Card to Credential binding (CD-P4) Card Holder to Card binding (CH-P1)	Card Holder to Credential binding occurs transitively due to Card Holder to Card and Card to Credential bindings.

VERY HIGH Confidence: The authentication process (BIO) described in the previous section, when carried out under the watch of an attendant (when the user is submitting fingerprints to a scanner especially) is classified under VERY HIGH Confidence assurance level. In addition, another authentication process called "Authentication using PIV Asymmetric Cryptography" (labeled as PKI) is specified under this level. This authentication process calls for the card to encrypt a challenge sent by the reader system using the private key of a private-public key pair the card holds (Challenge-Response Cryptographic protocol). This process therefore verifies the property that the card possesses a tamper-proof valid secret (CD-P2). To enable the card to perform this private key operation, the card holder is required to provide a PIN thus performing the Card Holder to Card binding verification using the CH-P1 approach. A key credentialing element FASC-N is embedded in the certificate that contains the public key that corresponds to the private key held by the card. Hence validation of the signature of the PKI certificate using the issuer's public key implicitly validates the origin and integrity of the credential, in addition to verifying the PKI certificate to credential binding. Further since the Card to PKI Certificate binding is established through the challenge response cryptographic protocol, we have transitively obtained the verification of Card to Credential binding through the verification approach CD-P4. The only property not directly verified in this process is the Card Holder to Credential binding but that property occurs transitively due to Card Holder to Card and Card to Credential bindings that have already been established.

5 COMPARISON WITH RELATED APPROACHES

Smart card-based authentication schemes appear in two categories of published literature. One category appears in various research papers in technical professional journals. The other category appears in technical specifications for large-scale smart card deployments. The central theme of the research papers has typically been to present new and novel schemes that are robust enough to withstand all types of known and potential attacks. Examples are: An improved scheme for asymmetric smart card authentication which is resistant to not only replay and active attacks but also hostile attacks [3], Password-based authentication schemes using smart card that are resistant to logic attacks [7,10], smart card-based biometric authentication schemes that provide assurance against replay attacks [4] and so on. Because the core focus of research community is on security robustness, certain other factors such as scalability, performance and usability may not be given their due consideration in their proposed schemes. On the other hand, the authentication schemes proposed in technical specifications relating to smart card deployments in industry or government, are generally chosen because they have some track record of earlier deployments and found to be usable with reasonable performance overheads. However, it was generally found that those technology choices are macro-level selections without an analysis of the core properties each of the mechanisms verify in the context of the entities participating in the authentication transactions. The purpose of this paper is to bring some formalism into the process of specifying authentication schemes for real-world smart card deployments by providing a framework to analyze authentication mechanisms in terms of some fundamental property verification approaches.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

6 BENEFITS AND CONCLUSIONS

The two main contributions of this paper are the following:

- Development of a new taxonomy called the SCBA taxonomy consisting of authentication classes and associated property verifications for analyzing the interactions involved in smart card-based authentication scenarios.
- Illustration of the taxonomy for analyzing the authentication profiles or authentication use cases specified in two real-world smart card deployment specifications.

The benefits of the SCBA taxonomy come from its flexibility to be used in two ways. They are:

- It can be used for analyzing authentication profiles/schemes chosen or selected in the smart card usage specifications. When used in this mode, it provides a formal approach to determine whether the assurance levels assumed for those profiles/schemes are realistic.
- The list of property verification approaches can be used to build a combination that is appropriate for a given smart card deployment scenario. In this usage mode, it provides a methodology for specifying authentication mechanisms that will meet the security requirements in various smart card deployment scenarios.

In addition to its flexibility of use, the SCBA taxonomy by itself is extensible. For example, when new threats are discovered for the three entities (Credential, Smart Card, Card Holder) or the binding between the entities, additional property verification approaches can easily be added. Also, when new entities are added to the authentication scheme, additional property verification approaches are to be added as well. For example, in our SCBA taxonomy, we have assumed that the card reader is an integral part of the authentication system or connected to the authentication system through a closed network connection. On the other hand, if a smart card-based authentication system involves remote readers connected through an open network to the authentication system, additional property verification approaches relating to integrity of communication between the readers and the authentication system, integrity of reader operation (as the readers may be tampered or compromised) must be developed and incorporated into the taxonomy.

REFERENCES

[1] Biometric Data Specification for Personal Identity Verification, SP 800-76, http://csrc.nist.gov/publications/nistpubs/800-76/sp800-76.pdf (July 27, 2007)

[2] Cryptographic Algorithms and Key Sizes for Personal Identity Verification, SP 800-78, http://csrc.nist.gov/publications/drafts/800-78-1/draft-SP_800-78-1-070306.pdf (July 27, 2007)

[5] FIPS 201 – Personal Identity Verification of Federal Employees and Contractors, http://csrc.nist.gov/publications/fips/fips201-1/FIPS-201-1-chng1.pdf (July 27, 2007)

[6] Interfaces for Personal Identity Verification. NIST Special Publication SP 800-73-1. http://csrc.nist.gov/publications/nistpubs/800-73-1/sp800-73-1v7-April20-2006.pdf (July 27, 2007)

[7] Kumar, M. New Remote User Authentication Scheme Using Smart Cards, IEEE Transactions on Consumer Electronics. Volume 50, Issue 2, May 2004 Page(s):597 - 600

[8] Securing e-business applications using Smart Cards, IBM Systems Journal, Vol 40, Number 3, 2001 - http://www.research.ibm.com/journal/sj/403/hamann.html (July 27, 2007)

[9] Technical Implementation Guidance: Smart Card-Enabled Physical Access Control Systems – Version 2.3, http://smart.gov/iab/documents/PACS.pdf (July 27, 2007)

[10] Wang, X.,Zhang, J.,Zhang, W.,Khan, M.K., Security Improvement on the Timestamp-based Password Authentication Scheme Using Smart Cards. In Proceedings of IEEE International Conference on Engineering of Intelligent Systems April 2006.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Conceptualising and analysing internet threats using a 4-dimensional hypercube

Jan van den Berg

Delft University of Technology,
Faculty of Technology, Policy and Management, Section of ICT
Jaffalaan 5, P.O. Box 5015, 2600GA Delft, The Netherlands
email: j.vandenberg@tudelft.nl

Abstract The current developments on the Internet show that the dynamic phenomena that are taking place in this continually growing virtual world bear more and more resemblance to occurrences in the real world. As a consequence, next to the good services the Internet offers, we observe a lot of potential security and human values' threats we are familiar with in the real world. Due to the international, distributed and dynamic character of the Internet it is hard to take appropriate action to deal with these threats. This concerns a problem that is enforced by the weak governance structure of the Internet. And even if we would be able to well organize this governance structure, we should have first developed an appropriate conceptualisation of all threats. In this paper, we propose to use a 4-dimensional hypercube to perform this conceptualisation. It is supposed to help mapping out current Internet developments and threats based on which the most relevant risks analyses can be performed and the best technical and institutional countermeasures designed.

Keywords the Internet, security, human values, threats, risk analysis, technical and institutional design.

1 BACKGROUND

Starting as medium for fault-tolerant communication [1], the current dynamically evolving Internet offers a variety of services related to information provision, communication, business and entertainment. It has resulted into a new extensive virtual world where many phenomena similar to those in the real world are taking place: people (or their representatives) can travel (e.g., in the world of 2nd life), meet and discuss with friends, buy and sell goods on virtual markets and via auctions, perform (financial) transactions, watch movies, build up social or business networks, do e-business, contact the administration, etc. As a consequence, similar problems are taking place in the virtual world of Internet like, to just name a few, unreliability of services, theft and other forms of criminal activities, reputation loss, plagiarism, social pressure, privacy violation and other types of personal life interference, addiction, uunwanted exposure to sexual material, and even (information) ware fare. In short one may say that the very dynamic virtual world of Internet offers many good things but, at the same time, it is full of security and human values' threats.

Looking at the ways the global society deals with the Internet, we observe a mixed picture. On the one hand, we perceive that, despite the lack of a strong centralized governance structure, technical developments are taking place at high speed and adoption of new standards at a worldwide scale occurs quite easily. In addition, by the work of thousands of Service Providers, the Internet can be used throughout the world, 24 hours a day. The activities at the technical level have also resulted in more or less effective ways of dealing with technical imperfections as exposed by computer viruses, the first ones of which appeared 25 years ago [i].

On the other hand, considering the security, privacy and other problems around the many available (easy-to-use and easy-to-create) Internet applications, we observe, also due to the above-mentioned lack of a centralized governance structure, that the contest with all these problems is certainly less well-organised and is taking place in rather incident-driven ways, with big differences between different countries. Examples include the differences in the accessibility to Internet sites (where governments sometimes try to protect users against 'undesirable information') and in the way personal privacy is understood and protected by the law in different countries. More generally, we notice differences between governments to organize tracing of and contest against cyber crime, which is a hard to tackle phenomenon in itself. We further observe all kinds of social and political commotion popping up after the emergence of new Internet phenomena (like in the virtual worlds of 2nd life, You Tube, and online gambling), around the Internet sale of narcotics, pharmaceutics and weapons, and after the occurrence of the thousand and first security incident. This commotion however is usually very temporary and soon fades away if new and other incidents take place and ask for attention. In short, at the moment, the global society does not seem to be well prepared to deal with the worse events taking place on the

Internet and seems to be more busy with combating against symptoms and individual incidents than to choose for a proactive, structured manner of tackling them.

Considering the difficulties society encounters to deal with the above-mentioned Internet problems, we must admit that they concern a very complex, worldwide matter. We also recognize that history has taught us that the resolution of problems at a worldwide scale is often thwarted by the lack of international governance structures having the authority to impose solutions. This however does not discharge us from the first task, namely, that of analysing what the main Internet problems are. With these observations and assumptions in mind, the goal of this article is to present a conceptualisation that enables the identification of the main security and human values' threats taking place on the Internet. Based on this identification, the most relevant risk analyses can be executed and countermeasures devised.

The rest of this paper is structured as follows. In the next section, we present the conceptualisation consisting of a 4-dimensional hypercube. In section three, we will show how this hypercube can be used to identify the most important Internet trends and threats. In section four, we sketch some ideas on how to perform the corresponding risk analyses and to design technical and institutional solutions. Next, as an illustration of our ideas around the applicability of the hypercube, a set of relevant research topics is presented in section five. We finalize the paper by presenting our main conclusions in section six.

2 CONCEPTUALISATION OF INTERNET THREATS

To come up with a conceptualization, we reformulate the above-sketched picture of the virtual world of Internet in just one sentence: the current global dynamic Internet (actually a network of networks) offering many different applications and services, being provided by a large groups of different enablers (stakeholders) from countries through over the world having different interests, and being in use by a large variety of users from al kinds of cultures, suffers from many security and human values' threats. This only one sentence reveals that four dimensions can be distinguished to conceptualize the world of Internet and its threats:

1. the Internet having a layered structure which can be decomposed in three levels:
   a. Basic Infrastructures (offering transmission services),
   b. Basic Communication Services (basic services like email, file transfer, VoIP, chat services),
   c. Information & Transaction Services (including online banking, Google services, e-auctions, e-business transaction services, web 2.0 applications, etc.);
2. its Users: Companies, Government, other Organizations, and Individuals in their role as end-users, from countries through over all world;
3. its Enablers/Stakeholders: International Internet organizations (ISOC, ICANN, and many more), Service Providers (of services mentioned above), Scientific Communities, Standardization Committees, Government, Individuals in their role as enablers/service providers/designers/policy makers/controllers/content providers, etc., again from countries through over the world;
4. the Security Requirements and Human Values that are possibly threatened:
   a. Security Requirements: Confidentiality, Integrity, Availability, Accountability (CIAA)
   b. Human Values: Privacy, Trust, Reputation, Autonomy, Sustainability, Sincerity, Reliability, Clearness, Social Cohesion, Safety, and many more.

Ad 1: The three level structure presented has been inspired by the OSI model [1] where the four layers 'physical layer', 'data link layer', 'network layer' and 'transport layer' together are here termed Basic Infrastructures offering (more or less secure) peer-to-peer transmission services. Next, the Basic Communication Services are offered by what is usually called the 'application layer' [1]. All other, higher level and often quite dedicated applications that make use of these Basic Infrastructures and Basic Communication Services and offer all kinds of services to the end-users, are simply put together in one set of what we term Information & Transaction Services.

Ad 2 and 3: A first important observation here is that organizations and even individuals (can) currently have two roles at the same time, namely, that of end-user and that of enabler/stakeholder. E.g., by available so-called Web 2.0 technologies [3] like weblogs, social bookmarking, wikis, podcasts, RSS feeds and other forms of many-to-many publishing, current Internet users are not only content consumers but often, they also behave as content providers. The easy-to-use character of these technologies has taken away the former thresholds for both Internet use and content creation. Or, stated in somewhat more general terms, we observe that the currently available Internet technologies strongly support active interaction between Internet users and organizations, in similar ways we are familiar with in the real world: in this sense, the virtual world of Internet actually concerns a(n extension of the) real world and it is equally easy to enter as the 'tradional' real world. This certainly has several consequences, including for the responsibilities of Internet users.

Actually, the users as well as the enablers/stakeholders concern a wide variety of organizations and individuals spread over the globe having a very limited governance structure. It is remarkable that neither the U.S. government nor other governments have used their rule-making powers to settle Internet issues (including difficult policy decisions) [ii] but decided 'to threw the responsibility back to the warring parties' [2]. As a consequence, the central governance of the current Internet is mainly limited to technical (and directly to that related) issues like Internet domain names and addresses.

Ad 4: It is remarkable that Internet research related to the mentioned security requirements CIAA is very extensive

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

while research related to the (may-be at least equally important) human values' threats (see, for example, [4]) still seems to be in its infancy. We will come back to these issues in more detail in the sections below.

3 USING THE 4-D HYPERCUBE

3.1 Some examples

The given conceptualisation can, at least in principle, be represented by a four-dimensional hypercube where groups of cells can be used as a starting point for describing how a specific group of end-users (e.g., the group of female teenagers in the European countries X and Y using their PC at home) make use of the Internet by means of certain applications (e.g., 2nd life), enabled by a group of enablers/stakeholders (among others, the 2nd life provider as well as the users who participate in 2nd life using their avatar) while being vulnerable to a specific set of human values' threats like, for example, social pressure, social isolation, and sexual intimidation.

Another, this time real and well-documented example concerns a money extortion case performed by the Russian mafia supported by a small group of computer hackers who were able to create a zombie network of around 1.5 million PCs. To create the zombie network, the hackers (probably developed and) distributed the 'W32.Toxbot' virus [5]. According to the news announcement [6], the suspected persons threatened several organizations to attack their computer systems by means of the created Zombie network. In August 2005, one organization yielded to the menace and next informed the Counsel for the Prosecution.

Still another analysis within a different region of the hypercube may reveal that security threats (related to, especially, the availability of Internet services) exist at the IP level caused by the opaqueness of the routing policies as applied by different service providers/companies/universities (owners of so-called Autonomous Systems [1]). At this level of the Internet protocol stack, it may become clear that also other phenomena need our attention like the integration of the current Internet with existing and new mobile networks.

In the wide field of e-business, an analysis may show (again, to mention just a few examples)

1. the existence of security risks of unauthorized access to transaction services (because of, e.g., 'phishing' and denial-of-service attacks) potentially causing reputation loss and other damages to the concerning financial institutions;
2. the unfair organization of e-auctions and e-market places such that large business partners have a privileged position compared to other competitors while tax inspectors discover that huge tax revenues are illegally defrauded.

3.2 How to use the 4D-hypercube in practice

Based on the above-given examples it is clear that, by analyzing all cells in the hypercube, we can get detailed insights into what is happening on the Internet. The conceptualisation offers a nice 'mind map' for positioning existing and new developments taking place around the Internet, both with respect to applications and underlying technical issues as well as to the role of users and stakeholders, and to corresponding potential security and human values' threats.

In practice however, it may be very hard to get a complete overview because of the combinatorial explosion of the number of cells, so it seems wise to relax the analysis' goals somewhat. A better challenge may be to confine oneself to those applications that suffer from the seemingly most severe security and human values'. A clustering of cells in the hypercube may help to identify these applications with associated users, threats, and enablers/ stakeholders.

It may turn out to be too difficult to solve even this less difficult identification problem. A severe complication is also present, namely, that of the high dynamics of the Internet: once an overview has been made, it is often already out-of-date. Therefore, taking these considerations into account, the best way to go may be to simply select a few applications having very different characteristics which are more or less representative for the complete set of available applications (e.g., a set related to web 2.0, a set related to current e-business practices, a set related to new mobile applications, and a set related to routing problems at the IP level), and to analyse the related security and human values' problems.

3.3 Some first tentative observations

Looking at the current discussions on responsibility for Internet, they seem to focus on the security and privacy at lower levels [iii]. However, security of the data bits being sent from A to B by e.g., secure email, secure ftp, secure msn, 2nd life, etc. does not guarantee a safe Internet! Current discussions are declining

- not to take into account what's going on in the higher layers of the OSI-model which is the responsibility of all kinds of content providers and other stake holders of high level application services (but, instead, just to concentrate on the security of low-level information exchange and corresponding governance structures [iv]) ;
- not to look at general human values' threats in the virtual community (global village) of Internet and the corresponding governance structures needed (but, instead, to confine the discussion to basic security and privacy issues only);
- not to be attended by all relevant governmental institutions (while instead, generally spoken, most governmental authorities play a waiting game).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4 TOWARDS SOLUTIONS

4.1 Risk analysis

Having selected a few, seemingly most relevant Internet applications (the output of the previous step), we can start an analysis in depth. This concerns a risk analysis in the first place where the security and human values' threats and their impact (expectation of the product of damages and probabilities of occurrence) are identified. We observe here that a lot of literature is already available with respect to the most important security (= CIAA) threats on the Internet (including unauthorised access, interception of communications, viruses, spam, malicious representation) while with respect to the analysis of human values' threats, much less research results are available. As a consequence, society has not been able to use the necessary research results to take appropriate countermeasures.

There exist several risk analysis methodologies with respect to basic security requirements CIAA like, e.g., SPRINT [8], which are often based on interview techniques. In addition, statistical and data mining [9] techniques may be useful to get a better insight in the probability of the various types of security incidents taking place [v]. Finally, existing literature from international security organizations offers a lot here. So, it seems to be that quite a lot of (scientific) sources and methods are available to get insight into the most relevant basic security risks, i.e., the risks related to the fundamental CIAA requirements. As a researcher, the most difficult part is probably to get in touch with the right information sources, i.e., data and expert knowledge, especially, if we wish to acquire a global, worldwide view.

With respect to human values' threats, there are some general descriptions (including [4]) available but it is a less clear how to estimate their impact. For several more specific threats (e.g., reputation loss, privacy breaches), there are literature sources available but only in a few cases, the impact is modelled in a quantitative manner. So, we might be forced to limit ourselves here to qualitative risk assessments, mainly based on interviews with experts/stakeholders in the field. It is our impression that risk analyses, i.e., evaluations of the possible impact of other human values' threats as occurring in all kind of new Internet applications like 2nd life, YouTube, e-auctions, social networks, can not yet be executed, simply, since we have not finalized the preceding step, that of determining what types of risks are actually taken. This makes it impossible to sketch here an overview of all relevant risk evaluation methodologies needed. On the other hand, we assume that in general, both business and private organizations are expected to be willing to make available their experience and loss data, provided a good 'incentive structure' has been chosen: actors (stakeholders) are more willing to contribute in case they recognize their personal (or organisational) profit of the effort.

Knowing the existing threats and their possible impact, we can try to deal with them to reduce the expected impact to a desired level (reduction to an expected impact level of zero is impossible since both 100% safety and 100% protection of human values does not exist). Here, several approaches can be adopted, both technical and institutional, some of which we sketch below.

4.2 Towards technical solutions

With respect to the technical approaches, both security and human values' enhancing techniques are of interest: both cryptographic tools [1] and so-called intelligent techniques [10] are supposed to be useful here. On the other hand, still much work has to be done especially if we consider the security at higher levels. E.g., Identity Management (consisting of Identification, Authorization and Access Control) is often badly organized in the virtual world. Users have to login many times, have to fill in personal details again and again and the idea of a universal 'e-passport' for identification purposes has not been elaborated. Actually, due to the increasing level of distribution of services (caused by developments like web services and service oriented architectures) Identity Management in distributed environments is becoming even more complicated everyday. Security enhancing technologies are needed to solve the problems mentioned.

With respect to privacy and other human values' enhancing technologies to support their protection and to calibrate all kinds of high level personalized services like location-based services [11] and dynamic informed consent decision making [12], we have just started to design and implement technical solutions. Due to the broad spectrum of the necessary solutions, we may conclude that this type of research is just starting and generally accepted solutions are hardly available at this moment.

4.3 Towards institutional solutions

Thinking about institutional solutions, we face many basic problems including (i) the high number of stakeholders with different responsibilities and interests and (ii) the international character of the Internet. However, this may not prevent us to take up our responsibility and to start all kinds of organisational, contractual and legislative initiatives to make the Internet a safer place, within any country of the world, within Europe, Asia, Australia, and the America's and ultimately, at global level.

Our conceptualisation using the hypercube makes clear that the responsibilities in relation to what's happening on the Internet are spread over many different organizations: we can discriminate between (worldwide) organizations that provide the physical infrastructure, (worldwide) organizations that together provide packet switched end-to-end communication [vi] (being competed by new developments like Bluetooth-, WIFI-, PICO-, and WIMAX-based Wireless Local Area Network for Multimedia Communication like PICO, WIFI and WIMAX networks [vii]), (worldwide) organizations that provide basic communication services like email, chatting and VOIP, and (worldwide) organizations that offer even more advanced services to shape the virtual world (web2.0 applications, role-based applications like 2nd

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

life, personal network services, e-auctions, e-markets, etc.). As a consequence, some basic ideas are that

- the responsibility for the Internet is a responsibility for all participating organizations and individuals [viii];
- (like in the real world) governments, the European Commission, etc. have their specific responsibility to orchestrate, i.e., to
  - decide on the governance structure (rules of the game for all stakeholders);
  - fix appropriate laws;
  - organize compliance control;
  - facilitate penalization and accusation of organizations and individuals that/who break the rules.

The orchestration role of the Government is certainly not a trivial affair simply since Internet is an extremely dynamic virtual world covering very many different phenomena where almost every day something new and unexpected pops up. At the same time, every day we are becoming more and more dependent on the same Internet since we make use of it on a daily or even almost permanent basis. So, there is actually no choice (like there is now choice for Government to orchestrate the real world phenomena in our society). According to this way of reasoning, Government should take the lead, in many different ways. Especially the threats that may have major impact at a regional or national scale (like, e.g., the unavailability of Internet services and the occurrence of severe human values' breaches) may be considered as a responsibility for central governmental authorities.

At the same time it seems clear that at the moment, governments struggle with their role as orchestrator and seem unable to define the precise 'rules of the game' (i.e., the appropriate legislation) for the virtual world of Internet. As a consequence, many other stakeholders have also difficulties to precisely define their roles. It is further clear that the various stakeholders have different opinions about their own role and that of the other stakeholders, which strongly thwarts effective actions. We also observed that, fortunately, very many organizations are willing to contribute. So, in principle, institutional solutions seem to be achievable to a certain extent.

With respect to the research issues touched in this paper, institutional approaches should focus on governance arrangements to cope with the identified security and human values' threats. The above-given hypercube offers inspiration how to partition the phenomena that are taking place on Internet. As said above, having selected these different areas and having analyzed in depth the security and human values' threats, one can try to define, next to technical solutions, appropriate governance structures. The basic idea here might be that different services on the Internet as enabled by different organizations need different governance structures. E.g., with respect to basic communication services within a certain country, the Ministry of Economic Affairs may have leading initiatives with respect to general legislation, the ISPs may be the most important enabling business partners, the existing association of consumers may take their role of defending the interests of individual end-users, and the dedicated supervision authority may be the organization with responsibilities for control. The precise governance structure should be defined. In addition to this, similar things should be organized at European and other international levels.

However, if we look at what's going at the higher levels of the Information & Transaction Services (like in 2nd life and in the world of e-business), quite different governance structures may be needed with different representatives from government, other stakeholders, other end-users etc. Organizing e-auctions and e-market places clearly has a lot to do with institutional economics and the Ministry of Economic Affairs seems to be a key player here, for example with respect to the canalisation of financial streams and corresponding tax structures. In addition, the Ministry of Justice may be busy with legislation concerning fraud detection methods, criminal procedures, and appropriate punitive measures. So, in general, all the ministries within all countries have something to do with the Internet and should start and coordinate activities like for what is happening in the (virtual = real) world of Internet.

4.4 Testing and Validation

Both technical and institutional solutions need to be designed. It is tempting to look for areas in the hypercube where both technical and institutional arrangements can be analysed in parallel in complementary ways. In other cases, the emphasis may be on either security issues or certain human values, depending on the problem being tackled.

As a matter of fact, proposed technical and institutional solutions should be tested and validated as much as possible using the usual scientific means. The above sketch of the methodology of our research (roughly consisting of (a) domain selection, (b) (risk) analysis, (c) technical and institutional design, (d) testing and validation) is according the principles of what is currently named Design Science, i.e., the 'science of the artificial' [16].

5 A POSSIBLE SET OF RELEVANT RESEARCH TOPICS

To illustrate our ideas, we here offer a few example research topics that may be started inspired by our conceptualization and analysis. They may be considered as sketches for several PhD research plans.

1. The impact of connecting the new Wireless Local Area Networks into a new critical Internet infrastructure (topic in the sub-area of critical infrastructures).

   Illumination: At the moment all kinds of initiatives are taking place around the setup of large Local Area Networks based on WIFI-, PICO-, and WIMAX-technology. The research focuses on the safeguarding of the security of communication services (how to guarantee availability, confidentiality, integrity, and accountability?) and on the organization of correspondent governance structure (who is responsible for what according to which rules (clear definitions are needed), which organization is controlling?).

Side effect: this research may/should also offer solutions for the governance structure around existing communication services (at, say, the IP-level).

2. The (im)possibility of an (inter)national multi-purpose e-passport (topic in the area of high-level authentication services)

   Illumination: Identification and Authentication are of high importance in the virtual world (like it is in the real world). For many services, we need to offer personal information the amount of which is strongly dependent on the actual context. (So actually, we need tools for dynamic consent decision-making). So, the question arises whether we can technically design a secure, general-purpose solution that also stands human values' threats like privacy, theft, and trust. Cryptographic and intelligent techniques are supposed to be strongly needed here. In addition it is clear that institutional design is needed: the issue of a passport is a complicated matters where the organization of 'chains of trust' is of key importance. Certification is supposed to be a relevant issue. A smart incentive structure might also be needed. The research should make clear to what extent an e-passport is feasible, i.e., for which sets of Internet applications it can be used trustworthily.

   (Microsoft has already implemented Windows Live ID [17], an e-passport that enables single sign-on to a variety of web services. However, this implementation cannot be considered is very successful yet.)

3. Security and human values' threats caused by the role-change of end-users into content providers and active Internet community participants (topic in the area of new, high level Internet phenomena)

   As mentioned above, the current Internet offers a rich virtual world where its participants adopt all sorts of activities. The basic idea of this research is to select a few sub-worlds (e.g., 2nd life, e-markets, YouTube) and to understand the various threats in the first place. Or, simply stated, it will focus on what's going in the virtual world of Internet and on what the major security and human values' threats are. This concerns social research among the group of participants that can be supported by monitoring and data mining techniques having as goal to describe patterns of behavior shown by participants. Based on the insights found, ideas for technical and institutional solutions may pop up. It is a challenging question for governments how to deal with these problems of the virtual world. The research performed is supposed to offer enough insights to provide concrete handles for law and rule setting.

6 CONCLUSIONS

In this paper we have introduced a 4-dimensional hypercube for creating some structure in the conceptualization of what is currently going on in the dynamic virtual world of Internet. Actually, we consider what is happening in society enabled by the so-called 'Information & Transaction Services' of the Internet as real things or, to be very clear: the virtual world of Internet has become real. This statement is underpinned by the fact that we observe that many phenomena that are taking place in the real world are also occurring on the Internet. As a consequence, we need to deal with the Internet as we deal with the real world. Or, in other words, we should organize the Internet world according the same principles where governments, business partners, private organizations, and individuals have and therefore should adopt their respective responsibilities. This concerns our main conclusion.

Like in the real world, safety is an important issue to guarantee on the Internet. In this paper, we have interpreted safety in terms of the well-known security requirements CIAA and human values like privacy, trust, reputation, autonomy, sustainability, sincerity, reliability, clearness, social cohesion (an exhaustive list has not been presented). Creating a safe Internet means that we should take (organizational, contractual, legislative, and other necessary) measures that try to guarantee the deference of these requirements and values.

To identify the right measures that should be taken by all stakeholders (including you and me), we need to analyze the security and human values' threats that are present on the Internet. These threats are present in very different areas of the Internet. By this observation some important research tasks are identified namely (i) the assignment to fully understand all phenomena in all different areas of the current Internet and (ii) the charge to perform a risk analysis. For helping to understand of what's happening on the Internet we have introduced the 4D-hypercube, for performing the risks analyses several methodologies are available we have referred to. To illustrate our observations and ideas, we have presented several topics of research areas. In this way we have tried to give some direction to the research efforts needed to create a safe Internet in the near future.

REFERENCES

1. Andrew S. Tanenbaum. (2003), Computer Networks, Prentice Hall, 4th edition, Upper Saddle River, NJ, USA..
2. Milton L. Mueller. (2004), Ruling the Root: Internet Governance and the Taming of Cyberspace, MIT Press, Cambridge, MA, USA..
3. 'Web 2.0', Wikipedia, http://en.wikipedia.org/wiki/Web_2, (July 25, 2007).
4. Batya Friedman and Peter H. Kahn, Jr., 'Human Values, Ethics, and Design'. (2002), in Human Factors And Ergonomics, pp. 1177-1201, The Information School, University of Washington.
5. 'W32.Toxbot', Symantec, http://www.symantec.com/security_response/writeup.jsp?docid=2005-031012-0442-99, (July 30, 2007) .
6. 'Dutch hackers worked for the mafia', http://www.elsevier.nl/nieuws/internet_en_gadgets/artikel/asp/artnr/72215/index.html, (in Dutch), (July 30, 2007).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

7. 'ECP NL, Platform voor eNederland', http://www.ecp.nl/, (July 30, 2007).
8. ENISA. (2006), 'Risk Management: Implementation principles and Inventories for Risk Management/Risk Assessment methods and tools', http://www.enisa.europa.eu/rmra/files/D1_Inventory_of_Methods_Risk_Management_Final.pdf, (July 30, 2007).
9. Ian H. Witten & Eibe Frank. (2005), Data Mining, Practical Machine Learning Tools and Techniques, 2nd edition, Morgann Kaufman, Burlington, MA 01803, USA..
10. D.B. Fogel & C.J. Robinson, editors. (2003), Computational Intelligence, The Experts Speak, IEEE Press, John Wiley & Sons, USA.
11. 'Location-Based Service', Wikipedia, http://en.wikipedia.org/wiki/Location-based_service, (July 30, 2007).
12. Amr Ali Eldin. (2006), Private Information Sharing Under Uncertainty: Dynamic Consent Decision-making Mechanisms, PhD thesis, Faculty of Technology, Policy, and Management, Technical University Delft, The Netherlands.
13. Bank for International Settlements, 'About the Basel Committee', http://www.bis.org/bcbs/, (July 31, 2007).
14. 'Complex adaptive system', Wikipedia, http://en.wikipedia.org/wiki/Complex_adaptive_system, (July 30, 2007).
15. 'Self-organisation', Wikipedia, http://en.wikipedia.org/wiki/Self-organization, (July 30, 2007).
16. 'DESRIST 2007, Links to Design Science' (2007), http://ncl.cgu.edu/desrist2007/desrist_links.htm, (July 31, 2007).
17. 'Windows Live ID', Wikipedia, http://en.wikipedia.org/wiki/Windows_Live_ID, (July 31, 2007).

ENDNOTES

i The idea we wish to express here is that advanced technical solutions have been developed to fight against viruses and other malware. We recognize at the same time that appropriate preventive countermeasures like an effective battle with the (criminal) distribution of viruses, are still missing sufficient success.

ii This strongly contrasts with the legislation that holds, in a very detailed way, for all kinds of 'classical' real world phenomena and that is based on many decades of experience.

iii At least this observation seems to hold for the discussions started in the Netherlands like those on the widely visited 'E-commerce platform for eNederland' [7]. Fortunately, we observe that a few exceptions are currently becoming visible: E.g., on September 20, 2007, a symposium takes place having as central theme "the future of the soft copy at home in the digital area" where generic solutions are discussed to cope with the loss of property rights' income caused by copying behaviour of consumers at home.

iv Once again, sending information in a secure way (taking CIAA into account) from A to B does not guarantee a safe Internet: the information has got context dependent semantics that is interpreted by its users. An evaluation at this higher level that relates to the content exchanged and the behaviour of users is also needed to judge whether the Internet works fine and safely

v The way financial risks are estimated as imposed by the Basel Committee on Banking Supervision [13] may also be a good source of inspiration here.

vi The IP-protocol is in the heart of the packet-switched communication technology: the IP-protocol is a nice example of a protocol where, in this case, the robustness of the communication is an emergent property that results from the local routing decisions of the very many routers within the global Internet. Many other, similar phenomena exist which are studied under the umbrella of the theory of Complex Adaptive Systems [14]. It is sometimes believed that this theory may also be of use to understand the success of Internet with respect to its governance structure.

vii The (interesting) question arises whether these wireless locally broadcasting LANs will be integrated into large basic substructures of the Internet and become a serious competitor of existing packet-switched infrastructures, which would be a new example of 'inverse infrastructures' showing again the principle of self-organisation [15].

viii Please remember that users have many different roles and that their ((in)secure, (in)sincere, …) behaviour may have its impact on the safety of all other Internet participants

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Towards incentive-based cyber trust

Russell Cameron Thomas

Meritology

Patrick Amon

Center for Interdisciplinary Research for Information Security, Ecole Polytechnique Federale de Lausanne

Abstract "Cyber trust" is the confluence of information security, privacy, digital rights, and intellectual property. Many problems in cyber trust exist at least partially because the people and institutions involved are not properly motivated to solve them. The incentives are often perverse, misaligned, or missing. By improving economic, social, and personal incentives, cyber trust can be significantly improved. The essential elements for incentive-based cyber trust include usability, risk information systems, risk communications, social knowledge, markets, and incentive instruments, along with enabling technology and a supporting legal/ regulatory/institutional framework. We describe an application example in the information supply chain for financial services sector to illustrate the potential benefits and research problems.

Keywords Internet security, privacy, intellectual property, digital rights, e-risk, e-trust, cyber law, information assurance, risk management, incentives, social and organizational aspects.

1 INTRODUCTION

The term "cyber trust" means the confluence of information security, privacy, digital rights, and intellectual property (IP) protection in pervasive communications and computing systems. From the socio-economic perspective of risk management, these information risks are interrelated and are becoming more so. A prime example of the cyber trust confluence is the case of Sony BMG Music Entertainment in 2005, who distributed a copy-protection scheme with music CDs that secretly installed a root kit on computers that played the CDs [1]. (A "root kit" can allow someone else to gain and maintain access to your computer system without your knowledge.) This case involved digital rights (ostensibly, Sony's original intent), information security, copyright infringement, and potential privacy violations.

The problem addressed by this paper is that cyber trust is currently deficient largely because of perverse, misaligned, or missing incentives at all levels. Thus, people and institutions involved are not properly motivated to solve cyber trust problems or do what they can to maximize social welfare [2] [3]. The Sony BMG case reveals conflicting incentives for various actors – media companies such as Sony, platform companies such as Microsoft, security companies such as Symantec and McAfee, and consumers [1] [4].

For individual users, security, privacy, and digital rights mechanisms are often hard to use and, therefore, are often not used as intended [5]. Consumers continue to be very worried about privacy violations and identity theft, yet they do not take action to protect their personal information on home computers [6] [7].

Organization incentives depend on mapping cyber trust to organization performance metrics and decision criteria. A recent survey by the Conference Board [8] found that "most security managers don't know how to map their priorities to business objectives, and most top managers don't understand how security fits into their business objectives." Another factor that makes rational decision-making more difficult is that cyber trust claims made by information and computing technology (ICT) and security vendors are frequently not verifiable or they do not stand up to scrutiny [9]. The result is that the marketplace does not sufficiently reward better security, leading to underinvestment [10].

The incentive-based approach shares the gains (benefits) of cyber trust outcomes in order to align the interests of all stakeholders and mobilize their collective capabilities – intelligence, initiative, agility, and creativity. This is approach requires sociological and economic innovations such as risk modelling and social knowledge pooling, among others.

The main argument of this paper is that an incentive-based approach to cyber trust will yield solutions that are substantially more efficient and effective than alternative approaches, and can also be used in conjunction with other approaches.

2 BACKGROUND

2.1 Terminology

- "Cyber trust" – an umbrella term we have borrowed from the National Science Foundation but we have expand its definition to include the confluence of information security, privacy, digital rights, and intellectual property (IP) protection in pervasive communications and computing systems as seen from the perspectives of all key stakeholders – individuals, organizations, technologists, governments, and society. In this usage, "Cyber" is short for "cyberspace". Therefore, "cyber trust" means trust in cyberspace, in all its forms.
- "Cyber trust risk" or "Cyber risk" – the socio-economic risks associated with cyber trust, from the viewpoint of all relevant stakeholders.
- "Incentive" – Our definition differs somewhat from the usual economic definition: "In economics, an incentive is any factor (financial or non-financial) that provides a motive for a particular course of action, or counts as a reason for preferring one choice to the alternatives." [11] Generally, the incentives we consider are tied to desired outcomes, so that they are a form of gain sharing or shared equity, including remunerative, moral, and personal incentives. We exclude negative or coercive incentives from this definition because we want to draw on and stimulate market forces, broadly defined (see section ⊠3.1).
- "Risk management" – a socio-economic approach to managing uncertain and uncontrollable outcomes, especially when faced with possible events that are hard to estimate and have very bad outcomes [12] [13]. The essence of the risk management approach is to estimate the likelihood and severity of uncertain events and then use these estimates in a rational decision-making framework to guide investments, contingency planning, and other decisions. The general spirit of risk management is to balance the expected value of losses with the costs for mitigating those losses. The sociological aspect of risk management incorporates ideas such as risk tolerance/aversion, bias, risk perception, and motivational dynamics [14].

2.2 Alternative Approaches

There are a variety of approaches to achieving cyber trust and controlling risk [15] [16], which may be used alone or in combination. To be clear, we do not argue that the incentive-based approach is the only approach that should be used nor is it always the best approach. It is unlikely that any of these approaches will be successful in isolation.

1. Technological approach – views cyber trust as a technical problem with primarily technical solutions. Information, communication, biometric, mechanical technology is the prime element in security solutions, with human actors either absent, secondary, or serving merely as users of the technology. In its purest form, there is little or no dependence on organization or social entities other than to permit the technology to be implemented. In essence, the technological approach says, "We can target and subdue the problem with our technology and tools". Its success depends on being able to create and deploy technology with sufficient power and sophistication to overcome the problem.
2. Mandates-based approach – views cyber trust as behaviour and policy control problem and attempts to create solutions involving explicit mandates emanating from centres of authority. Mandates could take the form of regulations, policies, procedures, rules, laws, codes of conduct, contracts, and the like. Centres of authority could include governments, organizations, leaders (formal or informal), administrators, asset owners, or the legal system. Mandates are mostly enforced through audits or inspections, and may or may not have penalties associated with non-compliance. In essence, the mandates-based approach says "Do this!", over and over again. One author puts it succinctly: "As nearly any serious security publication will tell you, security is about control." [17] The success of this approach depends on being able to define explicit mandates and instructions, and also to audit and enforce compliance in practice.
3. Penalty-based approach – views cyber trust as a problem of deviant behaviour and lack of will power to resist temptations to cheat or exploit. It attempts to create solutions that involve penalty or liability schemes that cause individuals or institutions pay a heavy price for actionable vulnerabilities or insecure products [18]. The penalty-based approach is often used in conjunction with mandates, but not always. Together, they say "Do this or else!".

   In cyber trust, the penalty-based approach has been used for many years in copyright and patent protection and, more recently, in enforcing digital rights to creative works. It is being used more and more by governments to protect consumer privacy. There have been some people who have advocated that product liability law should be applied to information and communications technology (ICT) vendors for information security flaws and vulnerabilities, though nothing has been implemented as yet [19] [20] [21] [22].

   To succeed, the penalty-based approach requires that we be able to clearly recognize and define negative outcomes, define injury magnitude, and assign clear responsibility, and to map cause-effect relationships between responsible parties and the negative outcomes. It also requires a system of meaningful and proportionate penalties, along with adjudication and enforcement mechanisms.
4. Political approach – views cyber trust as a problem of power relationships and collective interests. In essence, the political approach says, "Change the power structure, and good things will follow". Its success depends on knowing what is wrong with the current power structure and defining remedies that will make the environment better and not worse. Solutions offered in-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

clude alliances, coalitions, power-shifting actions (e.g. anti-trust law suits), countervailing actions or threats, reciprocal commitments, standardization efforts, and communications to influence public opinion. While the political approach is rarely used to remedy cyber trust at the level of individual incidents or breaches, it is often recommended for use at a societal level or institutional level to explain or remedy perceived root causes of cyber trust problems. Indeed, many experienced information technology (IT) security professionals and consumer advocates believe that a political approach is essential due to entrenched corporate or national interests and actors.

2.3 Limitations of Alternative Approaches

While each of these approaches has some advantages and apply to some circumstances, they have substantial limitations in the current environment and stakeholder needs.

If cyber trust involved only interactions between machines, then technical approaches alone might be sufficient. However, it is well know that cyber trust is a function of technology combined with policies, processes, organization strategy, and various human factors. Thus, improving cyber trust will involve improvements across this spectrum. Furthermore, there is always the question of whether decision-makers will actually invest in technology solutions, should they be available. Therefore, technology alone will not be sufficient to improve cyber trust.

If cyber trust were not so complex, context-dependent, and fast changing, it might be possible to implement command-and-control approaches (mandates, penalties, and/or politics) efficiently and effectively without much concern for incentives. For example, consider the fact that it takes months or years to make and implement command-and-control decisions (e.g. policies, procedures, penalties, laws, etc.). The wheels of bureaucracy turn slowly. Unfortunately, the cyber trust environment changes so fast that, by the time those decisions get put into practice, it almost impossible to avoid obsolescence or irrelevance. Another problem with command-and-control approaches is unintended consequences, especially in crisis situations [23].

The most significant limitation of the political approach is that it is too blunt an instrument. Yes, it could change the cyber trust economics, but there's no way of knowing that the new regime would be better than the current regime. It is very hard to fine-tune or refine.

3 THE INCENTIVE-BASED APPROACH

This approach defines cyber trust as a problem of motivation and action by individuals and institutions, especially their actions toward mutual support, protection, and cooperation. Motivations shape actions, and are in term shaped by perceptions of alternatives, payoffs, risks, and uncertainties. Solutions offered involve incentives and communication of incentives. Incentives may be tangible or intangible, monetary or non-monetary, fungible or non-tradable. Incentives can be embedded in products and services in the form of ease-of-use or help systems. They can be embedded in social systems in terms of social norms and group membership requirements.

In essence, the incentive-based approach says, "Give key actors a share of the potential gains of cyber trust, and thereby draw on the power of self-interest to drive the right actions." To succeed, the incentive-based approach requires that we have a good understanding of what motivates individuals and institutions, what they value, how they perceive cyber risks and rewards, and how to create incentives to shift those motivations in positive directions.

To date, the incentive-based approach has only been implemented on a limited basis in security and privacy. Outside of copyright, digital rights, and IP licensing, there has been little success in monetizing the value of cyber trust. Other forms of incentives have been implemented in an ad hoc fashion.

3.1 Should penalties be included as negative incentives?

Some people view penalties as negative incentives, and would advocated including both positive and negative incentives in an incentive-based approach. We disagree for three reasons.

First, negative incentives tend to promote avoidance behaviours, including shirking, blame shifting, and information hiding (both obscuring and misrepresentation), among other things. This is the opposite of what we are trying to encourage.

Second, there is almost no way to craft negative incentives in such a way to ensure or encourage the most desirable outcomes (i.e. optimization). At best, you can hope to avoid the worst categories of outcomes. This is not sufficient for the current cyber trust environment, where we need to encourage innovation and creative adaptation by all stakeholders.

Finally, our incentive-base approach is based on market systems. Much of the power of market systems comes from its capability to spawn new and complementary markets that share gains and risks. But market systems almost never "trade" negative incentives. For example, if a bank gets a huge fine for regulatory violation, there is no way for the bank to "share" that penalty with their stakeholders (key employees, partners, vendors, etc.), unless those stakeholders are also penalized by the same regulatory body or court. The same is true for criminal liability or stigma/shame. These adhere to specific individuals who have no way to "share" with the organization they work for. There is no practical way to share negative incentives, especially if you are trying to guide collective behaviour toward some global optimum.

The economic impact of potential penalties can be incorporated into the incentive-based approach through the Total Cost of (In)security framework, discussed in section 5.1, below. However, if incentive instruments need to include

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

non-monetary impacts (e.g. stigma, reputation loss, personal consequences), this would require separate treatment and modelling.

3.2 Advantages of the incentive-based approach

Simply put, the incentive-based approach will be more effective than the alternative approaches in the current fast-changing cyber trust environment. While the incentive-based approach is not purely based on market prices and payments, the argument in favour of its relative efficiency and effectiveness reason is basically same as the argument in favour of markets over command-and-control in the modern world economy [24].

A significant advantage of the incentive-based approach is that it does not require the existence of a single global system of incentives or risk measures. To get started and to develop, there is no pre-requisite for industry-wide, country-wide, or international standards or systems. All that is needed to get started is for two parties to have some measure of their relative cyber risk across decision alternatives and how relative cyber risk is driven by observable metrics. This can form the basis of incentive instruments that are mutually agreeable, fair, reward the right behaviour, and aren't easily cheated.

4 ESSENTIAL ELEMENTS

In this section we describe the essential elements for the incentive-based approach and explain why they are necessary.

4.1 Usability

Personal incentives are the foundation for any incentive-based approach to cyber trust. If personal incentives are missing or are in conflict with other incentives, we should expect principal-agent problems (i.e. individuals and organizations may be inclined to bypass or avoid good cyber trust practices, and "principals" incur monitoring and enforcement costs to protect their interests. Efficiency and social welfare both suffer.).

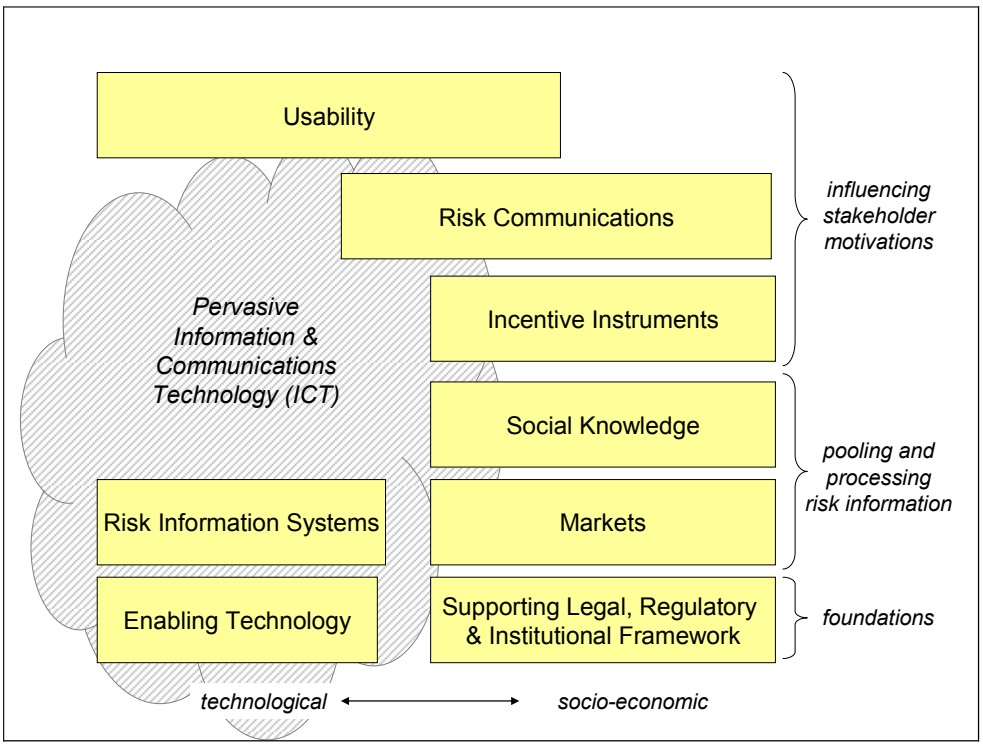
In a sense, we can say that personal incentives are embedded in the design of information and communication systems, and specifically in the usability of their cyber trust features. These personal incentives include making it easy to do the right things, hard to do the wrong things, and making it clear what the risk consequences are of possible actions. Usability includes technology, people, and processes. Poor usability can undermine all the other incentive elements.

There has been considerable interest and research activity recently regarding usable security and privacy [5] [25] [26]. However, this research is not yet well integrated with other elements of incentive-based cyber trust.

4.2 Risk information systems

It will be necessary to have systems to continuously collect and aggregate operational cyber trust information. Without it, it will be impossible to create efficient and effective incentive systems. There have been many calls for information collection and sharing [27] [28], and various organizations and institutions have been set up for this purpose, including CERT, Information Sharing and Access Centres (ISACs) and others. However, these mechanisms almost exclusively focus on operational and technical aspects of cyber trust

Figure 1. Essential Elements for Incentive-based Cyber Trust



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

(vulnerabilities, mitigation, remediation, etc.) and not on the risk management aspects. There is very little empirical data on the social and economic aspects of cyber trust, either for academic researchers or for practitioner in industry or government [29].

Once data is collected, it is necessary to analyze it to discover cause-effect relationships between operational metrics and stakeholder value. Models are needed to help stakeholders make forward-looking, value-based decisions based on risk scenarios and trade-offs. Models will have to cope with many forms of ignorance and uncertainty – an area of active research [30] [31] that has not yet been applied to incentive-based cyber trust. Finally, models need to be structured to fit corporate spending and strategy decisions, which means that the results need to map to existing accounting and budgeting information.

4.3 Risk communication

Incentives have to be presented to actors in a way that is meaningful and actionable, otherwise they won't work. Risk communication includes a range of activities from simple disclosures to sophistication visualizations. Current research on risk perception [14] [32] and risk communication [33] has defined the following challenges that incentive-based cyber trust solutions must address:

- "Risk" has different meanings at an individual level, organization level, and societal level [16] [34].
- Risks and risk perception are usually very specific to context and systemic performance.
- To influence individual behaviour, it is best to give feedback in real-time.
- Risks and risk factors are very interdependent, making the cause-effect relationships very complicated.
- Much of cyber trust knowledge is contingent, tentative, vague, ambiguous, and even contradictory.
- Risk cannot always be measured by a simple numerical scale or value system such as money.
- Prior perceptions and mental models are critical to successful communication and to influence behaviour.
- It is hard to avoid diving into technical details that most people find befuddling and taxing.
- There are many social and political obstacles to disclosing information about cyber trust and risks. No business decision-maker wants to look bad or untrustworthy, so there is a natural inclination to avoid disclosing or even learning about breaches of cyber trust.

4.4 Social knowledge

Mobilizing social knowledge will be critical to incentive-based cyber trust for two reasons. First, knowledge about cyber trust – vulnerabilities, exposures, incidents, losses, mitigation, cost, and forward-looking estimates and perceptions – are all widely distributed. Cyber trust is very dependent on context. Therefore, only the people in that specific context have the necessary information and perspectives to make proper judgments. Second, cyber trust involves both perceptions and forward-looking estimations of risk and these are social processes. Finally, there may be some elements of incentive-based cyber trust that can only be produced by the "wisdom of the crowds", including valuation of hard-to-estimate risks and best practices.

There has been considerable research on social knowledge systems, and also use in practice, with mixed results. Examples include reputation systems [35], peer-to-peer information sharing [36], pooling expert assessments in the face of uncertainty, bias, and weak signals [37] and other mass collaborations [38]. It also includes certification [39] and other products of trusted third parties (TTPs). However, social knowledge systems have only had a limited effect on improving cyber trust, either because they served a limited community (information sharing) or because the information they produced (certifications) was an erroneous signal for cyber trust [40]. Furthermore, social knowledge systems to date have not been integrated with other incentive-based systems.

4.5 Markets

It has been widely recognized that one of the core economic problems of cyber trust is incomplete markets [2]. Because the economic value of cyber trust is not priced and traded, economic actors can not make rational trade-off decisions, leading to inefficient allocation of resources and less-than-optimal results. (By "markets" we mean trading systems that allow buyers and sellers to exchange goods and/or services, including information.)

Of course, primary markets for cyber trust include the real-world commercial markets where customers pay money to suppliers for security products and services. However, it is clear that these markets are far from complete or even sufficient. For example, there are markets for information security products and services, but these are rarely "value priced" in the sense that buyers do not know what improvements cyber trust they are getting when they buy each product or service.

But the range of possible markets also includes synthetic and simulated markets that are created specifically to discover prices [41], to draw out the "wisdom of the crowds" (e.g. prediction markets [42]), to rectify "Tragedy of the Commons" problems due to externalities (e.g. "cap and trade" such as pollution rights markets) [43], markets for private information [44] and to draw out information directly related to cyber trust (e.g. "Zero-day" vulnerability auctions [45]). There has been a significant amount of research lately on artificial markets in general, including these examples: artificial trading markets [46], derivative markets for trading macro risks [47] [48], and artificial markets with intelligent agents [49]. Also relevant is the research into pricing non-marketed assets [50] [51] and non-market methods for eliciting value and preferences [52], which bridges the domains of risk information, risk communication, and markets.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4.6 Incentive instruments

We define "incentive instruments" as any social or economic device, mechanism, process, or agreement that explicitly ties payoffs for actors to desirable future states of the world so that those actors are motivated to help bring about those states. A "payoff" could be monetary, near-monetary (e.g. a tradable good or service), or non-monetary-but-valuable (e.g. offer of mutual assistance). The reason incentive instruments are essential is that they put the value proposition of cyber trust front-and-centre for each stakeholder. They also open the possibility of side payments, compensation, and other balancing transactions to align the interests of stakeholders.

Examples cyber trust incentive instruments that have been implemented or extensively researched include cyber insurance [53] [54] [55] [56], risk-sharing contracts [57], and "bug bounties" [58]. Since the risks associated with cyber trust are frequently either not insured or are not insurable [59] [60] [61], other risk finance and incentive instruments are worth exploring. Outside of the domain of cyber trust, there has been considerable research on risk sharing pools in developing countries [62] [63] [64], risk-based payments and contracts in supply chain management [65] [66] [67], decision insurance (internal to an organization) [68], and risk sharing in other contexts [69] [70]. Those methods and research results should be applied to cyber trust.

New methods are required for digital rights licensing in an era where it is difficult or impossible to prevent unauthorized copying and distribution [71]. The "Street Performer Protocol" [72] and variants are particularly interesting, since they provide for payment to authors/creators prior to distribution. Other interesting variants include the software completion bond [73] and "Voted Compensation" [74] [75]. With some imagination, these might be applied to cyber trust. For example, the Street Performer Protocol might be applied to the market for vulnerability information, fulfilling some of the same objectives as an auction market without some of the negative aspects. Rights-based licensing could be also applied to privacy, where each person retains some rights over their personal information. This requires a new legal framework [76] and appropriate rights management collectives [77].

4.7 Enabling technology

It's obvious that any incentive-based cyber trust scheme would need support from technology. While a detailed discussion of enabling technology is beyond the scope of this paper, we want to make two comments about its required characteristics. First, the incentive systems should be widely distributed and embedded in the pervasive computing and communication systems. Second, the enabling technology needs to present incentive signals to actors at the right times, i.e. when it will have the most effect on behaviour and performance.

4.8 Supporting legal, regulatory, and institutional framework

In addition to enabling technology, it is necessary to have a supporting framework of laws, regulations, and institutions. The best model to draw on is the existing framework for modern financial markets. The laws, regulations, and supporting institutions are set up to facilitate fairness and trust primarily through self-regulation and transparency. Oversight by regulators is essential to make sure that the spirit of laws and regulations are carried out in changing market circumstances. Finally, the day-to-day functioning of the market is carried out by a network of trusted intermediaries (exchanges, clearinghouses, and licensed broker/dealers) and trusted third parties (rating agencies, public accountants, etc.). While these intermediaries and third parties are private institutions, they have a quasi-legal role and have a degree of governmental sanction and oversight. We expect to see a similar framework evolve to support incentive-based cyber trust, perhaps even drawing on the existing framework for financial markets.

In comparison to the legal, regulatory, and institutional framework required by the other approaches, this free market framework is more efficient and more agile to adapt to changing circumstances. It scales better, both in size and across geographic and jurisdictional boundaries. Finally, it is much more likely to foster innovation and avoid unproductive stakeholder conflict.

5 INCENTIVE-BASED CYBER TRUST IN ACTION: FINANCIAL SERVICES

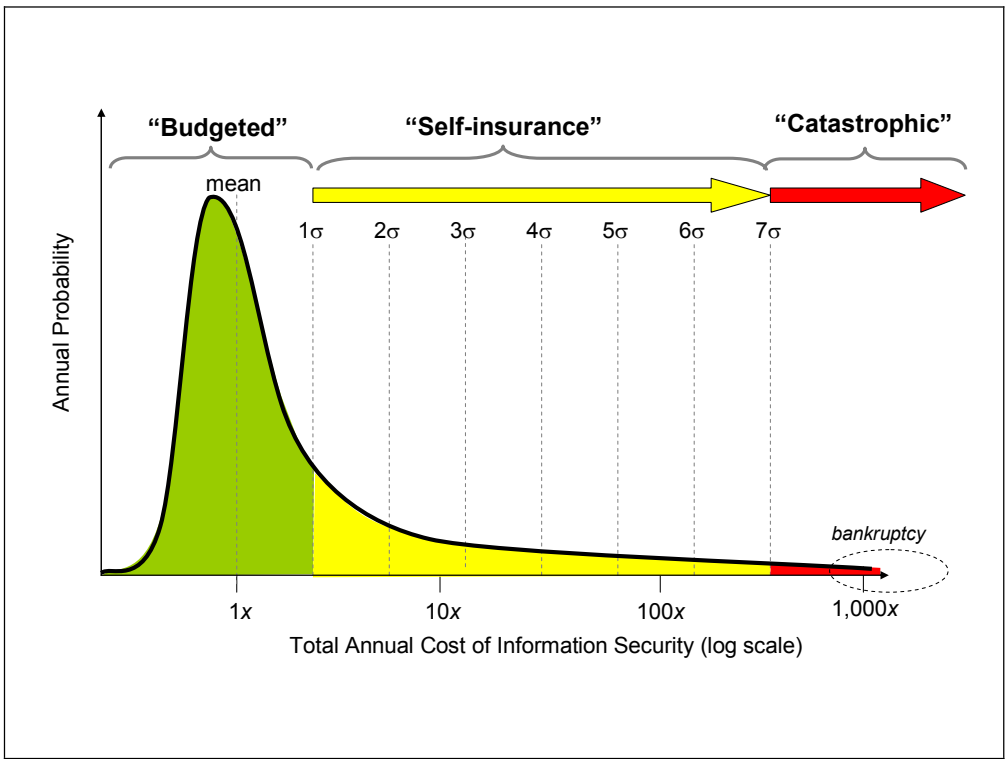
This section provides an illustrative example of how incentive-based cyber trust could be applied to the information supply chain in the financial services (FS) sector.

The FS sector is one of many sectors that face significant cyber trust challenges. In information security, it faces major threats of insider abuse and fraud, external targeted attacks and fraud (e.g. "phishing", money laundering, etc.), misuse of customer's private data (including identity theft), threats to confidential information (insider trading, improper disclosure, etc.), and threats to intellectual property (computer software, proprietary financial data, and even patented business processes). The only element of cyber trust not significant in FS is digital rights, but that might be changing as more financial information and services are delivered digitally and on-demand to be bundled with other digital content.

Furthermore, the FS sector is part of the "critical information infrastructure". (For an excellent overview of the FS sector, its critical information infrastructure, and information risks, see the '2006 Annual Report' of the Financial Services Sector Coordinating Council [78].) The key challenge is "how to steer multi-actor decision making toward an adequate performance of the integrated system with respect to long-term public interests" [79].

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Idealized Probability Distribution for a firm's Total Cost of (In)security



We believe that the FS sector has good potential to be an early adopter for innovative incentive-based solutions for cyber trust. Over the last 20 years, there has been a revolution in quantitative risk management in the FS sector, culminating in the Basel II Accord to promote stability in the financial system through, among other things, market discipline (i.e. incentives). This has led most large banks and insurance companies to take a more comprehensive approach to risk, including developing sophisticated models of operational risk, which includes cyber risk.

We focus our example on cyber risk associated with key links in the FS "information supply chain" – defined as the network of trading and service relationships between firms and between firms and individuals that collectively provide information processing, data, and communication services necessary to support economic value creation.

5.1 Intra- and Inter-firm cyber risk

Problem: One of the main challenges facing information technology (IT) managers and business executives is how to map security metrics and performance to business metrics and performance [8]. This is necessary to align business goals and investments with security requirements, and to balance risks against costs and rewards. Because the benefits of security are the avoidance of uncertain losses, applying traditional cash flow return on investment (ROI) techniques would be inappropriate and misleading. Furthermore, the domain is rife with "unruly uncertainty" (ambiguity, incomplete information, contradictory information, intractability, unknown-unknowns, etc. [14] [80] [81] [82] [83]) which makes it difficult or impossible to reliably estimate annualized loss expectation (ALE) or other probabilistic estimates of expected losses for given incident types.

In large FS firms, they generally have a Risk Management department that measures "operational risk", which is a broad category that includes cyber trust risks. Unfortunately, these operational risk measures are usually aggregated across all risk types (including fraud, process errors, etc.), which makes them less relevant to managing cyber risk through incentives.

Solution: "Total Cost of Cyber (In)security" Framework. Essentially, this would consist of managerial accounting methods and decision support tools to measure the total cost of security (or "insecurity", since loss event costs are included). Here's how it might work:

Divide security-related or cyber trust costs into three categories: "Budgeted", "Self-insured", and "Catastrophic" (Figure 1). Basically, this approach divides the aggregate cost probability distribution into three sections. The fat part of the curve near the mean is "budgeted". The tail section up to some threshold (95%, 99%) is "self-insured". The very far end of the tail is "catastrophic". Therefore, any given incident type, vulnerability, or threat could contribute costs into any or all of these categories.

- "Budgeted costs" are defined to be costs that are predictable and likely within the budget year. This includes all direct spending on security, plus indirect costs, plus the expected value of all high frequency losses and some small mix of lower frequency losses. It also includes the opportunity costs – business activities that are prevented or inhibited by security.
- "Self-insured costs" are less predictable and/or much lower probability within the budget year. Loss magnitudes are potentially big enough to bust the budget (i.e. material to quarterly earnings statements) and even threaten the firm's credit rating, but not neces-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

sarily threaten firm survival. Formally, the loss exposure to be self-insured is the difference between the Budgeted costs and the upper limit for self-insurance, defined by firm strategy. (Note that this scheme does not depend on actually having a self-insurance fund in the accounting sense.)

- "Catastrophic costs" are very unlikely and/or very unpredictable, but could threaten firm survival or even more widespread systemic losses.
- Cost models would be built for each category, drawing on operational security metrics, business process metrics, and estimates of asset value and other values at risk. But the models for each category will be very different.
- Budgeted costs would be modelled using fairly conventional cost-driver models (i.e. linear relationships between operational metrics and indirect or overhead costs, etc.).
- Self-insured costs would be modelled using rank order or order-of-magnitude approaches, possibly combining stochastic methods with inferential reasoning (see Section 5.5, below).
- Catastrophic costs would be modelled using scenario analysis and ordinal or nominal scales. Here, the precision of cost estimate is much less important than its value as a guide strategy and business continuity planning, for example.

This solution would work for any type of security risk or, more broadly, cyber risk. If the loss distribution estimate happens to be normal distribution with relatively modest variance, then it would all fall into the "budgeted" category, and thus could be managed using traditional budget and cash flow methods. On the other hand, if the loss distribution has a "fat tail", then the three-part approach becomes very useful to distinguish between what we know with confidence and what we know with less confidence or don't know at all.

This solution makes the most of existing information, aligns with decision-making processes, and avoids the problem of conflating reliable and unreliable estimates. It requires innovations from Enterprise Risk Management, Activity-based Costing, and qualitative reasoning. The approach is roughly analogous to the Total Cost of Quality concept that helped motivate the Total Quality Management movement [84]. In addition to helping with security cost and performance management, this approach highlights the importance of organization learning and discovery.

Another advantage of this method is that it is compatible with existing methods for enterprise investment and performance management, including "Risk-adjusted Return on Capital" (RAROC) in financial services and "Economic Value-added" (EVA) across various industries. In essence, "self-insurance" adds to the capital required by a project or business unit. Higher levels of cyber risk mean a larger "self-insurance" pool is required, which lowers return on capital, and vice versa [85].

It may be possible to standardize these methods with industries and organization types to allow, for the first time, meaningful aggregation of cyber trust cost information to guide government policy and vendor product development decisions. It would also allow meaningful public disclosure of cyber trust risks and risk tolerance in stakeholder reports and regulatory filings.

Problem: A significant source of cyber risk in the FS firms is the information links it has with trading and outsource partners. This is especially true from the perspective of the financial system as a whole, i.e. systemic risk. It's common practice for partners to have contracts that govern their relationship and transactions, including clauses for information security practices and requirements. These clauses usually define mandates and, sometimes, penalties. Incentive contracts have been used in a few cases to manage supply chain risk [86] [87], especially to build trust and commitment [88]. However, it has not generally been used to manage cyber risk specifically.

Solution: Contingent Payments. It should be possible to define contingent payments tied to specific information security goals, measured by existing operational metrics or scorecards [89]. The contingent payment amounts would be negotiated. While these instruments would provide relatively crude incentives, it could be more efficient than a purely mandates + penalties approach, especially if it promotes creative solutions and information sharing to reduce mutual risks. The risk information system for these contingent payments would be a "Total Cost of Cyber (In)security" model for each firm, or equivalent. They need not be identical.

5.2 ICT vendor cyber risk

Problem: Information and communication technology (ICT) vendors are a special class of trading partners for FS firms, since they provide the very foundation products and services for cyber trust capabilities. However, FS firms often feel as though both operating costs and risk of cyber trust are being dumped on them by vendors through license contracts, service contracts, pricing, and vendor testing and patch release practices. For example, one industry group estimates that the US financial services industry spending on vulnerability and patch management approaches $1B per year [90]. Furthermore, no party in the value chain is disclosing or sharing enough information about vulnerabilities in ICT products, which essentially means that all parties are making decisions in relative darkness. What's missing is compelling incentives for the ICT vendors and buyers to share cyber trust information and work together to implement cost-effective solutions.

It's widely recognized that emergent forms of value for ICT in use (e.g. quality, security, and availability) are jointly created by ICT vendors and their customers. Therefore, cyber trust outcomes should be managed as a joint responsibility. However, current payment and relationship structures don't reflect these facts. No one has figured out how to charge more for higher quality or more secure software due to the "lemon's market" effect (i.e. systemic under-pricing in the

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

used car market due to information asymmetries about quality and post-purchase costs). [2]

Solution: Risk/reward sharing instruments. This financial and contractual instrument between IT vendors and their customers would effectively create risk-adjusted pricing and gain sharing, plus incentives for information disclosure and learning [91] [92]. Here's how it might work:

- The instrument(s) would be some form of forward contract on predefined cash flows from both ICT vendors and customers, approximating a portion of the self-insurance pool for each party associated with their joint cyber trust risks.
- The cash flows would be calculated through activity-driven models using observable quality, reliability, availability, and security metrics, similar to those that are the foundation of Service Level Agreements (SLAs) [93].
- Both vendors and customers would regularly feed metrics information to a trusted third party, who would use simulation models to estimate the expected cash flows and then publish the results. Periodic audits and comparison with public financial statements would be used to validate the output of the activity-driven cash flow models.
- The cost of externalities (i.e. systemic risk) could be included in the models in a variety of forms.
- Based on simulated performance driven by actual operational results, vendors and customers either share the gain (better-than-expected), or loss (worse-than-expected), according to pre-agreed formulas or triggers. (Similar approaches in financial risk management are called "mark-to-model" and "mark-to-future". [94])
- Because they represent cash flows, these instruments could be bundled, repackaged, sold on secondary markets, or tied to subordinated debt to provide liquidity and/or market prices for risk.
- The resulting risk prices could serve the same incentive and signal effect as insurance premiums for traditional property/casualty.

This solution could make a revenue contribution to ICT vendors because any time you can optimize the pricing/packaging/placement of a product or service to better fit what the customers really want, you have the potential to increase customer satisfaction, market share, "share of wallet", or to open up new segments that were not previously economical.

5.3 Consumer risk

Problem: Consumers and individual ICT users generally do not have sufficient understanding or enough information to make good risk/reward decisions regarding cyber trust. This is true not only for major decisions (e.g. purchase, configuration, update, or upgrade) but also for moment-by-moment usage decisions (e.g. visit a web site, enter personal information, use a public WiFi access point, use peer-to-peer file sharing, etc.). As a result, consumers and individual ICT users are both too cautious and too lax in their practices. At best, this leads consumers to worry and feel discomfort; at worst, loss of tangible or reputation. It also creates significant external costs for other individuals and institutions.

Solution: Real-time risk dashboard. It is basically a meter or animated display that provides risk feedback in real-time as the consumer or individual is making use of the ICT devices and services. Microsoft's Internet Explorer (IE) 7 comes with a simplified version of this solution, to warn users about known phishing web sites. Also, Symantec has released a free to download Symantec Internet Threat Meter, based on Yahoo! widgets platform [95]. It displays a qualitative risk index rates the four main online activities, including e-mail, web activities, instant messaging and file sharing on a low, medium or high risk level based on general conditions on the internet, but not on a particular user's system or related to their specific activities. What we are suggesting is much more complete and compelling for the consumer. Here's how it might work:

- It would need to be fed by a knowledge base of considerable depth and sophistication, preferably pooling the knowledge of many users in similar circumstances. Peer-to-peer data and knowledge sharing models could be appealing, with appropriate mechanisms for preserving anonymity and protection against gaming the system.
- Sophisticated modelling would be required to characterize the user's configuration, assets at risk, normal and abnormal activity patterns, risk tolerance, and to map these factors to threats. However, considerable modelling and data complexity can be avoided through abstraction, pattern recognition, and inferential reasoning.
- Prediction markets for estimating or forecasting key parameters could be useful. Participants could include ICT vendors, security and privacy experts, risk management professionals, and even (by proxy) consumers themselves.
- The most important information to give the consumer/user is relative expected value changes for alternative courses of action (e.g. visit the site vs. not). While it is tempting to put this into a rigorous decision-theoretic framework using money values, that may not be necessary or even the most useful way to model or convey the information.
- It would be useful to incorporate real-time pricing for identity theft insurance. Currently in the US, several companies provide identity theft insurance. While many policies cover both credit losses and lost wages, this insurance doesn't cover the largest potential cost – destruction of consumer's credit rating [96]. Furthermore, premiums for identity theft do not reflect the relative risk of policy holders. Identity theft insurance could be a more effective cyber trust incentive instrument if even simple methods were used to value expected drop in credit rating vs. income level, and also rating the risk exposure and reduction practices of policy holders. This would provide risk pricing information to consumers and might improve their risk mitigation behaviour.
- Whatever information is chosen for display, it is critical that it is displayed in a meaningful, compelling,

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

and comfortable way. Perhaps there is some middle ground between the static or animated icons now used on browsers and the animated cartoon Office Assistant by Microsoft, which was engaging but entirely uninformative.

This solution might be offered as an independent product or service, or it might be bundled with existing or new products or services, which might speed adoption and enhance the value proposition for both consumers and vendors. For example, if this solution were linked with a consumer risk sharing pool, then it might be possible to display their real-time "insurance premium", "coverage limit", or other related self-insurance or mutual assurance value (either monetary or in-kind value).

6 CONCLUSION

We believe that the incentive-based approach to cyber trust will yield solutions that are substantially more efficient and effective than alternative approaches alone. It can also augment technical solutions and mandate-plus-penalty systems to make them more effective. There are many unresolved theoretical and empirical questions, including:

Is it theoretically possible to model cyber trust risks and incentives in a unified, forward-looking valuation framework? What are the fundamental limits [15] [97] [98]?

If analytic, quantitative models are not feasible, is it possible to devise coarse-grained or qualitative models that are robust and usable in practice (e.g. rating or ranking schemes) as the basis for incentive instruments?

REFERENCES

1. Schneier B. (2005), 'Sony's DRM Rootkit: The Real Story', Schneier on Security (blog), Nov.17, 2005, http://www.schneier.com/blog/archives/2005/11/sonys_drm_rootk.html (1 Sept. 2007).
2. Anderson R. (2001), 'Why information security is hard - an economic perspective', in Proceedings of the 17th Annual Computer Security Applications Conference, 10-14 Dec. 2001, New Orleans, LA, IEEE Computer Society, Washington, DC.
3. Spafford E. (2003), 'Four grand challenges of trustworthy computing', Computer Research Association, www.cra.org/Activities/grand.challenges/security/ (1 Sept. 2007).
4. Krebs B. (2005), 'Study of Sony anti-piracy software triggers uproar', Washington Post, 2 Nov. 2005, http://www.washingtonpost.com/wp-dyn/content/article/2005/11/02/AR2005110202362.html (1 Sept. 2007).
5. Cranor L. and Garfinkel S, (2005), Security and Usability - Designing Secure Systems that People Can Use, O'Reilly Media, Sebastopol, CA.
6. _____ (2005), 'Internet security voter survey', Computer Security Industry Alliance, https://www.csialliance.org/resources/pdfs/CSIA_Internet_Security_Survey_June_2005.pdf (1 Sept. 2007).
7. _____(2006), 'Global consumer attitudes toward data protection', Visa and Harris Interactive, www.corporate.visa.com/pd/pdf/Consumer_Global_Research_Backgrounder.pdf (1 Sept. 2007).
8. _____ (2006), 'Navigating risk—the business case for security', Conference Board, http://www.conference-board.org/publications/describe.cfm?id=1231 (1 Sept. 2007).
9. Schneir B. (2007), 'How security companies sucker us with lemons', Wired, April 7, 2007, http://www.wired.com/politics/security/commentary/securitymatters/2007/04/securitymatters_0419 (1 Sept. 2007).
10. Garcia A. and Horowitz B. (2007), 'The potential for underinvestment in internet security: implications for regulatory policy', Journal of Regulatory Economics, Springer, vol. 31(1), pages 37-55.
11. _____ (2007), 'Incentive', Wikipedia, Wikimedia Foundation, http://en.wikipedia.org/wiki/Incentive (1 Sept. 2007).
12. Crouhy M., Galai D., and Mark R. (2006), The Essentials of Risk Management, McGraw-Hill, New York, NY.
13. Lam, J. (2003), Enterprise Risk Management: From Incentives to Controls, John Wiley & Sons, Hoboken, NJ.
14. Adams, J. (2001), Risk: the policy implications of risk compensation and plural rationalities, Routledge, London, UK.
15. Cibrorra C.(2004), 'Digital technologies and the duality of risk', London School of Economics, Economic & Social Research Council (ESRC), Discussion Paper #27, http://www.lse.ac.uk/collections/CARR/pdf/Disspaper27.pdf (1 Sept. 2007).
16. Starr C. and Whipple C. (1980), 'Risks of risk decisions', Science, Volume 208, Issue 4448, pp. 1114-1119,

http://www.sciencemag.org/cgi/content/abstract/208/4448/1114 (1 Sept. 2007).

17. Jaquith, A. (2007), Security Metrics – Replacing Fear, Uncertainty, and Doubt, Addison-Wesley Upper Saddle River, NJ.
18. Schneier B. (2007), 'Information security and externalities', ENISA Quarterly (European Network and Information Security Agency), January 2007, http://www.schneier.com/essay-150.html (1 Sept. 2007).
19. Shostack A. (2005), 'Avoiding liability: an alternate route to more secure products', at 4th Workshop on the Economics of Information Security (WEIS 2005), Cambridge, MA, 2-3 June 2005, http://infosecon.net/workshop/pdf/44.pdf (1 Sept. 2007).
20. Schneier B. (2004), 'Security and compliance', IEEE Security & Privacy,IEEE, vol. 2 num. 3, p. 96.
21. Kim B. C., Chen P.-Y., and Mukhopadhyay T (2004), 'Monopoly, software quality and liability', at Web Information Systems Engineering

(WISE 2004), 22-24 Nov. 2004, Brisbane, Australia, http://opim.wharton.upenn.edu/wise2004/sat621.pdf (1 Sept. 2007).

22. Crews C. W. (2005), 'Cybersecurity finger pointing – regulation vs. markets for software liability, information security, and insurance', Issue Analysis, Competitive Enterprise Institute, Washington, DC, no. 7, http://www.cei.org/pdf/4569.pdf (1 Sept. 2007).
23. Hennessy, J. , Patterson D., and Lin H., (editors) (2003), Information Technology for Counterterrorism: Immediate Actions and Future Possibilities , Committee on the Role of Information Technology in Responding to Terrorism, National Research Council, National Academies Press, Washington, DC, http://www7.nationalacademies.org/cstb/pub_counterterrorism.html (1 Sept. 2007).
24. Hayek F. A. (1945), 'The use of knowledge in society', The American Economic Review, Vol. 35, No. 4, pp. 519-530.
25. _____ (2006), Proceedings of the second symposium on Usable privacy and security, 12-14 July, Pittsburgh, PA (SOUP '06),

ACM Press, New York, NY, http://cups.cs.cmu.edu/soups/ (1 Sept. 2007).

26. Egelman S. and Kumaraguru P. (2005), 'Report on the DIMACS workshop and working group meeting on usable privacy and security software', Center for Discrete Mathematics & Theoretical Computer Science (DIMACS), Rutgers University, Piscataway, NJ, 7-8 July 2004 http://dimacs.rutgers.edu/Workshops/Tools/dimacsrpt.pdf (1 Sept. 2007).
27. _____ (2006), 'Information sharing/critical infrastructure protection task force report', President's National Security Telecommunications Advisory Committee, National Communication System (NCS), a division of the US Department of Homeland Security, http://ncs.gov/nstac/reports/2000/ISCIP-Final.pdf (1 Sept. 2007).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

28. Cukier K., Mayer-Schoenberger V., and Branscomb L. (2005) 'Ensuring (and insuring?) critical information infrastructure protection', 5th Conference on Information Law and Policy for the Information Economy, Harvard Kennedy School of Government, Rueschlikon, Switzerland, 16-18 June 2005, http://bcsia.ksg.harvard.edu/BCSIA_content/documents/rwp_05_055_viktor_branscomb.pdf (1 Sept. 2007).
29. Pfleeger S., Rue R., Horwitz J., Balakrishnan A., (2006), 'Investing in cyber security: the path to good practice', in Cyber Security: Strengthening Corporate Resilience, Cutter Consortium, Arlington, MA.
30. Ghirardato P. (2001), 'Coping with ignorance: unforeseen contingencies and non-additive uncertainty', Economic Theory 17, Springer-Verlag, p 247-276.
31. Halpern J. (2003), Reasoning about Uncertainty, MIT Press, Cambridge, MA.
32. Slovic P. (2000), The Perception of Risk, Earthscan, London, UK.
33. Morgan, M.G., Fischhoff, B., Bostrom, A., and Atman, C. (2002), Risk Communication – A mental models approach, Cambridge University Press, Cambridge, UK.
34. Thompson P. and Dean W. (1996), 'Competing Conceptions of Risk', Risk, Risk Assessment & Policy Association (Franklin Pierce Law Center), Concord, NH, vol. 7 p. 361, http://www.piercelaw.edu/risk/vol7/fall/thompson.htm (1 Sept. 2007).
35. Resnick P., Zeckhauser R., Friedman E., and Kuwabara K. (2000) 'Reputation Systems', Communications of the ACM, 43(12), p 45-48.
http://www.si.umich.edu/~presnick/papers/cacm00/index.html . Also see: http://web.si.umich.edu/reputations/bib/bib.html
36. Verma, D. (2004), Legitimate Applications of Peer-to-Peer Networks, Wiley-Interscience, Hoboken, NJ.
37. Bouissou M. and Thuy N. (2002), 'Decision-making based on expert assessments: use of belief networks to take into account uncertainty, bias, and weak signals", in Proceedings of European Safety & Reliability International Conference (ESREL, Lyon, France, 18-21 March 2002), http://perso-math.univ-mlv.fr/users/bouissou.marc/ExpertsAndBN_ESREL02.pdf (1 Sept. 2007).
38. Tapscott D. and Williams A. (2006), Wikinomics: How Mass Collaboration Changes Everything, Portfolio (Penguin Group), New York, NY.
39. Ross R., et. al. (2004), 'Guide for the security certification and accreditation of federal information systems', National Institute of Standards and Technology, http://csrc.nist.gov/publications/nistpubs/800-37/SP800-37-final.pdf (1 Sept. 2007).
40. Edelman, Benjamin (2006), 'Adverse selection in online 'trust' certifications', at 5th Workshop on the Economics of Information Security (WEIS 2006), Cambridge, UK, 26-28 June 2006, http://weis2006.econinfosec.org/docs/10.pdf (1 Sept. 2007).
41. Scalas E., Bianchetti M., Mainardi F., Roman H. E. and Vivoli A. (2003), 'Synthetic markets', http://www.mfn.unipmn.it/~scalas/wehia2003sm/wehia2003sm.html (1 Sept. 2007).
42. Wolfers J. and Zitzewitz E. (2004), 'Prediction markets', Journal of Economic Perspectives, American Economics Association, vol. 8 issue 2, p. 107-126.
43. Camp L. J. and Wolfram C. (2004), 'Pricing security: vulnerabilities as externalities', Economics of Information Security, Vol. 12, Spring.
44. Gopal R., Garfinkel R., Nunez M., and Rice D. (2006), 'Electronic markets for private information: economic and security considerations', in Proceedings of the 39th Hawaii International Conference on System Science (HICSS'06), IEEE Computer Society, Los Alamitos, CA, vol. 6, issue 04-07, page 113a.
45. Böhme R. (2005), 'Vulnerability Markets – what is the economic value of a zero-day exploit', at 22nd Chaos Communications Congress (CCC 2005, Berlin, Germany), http://events.ccc.de/congress/2005/fahrplan/attachments/542-Boehme2005_22C3_VulnerabilityMarkets.pdf (1 Sept. 2007).
46. LeBaron B. (2002), 'Building the Santa Fe artificial stock market', Working Paper, Brandeis University, June 2002. www.econ.iastate.edu/tesfatsi/blake.sfisum.pdf (1 Sept. 2007).
47. Shiller R. (1998), Macro Markets – Creating Institutions for Managing Society's Largest Economic Risks, Oxford University Press, Oxford, UK.
48. Shiller R. (2003), The New Financial Order: Risk in the 21st Century, Princeton University Press, Princeton, NJ.
49. Karacapilidis N. and Moraïtis P. (2001), 'Intelligent agents for an artificial market system' , in Proceedings of the Fifth international Conference on Autonomous Agents (AGENTS '01), Montreal, Canada. 28 May – 1 June 2001, ACM Press, New York, NY, p. 592-599.
50. Luenberger D. (2001), 'Projection pricing', Journal of Optimization Theory and Applications, vol. 109, num. 1, p. 1-25(25).
51. Luenberger D. (2002), 'A correlation pricing formula' (working paper), Stanford University, www.stanford.edu/dept/MSandE/people/faculty/luenberger/RealOptions/CPFWeb.pdf (1 Sept. 2007).
52. Cameron T., Poe G., Ethier R., and Shulze W. (2002), 'Alternative non-market value elicitation methods – are the underlying preferences the same?', Journal of Environmental Economics and Management, Association of Environmental and Resource Economists (Elsevier), vol. 44, p. 391-425.
53. Kesan J., Majuca R., and Yurcik W. (2004), 'Economic case for cyberinsurance' (working paper), University of Illinois College of Law, 2004-2.
54. Gordon L., Loeb M., and Sohail T. (2003), 'A framework for using insurance for cyber-risk management'. Communications of the ACM, ACM Press, New York, NY, vol. 46, num. 3, p. 81-85.
55. Böhme R. (2005), 'Cyber insurance revisited', at 5th Workshop on Economics of Information Security (WEIS 2005), Cambridge, MA 2 - 3 June 2005, http://infosecon.net/workshop/pdf/15.pdf (1 Sept. 2007).
56. Baer W. (2003), 'Rewarding IT security in the marketplace', RAND Corporation, http://tprc.org/papers/2003/190/BaerITSecurity.pdf (1 Sept. 2007).
57. Kim B. C., Chen P-Y., and Mukhopadhyay T. (2005), 'An economic analysis of a software market with risk-sharing contracts', in Proceedings of the International Conference on Information Systems (ICIS), Las Vegas, 11-14 Dec. 2005.
58. Granick J. (2006), 'Bug bounties exterminate holes', Wired, 12 Apr. 2006, http://www.wired.com/news/columns/0,70644-0.html (1 Sept. 2007).
59. Rossi M. (2002), 'Insuring first-party cyber risk for Fortune 1000 companies—a worthwhile endeavor or boondoggle?', International Risk Management Institute, Nov. 2002,
http://www.irmi.com/Expert/Articles/2002/Rossi11.aspx (1 Sept. 2007).
60. Wood L. (2004), 'When all else fails, there's cyberinsurance', Information Security, August 2004, http://infosecuritymag.techtarget.com/ss/0,295796,sid6_iss446_art920,00.html (1 Sept. 2007).
61. Holmes T. (2004), 'Cyber insurance: weighing the costs with the risks', National Federation of Independent Business, 20 Oct. 2004, www.nfib.com/object/IO_18562.html (1 Sept. 2007).
62. Dubois P., Jullien B., Magnac T. (2006), 'Formal and informal risk sharing in LDCs', submitted to Econometrica, http://www.toulouse.inra.fr/centre/esr/CV/dubois/djm.pdf (1 Sept. 2007).
63. Briys E. and Célimène F. (2004), 'Globalisation and risk sharing: debunking some common pitfalls', at Colloque International sur les Impacts Economiques et Politiques de la Mondialisation, Port-au-Prince , Haiti, September 2004, http://cyberlibris.typepad.com/blog/files/Haiti2.doc (1 Sept. 2007).
64. Bramoullé Y., Kranton R. (2004), 'Risk sharing networks', Centre Interuniversitaire sur le Risque, Les Politiques Économiques et l'Emploi, Working Paper 05-26, Sep. 2005.
65. Narayanan V. G. and Raman A. (2004), 'Aligning incentives in supply chains', Harvard Business Review, Nov. 2004, p 94-102.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

66. Gulati R. (2001), 'Increasing the odds: creating and managing intelligent alliances' (presentation), Kellogg Graduate School of Management, Northwestern University.

67. Cruz J., Nagurney A., Wakolbinger T. (2006), 'Financial engineering of the integration of global supply chain networks and social networks with risk management', Naval Research Logistics 53, p 674-696.

68. Hogg T. and Huberman B. (2006), 'Taking risk away from risk taking: decision insurance in organizations', HP Labs working paper, 13 April 2006, www.hpl.hp.com/research/idl/papers/insurance/insurance.pdf (1 Sept. 2007).

69. Kunreuther H. and Heal G. (2003), 'Interdependent security', Journal of Risk and Uncertainty 26, p 231-249.

70. Krueger D. and Uhlig H. (2005), 'Competitive risk sharing contracts with one-sided commitment', CFS Working Paper Series 2005/07, Center for Financial Studies.

71. Wallenberg F. (2002), 'Aligning incentives in copyright – a soft approach to fair use', UC Berkeley working paper, May 9, 2002. www.ischool.berkeley.edu/~fredrik/research/papers/AligningIncentives.pdf (1 Sept. 2007).

72. Kelsey J. and Schneier B. (1998), 'Electronic Commerce and the Street Performer Protocol', in The Third USENIX Workshop on Electronic Commerce Proceedings, 31 Aug.-3 Sept.1998, Boston, MA, USENIX Press, November 1998.

73. Rasch C. (2001), 'The Wall Street performer protocol: using software completion bonds to fund open source software development', First Monday 6-6, , http://www.firstmonday.org/issues/issue6_6/rasch/index.html (1 Sept. 2007).

74. Dorrell P. (2005), 'Published digital information is a public good: the case for voted compensation', Mar. 2, 2005. http://www.1729.com/ip/PublicGood.html (1 Sept. 2007).

75. Dorrell P. (2006), 'Looking for a win/win solution to the war between 'premium content' and digital freedom', Dec. 26, 2006, http://www.1729.com/blog/LookingForAWinWin.html (1 Sept. 2007).

76. Bergelson V. (2003), 'It's personal but is it mine? Toward property rights in personal information', Rutgers Law School (Rutgers), Faculty Papers, Paper 33, http://law.bepress.com/rutgersnewarklwps/fp/art33 (1 Sept. 2007).

77. Besen S., Kirby S. N. (1989) 'Compensating creators of intellectual property', RAND Corporation, http://www.rand.org/pubs/reports/R3751/ (1 Sept. 2007).

78. _____ (2006), 'Annual report 2006', Financial Services Sector Coordinating Council, https://www.fsscc.org/reports/2006/annual_report_2006.pdf (1 Sept. 2007).

79. Weijnen M.P.C., Bouwmans I., de Vries L.J. (2005), 'Coping with critical infrastructures – steering multi-actor decision making', at General Conference of The International Risk Governance Council: Implementing a Global Approach to Risk Governance, Beijing, China, 20-21 Sep. 2005.

80. Acquisti A., Grossklags J., 'Uncertainty, ambiguity, and privacy', at 4th Annual Workshop on the Economics of Information Security (WEIS 2005), Cambridge, MA, 2-3 June, http://infosecon.net/workshop/pdf/64.pdf (1 Sept. 2007).

81. Syverson P. (2003), 'The paradoxical value of privacy', at 2nd Annual Workshop on Economics and Information Security (WEIS 2003) , College Park, MD, 29-30 May 2003, http://chacs.nrl.navy.mil/publications/CHACS/2003/2003syverson-privcost.pdf (1 Sept. 2007).

82. Ryan J. (2003), 'The use, misuse, and abuse of statistics in information security research', at American Society of Engineering Management National Conference (ASEM 2003), St. Louis, MO, 16-18 Oct. 2003, http:// www.seas.gwu.edu/~jjchryan/asem03.pdf (1 Sept. 2007).

83. Mandelbrot B. and Taleb N. (2006), 'A focus on the exceptions that prove the rule', Financial Times, March 23 2006, http://www.ft.com/cms/s/5372968a-ba82-11da-980d-0000779e2340,dwp_uuid=77a9a0e8-b442-11da-bd61-0000779e2340.html (1 Sept. 2007).

84. Crosby P. (1979), Quality Is Free: The Art of Making Quality Certain, New York: McGraw-Hill, 1979.

85. Thomas R. (2007), 'Total cost of cyber (in)security', at Mini-Metricon 6 Feb., San Francisco, http://meritology.com/resources/Total%20Cost%20of%20Cyber%20(In)security.ppt (1 Sept. 2007).

86. Kam B. (2005), 'Managing outsourcing risks in the global supply chain: an exploration of approaches', presented at International Trade and Logistics, Corporate Strategies and the Global Economy, Le Havre, France, 28-29 Sept. 2005.

87. Bryson K.-M. and Sullivan W. (2003), 'Designing effective incentive-oriented contracts for application service provider hosting of ERP systems', Business Process Management Journal, vol. 9 no. 6, 2003, p. 705-721, http://www.ituniv.se/program/sem_research/Publications/2003/Sul03/BPMJ9-6Final.pdf (1 Sept. 2007).

88. Goo J. and Nam K. (2007), 'Contract as a source of trust – commitment in successful IT outsourcing relationships: an empirical study', in Proceedings of the 40th Hawaii International Conference on System Sciences,Waikoloa, HI, 3-6 Jan. 2007, www.hicss.hawaii.edu/hicss_40/decisionbp/09_11_01.pdf (1 Sept. 2007).

89. _____ (2004), 'Kalculator: BITS key risk measurement tool for information security operational risks', BITS (Financial Services Roundtable, http://www.bitsinfo.org/downloads/Publications%20Page/BITS%20Kalculator/bitskalcnarrative.pdf (1 Sept. 2007).

90. _____(2004), 'Software providers should accept responsibility for their role in supporting us financial institutions and critical infrastructure', BITS/Finanical Services Roundtable press release, April 27, 2004, www.bitsinfo.org/downloads/Misc/bitssoftsecuritypolicy-apr04.pdf (1 Sept. 2007).

91. Osei-Bryson K.-M. and Ngwenyama O. (2000), 'Structuring IS outsourcing contracts for mutual gain: an approach to analyzing performance incentive schemes', Journal of the Association for Information Systems, Vol. 1, November 2000.

92. Osei-Bryson K.-M. and Ngwenyama O. (2006), 'Managing risks in information systems outsourcing : An approach to analyzing outsourcing risks and structuring incentive contracts', European Journal of Operational Research, Vol. 174, No 1, 2006, pp. 245-264.

93. Goo J., Kim D., and Cho B. (2006), 'Structure of service level agreements (SLA) in IT outsourcing: the construct and its measurement', in Proceedings of the 12th Americas Conference on Information Systems (AMCIS), Acapulco, Mexico, 4-6 Aug., 2006.

94. Dembo, R., Aziz A., Rosen D., and Zerbs M. (2001), Mark-to-Future – A framework for measuring risk and reward, Toronto: Algorithmics Publications, May 2001, http://www.algorithmics.com/research/MarktoFuture/toc.shtml (1 Sept. 2007).

95. _____(2006), 'Symantec internet threat meter', Symantec, Cupertino, CA, http://www.symantec.com/norton/themes/threatmeter/index.jsp (1 Sept. 2007).

96. _____ (2004), 'Identity Theft Insurance', http://www.targetwoman.com/articles/identity-theft-insurance.html (1 Sept. 2007).

97. du Toit B., 'Risk, theory, reflection: limitations of the stochastic model of uncertainty in financial risk analysis', June 2004, www.riskworx.com/insights/theory/theory.pdf (1 Sept. 2007).

98. Nilsen T. and Aven T. (2003), 'Models and model uncertainty in the context of risk analysis', Reliability Engineering and System Safety, vol. 79, no 3, 1 March 2003, pp. 309-317(9), http://risikoforsk.no/Publikasjoner/ModUsRESS2.pdf (1 Sept. 2007).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Language-based security policy enforcement

George S. Oreku
Department of Computer Science and Engineering,
Harbin Institute of Technology,
Nangang District, Harbin 150001 China
A13 Room 601, P.O. Box 773, 92 Xi Dazhi Street
gsoreku@yahoo.com

Jianzhong Li
Department of Computer Science,
Harbin Institute of Technology
92 West Dazhi Street, Nangang District,
Harbin 150001, China
lijzh@hit.edu.cn

Fredrick J. Mtenzi
Dublin Institute of Technology, Faculty of Science,
Kevin Street, Dublin 8, Ireland
Fred.mtenzi@dit.ie

Abstract Languages-based security promises to be a powerful tool with which provably secure routing applications may be developed. Programs written in these languages enforce a strong policy of non-interference, which ensures that high-security data will not be observable on low-security channels. The information routing security proposed aim to fill the gap between representation and enforcement by implementing and integrating the divers security services needed by policy. Policy is enforced by the run-time compiler and executions based mechanism to information violating routing policy and regulation of security services. Checking the routing requirements of explicit route achieves this result for statements involving explicit route. Unfortunately, such classification is often expressed as an operation within a given program, rather than as part of a policy, making reasoning about the security implications of a policy more difficult. We formalize our approach for a C++-like language and prove a modified form of our non-interference method. We have implemented our approach as an extension to C and provide some of our experience using it to build a secure information routing

Keywords Security, Policy enforcement, Compiler, Execution

1 INTRODUCTION

Works in communication security policy have recently focused on general-purpose policy languages and evaluation algorithms. However, because the supporting frameworks often defer enforcement, the correctness of a realization of these policies in software is limited by the quality of domain-specific implementations. The term security policy has been used to represent many aspects of computer security as can be seen in McLean (1990), (Woo and Lam 1991), Sandhu (1993), Blaze et al., (1996), Bellovin (1999), Bartal et al., (1999) McDaniel (2002).

A security policy is enforced when its semantics are realized by software behavior. Enforcement can be as simple as the dropping of a packet by a firewall, or as complex as the execution of a leader election protocol in a secure group. How an application or service enforces policy has a direct affect on the security and efficiency of its operation. Communication security policies have historically been crafted for the specific systems they support according to Ylonen (1998) and Kent et al., (1998). The architects of these systems explicitly define the range of security behaviors desired. Therefore, security is addressed only in as much as the architects foresee the needs of its future users.

Recent efforts within the security and policy communities have investigated general-purpose representations that vastly increase the scope of policy Blaze et al.,(1996), Ryutov and Neuman (2000), Durham et al.,(2000), Patz et al.,(2001), and hence address the needs of a much larger constituency. While these efforts have achieved many of the stated goals, they have not yet considered general-purpose policy enforcement. This paper introduces an approach to language-based security policy enforcement. General purpose of the lan-

guage-based security policy enforcement is to fills the gap between general-purpose representations and enforcement by defining a proof concept in which the diverse services required by policy can be easily implemented and integrated.

An information routing policy is a security policy that describes the authorized paths along which that information can route. Each model associates a label, representing a security class, with information and with entities containing that information. Each model has rules about the conditions under which information can move throughout the system. Historically communication security policies have been always crafted for the specific systems they support. According to Ylonen (1998) and Kent et al., (1998) we find that language provided the rudimentary tools to achieve low-level security goals and its extension were necessary to formulate and enforce application policy.

These languages provide a means of provably enforcing a security policy in a broader sense. A current technique for enforcing security routing relies on so called best practices like it has been looked by Montgomery and Murphy (2006) which include simplistic techniques (such as password, TCP, Authentication, rout filter, and private addressing) to mitigate the most rudimentary vulnerabilities and threats. Theoretical models for security-typed languages have been actively studied and are continuing to evolve Volpano et al., (1996). For example, researchers are extending these models to include new features, such as exceptions, polymorphism, objects, inheritance, side-effects, threads, encryption, and many more as presented by Sabelfeld and Myers (2003).

A current techniques within the security and policy communities have investigated general purpose representations that vastly increase the scope of policy McDaniel (1996), Ryutov and Neurman (2000), Durham et al.,(2000) Patz et al.,(2001), and hence address the needs of a much larger constituency. While these efforts have achieved many of the stated goals, they have not yet considered general purpose policy enforcement

The threats from malicious attack are both real and serious. All routing protocols currently deployed on the internet are vulnerable to several classes of attack. The simplest, and perhaps most threatening, is the compromise and control of valid routers. Some reports suggest that would-be attackers can gain access to hundreds of BGP-speaking routers on the black market for a single stolen credit-card number Montgomery and Murphy (2006). Consequences include loss of connectivity (black holes and partitions, for example), eaves dropping (routing traffic through malicious nodes), suboptimal routing (using congested, delayed, or unstable paths), and routing system disruptions (causing churn and instability in the routing protocols themselves, for instance).

To address this lack of practical experience, to date some works have been going on, i.e. Secure Protocols for the Routing Infrastructure [SPRI] project available at: www.cyber.st.dhs.gov/spri.html.Internet Engineering task Force (IETF) is striving to understand existing routing protocols threat addressing the practical requirements and constraints of today's operational environments. K.Butler et al (2004) are proposing secure BGP to strike different balance between security and performance. Promising approach such as understanding practical application developments in security-typed languages implementations in real-world systems and policy using security-typed languages have been discussed in Hicks et al., (2006) while specific applications such as security policy enforcement in the Antigone system to present the Antigone architecture, and demonstrate non-trivial applications and policies presented in McDaniel and Prakash (2002) gives a far-reaching vision of policy enforcement.

Our extended view of policy allows us to consider new ways of using context. Security-typed programming language allows the issuers of policy to augment applications through policy specification. We sought to discover whether this tool for secure programming could hold up to its promise of delivering real-world applications with strong security guarantees. In practice, the security policies enforced by program monitors grow more complex both as the monitored software is given new capabilities and as policies are refined in response to attacks and user feedback. This is best illustrated by examples proposed dealing with policy complexity by organizing policies in such a way as to make them composeable. We present a fully implemented Compiler and execution-based mechanism that allows security engineers to specify and enforce composeable policies on C++ applications. We also formalize the central workings by defining an unambiguous semantics for our applied language.

1.1 Security Challengies,Requirements and Goals

The security policy we defined at the outset is driven by a range of security goals and requirements, Confidentiality, Integrity and Availability (CIA). Based cryptographic traditional security mechanism, such as authentication protocols, digital signature and key management which responsible to keep track of binding keys and assist on establishing mutual trust and secure communications are posing both challenges and opportunities of archiving security goals. Cryptographic in routing protocols gives challenges of difficulties on time synchronizations, dependence complexity of techniques as routing service need to bootstrap themselves (i.e. directories, basic startup operations of management system).Consequence of potential nor loss on investment have been encouraging Commercial entities to devote and deploy more secure infrastructure.

There is no standardized security solution for most routing technologies to date. Designing extended security or new protocols is extremely difficult. In a long run no single security solutions can address all routing protocols since routing protocols differ in their design even within single routing protocols different security might be required. Platform in which routing protocols are operating is another challenge i.e. More than three orders of magnitude have different exit in the control, different data plane's processing capabilities.

The complexity of and requirements imposed on routing technologies continue to escalate and this will increase

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

the potential vulnerabilities to and consequence of focused routing system attacks Montgomery and Murphy (2006). Internet Engineering Task Force (IETF) routing protocols security requirements working group gives more discussion on this, available at www.ietf.org/html.charters/rpsec-charter.httm

2 INFORMATION ROUTING POLICY

Information routing policies define the way information moves throughout a system. Typically, these policies are designed to preserve confidentiality of data or integrity of data. In the former, the policy's goal is to prevent information from routing to a user not authorized to receive it. In the latter, information may route only to processes that are no more trustworthy than the data.

Any confidentiality and integrity policy embodies an information routing policy.

Example: The Model describes a lattice-based information routing policy. Given two compartments *A* and *B*, information can route from an object in *A* to a subject in *B* if and only if *B* dominates *A*. Let *x* be a variable in a program. The notation $\underline{x}$ refers to the information routing class of *x*

Example: Consider a system that uses the Model above. The variable; which holds data in the compartment (TS, {NUC, EUR}), is set to 3. Then *x* = 3 and *x* = *(TS,{NUC,EUR})*.

Intuitively, information routing from an object *x* to an object *y* if the application of a sequence of commands *c* causes the information initially in *x* to affect the information in *y*.

Definition 1. The command sequence *c* causes a routing of information from *x* to *y* if, after execution of c_1 some information about the value of *x* before *c* was executed can be deduced from the value of *y* after *c* were executed. This definition views information routing in terms of the information that the value of allows one to deduce about the value *y* in *c*. For example, the statement

```
y := x;
```

reveals the value of *x* in the initial state, so information about the value of *x* in the initial state can be deduced from the value of *y* after the statement is executed. The statement

```
y := x / z;
```

reveals some information about *x*, but not as much as y := x statement. The final result of the sequence *c* must reveal information about the initial value of *x* for information to route. The sequence

```
tmp := x;
y := tmp;
```

has information routing from *x* to *y* because the (unknown) value of *x* at the beginning of the sequence is revealed when the value of *y* is determined at the end of the sequence. However, no information routing occurs from trap to x, because the initial value of trap cannot be determined at the end of the sequence.

Example: Consider the statement

```
x := y + z;
```

Let *y* take any of the integer values from 0 to 7, inclusive, with equal probability, and let *z* take the value i with probability 0.5 and the values 2 and 3 with probability 0.25 each. Once the resulting value of *x* is known, the initial value of *y* can assume at most three values. Thus, information routes from *y* to *x*. Similar results hold for *z*.

Example: Consider a program in which *x* and y are integers that may be either 0 or 1. The statement

```
if x = 1 then y := 0;
else y := 1;
```

does not explicitly assign the value *x* of to *y*.

Assume that x is equally likely to be 0 or 1. Then $H(x_S)=1$. But $H(x_S \; I \; y_f)=0$, because if *y* is 0, *x* is 1, and vice versa. Hence,

$H(x_S I y_f)=0 < H(x_S I y_S) = H(x_S)=1$ Thus, information routes *x* from to *y*.

Definition 2. An implicit routing of information occurs when information flows from *x* to *y* without an explicit assignment of the form *y:=f(x)*, where *f(x)* is an arithmetic expression with the variable *x*.

The routing of information occurs, not because of an assignment value of x, but because of a routing control based on the value of x. This demonstrates that analyzing programs for assignments to detect information routing is not enough. To detect all routing of information, implicit routing must be examined

3 EXECUTION BASED MECHANISM

Bulleted lists The goal of an execution-based mechanism is to prevent an information routing that violates policy. Checking the routing requirements of explicit route achieves this result for statements involving explicit routings. Before the assignment

$y=f(x_1, \ldots,x_n)$ is executed, the execution-based mechanism verifies that

$lub(\underline{x}_1, \ldots,\underline{x}_n) \leq \underline{y}$

If the condition is true, the assignment proceeds. If not, it fails. A naive approach, then, is to check information routing conditions whenever an explicit routing occurs.

Implicit routing complicates checking

Example: Let x and y be variables. The requirement for certification for a particular statement *y op x is that* $\underline{x} \leq \underline{y}$.The conditional statement

```
if x = 1 then y := a;
```

causes a routing from x to y. Now, suppose that when $x \neq 1$, $\underline{x} \leq High$ *and* $\underline{y} \leq Low$. If routing were verified only when explicit, and $x \neq 1$, the implicit routing would not be checked . The statement may be incorrectly certified as complying with the information routing policy.

3.1 Variables Classes

The classes of the variables in the examples above are fixed. This suggests a notion of dynamic classes, wherein a variable can change its class. For explicit assignments, the change is straight forward. When the assignment

$y := f(x_1, \cdots, x_n)$

occurs, y's class is changed to $lub(x_1, \ldots, X_n)$. Again, implicit routing complicates matters.

Example: Consider the following program (which is the same as the program in the example for the Data Mark Machine adopted from Denning (1982)

```
proc copy (x: integer class {x};
var y: integer class {y});
var z: integer class variable {Low};
begin
y := 0;
z := 0;
if x=0 then z := 1;
if z=0 then y := 1;
end;
```

In this program, $\underline{z}$ is variable and initially *Low*. It changes when something is assigned to z.Routings are certified whenever anything is assigned to y. suppose $\underline{y} < \underline{x}$.

If $x=0$ initially, the first statement checks that $Low \leq \underline{y}$ (trivially true). The second statement sets z to 0 and $\underline{z}$ to *Low*. The third statement changes $\underline{z}$ to1 and $\underline{z}$ to $lub(Low, x) = \underline{x}$. The fourth statement is skipped (because $z=1$). Hence, y is set to 0 on exit.

If $x=1$ initially, the first statement checks that $Low \leq \underline{y}$ (trivially true). The second statement sets $\underline{z}$ to 0 and z to *Low*. The third statement is skipped (because $x=1$). The fourth statement assigns 1 to y and checks that $lub(Low, \underline{z}) = Low \leq \underline{y}$ (again, trivially true). Hence, y is set to 1 on exit.

Information has therefore routed from toeven though. The program violates the policy but is nevertheless certified.

4 COMPILER, BASED MECHANISM

Compiler-based mechanisms check that information routing throughout a program are authorized. The mechanisms determine if the information routing in a program could violate a given information routing policy. This determination is not precise, in that secure paths of information routing may be marked as violating the policy; but it is secure, in that no unauthorized path along which information routing will be undetected.

Definition 3. A set of statements is certified with respect to an information routing policy if the information routing within that set of statements does not violate the policy.

Example: Consider the program statement

```
if x = 1 then y := a;
else y := b;
```

By the rules discussed earlier, information routes from x and a to y or from x and b to y, so if the policy says that, $\underline{a} \leq \underline{y}$ $\underline{b} \leq \underline{y}$, and $\underline{x} \leq \underline{y}$ then the information routing is secure. But if $\underline{a} \leq \underline{y}$ only when some other variable $z=1$, the compiler-based mechanism must determine whether $z=1$ before certifying the statement. Typically, this is infeasible. Hence, the compiler-based mechanism would not certify the statement. The mechanisms described here follow those developed by Denning (1982).

4.1 Declarations

For our discussion, we assume that the allowed routing is supplied to the checking mechanisms through some external means, such as from a file. The specifications of allowed routing involve security classes of language constructs. The program involves variables, so some language construct must relate variables to security classes. One way is to assign each variable to exactly one security class. We opt for a more liberal approach, in which the language constructs specify the set of classes from which information may route into the variable. For example,

```
x: integer class { A, B }
```

states that x is an integer variable and that data from security classes A and B may route into x. Note that the classes are statically, not dynamically, assigned. Viewing the security classes as a lattice, this means that x's class must be at least the least upper bound of classes A and B that is, $lub\{A,B\} \leq \underline{x}$.

Two distinguished classes, *Low* and *High*, represent the greatest lower bound and least upper bound, respectively, of the lattice. All constants are of class *Low*. Information can be passed into or out of a procedure through parameters. We classify parameters as *input parameters* (through which data is passed into the procedure), *output parameters* (through which data is passed out of the procedure), and *input/output parameters* (through which data is passed into and out of the procedure). Consider the following program which is the same as the program in the example in Bishop (2004).

```
(* input parameters are named i_s; output parameters, o_s; *)
(* and input/output parameters, io_s, with s a subscript *)
proc something(i_1, ..., i_k; var o_1, ..., o_m, io_1, ..., io_n);
var l_1, ..., l_j; (* local variables *)
begin
        S; / * body of procedure *)
end;
```


Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

The class of an input parameter is simply the class of the actual argument:

i_s: *type* **class** { i_s }

Let r_1, ..., r_p be the set of input and input/output variables from which information routing to the output variable os. The declaration for the type must capture this:

o_s: *type* **class** { r_1, ..., r_p }

(We implicitly assume that any output-only parameter is initialized in the procedure.) The input/output parameters are like output parameters, except that the initial value (as input) affects the allowed security classes. Again, let r_1, ..., r_p be defined as above. Then:

io_s: *type* **class** {r_1, ..., r_p, io_1, ..., io_k }

Example: Consider the following procedure for adding two numbers.

```
proc sum(x: int class { x };
var out: int class { x, out });
begin
        out := out + x;
end;
```

Here, we require that $x \leq out$ and $out \leq out$ (the latter holding because ≤ is reflexive). The declarations presented so far deal only with basic types, such as integers, characters, floating point numbers, and so forth. Nonscalar types, such as arrays, records (structures), and variant records (unions) also contain information. The rules for information routing classes for these data types are built on the scalar types.

Consider the array

a: array 1 .. 100 of int;

First, look at information routing out of an element *a[i]* of the array. In this case, information routing from *a[i]* and from *i*, the latter by virtue of the index indicating which element of the array to use. Information routing into *a[i]* affect only the value in *a[i]*, and so do not affect the information in *i*. Thus, for information routing from a[i], the class involved is $lub\{\underline{a[i]},\underline{i}\}$; for information routing into *a[i]*, the class involved is $\underline{a[i]}$.

5 PROGRAM STATEMENTS

A program consists of several types of statements some of them typically are conditional statement, Goto statement and procedure calls. We use the same statements for our compiler based approach.

5.1 Conditional Statements

A conditional statement has the form

```
if f(x1, ..., xn) then
        S1;
else
        S2;
end;
```

where $x_1, \ldots, x_n$ are variables and f is some (boolean) function of those variables. Either S_1 or S_2 may be executed, depending on the value of f, so both must be secure. As discussed earlier, the selection of either S_1 or S_2 imparts information about the values of the variables x_1, ..., x_n, so information must be able to route from those variables to any targets of assignments in S_1 and S_2. This is possible if and only if the lowest class of the targets dominates the highest class of the variables x_1, ..., x_n. Thus, the requirements for the information routing to be secure are:

- S_1 secure
- S_2 secure
- $lub\{x_1, ..., x_n\} \leq glb\{ y \mid$ y is the target of an assignment in S_1 and S_2 $\}$

As a degenerate case, if statement S_2 is empty, it is trivially secure and has no assignments.

Example: Consider the statements

```
if x + y < z then
        a := b;
else
        d := b * c - x;
end;
```

Then the requirements for the information routing to be secure are $\underline{b} \leq \underline{a}$ for S_1 and $lub\{\underline{b},\underline{c},\underline{x}\} \leq d$ for S_2. But the statement that is executed depends on the values of *x*, *y*, and *z*. Hence, information also routes from *x*, *y*, and *z* to *d* and *a*. So, the requirements are $lub\{\underline{y},\underline{z}\} \leq \underline{x},\underline{b} \leq a$, and $lub\{\underline{x},\underline{y},\underline{z}\} \leq glb\{\underline{a},\underline{d}\}$.

5.2 Go to Statements

A *goto statement* contains no assignments, so no explicit routing of information occurs .Implicit routing may occur; analysis detects these routing.

Definition 4. A basic block is a sequence of statements in a program that has one entry point and one exit point.

Example: Consider the following code fragment from Bishop (2004) adopted for our method.

```
proc transmatrix (x: array [1..10] [1..10] of int class{x});
var y: array [1..10][1..10]of int class{y}};
var i , j : int class {tmp}
begin
        i : =1                  {*b1*}
12:     if i>10 goto 17 ;       {*b2*}
        j : =1;                 {*b3*}
14:     if j>10 then goto 16;{*b4*}
        y[j][i] : x[i][j];      {*b5*}
        j:=j+1;
        goto 14;
16:     i:=i+1;                 {*b6*}
        goto 12;
17:     {*b7*}
end;
```


Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

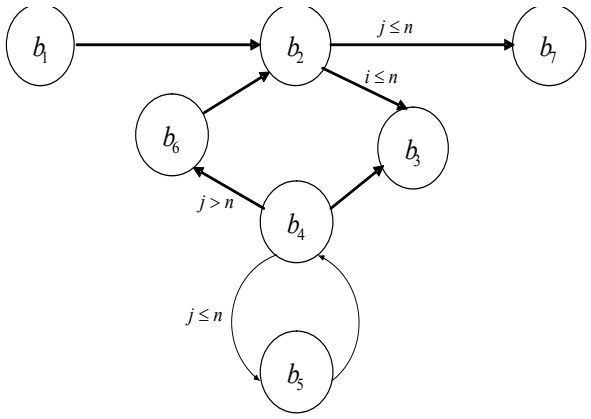
Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

There are seven basic blocks, labeled b_1 through b_7 and separated by lines. The second and fourth blocks gave two ways to arrive at the entry either from a jump to the label or from the previous line .They also have two ways to exit either by the branch or by falling through to the next line .The fifth block has three lines and always ends with a branch .The sixth block has two lines and can be entered either from a jump to the label or from the previous line .The last block is always entered by a jump.

Control within a block routing from the first line to the last. Analyzing the routing of control within a program is therefore equivalent to analyzing the routing of control among the program's basic blocks .Figure 1 shows the routing of control among the basic blocks of the body of the procedure Tran matrix.

The basic blocks are labeled b_1 through b_7.The conditions under which branches are taken are shown over the edges corresponding to the branches.

Figure1. The control routing graph of the procedure transmatrix



When a basic block has two exit paths ,the block reveals information implicitly by the path along which control routing .When these paths converge later in the program ,the (implicit)information routing derived from the exit path from the basic block becomes either explicit (through an assignment) or irrelevant .Hence, the class of the expression that causes a particular execution path to be selected affects the required classes of the blocks along the path up to the block at which the divergent paths converge .

Definition 4. An immediate forward dominator of a basic block b (written IFD (b)) is the first block that lies on all paths of execution that pass through b.

Example: In the procedure transmatrix, the immediate forward dominators of each block are $IFD(b_1)=b_2$, $iFD(b_2)=b_7$, $iFD(b_3)=b_4=b_6 IFD(b_5)=b_4$,

and $IFD(b_6)=b_2$,

Computing the information routing requirement for the set of blocks along the path is now simply applying the logic for the conditional statement. Each block along the path is taken because of the value of an expression .Information routing from the variables of the expression into the set of variables assigned in the blocks. Let be B_i the set of blocks along an execution path from b_1 to $IFD(b_1)$,but excluding these endpoints. Let X_{i1}, ..., X_{in} be the set of variables in the expression that selects the execution path containing the blocks in B_i.The requirements for the program's information routing to be secure are : All statements in each basic block secure $lub\{\underline{x}_{i1}, ..., \underline{x}_{in}\}\{y \mid y$ is the target of an assignment in $B_i\}$

Example: Consider the body of the procedure transmatrix .We first state requirements for information routing within each basic block:

b_1: $low \le \underline{i} \Rightarrow secure$

b_1: $low \le \underline{j} \Rightarrow secure$

b_5: $lub\{\underline{x[i][j]}, \underline{i}, \underline{j}\} \le \underline{y} \le \underline{y[i][j]}$; $\le \underline{j} \le \underline{j} \Rightarrow lub\{\underline{x[i][j]}, \underline{i}, \underline{j}\} \le \underline{y[i][j]}$

b_6: $lub\{low, \underline{i}\} \le \underline{i} \Rightarrow secure$

The requirement for the statements in each basic block to be secure is ,for $i=1,...,n$ and $j=1,...,n$, $lub\{\underline{X[i][j]}, \underline{i}, \underline{j}\} \le \underline{y[j][i]}$. By the declarations, this is true when $lub\{\underline{X}, \underline{i}\} \le \underline{y}$. In this procedure, $B_2=\{b_3,b_4,b_5,b_6\}$ and $B_4=\{b_5\}$,Thus, in B_2, statements assign values to i,j *and* $y[i][j]$.In B_4,statements assign values to j *and* $y[i][j]$.The expression controlling which basic blocks in B_2 are executed is $i \le 10$;the expression controlling which basic blocks in B_2, are executed is $j \le 10$.Secure information routing requires that $i \le glb\{\underline{i},\underline{y}\}$ *and* $i \le glb\{\underline{i},\underline{y}\}$, *or* $\underline{i} \le \underline{y}$. Combining these requirements ,the requirement for the body of the procedure to be secure with respect to information routing is $lub\{\underline{X}, \underline{i}\} \le \underline{Y}$.

5.3 Procedure Calls

A procedure call has the form

```
proc procname(i1, ..., im : int; var o1, ..., on : int);
begin
        S;
end;
```

where each of the ij's is an input parameter and each of the oj's is an input/output parameter. The information routing in the body S must be secure. As discussed earlier, information routing relationships may also exist between the input parameters and the output parameters. If so, these relationships are necessary for S to be secure. The actual parameters (those variables supplied in the call to the procedure) must also satisfy these relationships for the call to be secure. Let x_1, ..., x_m and $y1$, ..., y_n be the actual input and input/output parameters, respectively. The requirements for the information routing to be secure are

S secure

For j = 1, ..., m and k = 1, ..., n, if $ij \le o_k$ then $x_j \le y_k$

For j = 1, ..., n and k = 1, ..., n, if $o_j \le o_k$ then $y_j \le y_k$

Example: Consider the procedure transmatrix from section 5.2. As we showed there, the body of the procedure is secure with respect to information routing when $lub\{\underline{x}, \underline{tmp}\} \leq y$. This indicates that the formal parameters x and y have the information routing relationship $\underline{x} \leq \underline{y}$. Now, suppose a program contains the call

transmatrix (a, b)

The second condition asserts that this call is secure with respect to information routing if and only if $\underline{a}$.

6 CONCLUSION

This paper focus on the language based information routing security. We have separately presented complier based and execution based mechanism to specify and enforce security policies with c++ language. We have demonstrated that it is possible to implement security policy using security-typed languages. However, further investigation of the language based support for policy enforcement is necessary before they can fulfill their considerable promise of enabling more secure routing.

Our work in language based (C++) policy enforcement also uncovered three central deficiencies. First aspects of information routing are the amount of information routed and the way in which it is routing. Given the value of one variable, entropy measures the amount of information that one can deduce about a second variable. Second the routing can be explicit, as in the assignment of the value of one variable to another, or implicit, as in the antecedent of a conditional statement depending on the conditional expression. Third traditionally, models of information routing policies form lattices. Should the models not form lattices, they can be embedded in lattice structures. Hence, analysis of information routing assumes a lattice model.

The concept language based security lies in policy. The current security requirements and future environments are unable to get along with the largely fixed security models embodied in existing software systems or infrastructure as there is little or no infrastructure to formulate policy. Our approach addresses incongruity insecurity by allowing flexibility to environment applied as security requirements are as diverse as the environments in which systems exist, support for flexible policy- defined security is desirable.

Our policy compiler and execution-based mechanisms definitely to be in practice in a larger sense security-typed language, more research before their promise is met is needed. We take this work as another milestone in that achievement

REFERENCES

1. Available at: www.cyber.st.dhs.gov/spri.html (Accessed date: 16 March 2006).
2. Available at: www.ietf.org/html.charters/rpsec-charter.httm (Accessed date: 16 March 2006)
3. Available at: www.patrickmcdaniel.org/pubs/td-5ugj33.pdf. (Accessed date: 8 May 2007)
4. Bishop M., Introduction to Computer Security Prentice Hall PTR, October 2004.
5. Butler K. et al., (2004) A Survey of BGP Security, tech. report TD-5UGJ33, AT&T Labs- Research.
6. Blaze M., Feigenbaum J., and Lacy J., (1996) Decentralized Trust Management. In Proceedings of the IEEE Symposium on Security and Privacy, pages 164–173, Los Alamitos.
7. Bellovin S., Distributed Firewalls. ; Login: pages 39–47, 1999.
8. Bartal Y., Mayer A. J., Nissim K., and Wool A., (1999) Firmato: A novel firewall management toolkit. In IEEE Symposium on Security and Privacy, pages 17–31.
9. Denning D., Cryptography and Data Security, Reading, MA Figure 5.5, p. 285 Addison-Wesley 1982.
10. Durham D., Boyle J., Cohen R., Herzog S., Rajan R., and Sastry A..(2000) RFC 2748, The COPS (Common Open Policy Service) Protocol. Internet Engineering Task Force.
11. Hicks B., King D., McDaniel P., and Hicks M., (2006) Trusted declassification: High-level policy for a security-typed language. In Proceedings of the 1st ACM SIGPLAN Workshop on Programming Languages and Analysis for Security (PLAS '06), Ottawa, Canada, ACM Press.
12. Kent S. and Atkinson R., (1998) Security Architecture for the Internet Protocol. Internet Engineering Task Force, RFC 2401. Available at : www.ietf.org/rfc/rfc2401.txt (Accessed date: 8 May 2007)
13. McDaniel P., (1996) Proceedings of the 1996 IEEE Symposium on Security and Privacy, Los Alamitos, pages 164–173.
14. McDaniel P. and Prakash A., (2002) Methods and Limitations of Security Policy Reconciliation. In IEEE Symposium on Security and Privacy, (IEEE, CA) pages 73–87.
15. McLean J., (1990) The Specification and Modeling of Computer Security. IEEE Computer, 23(1):9–16, January.
16. Montgomery D. and Murphy S.,(2006) "Towards Secure Routing Infrastructures" IEEE Security & Privacy, Volume 4, no.5 (Sept. / Oct.),pp 84-87.
17. Patz G., Condell M., Krishnan R., and Sanchez L., (2001) Multidimensional Security Policy Management for Dynamic Coalitions. In Proceedings of Network and Distributed Systems Security 2001, Internet Society, San Diego, CA, (to appear)
18. Ryutov T. and Neuman C., (2000) Representation and Evaluation of Security Policies for Distributed System Services. In Proceedings of DARPA Information Survivability Conference and Exposition, pages 172–183, Hilton Head, South Carolina, DARPA.
19. Sabelfeld and Myers A. C., (2003) Language-based information flow security, IEEE Journal on Selected Areas in Communications, 21(1):5–19.
20. Sandhu R. S., (1993) Lattice-based access control models. IEEE Computer, 26(11):9–19.
21. Volpano D., Smith G., and Irvine C., (1996) A sound type system for secure flow analysis, JCS, 4(3):167–187.
22. Woo T. and Lam S., (1993) Authorization in Distributed Systems; A New Approach. Journal of Computer Security, 2(2-3):107–136,
23. Ylonen T., (1996) SSH - Secure Login Connections Over the Internet. In Proceedings of 6th USENIX UNIX Security Symposium, pages 37–42. USENIX Association, June. San Jose, CA.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Korea prepares for the upcoming ubiquitous society

Byung Joo Jeong

Abstract This paper aims at introducing Korean government's ubiquitous IT strategy and pilot projects, and Korea's hopes for the future information society. Korean government has been exerting best efforts to prepare for the upcoming ubiquitous society, which pursues a human oriented society by making people's life more comfortable and affluent. As part of its efforts, the government proposed a blueprint for future information society named the 'u-Korea Master Plan'. Its goal is to create the world's first ubiquitous society and transform Korea into an advanced society. Everything dubbed "ubiquitous" - ranging from u-health, u-defense, u-city to u-payment - means an environment where people can enjoy access to high-speed networks and advanced communication services anywhere and anytime through a ubiquitous computing network. Korea's strategy especially emphasizes RFID (Radio Frequency Identification) and USN(Ubiquitous Sensor Network) application pilot projects, to create demand and activate ubiquitous IT industry. Because it is a field which can bring about a huge change to industry as a whole. Now, Korea is confronted with the opportunity and challenge of future information society. The successful implementation of ubiquitous IT strategy and pilot projects will be the touchstone to benchmark for future information society.

1 KOREAN GOVERNMENT'S UBIQUITOUS IT STRATEGY

Since the late 1990s, Korea's information technology sector has achieved remarkable growth, and has been the most important industry driving the country's economic growth. IT sector's contribution to the country's gross domestic production averages at 38.4 percent per year. IT exports took up 36 percent of the country's total exports, or $102.3 billion in 2005.

To keep the IT growth alive, Korean government suggested a new IT roadmap aimed at ushering in a ubiquitous society. Because the world ubiquitous IT industries are expected to emerge as the strategic next generation areas. Korea has carried out research to establish developmental strategy in coping with the future information society paradigm shift to the ubiquitous society.

As a part of these efforts, the government has established a vision in which the 'ubiquitous revolution' provides a momentum for further national development, enabling the country

Figure 1. u-Korea vision & goals

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 1. Current status of informatization in Korea

Category	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007.6
Broadband internet subscriber(10,000 households)	1.4	37.4	401	781	1,041	1,118	1,192	1,219	1,404	1,444
Internet user(10,000 persons)	310	1,080	1,904	2,438	2,627	2,922	3,158	3,301	3,412	3,443
Internet usage rate(%)	-	-	44.7	56.6	59.4	65.5	70.2	72.8	74.8	75.5
Mobile Phone Subscriber(10,000 persons)	-	2,344	2,682	2,905	3,234	3,359	3,659	3,834	4,020	4,232
Internet Banking Subscriber(10,000 persons)	-	-	409	1,131	1,771	2,275	2,427	2,674	3,591	4,011

* Korea population (2005) : 4,704(10,000 persons)
* Source : NIA, 2007)

Table 2. u-Korea Master Plan Goals

Goals		Contents
Advancement Goals for Five Areas	Friendly Government	Pursue a friendly government using ubiquitous IT and evolve from supplier-oriented administrative services to user and site-oriented services.
	Intelligent Land	Establish the 'Intelligent Land' using ubiquitous IT features of status identification and autonomous response.
	Regenerative Economy	Improve transparency in economic systems using various technologies such as RFID/USN, mobile Internet, and intelligent robots, digital TVs, etc., and revitalize the national economy and enhance growth potential by invigorating traditional industries
	Secure & Safe Social Environment	Establish a safe and clean society by building a preventive environment against disasters and diseases through real-time data collection enabled by diverse new technologies such as bio-sensors, high performance computers, satellite communication technology, RFID/USN, etc.
	Tailored u-Life Services	Provide convenient and affluent living conditions for individuals by delivering services customized for individual taste and environment through a home-network, intelligent robots, etc.
Optimization Goals for Four Engines	u-Globalization Engine: Balanced Global Leadership	Strengthen IT trade and global cooperation to become the u-IT leader, and vigorously lead global standardization activities.
	u-Industry Engine: Ecological Industrial Infrastructure	Promote u-Cluster, establish test-beds for developing core technologies; vitalize industry, and build integrated u-infrastructure for bringing diverse ubiquitous services together.
	Streamlining Social Infrastructure	Create public consensus by expanding opportunities to experience u-services, establish ubiquitous social systems through consolidating legal systems, build safe and reliable policies for the ubiquitous society, and strengthen policies for privacy protection.
	Transparent Technological Infrastructure	Establish ubiquitous network accessible anywhere and anytime, develop application technology for promoting u-Korea, create standardized environment to enhance global market competitiveness, and continue to promote IT839 Strategy as the key engine for u-Korea.

* Source : NIA, 2007

to establish its position as the world's IT hub. As part of its efforts, the government proposed a blueprint for future information society named the 'u-Korea Master Plan(2006)'. The Master Plan provides a blueprint that guides how to use IT to deal with the new social and economic demands and carry out nationwide innovation to become the world's top in terms of IT in the ubiquitous society. The Vision of the 'u-Korea Master Plan' is to achieve an advanced Korea by realizing the world's FIRST u-Society based on the world's BEST u-Infrastructure.

Under the vision, the Plan provides advancement goals for five areas – government, land, economy, social environment, and individual life; and optimization goals for four engines – globalization, industrial infrastructure, social infrastructure, and technology development.

The ultimate goal of the 'u-Korea Master Plan' is to achieve a society where all people can benefit from a safer ubiquitous society (4U: Universal, Usable, Unisonous, Upgraded) through advancement of the five areas and optimization of the four engines. (MIC, 2006)

Its goal is to create the world's first ubiquitous society and transform Korea into an advanced society. Under the five-year master plan, the government aims at advancing Korea's global ranking to 15th in state competitiveness and 25th in quality of life until 2010. (According to the IMD, South

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 3. RFID/USN projects

RFID Pilot Projects	Carry out pilot projects in 12 areas including procurement, national defense, and environment from 2004 Carry out 6 pilot projects in 5 areas such as medicine, food, air freight, u-Fishfarm, and mobile in 2006
RFID Full-scale Projects	Carry out 4 Full-scale projects such as environment, national defense, port logistics, and unification that has ripple effects in the industry among service model verified through pilot projects
USN Field Tests	Carry out 9 Field Tests on such as marine environment, agricultural product cultivation environment, and bridge monitoring

Table 4. 7 leading service models

Service Name	Contents
u-Health monitoring service	Assessing bionic information simply with remote health assessing equipment in home and set doctor provides proper health managing information according to the record. In urgent situation, by picture communication equipment in an ambulance and remote medical equipment, bionic signal monitoring is supported.
u-Remote medical service	Providing remote medical service between nurse in nurse based facilities In urgent situation, by picture communication equipment in an ambulance and remote medical equipment, bionic signal monitoring is supported. u-Health based united supporting system for household management/convalescence of discharged patients and patients of chronical disease.
Glycosuria managing service	Provides risk management like figure change analysis or test for a compilation based on user personal information like fatness or high-cholesterol. Provides proper Glycosuria related information such as diet/exercise program, medicine or insulin dosing education materials for managing blood sugar.
Hypertension managing service	Provides risk management like figure change analysis or test for a compilation based on user personal information such as accompanied disease like Cardiac muscle blocking. Provides proper Hypertension related information such as diet/exercise program, medicine dosing education materials for managing blood pressure.
Chronical raspiratory disease managing service	Provides risk management fit to user's basic information, health condition and environmental change like disease seriousness Provides proper respiratory disease related information such as rehabilitation program, medicine dosing education materials for managing respiratory disease.
Musculoskeletal disease managing service	Provides exercise program, disease related information, living environment managing information for curing, rebelling musculoskeletal diosease.
Remote medical service for isle/mountain village	Nurse visits a patient of chronical disease lives in island/mountain village so that he can't go to hospital himself easily, measures a patient's health condition and sends record instantly to a doctor of a related medical organization and perform treatment according to the doctor's guide. In urgent situation, by picture communication equipment in an ambulance and remote medical equipment, first aid supporting system is provided.

* Source : NIA, 2006

Korea ranked 29th in state competitiveness and 41st in quality of people's life in 2005.) During the 2006-2010 period, Korea will make a strong push for u-city, u-health, u-transportation, environment and u-logistics.

Government has been exerting best efforts to prepare for the upcoming ubiquitous society, which pursues a human oriented society by making people's life more comfortable and affluent.

2 RFID PILOT PROJETS

Korea's strategy especially emphasizes RFID (Radio Frequency Identification) and USN(Ubiquitous Sensor Network) application pilot projects, to create demand and activate ubiquitous IT industry. Because it is a field which can bring about a huge change to industry as a whole.

Many new pilot projects with new technologies applied were explored and implemented in 2006 in order to seek for strategic project development that will lead the RFID/USN fields in the future. Also upon evaluating the performance of RFID pilot projects since 2004 and the expected benefits of the projects when expanded, four main projects were selected and implemented in full-scale from 2006.

Especially u-city project to have the ubiquitous computing based RFID/USN industry integrated with every part of our society. U-city project intends to give residents a convenient and safe lifestyle.

Select & carry out tasks in promising areas that can lead the increase of RFID demands among tasks verified with business potential through pilot projects and ISP. In 2007, expected to support 7 tasks including national commodities, medicine, and air baggage mgmt. etc.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

And Select & carry out promising field with big impact on the public that can create USN market for promoting the commercialization of USN service. In 2007, expected to support 7 tasks including underground water, marine safety, highway facility mgmt. etc.

3 UBIQUITOUS HEALTHCARE PILOT PROJETS

Everything dubbed "ubiquitous" - ranging from u-health, u-defense, u-city to u-payment - means an environment where people can enjoy access to high-speed networks and advanced communication services anywhere and anytime through a ubiquitous computing network.

The quality of Koreans' life stood at the lowest level among OECD member countries and the country is now facing new problems such as aging society. And risk of chronical disease, such as glycosuria, hypertension or high cholesterol, is increasing according to the change of living style accompanied with the growth of living level.

Accordingly, Korea government plans to solve such problems through the u-health strategy. U-health service will be provided in which wearable computers will be used to monitor the health of elderly people living alone and patients with chronic diseases. The devices, in the form of shirts, alert medical staff in real time when an emergency occurs. Daegu City will distribute the bio-shirts to some 100 elderly people and patients with chronic diseases later this year. The shirts have embedded sensors that register vital signs and send the information to medical centers through the network. They also permit self-diagnosis, distance monitoring, emergency care and medical consultation for users in an environment of ubiquitous connectivity. u-Health pilot projects enable the public to enjoy benefits of u-IT which improves quality of life.

4 CONCLUSION

In order to actively respond to the fast-changing IT trends, and to successfully enter the society of knowledge-information and the new economy, Korea has continuously promoted informatization with strong determination. With the constant efforts by the government and the private sector for informatization during the last two decades, Korea has built the world's best quality infrastructure and is becoming a benchmark target for the world.

In recent years, as the 'ubiquitous society' is being anticipated, which is differentiated from the existing Internet-based knowledge information society, the rapid transference of different sectors in such a society has already begun. In order to maintain and enhance Korea's status as the world's strong IT leader, the IT industry is being fostered as the key engine for economic growth.

Now, Korea is confronted with the opportunity and challenge of future information society. The successful implementation of ubiquitous IT strategy and pilot projects will be the touchstone to benchmark for future information society.

REFERENCES

IMD(2005), '2005 World competitiveness yearbook'.

Ministry of Information and Communication(MIC, 2006), 'u-KOREA Master Plan-To achieve the world's first ubiquitous society', Republic of Korea.

National Information Society Agency(NIA, 2006), 'u-Health pilot project result paper', Republic of Korea.

National Information Society Agency(NIA, 2007), 'Informatization White Paper', Republic of Korea.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

i·Society 2007

Security decadence in electronic voting

Cyril E. Azenabor, Charles A. Shoniregun

School of Computing & Technology
University of East London
Docklands Campus, 4-6 University Way, London E16 2RD,United Kingdom
(cyrillicehi@yahoo.com and c.shoniregun@uel.ac.uk)

Abstract The Electronic voting (e-voting) encompassing several types of voting, embracing both electronic means of casting a vote and electronic means of counting votes. The e-voting technology is what governments are increasingly adopting and is gradually gaining grand in most of the e-Government practising countries. The e-voting includes electronic counting schemes combined with traditional paper ballots, touch-screen voting kiosks, Internet voting, interactive voice response (IVR), landline telephone voting, SMS text message voting, digital television voting, e-register enabled polling station and postal ballots. The e-voting has its peculiar problem of creating fraud and the most complex security issues in electronic government (e-Government) practise. The e-voting machine potentially makes electoral fraud unprecedentedly simply and the proliferation of similar programmed e-voting systems invites opportunities for large-scale manipulation of elections and the electronic machines were supposed to solve the problem of election malpractices and to aid easy casting of votes, the outcome is opposite. Though, the current development of software and hardware cannot fully provide adequate and acceptable level of security for this kind of application due to the high level of its vulnerabilities. In this research, we carefully proposed a model of Cyril Azenabor and Shoniregun (CAS) e-voting Security Model to solve the problems associated with e-voting.

Keywords Vulnerabilities, e-voting, Fraud, e-Government, e-Election, Ballot paper, Ballot-stuffing, Biometrics National Identity Card (BNIDC), Cyril Azenabor and Shoniregun (CAS) e-voting Security Model

1 INTRODUCTION

Electronic voting (e-voting) has its peculiar problem of creating fraud and the most complex security issues in electronic government (e-Government) practise. In some countries where it has been tested like in the U.S. and the U.K. and other parts of the world, e-voting has serious security breach hence, e-voting has a delicate set of security requirements. Not withstanding either e-voting or manual voting that a country decides to practise, perfect voting system does not exist. The case with e-voting is that it is more vulnerable than the manual voting systems. Online voting was first tested in March (2000) when Arizona Democratic Party in the U.S. allowed for the first time remote voting in its presidential preference primary after which political leaders and policymakers worldwide have since been investigating the viability of using the Internet for public elections in their own countries (Mohen and Glidden, 2001). The current development of software and hardware cannot fully provide adequate and acceptable level of security for this type of application. The e-voting encompassing several types of voting, embracing both electronic means of casting a vote and electronic means of counting votes. The e-voting technology is what governments are increasingly adopting and is gradually gaining grand in most of the e-Government practising countries. Despite all the technological advances everywhere it is being used, there has not been a complete secure e-voting solution. The e-voting includes electronic counting schemes combined with traditional paper ballots, touch-screen voting kiosks, Internet voting, interactive voice response (IVR), landline telephone voting, SMS text message voting, digital television voting, e-register enabled polling station and postal ballots. To large extents, the different technologies were imperfect in their ability to count votes. Although e-Government is associated with making services and information more accessible; there are limited risks with this approach, and failure in service may create an inconvenience for the individual citizen, but it does not pose fundamental risks for the government. However, the failure of e-voting technology has profound consequences for the reliability of and public confidence in our electoral system. The consequences of a failed election are much greater, and the adoption of e-voting has increased the risk that such failure will occur and there are fear in the mind of voting experts and many computer scientists that the elections were at risk due to the use of the e-voting machines and has failed because it does not serve its purpose (Moynihan, 2004).

2 ELECTRONIC VOTING MACHINE

The e-voting machine potentially make electoral fraud unprecedentedly simply and the proliferation of similar programmed e-voting systems invites opportunities for large-scale manipulation of elections, while the use of direct-recording electronic voting machines (DREs), or more gen-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007

erally, any electronic means of vote tabulation and reporting, raises the concerned that a single, simple, subtle fraudulent change to the system software can effect everywhere these machines are deployed (Di Franco et al., 2004). Electronic machines were supposed to solve the problem of election malpractices and to aid easy casting of votes, the outcome is opposite. It has been proved by researchers that using electronic machines was a calculated risk and the election system, in its entirety, exhibits shortcomings with extremely serious consequences, especially in the event of a close election. The e-voting machines could be rig or manipulated, the people who designed the system are in position to do so in such a way so that they would be able to control the outcome if wanted to. Therefore, with the help of malicious code, votes could be stolen from a machine undetectable; modifying all records, logs and counters to be consistent with the fraudulent vote count it creates and sometimes machines may not record the votes (Thompson, 2006). Other risks pertain to the fact that these are computer-based systems, and may have been poorly designed and inadequately tested, and may have security holes which can be exploited (Pitt et al., 2006). The replacement of outdated voting machines to recent and newly designed machines could not help in the security lapses of e-voting systems. The failure of the e-voting machines are shown in the lapses it has created every where the systems have been adopted, especially in the U.S. where it has been use in several elections and it has drawn criticism from academics, election officials and concerned citizens.

The e-voting machines are vulnerable to manipulation and fraud and cannot be relied on entirely. One of the flaws of electronic voting machines was during an election in Boone county 2003, Indiana, electronic machines initially registered 144,000 votes in the county with about 19,000 registered voters, and of those, only 5,532 actually voted and the 2000 presidential election indicated that electronic machines failed to record nearly 700 votes in new Mexico, a state Al gore won by only 366 votes and with electronic machines, however, a rogue programmer who slipped an unnoticed trapdoor into the software or exploited a flaw in the code for the operating system could potentially change the outcomes on many machines at once (Seife 2004, Simons 2004). Concerning design, researchers have shown, and experience has confirmed, that e-voting machines do not meet reasonable expectations for correctness, availability, accessibility, and security (Barr et al., 2007). In some developed countries where large percentage of their votes are done electronically, there are fear in the minds of voting experts and many computer scientists that the elections were at risk due to the use of those machines. Paperless voting machines threaten the integrity of democratic process by what they don't do, and a computer can easily display one set of votes on the screen for confirmation by the voter while recording entirely different votes in electronic memory. Therefore, there is no way to check whether the votes were accurately recorded once the voter leaves the booth; consequently, the integrity of elections rests on blind faith with the vendors, their employees, inspection laboratories, and people who may have access – legitimate or illegitimate – to the machine software (Dill et al., 2003). The Independent National Electoral Commission (INEC), the body controlling elections in Nigeria decides to adopt electronic registration in 2006 towards the 2007 general elections. At the end of the programme, the electronic registration embarks on by INEC failed because little did they know about the problems associated with electronic machines. The electronic machines packed up for lack of power a problem associated with third world countries and inadequacy of the data capture machines and lack of training on the parts of staff deployed to use the machines, and it resulted in failure to register half of the eligible voters at the stipulated time. The electronic registration was extended for weeks and when eligible voters were satisfactorily registered by INEC and the voters' lists were displayed across the country, more than two third of people who registered could not find their names on the lists, this is because of the short-coming of e-voting machines(Daily Independent Nigeria, January 26 2007).The failure of the e-voting machines gives room for the security issues and is always regarded as crucial factors and is causing great concern to the global democracy.

3 SECURITY ISSUES IN E-VOTING

The human interaction in electronic system makes it vulnerable to so many attacks. Record shows that electronic democracy (e-Democracy) applications which has to do with e-Government are highly at security risks and most vulnerable. The e-voting pilot carried out in the U.K. in (2003) elections shows concern mainly on security lapses which is not different from the same security threat facing e-voting system anywhere in the world where it has been practised (Xenakis and Macintosh, 2004). The e-voting was the cause of the problem in Florida in the U.S. 2000 presidential election which almost marred the race between Bush and Al Gore because punch-card voting machines were used, because the machines were error-prone. Hence federal election initiatives, such as Help America Vote Act (HAVA) was written into law as direct result of the Florida 2000 presidential election controversy and the subsequent malfunction of new election equipment in that state during 2002 (Mercuri and Camp, 2004). The cost of protection in electronic voting is very high such as the use of biometrics-based voter registration which is use to prevents voter fraud, and still it is not 100% secured. Though, the current development of software and hardware cannot fully provide adequate and acceptable level of security for this kind of application. In the past there has been multiple-channel voting and there are still occurrences of these across the globe where Internet or electronic voting systems are being practised, and technological advancement have not been able to provide a completely secure e-voting solution. The insecure e-voting systems could undermine democracy anywhere in the world it is being practise and lack of confidence in the electronic systems due to security issues will not allow it to work. Grove 2004, in ACM statement on e-voting systems declared that virtually all voting systems in use today (punch-cards, lever machines, hand-counted paper ballots, among others) are subject to fraud and error, including electronic voting systems, which are not without their own risks and vulnerabilities. In particular, many electronic voting systems have been evaluated by independent, generally recognised experts and have been found to be poorly designed; developed using inferior software en-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

gineering processes; designed (or with very limited) external audit capabilities; intended for operation without obvious protective measures; and deployed without rigorous, scientifically designed testing. The e-voting system creates doubts in the minds of voters and candidates.

Xenakis and Macintosh (2004) outline some cases of procedural security lapses in voting which have been documented and grouped into the following generic area:

- The lack of procedures to control the activities of commercial vendors and government officials before and during the election, providing an audit trial of their actions.
- Existing measures of procedural security, which are inadequate to cover all aspects of the electoral process such as the verification of voter providing data, the secure dissemination of voter credentials and the prevention of double voting through multiple channels.
- The lack of agent compliance existing measures of procedural security.

Mote (2001) report of the National Workshop on Internet discussed some security issues adopting e-voting which says that remote Internet voting systems pose significant risk to the integrity of the process, and should not be fielded for use in public elections until substantial technical and social science issues are address. The security risks associated with these systems are both numerous and pervasive, and in many cases, cannot be resolved using today's most sophisticated technology. Internet-based voter registration poses significant risk to the integrity of the voting process, and should not be implemented until adequate authentication infrastructure is available and adopted. On-line registration without the appropriate security infrastructure would be at high risk for automated fraud, that is, the potential undetected registration of large numbers of fraudulent voters. This is because computer-based voting systems as well as other distributed computing systems are vulnerable to attack at three main points, the server, the client and the communication path. The current hardware and software architectures, a malicious payload on a voting host can actually change a voter's vote without the voter or anyone else noticing, regardless of the encryption or voter authentication in place, because the malicious code can do its damage before the encryption and authentication is applied to the data, and the malicious module can then erase itself, so no evidence of fraud is left to correct or even detect (Rubin, 2002). We cannot reliable inspect a programme to determine that it does contain hidden functionality, and it is impossible to guaranty detection of all hidden functionality by black-box testing (Jones and Neumann, 2006). With the security issues in e-voting, the outcome of such result cannot reflect voters' intentions and ensure public confidence. Therefore it declines the sanctity of such elections and the legitimacy of the governing regime. Despite the risks involve in e-voting, governments are beginning to introduce electronic elections (e-Elections) to the public.

4 ELECTRONIC ELECTIONS

The web will unavoidably become the infrastructure of our democratic processes, as we move to e-Elections and similar technological-based means of eliciting democratic representation (Zwass, 2006).With the adoption of e-Elections, the electorate are able to cast their ballots from the location of their choice, whether home, work, a public library or a traditional polling place electronically via a web browser (Mohen and Glidden 2001, Juels et al., 2005). The adoption of Internet voting has created apathy by election officials and voting-rights advocates due to risk but it has the potential to increase voter participation and access for all communities. In this process, voting technology can be diffused in a democracy whereby government parties may have strong and vested interests in the reliability of the technology, or alternatively may which to ensure that the technology can be allowed to be subverted to favour certain groups or individuals (Mercuri and Camp, 2004). It would not be legally, practically, or fiscally feasible to develop a comprehensive remote Internet voting system, because of it's fraught with risks to the integrity and security of elections (Phillips and Von Spakovsky, 2001). Electronic interaction use in the process of election has serious consequences on the validity of the outcome, in that it is subject to manipulation more than the traditional voting method. The operating system and software also cause severe damage to the e-voting application wherever it is being practised. The hope and believed that e-voting would stop the problems encounter by traditional voting system has failed and the technology has not been able to meet the impossible promises. Unlike the traditional voting system, if there is dispute the individual ballot papers can be recounted. In the case of e-voting system the votes cast in each station cannot be inspected – the system simple records the running total for each candidate. For this reason, the task concluded that in the foreseeable future, Internet voting should be seen as a supplement for traditional paper-based voting (Financial Times London June 20, 2001).

Studies have revealed that there are problems in developing e-voting infrastructure for elections. In some cases where there is a software bug – or malicious hacker has the ability to effect changes to the code, so that such votes are counted incorrectly and being influenced by transferring to a particular candidate. The Internet is known for its security lapses in that Internet voting systems would make unchangeable target for hackers or darker forces seeking to influence the democratic process, therefore some class of thought believed that Internet voting should not be encourage due to privacy and security issues. The advent of computerised election administration systems incorporating e-voting machines highlights concerns about security, verification and general public response and the implementation of computerised systems as centralised processing systems subverts this build-in control feature and leaves the system open to problems of fraud, corruption and catastrophe (Kassicieh et al., 1988). Therefore, the adoption of e-voting is highly vulnerable to attacks and fraud and cannot be relied on absolutely. Despite security issues and vulnerabilities that characterised e-voting systems, some countries who have adopted full e-voting method is

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

bound to have hindrances in acceptance by the citizenry because most households in an e-Government nations cannot boast of half of its citizens having access to the Internet.

5 THE E-VOTING VULNERABILITIES

The counting of e-voting is subject to fraud due to its high vulnerabilities and it receives many attacks from invaders. There are tendencies of multiple voting in the systems without being detected because the e-voting process is open to such short-comings. The Internet is independent of national boundaries, an election held over the Internet is vulnerable to attacks from anywhere in the world, and Direct-recording elections (DRE) voting systems have been widely criticised for various deficiencies and security vulnerabilities: that the software undergoes insufficient scrutiny during qualification and certification; that the DREs are especially vulnerable to various forms of insider (programmer) attacks (Jefferson et al., 2004). The e-voting is also vulnerable in other ways like in the case of cyber-attacks; such as denial-of-service attacks, spoofing, viral attacks on voter PCs, any of these could be catastrophic to genuine electoral process. It is difficult to detect and neutralise these type of attack, and could have a devastating effect on the public confidence in elections. Large scale vote buying and selling can be automated via the Internet and also in Internet voting, electoral authorities do not have control over the use of equipments by voters. Though it is generally not feasible to remove fraudulent ballots from an election, Internet voting servers may be subject-of-service attacks and other security threats (Hoffman and Cranor, 2001). Armen and Morelli 2005 outlined some of the vulnerabilities associated with electronic machines. With the punch card systems, incompletely punched holes in the form of dimples or hanging chads make the card unreadable, this is known as undervote. If the voter inadvertently punches too many holes for a given office, the overvote will also make the card unreadable. For optical scan systems, an undervote may be caused when the voter's marks are illegible and an overvote may be caused when the voter makes too many marks or if the paper gets smudged in the wrong place. Diebold machines are so vulnerable to hacking that someone could easily wipe out all the data at all the machines in one precinct (Sauer, 2006). Research shows that all voting technologies such as punch cards, DREs and others are susceptible to fraud and a threat posed by DREs is undetectable to fraud and are fundamentally flawed in design, as well as being poorly implemented in many cases hence their vulnerabilities.

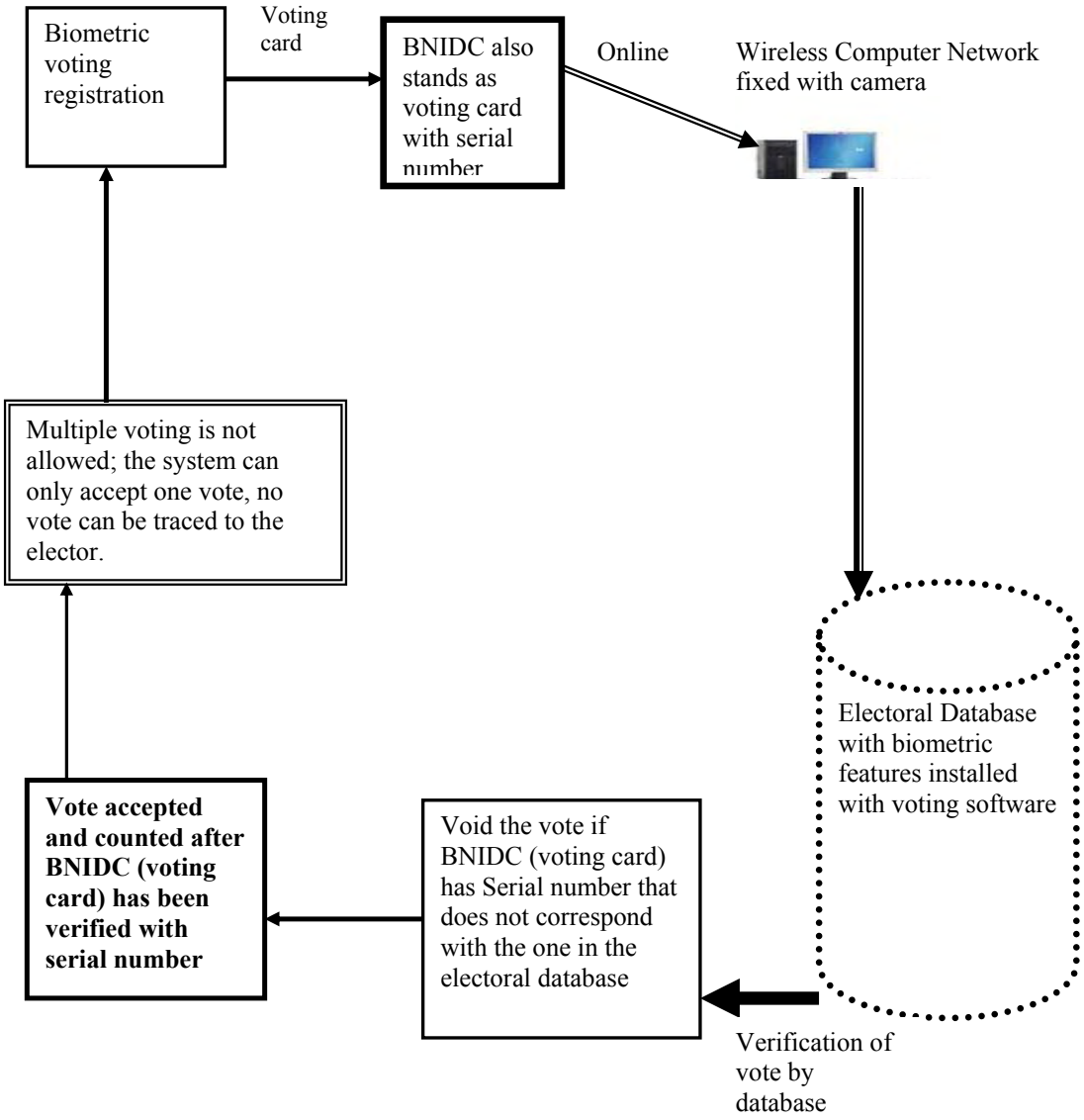
6 DISCUSSION

If electoral officers properly manage the e-voting system and there are genuine scientific method adopted, online voting will be less vulnerable and the outcome of the election results will be close to perfection. In e-voting the election becomes inaccessible to the voter who can no longer scrutinize it and transparency as a basic of principle of democracy is eliminated and it is argued that whoever controls the voting machines can control who wins the votes, with this fraud can hardly be eliminated in this type of voting systems. The first country to pioneer e-voting in all arms of the government is Brazil in 2002 general elections without physical ballot and was not error free, it was usual of all the problems associated with electronic voting systems and electronic machines were used. Less than 10 percent of the population in Brazil has access to the Internet, the adoption of e-voting technology is uniquely pleasing the interests of corporate actors; this is exacerbating digital divide and other social divisions (Filho et al., 2006). One would wonder why a developing country like Brazil decides to adopt e-voting in all arms of the government in the general elections in 2002 knowing the risks involved while more advance countries like the U.S. and the U.K. with strong democratic tradition are not yet using e-voting systems intensively, due to the concern for and emphasis on security. The e-voting can alter democratic institutions when it is entirely depended upon by any country because it is one of the most controversial aspects of election process. The Internet is an open system hence its vulnerability to attacks from anywhere around the world because it is independent of national boundaries. Therefore, there is no free error e-voting system anywhere it has been practised because the existing technology does not provide a completely e-transaction environment.

Figure 1. Biometrics National Identity Card (BNIDC) by Azenabor and Shoniregun (2007)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Cyril Azenabor and Shoniregun (CAS) e-voting Security Model



We believed that when appropriate scientific steps are taken, a less vulnerable and generally accepted e-voting systems could be achieved anywhere in the world. Azenabor and Shoniregun (2007), proposed a model of Biometrics National Identity Card (BNIDC), to solve many security problems in a country and this model is also to serve as a voting card. See Figure 1 below:

In this process the BNIDC could be use as a voting card in e-voting election because of its biometric features, which cannot be replicated. Each polling booth is to be connected with a wireless computer network fixed with camera and is regulated by electoral database with biometric features installed with voting software. After the voting card is swap through a mechanism, the camera is to focus on the facial thermogram and picture the biometric features such as the iris, retina and the thumbprint, to ascertain if they correspond with the details in the regulated electoral database system. If they are differ from the electoral database you cannot vote, but when the biometric features are the same the software will pop-up the soft ballot paper and you can vote by thumb-printing on the candidate of your choice, it will be a paperless ballot and this will help in preventing the printing of counterfeit paper ballot and ballot-stuffing. There will be constraint so that there will be no room for multiple voting because the voting card (BNIDC) will be embedded with a serial number for every eligible voter which is due for verification by the electoral database. Once the vote has been verified and accepted it will be counted, a voter who voted for a particular candidate will not be know, that is no vote can be traced to the elector. Figure 2: below shows Cyril Azenabor and Shoniregun (CAS) e-voting Security Model

7 CONCLUSION

The vulnerabilities of e-voting system cannot be underestimated due to several reports and evidence of its lapses everywhere it has been used. The e-voting machines and all the technologies employed to improve the voting systems and count the votes are imperfect in their inabilities to solve problems without leaving behind huge errors. We understand that there are no free and fair elections without its short-comings anywhere in the world whether e-voting or manual voting is adopted, but the security issues created by e-voting is alarming and if proper precaution is not taken, it will enhance the enthronement of un-popular government. Apart from security issues facing the e-voting systems,

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

fraud could cause another setback, which would prevent the system from working perfectly due to attitudes of desperate politicians who would do anything to win an election. If Cyril Azenabor and Shoniregun (CAS) e-voting Security Model in Figure 2: above is properly implemented, it will help to rectify so many lapses and security issues threatening the world's democracy. It will also help to solve the problems associated with e-voting and ease the issue of paper elections and therefore bring sanctity into the electoral systems.

REFERENCES

Armen, C., and Morelli, R., 2005, 'E-voting and Computer Science' http://delivery.acm.org/10.1145/1070000/1067508/p227-armen.pdf?key1=1067508&key2=3742025811&coll=GUIDE&dl=GUIDE&CFID=24665994&CFTOKEN=88679402 (Accessed date: 15/05/2007)

Azenabor, C.E., and Shoniregun, C.A., 2007, 'Electronic Government Security Measures', Accepted paper presented at Hawaii International Conference on Business, May 24-27, 2007, Waikiki Beach Marriott Honolulu, Hawaii, USA

Barr, E., Bishop, M., and Mark, G., 2007, 'Viewpoint: Fixing federal e-voting standards' Communications of the ACM: Vol. 50, No. 3, Pgs. 19-24

Daily Independent Nigeria January 26, 2007, http://www.independentngonline.com/?c=75&a=19786 (Accessed date: 07/06/2007)

Di Franco, A., Petro, A., Shear, E., and Vladimirov, V., 2004, Small Vote Manipulations: Can Swing Elections' Communications of the ACM: Vol.47, No. 10, Pgs. 43-45

Dill, D.L., Schneier, B., and Simons, B., 2003, 'Voting and Technology: Who Gets to Count Your Vote' Communications of the ACM: Vol. 46, No. 8, Pgs. 29-31

Filho, R.J., Alexander, J.C., and Batista, C.L., 2006, 'E-voting in Brazil – The Risks to Democracy'

Financial Times London June 20, 2001, http://proquest.umi.com/pqdweb?did=74235582&sid=3&Fmt=3&clientId=13314&RQT=309&VName=PQD (Accessed date: 07/06/2007)

Grove, J., 2004, 'ACM STATEMENT ON VOTING SYSTEMS' Communications of the ACM: Vol. 47, No. 10, Pgs. 69-70

Hoffman, J.L, and Cranor, L., 2001, 'Internet voting for public officials' Communications of the ACM: Vol. 44, No.1, Pgs 69-71

http://fl1.findlaw.com/news.findlaw.com/hdocs/docs/election2000/nsfevoterprt.pdf (Accessed date: 18/03/2006)

http://www.monitortv.at/monitortv/website/evotingconference06/pdfs/E-Voting_in_Brazil.pdf (Accessed date: 07/05/2007)

Jefferson, D., Rubin, D.A., Simon, B., and Wagner, D., 2004, 'ANALYSING INTERNET VOTING SECURITY' Communications of the ACM: Vol.47, No.10, Pgs 59-64

Jones, W.D. and Neumann, G.P., 2006, 'Does technology help or hinder election integrity?' http://delivery.acm.org/10.1145/1190000/1180188/p16-o_hanlon.pdf?key1=1180188&key2=8183025811&coll=GUIDE&dl=GUIDE&CFID=24668068&CFTOKEN=77807673 (Accessed date: 03/06/2007)

Juels, A., Catalano, D., and Jakobsson, M., 2005, 'Coercion-Resistant Electronic Elections' http://www.informatics.indiana.edu/markus/papers/WPES10-juels.pdf (Accessed date: 18/04/2007)

Kassicieh, S.K, Kawaguchi, G.H., and Malczynski, L., 1988, 'Security, Integrity and Public Acceptance of Electronic Voting' Journal of Systems Management: Vol. 39. No. 12, Pgs 5-6

Mercuri, R.T., and Camp, J.L., 2004, 'The code of elections' Communications of the ACM: Vol. 47, No. 10, Pgs. 52-57

Mohen, J., and Glidden, J., 2001, 'The case for Internet voting', Communications of the ACM: Vol. 44, No. 1, Pgs. 72-85

Mote, C.D. Jr., 2001, 'Report of the National Workshop on Internet Voting: Issues and Research Agenda'

Moynihan, D.P., 2004, 'Building Secure Elections: E-Voting, Security, and Systems Theory' ABI/INFORM Global: Vol. 64, No. 5, Pg. 515

Phillips, M. D., and Von Spakovsky, A.H., 2001, 'GAUGING THE RISKS OF INTERNET ELECTIONS' Communications of the ACM: Vol. 44, No.1

Pitt, J., Kamara, L., Sergot, M., and Artikis, A., 2006, 'Voting in Multi-Agent Systems', THE COMPUTER JOURNAL: Vol. 49, No. 2

Rubin .D. A., 2002, 'Security Considerations for Remote Electronic Voting, Communications of the ACM, Vol. 45, No. 12

Sauer, M., 2006, 'Electronic Voting Machines Still Very Vulnerable to Hack' http://blogs.abcnews.com/theblotter/2006/05/report_electron.html (Accessed date: 27/05/2007)

Seife, C., 2004, 'A vote of very little confidence: ELECTRONIC BALLOTS' http://proquest.umi/pqdweb?did=725727941&sid=3&Fmt=3&clientId=13314&RQT=309&VName=PQD (Accessed date: 15/05/2007)

Simons, B., 2004, 'Electronic Voting Systems: the Good, the Bad, and the Stupid' http://delivery.acm.org/10.1145/1040000/1035606/p20-simons.pdf?key1=1035606&key2=2454025811&coll=GUIDE&dl=GUIDE&CFID=29370819&CFTOKEN=11738353 (Accessed date: 15/05/2007)

Thompson, E., 2006, 'Election 2006: Penetrating The Voting Vortex' http://www.indypendent.org/?p=598 (Accessed date: 01/06/2007)

Xenakis, A., and Macintosh, A., 2004, 'Procedural Security Analysis of Electronic Voting' http://csdl2.computer.org/comp/proceedings/hicss/2004/2056/05/205650116c.pdf (Accessed date: 08/11/2007)

Zwass, V., 2006, 'The web-internet compound as the infrastructure of digital government' Business Process Management Journal: Vol. 12, No.1, Pgs. 6-7

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

i·Society 2007

Democracy development trends as a framework for edemocracy

Rui Pedro Lourenço
João Paulo Costa

Faculdade de Economia da Universidade de Coimbra INESC - Coimbra
Instituto de Engenharia de Sistemas e Computadores , Portugal
ruiloure@fe.uc.pt, jpaulo@fe.uc.pt

Abstract Citizens living in contemporary democratic societies show an ever growing apathy towards the political system, which is reflected on lower turnouts in voting occasions. Information and Communication Technologies have the potential to help reduce this disengagement from public life. However, technology driven solutions might not be the best approach. The purpose of this paper is to analyze trends in political theory that might be considered as a framework to develop technological solutions to support democratic transformation and reduce the gap to the democratic ideal. More direct citizen participation is now demanded by both political scientists and citizens who should not confine their political role to periodically casting a vote. This participation should not be restricted to the traditional political sphere; on the contrary, democracy should be a daily practice in the whole sphere of society. Finally, deliberation is now gaining emphasis with respect to preference aggregation. The relationship between citizens and other *players* of the political system (political parties, representatives and public administration) is considered, as well as the possibility that citizens might influence policy making without intermediaries. From this analysis it is possible to extract a set of requirements to be considered when designing technology support for democracy transformation

Keywords Participatory democracy; Deliberative democracy; e-Democracy

1 INTRODUCTION

The concepts and practice of democracy have evolved continuously since ancient Greece to the present day. When it emerged as a political system, around the fifth century B.C., democracy was based on the idea that each citizen was morally obliged to personally participate on the government of society, and therefore representation was not even considered. Citizens were expected to meet regularly to deliberate about all common affairs with equal voting power and full liberty of expression. Despite the fact that not every member of society had political rights, this ideal ("rule by the people") was powerful enough to survive until the present day. This plebiscitary system, based on direct participation, is also referred to as *assembly politics*. Almost two thousand years after, these central ideas were still advocated by Jean-Jacques Rousseau in his influential *Du Contrat Social* (1762). However, even Rousseau admitted that this kind of democratic system was only applicable to small city-states. From then on, two major transformations occurred that would change profoundly the democratic system. Democracy would have to cope with national states with much wider territory than that of city-states and consequently with an increasing number of citizens. At the same time, more and more members of society were starting to have citizenship rights and today universal suffrage means that almost every adult is entitled to participate in the democratic process. These changes in scale, together with an increased complexity in the governance of society, almost put aside the idea of *assembly politics* and *direct participation* in contemporary democratic societies. Representative institutions and political parties now play an important role in the democratic political system. Periodically voting to choose between competing candidates or political parties became almost the only citizen participation required by the political system. Control over representatives and their actions is becoming elusive and citizens no longer feel that they have any real influence on the governance of society in which they live in. This apathy towards the political system can be observed even in the ever lowering turnout in general elections. The adoption of Information and Communication Technologies (ICTs) promised to reduce this apparent disengagement from public life and gave origin to a new concept, *e-democracy*, that may be understood as "a collection of attempts to practice democracy without the limits of time, space and other physical conditions, using ICT or CMC [Computer-Mediated Communications] instead, as an addition, not a replacement for traditional 'analog' political practices." [1] If we analyze this definition it is possible to recognize a direct reference to the usage of ICTs to overcome some of the constraints that prompted the abandon of the Greek ideal of assembly politics and direct participation: "the limits of time, space and other physical conditions". Another important aspect to consider is that technological

supported initiatives are not meant to replace existing *analog* political practices, and therefore technology should not become a source of further political inequality (see discussion about the *digital divide* below). What remains unspecified is the nature and degree of political transformation that the adoption of technology may induce. The introduction of electronic voting, for instance, is regarded by many as a way to increase voting turnout and therefore reduce the general apathy towards the political system. This may attract, in the short term, some voters due to the novelty of the technical apparatus and allow for more efficiency in the electoral process. But does it change in a substantial way the nature of the democratic system? In order to promote transformations on substantial (rather than procedural) issues of the democratic political system, technology adoption must be considered in the framework of a serious reflection on political (democratic) theory: as Robert Dahl stated, "among the complex historical factors that contribute to democratic stability, breakdown, and transformation, a body of democratic theory reposing on reasonable assumptions is by no means of trivial importance." [2] The democratic ideals from ancient Greece and Rousseau, which inspired participatory political theorists such as Carole Pateman [3] and Benjamin Barber [4], may now be reconsidered: "… the simple juxtaposition of participatory with liberal representative democracy is now in flux given developments in information technology …. The merits of participatory democracy have to be re-examined now its technical feasibility is closer at hand." [5] Also, the absolute centrality of voting in the democratic processes and its exclusiveness as participatory mechanism is being challenged by deliberative democrats inspired by political philosophers such as John Rawls [6] and Jürgen Habermas [7]. These and other developments in political theory constitute therefore a major source of inspiration on the use of ICTs to transform democracy, not just in making its formal procedures more efficient, but to make it substantially more democratic. Nevertheless, we must keep in mind that "finding ways to reincorporate technology into a strong democratic strategy will depend not on the technologies themselves, which remain protodemocratic in many of their aspects, but on political will." [4].

The remaining of this paper is organized as follows. Section 2 presents a brief discussion about the contemporary democratic practices on western countries to highlight their most prominent characteristics. Building on the limitations of these practices, section 3 presents some political theory trends which can be considered when designing technology support for democracy transformation. These transformation efforts may be applied to the relations between citizens and political intermediaries identified on section 4. Section 5 analyses some of the requirements these transformations pose on technology and some of the research efforts being made in this area. A particular condition for e-democracy success, the reduction of the digital divide, is analysed in section 6 and some final remarks are presented on section 7.

2 LIBERAL REPRESENTATIVE DEMOCRACIES

How can we describe contemporary democracies? According to Robert Dahl, "there is no democratic theory – there are only democratic theories" [8], and therefore our approach will be to "consider as a single class of phenomena all those nation states and social organizations that are commonly called democratic by political scientists, and … discover, …, the distinguishing characteristics they have in common …" [8]. These nation states are usually called *western democracies* and their distinguished characteristics of governance constitute the *liberal representative democracy model*, which is taken to be the dominant contemporary form of democracy [4].

Modern liberal thought, which is at the core of the contemporary model of democracy, was largely influenced by the work of John Stuart Mill [5]. At its core is the notion of individual liberty, which "is that of pursuing our own good in our own way, so long as we do not attempt to deprive others of theirs, or impede their efforts to obtain it." [9] Any interference with an individual's liberty of action, either by another individual or by the collectivity, can only be justifiable as self-protection or "to prevent harm to others." [9] Since individuals are mostly motivated by self-interest, the role of liberal politics is then to reconcile and aggregate those predetermined interests when market economy, in the civil society sphere, fails to do so. The main concern of Mill was that, in an attempt to deal with complex problems, the state would develop so rapidly and extensively that would infringe on the liberty of citizens [5]. Despite his preoccupation with the non-interference of the state in the civil sphere, Mill considered that "nothing less can be ultimately desirable, than the admission of all to a share in the sovereign power of the state." However, due to the impossibility of every citizen to participate personally in the management of public affairs, "the ideal type of a perfect government must be representative." [10] Mill also acknowledged a "radical distinction between controlling the business of government, and actually doing it" [10], the former being the responsibility of "the representatives of the Many" and the latter reserved to the "specially trained and experienced Few" [10]. The meaning of *representative government* lies in the possibility that the entire body of citizens have to periodically elect deputies to form a *representative assembly* and, by that way, possessing the ultimate controlling power over the affairs of state [10].

More recently, Joseph Schumpeter developed a theory of *democracy* that highlights many recognizable features of contemporary democracies: the competitive struggle between parties for political power, the important role of public bureaucracies and the significance of political leadership. According to Schumpeter's vision, democracy is a political method, designed to arrive at political – legislative and administrative – decisions [11], whereby certain individuals are granted the power to decide by competing for the people's vote [11]. This vision of democracy, which is fairly common in contemporary societies, reduces the citizen's role in democracy to periodically casting their vote to choose their representatives considering the proposals put forward by political parties. This procedural account for democracy is bet-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

ter described as *thin democracy* "…which reduces decision-making to voting for elected representatives and relies on the institutions of majoritarianism and adversary politics." [4] Under this model citizenship becomes primarily a legal concept, defined in constitutions and legal systems regarding the citizens' rights and obligations [12]. Citizen's individual preferences are considered pre-determined and therefore there is no purpose on trying to transform them through reasoned discussion. They simply need to be aggregated and once again elections provide the main transmission mechanism from public opinion to governmental action [13].

Despite the fact that this vision of democracy is now sometimes confounded with the concept of democracy itself, many political theorists assume that the liberal representative model, and the way it is implemented on western societies, still falls short of the democratic ideal. Moreover it fails to fully respond to many of today's societal complex problems and leaves many room for change. Apart from some radical voices, political theorists proposals do not aim at replacing representative institutions (namely by direct participation ones). Instead, their proposals must be regarded as a roadmap to improvement, a process in which technology may play an important role.

3 DEMOCRACY DEVELOPMENT: SOME TRENDS IN POLITICAL SCIENCE

In this section we will present current democracy development trends: participatory democracy and its extent beyond the political sphere and the increased emphasis on deliberation.

3.1 Democracy beyond the political sphere

Within contemporary democratic countries there is a sense that democracy has attained some degree of development and stability: political democratic institutions are now in place, civil liberties (such as the freedom of speech and assembly) are recognized and political rights were extended up to universal suffrage. No longer is the number of people with the right to vote a unique indication of a country's democracy development stage [14]. Democratic development is now related with the 'quality' of effective participation, and participatory theories of democracy emphasise the educative effect of participation: it diminishes tendencies toward non-democratic attitudes in the individual; increases the individual's sense of and actual freedom; fosters human development; enhances a sense of political efficacy; reduces the sense of distance from the power centres; increases the feeling among individual citizens that they belong in their community (integrative function); enables collective decisions to be more easily accepted by the individual; and contributes to the formation of an active and informed citizenry with a renewed interest in government affairs [3]. Since any ordinary citizen would always be more interested in things related to his/her daily life, there is the need to extend the sphere of democratic participation to other domains beside national politics [3; 14]. One important democratic trend is therefore the transfer of democracy from the political sphere to the social sphere [14]. Citizens possess many facets which force them to interact with several institutions in the civil society, considered here as the "areas of social life which are organized by private or voluntary arrangements between individuals and groups outside the direct control of the state." [5] A single individual may be considered simultaneously as a patient, a student, a neighbour and a professional and, on those capacities, may have to deal with health and education institutions, neighbour associations and professional guilds. All these interactions should give him/her the opportunity to democratically participate in the management of these institutions that so closely affect his/her well-being.

3.2 The emphasis on deliberation/ assembly democracy

One major criticism made to the liberal democracy model is directed to the emphasis on preference aggregation through voting mechanisms. It is raised by Social Choice Theorists who claim that all aggregation mechanisms are vulnerable to strategic manipulation. This claim is substantiated by the work of Kenneth Arrow who proved that it is impossible for any mechanism for the aggregation of individual preferences into collective choices to simultaneous satisfy five desirable criteria: unanimity, non-dictatorship, transitivity, unrestricted domain, and independence of irrelevant alternatives [15]. Another important result for aggregative democratic theory comes from the Gibbard-Satterthwaite Theorem [16; 17] that demonstrates that under any voting scheme there exists the possibility for a single individual to manipulate the process. These results seem to be at odds with the popular idea that voting is at the core of democratic practice and it is intrinsically *democratic*. However, we must remember that "democracy involves both voting and discussion, and discussion is obviously at least as important to democracy, descriptively and normatively, as voting." [18] This idea is particularly pursued by deliberative democrats, who play down the role of interest aggregation and state that "the essence of democracy itself is now widely taken to be deliberation, as opposed to voting, interest aggregation, constitutional rights, or even self-government." [13] Deliberation as a social process is then a type of communication process that involves the careful and serious weighing of reasons for and against some proposition [19] and whereby deliberators are willing to change their judgments, preferences and views [13]. Collective choice may be obtained through reasoned agreement, particularly in locality-specific disputes and problems with a relatively small number of identifiable participants who can meet in face-to-face interaction. Another approach (more suitable for large-scale complex issues) elaborates on Jürgen Habermas's theory of deliberative democracy and proposes a discursive, and specifically rhetorical, transmission from the public sphere to the administrative state. This emphasis on public participation through informed deliberation is one of the most important characteristics of recent reflection about the nature and practice of democracy. It represents "a renewed concern with the authenticity of democracy: the degree to which democratic control is substantive rather than symbolic, and engaged by competent citizens." [13] It

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

is perhaps a return to the ancient ideal of assembly politics which was the cornerstone of Greek democracy.

4 THE PLAYERS OF DEMOCRACY

The trends identified on the previous section may be considered in the context of the existing *players* of democracy and their relation with the citizenry. Political parties, representatives and public administration organizations are perhaps the most important intermediaries between citizens and the exercise of power. Participatory approaches, particularly through deliberative means, should be considered when designing technological support for democracy development. Also, citizens themselves (without intermediaries) may influence decision processes by deliberating on the public sphere.

4.1 Political parties

Political parties are considered a key institution in contemporary democratic systems as "they perform the functions of selecting, aggregating and transmitting demands originating from the civil society and which will become objects of political decision." [14] Even those who advocate a more deliberative democracy recognize that it is not possible to deliberate on all issues and therefore "the limited pluralism of parties is required for effective participation." [20] However, political parties are, at least in part, oligarchical and highly hierarchical. In order to maintain the democratic character of the whole political system, it is necessary that political parties would be formed from the mass of citizens, who would select relevant issues, deliberate on them and propose the results for wider deliberation among the whole society. This process needs to be open to the contribution of any ordinary citizen who wishes to participate and transparent enough to allow for effective control by those ordinary citizens [20]. This view is also shared by political theorist C. B. Macpherson who defended a participatory political system based on parties democratized according to the principles and procedures of direct democracy, operating within a parliamentary structure complemented and checked by fully self-managed organizations in the local community [5].

4.2 Political representatives

With the adoption of universal suffrage, and the consequent increase in the number of potential candidates for public offices, political parties are also dedicated to the organization of representation [5]. Individuals who support party proposed policies are selected and put forward by those parties for election by the whole citizenry. While it is true that citizens choose among individuals competing for the legitimacy to govern, the political programs they support also play a role on that choice [20]. Therefore, once representatives are elected, a dilemma occurs: what/whom do they represent and to whom should they be accountable?

A famous statement by Edmund Burke illustrates the idea that representatives should guide their actions, ultimately, by their own reason and conscience:

"Your representative owes you, not his industry only, but his judgement; and he betrays, instead of serving you, if he sacrifices it to your opinion." (Edmund Burke, *Address to the Electors of Bristol*, 1774)

Norberto Bobbio stresses this view by arguing that only 'independent' representatives can pursue the general interest, instead of the sectional interests of whom they represent [21]. According to this view, the electorate can only exercise some influence over their representatives through the anticipation of retrospective control. Representatives are expected to take into account, at the moment they make a decision, the future evaluation that the electorate will make, retrospectively, of that decision. The possibility of non re-electing their representatives, at voting time, is the only means available to the electorate to exert some influence over them [22]. This is certainly a very weak control mechanism, particularly if we consider that an ordinary citizen, when voting on a certain candidate or political party, is selecting a package of policies without necessarily agreeing with all of them. Also, his/her opinion over an issue might change afterwards in face of new arguments or new circumstances. Finally, certain decisions made by representatives were not even considered at the time the representative was elected. According to these considerations citizens may feel that they are being misrepresented and their representatives completely escape their control.

On the other hand, the idea of prospective control of representatives by the electorate, through bound mandates, is in fact a replacement of representative democracy by direct (plebiscitary) democracy: before taking any decision each representative would have to assert his constituents opinion on the issue and decide accordingly. This would be impossible due to the amount of decisions to make and the number of citizens in modern societies. In practice, however, since each representative is elected in association with a political program put forward by political party, his/her mandate often becomes bound to that program. Political parties sometimes require vote discipline and may even punish the representative by revoking his/her mandate when voting discipline is broken [14].

So, between 'independent' representatives and representatives with bound mandates how can ordinary citizens influence their representatives other than by threatening not to re-elect them or revoking their mandates? The answer lies ultimately in the way representatives, as well as political parties, view their relationship with those who elect them. They need to continuously consider the points of view of those who have elected them, particularly between electoral moments, without surrendering their independence. Furthermore, they need to fully explain their options by publicly advertising their arguments.

4.3 Public administration and the expertise of civil society

Contemporary societies are characterized by complex societal problems, defined by their dynamic character, the many phenomena included, the many actors involved and the im-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

pact they have on society [23]. They are usually considered to be "wicked" or "ill-structured" problems [24], multi-defined, hard to analyze and to handle [25]. Furthermore they present a high degree of uncertainty, particularly with respect to the consequences of possible actions and decision making guiding values. Due to the growing complexity and number of problems to be dealt with, modern states organized administrative apparatuses, a vast network of organizations run by appointed officials with a professional administration and specialized officialdom which execute public policy (defined by representative institutions) under strict observance of the law. Public administrations, being responsible for running decision making processes which directly affect the quality of life of ordinary citizens, "become an important focal point, and some would say battleground, in the discussions over public involvement."

[26] Administrative theory and practice in the last few decades advocates the participation of stakeholders to increase the quality of decision analysis and support for decision making. These stakeholders may be defined as "organisations and individuals whose interests are affected by the policy under discussion" and it is assumed that they may provide important high quality information to complement the use of scientific data [27]. Other authors consider that besides expert/scientific and stakeholder information, a public participation process should include the views of *ordinary* citizens [28], considered here as "… those not holding office or administrative positions in government" [26]. This ordinary citizen participation is crucial for a number of reasons. First, it allows to overcome the shortfalls associated with stakeholder representation in deliberative institutions [29]. It is not often easy to identify all interests to consider and find a suitable representation for them. Even then, some citizens may consider themselves misrepresented by those who act as stakeholder representatives on behalf of his/her interests. Also, ordinary citizens may prove to be experts in some field where they have experience and/or knowledge at least as relevant as the *official* expertise [30; 26]. Their potential contribution (such as ideas, comments and proposed solutions) is simply disregarded if they are excluded from the decision making process. Ultimately, not only the success of implementing the outcome of the process depends on the acceptance by the citizens involved, but also, it is the cornerstone of democracy that they should influence that outcome [24]. The relation between public administrations and citizens has the potential to be transformed with the support of ICTs.

4.4 The constellations of discourses in the public sphere

The work of Jürgen Habermas, particularly his emphasis on deliberation in public spheres, has influenced many theorists of deliberative democracy. Among them, John Dryzek [13] proposes a more critical type of deliberative democracy, termed *discursive democracy*, that emphasizes the contestation of discourses in the public sphere. Dryzek defines *discourse* as "a shared way of comprehending the world embedded in language", having at its centre "a story line, which may involve opinions about both facts and values" and featuring "particular assumptions, judgments, contentions, dispositions, and capabilities" [31]. The public sphere is then at any time home to *constellations of discourses* and the role of deliberation is to promote reflective choice across them. This process of *contestation of discourses* in the public sphere influences the content of public policy according to the relative weight of these discourses at a given time and place. Therefore, Dryzek proposes to re-conceptualize public opinion as the "provisional outcome of the contestation of discourses in the public sphere as transmitted to the state (or transnational authority)". He proposes that such transmission can be accomplished by a number of different means, including the deployment of rhetoric through the alteration of the terms of political discourse, by creating worries about political instability, and by arguments being heard by public officials. This type of deliberative democracy gives citizens the possibility to influence policy adoption and execution without the interference of any political institution (such as parties and representatives) and constitutes therefore another trend to take into account when considering the role of ICTs in the transformation of democracy.

5 TECHNOLOGY AS A TOOL TO SUPPORT THE TRANSFORMATION OF DEMOCRACY

Information and Communication Technologies (ICTs) are increasingly influencing society and, over the last decade, the Internet has undoubtedly been the most influential and pervasive of those technologies. The recognition of the Internet potential to support democracy innovation derived from the suggestion that it could help to overcome same time and same place constraints posed by traditional face-to-face participatory initiatives. This initial rationale for the use of ICTs has been replaced by a more structured analysis about their potential contributes to the democratic process. The sections above provided an overview of some of the democratic areas in which that contribution may be applied, and put forward some requirements beside to overcome same time and same place constraints: help to foster communication between citizens and other important *players*: political parties, representatives, public administration organizations; allow access to relevant information; support collective deliberation; provide easy access to and experimentation with decision models.

When considering the basic Internet tools it is possible to recognize four different ways by which ICTs can support and promote citizen participation [32]: by providing information on a problem and its background, by supporting communication processes, by structuring debates, and by directly supporting decision processes (e.g. through electronic voting). A different taxonomy is proposed to characterize *e-democracy* initiatives [33]: *e-enabling* (supporting those who would not typically access the internet and take advantage of the large amount of information available), *e-engaging* (consulting a wider audience to enable deeper contributions and support deliberative debate on policy issues), and *e-empowering* (supporting active participation and facilitating bottom-up ideas). Finally, it is possible to outline four possible scenarios for technology supporting democracy [30]: by supporting direct democracy, by supporting civic communities (online

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

communities), using surveys and opinion polls to gauge public opinion, and engaging citizens in policy deliberation, emphasizing the deliberative element within democracy.

The World Wide Web (WWW) is perhaps the most important basic service provided by the Internet infrastructure. In its simplest form web sites may be used as a repository of information necessary to induce self-reflection and preference formation, two pre-conditions for deliberation. More recently, a special type of web site seems to proliferate on the Internet. Weblogs (or blogs) act like a personal journal of an individual or group of individuals and try to mimic the role of the traditional media (particularly newspapers) and bring into the public sphere topics and point of views that otherwise would pass unnoticed. Another interesting development of the traditional web sites is illustrated by the Wiki Wiki Web project

[34] that gives virtually anyone the possibility to edit the web site. However, this type of support for collaborative effort lacks the structure and coordination capabilities to ensure the production of agreed documents that can act as representative of the different discourses.

In its simplest form web sites provide one-way information flow but it is common to see them combined with other technologies such as chats, forums and online opinion polls to provide a two-way communication channel. Chats provide an interesting communication channel that may be useful to allow individual citizens to discuss among them, with or without the possibility to include experts and/or public officials.

Online opinion polls are becoming very popular on the Internet as a tool to quickly and cost effectively collect general public opinions. However, common opinion polls tend to collect instantaneous, non-reflexive opinions. An interesting improvement is proposed by James Fishkin under the designation of *deliberative opinion polls* [35]. By creating the necessary condition to promote deliberation before the opinion poll takes place, Fishkin proposes to discover what the public would think if it had more opportunity to think about the questions and more information about the issues. Internet is used to distribute relevant and balanced information to participants, support the deliberation process through synchronous audio conferencing and conduct the opinion collection. Despite the major improvement over traditional opinion polls, they still remain limited to deliberation about pre-determined options.

Discussion forums constitute another very common technology typically taking one of two forms [33]:

- Issue-based forums, i.e. organized around policy issues that have been formulated by policy-makers, interest groups or 'experts', and presented as the heading of one or more discussion 'threads';
- Policy-based forums, i.e. organized around themes/issues that relate directly to a draft policy that is meant to address these, and where discussion threads are intended to solicit responses from those affected.

Current discussion forums do not properly support deliberation and informed debate since the discussion is structured with links to previous messages, providing an unsorted collection of vaguely associated comments. Computer Supported Argument Visualization (CSAV) [36] aim at shifting from these current online forums to forums designed constructively to visualize arguments and counter arguments, thus enhancing the deliberation potential of these very popular systems. Such deliberation support could be based on Issue Based Information Systems (IBIS), a language and graphical representation scheme for visualizing argumentation.

Geographical Information Systems (GIS) research is looking into participatory approaches to local and regional spatial planning and has proposed new types of systems such as Public Participation Geographical Information Systems (PPGIS)

[37] and Web-based Public Participation Systems (WPPS) [38]. Their functionality includes allowing web browsing of documents and static map images, providing communication channels for discussion and voting, allowing interactive map-based queries, scenario building and on-line commenting. The main rationale behind these proposals is that an important part of local policy decision making has strong geographical references. The main limitations of these systems lie on the cognitive demand manipulating GIS systems pose on the common citizen and also on the fact that not all policy problems are geographically related.

Efforts are also being made to use Multi-Criteria Decision Making methods to support e-democracy initiatives [39], usually through Web-enabled Decision Support Systems. An e-negotiation system is being proposed [40] under the TED project (Towards Electronic Democracy, http://bayes.escet.urjc.es/ted). The Decisionarium site (http://www.decisionarium.hut.fi -[41]) proposes a set of interactive multicriteria decision support tools that can be used in public participation processes [42]. Again, the systems proposed so far demand a high cognitive effort from the common citizen and seem focused on providing scientific information (improving communication between experts and the public). Furthermore, this approach may not be suitable to deal with 'wicked nature' that characterize societal decision problems [25] and the emphasis on a formal-rational approach is contrary to the discursive way proposed by deliberative democrats. A more discursive proposal comes from Murray Turoff (and others) who propose the development of a Social Decision Support System (SDSS) to "support the investigation by large groups of complex topics about which many diverse and opposing views are held" [43]. Contributions to the debate would have to be expressed as an issue, option, comment or relationship between one of the above. A continuous dynamic voting system would help to filter and organize the submitted contributions. Despite the obvious improvement with regard to contribution organization, such a system would have to depend on the citizen ability to post each contribution under the 'correct' label. The danger would be of transforming the debate into a meta-debate about the correct label for each contribution ("Is it an option or simply a comment?").

Efforts are also being made to structure discussion and support public consultation (including opinion collection) about predetermined issues and respective options. However, support for genuine deliberation is yet far from being achieved and require more than a simple combination of existing Internet tools. "Deliberation is more than merely a discussion of the issues. Emphasis is also given to the product that arises from discussion (e.g., a decision or set of recommendations), and the process through which that product comes about" [44]. Efforts to provide such deliberative support include collaborative writing systems [45] aimed at engaging large number of citizens in producing as many documents as needed to express the *constellation of discourses* advocated by Dryzek. These documents may then be used to influence particular public decision making processes or long run policy discussions.

Having presented the trends and areas in democratic practice that can take advantage of the technology potential, as well as the research efforts being made in that sense, we must not neglect the fact that citizens must possess the necessary skills and conditions in order to use it. Since democracy is intrinsically associated with equality among citizens, this issue cannot be overlooked.

6 THE DIGITAL DIVIDE

Ancient Greek society was proud of the equality it provided for their citizens namely, *isocracy* (possession of equal power by all), *isegory* (equality in freedom of speech) and *isonomy* (equality of legal rights). Jean-Jacques Rousseau advanced the notion of equality between citizens does not mean absolute economic equality:

"...by equality, we should understand, not that the degrees of power and riches are to be absolutely identical for everybody; but that power shall never be great enough for violence, and shall always be exercised by virtue of rank and law; and that, in respect of riches, no citizen shall ever be wealthy enough to buy another, and none poor enough to be forced to sell himself." [46, bk. II, ch. XI]

The liberal ideal that individuals are "free and equal" to pursue their own goals and justify their own actions [5] is an important characteristic of contemporary societies which recognize an array of formal citizenship rights including freedom of speech, freedom of assembly and universal suffrage. However, these formal rights do not assure that individuals effectively possess equal opportunity to participate in public reasoning and decision. In practice, economic and social inequalities still deter many to fully exercise their rights and limit the access to *analog* media thus making it very difficult to some citizens to make themselves heard [47]. ICTs, and their pervasiveness in contemporary society, may contribute to further deepen these inequalities and further "amplify the voices of the digital 'haves' at the expense of the 'have-nots'."

[30] The problem does not lie exclusively in the lack of access to technology, such as broadband Internet, but also (and perhaps more importantly) in the lack of knowledge, training and willingness to use it as a democratic tool. Particular attention must be paid to the interface and its usability in order to avoid unnecessary cognitive efforts.

However this *digital divide* should not mean that the use of ICTs to support the transformation of democracy should be disregarded. Instead, provisions must be taken to ensure that alternative channels of participation (*analog* ones) are available and social and economic inequalities are attenuated.

7 FINAL REMARKS

Information and Communication Technologies (ICTs), particularly the Internet, are now increasingly part of our lives in contemporary societies. It is only natural that this pervasiveness included the political system in an attempt to bring democracy closer to the original "rule by the people" Greek ideal. But to achieve more than a gain in electoral efficiency it is necessary to avoid technology driven initiatives. Simple technological solutions, such as going online and/or providing email addresses, ignore the huge potential ICTs have to help reducing the gap between contemporary practices and the democratic ideal. We believe a more structured approach is needed and it is therefore necessary to consider democracy development trends in political theory and use them as a framework for technology adoption. Such trends include a demand for more direct citizen participation, not only in the traditional political sphere, but also in all aspects of citizen's daily societal life. Due to their many facets, citizens interact every day with numerous civil society institutions which would benefit from the expertise and points of view ordinary citizens can provide. On the other hand, the fact that citizens have the opportunity to *practice democracy* in a *close context* is perhaps the best way to increase participation in a more *remote context* of national politics.

"The future of democracy lies with strong democracy - with the revitalization of a form of community that is not collectivistic, a form of public reasoning that is not conformist, and a set of civic institutions that is compatible with modern society." [4]

The nature of participation is also experiencing an emphasis shift from participation as preference aggregation (through voting mechanisms) to a more deliberative approach that values informed discussion and preference transformation.

It is also important to consider these trends in the context of the relations between citizens and other *players* of the political system: the nature and mechanisms of citizens' influence on political parties and representatives, and the relation between public administration organizations and the citizens affected by their decision processes.

Particular attention must be paid to the *digital divide* which threatens to limit the efficacy of *e-democracy* initiatives and add to the socio-economic inequalities that undermine political equality, a cornerstone of democracy.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Finally, there is no need for a new legal framework, complete institutional revolution or the abandon of the representative model of democracy. The success of these transformations (particularly the use of ICTs on democratic initiatives) depends ultimately on the political will of those involved in the political system and, above all, on citizens' motivation to participate more actively in all aspects of public life.

ACKNOWLEDGEMENTS

This paper was supported by FCT POCI/EGE/58828/2004.

REFERENCES

1. Hacker, K. L. and J. van Dijk (2000), What is digital democracy? In K. L. Hacker and J. van Dijk (eds.), Digital Democracy: issues of theory and practice, pp. 1-9, SAGE Publications, London.
2. Dahl, R. A. (1996), Democratic theory and democratic experience. In S. Benhabib (ed.), Democracy and difference: contesting the boundaries of the political, pp. 336-339, Princeton University Press, Princeton.
3. Pateman, C. (1970), Participation and Democratic Theory, Cambridge University Press, Cambridge.
4. Barber, B. R. (1984), Strong democracy: Participatory politics for a new age, University of California Press (Berkeley).
5. Held, D. (1997), Models of Democracy, Polity Press, Cambridge.
6. Rawls, J. (1993), Political Liberalism, Columbia University Press, New York.
7. Habermas, J. (1996), Between facts and norms: contributions to a discourse theory of Law and Democracy, MIT Press, Cambridge, Massachusetts.
8. Dahl, R. (1956), A preface to democratic theory, University of Chicago Press, Chicago.
9. Mill, J. S. (1974), On Liberty. In On Liberty; Representative Government; The Subjection of Women. Three essays by John Stuart Mill, pp. 1-142, Oxford University Press, London.
10. Mill, J. S. (1974), Considerations on Representative Government. In On Liberty; Representative Government; The Subjection of Women. Three essays by John Stuart Mill, pp. 143-423, Oxford University Press, London.
11. Schumpeter, J. (1943), Capitalism, Socialism and Democracy, Unwin University Books, London.
12. Cooper, T. L. (1984), 'Citizenship and professionalism in public administration', Public Administration Review, March/143-149.
13. Dryzek, J. S. (2000), Deliberative democracy and beyond: liberals, critics, contestations, Oxford University Press, Oxford.
14. Bobbio, N. (1997), Democracy and Dictatorship: The Nature and Limits of State Power, Polity Press, Cambridge.
15. Arrow, K. (1951), Social Choice and Individual Values, Wiley, New York.
16. Gibbard, A. (1973), 'Manipulation of voting schemes: a general result', Econometrica, 4/41/587-601.
17. Satterthwaite, M. (1975), 'Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and Social Welfare Functions', Journal of Economic Theory, 2/10/187-217.
18. Mackie, G. (1998), All men are liars: is democracy meaningless? In J. Elster (ed.), Deliberative Democracy, pp. 69-96, Cambridge University Press, Cambridge.
19. Fearon, J. D. (1998), Deliberation as discussion. In J. Elster (ed.), Deliberative Democracy, pp. 44-68, Cambridge University Press, Cambridge.
20. Manin, B. (1987), 'On legitimacy and political deliberation', Political Theory, 3/15/338-368.
21. Keane, J. (1997), Introduction: Democracy and the decline of the Left. In Democracy and Dictatorship: The Nature and Limits of State Power, pp. vi-xxviii, Polity Press, Cambridge.
22. Elster, J. (1998), Introduction. In J. Elster (ed.), Deliberative Democracy, pp. 1-18, Cambridge University Press, Cambridge.
23. DeTombe, D. J. (2001), 'Compram, a method for handling complex societal problems', European Journal of Operational Research, 2/128/266-281.
24. Geurts, J. L. A. and C. Joldersma (2001), 'Methodology for participatory policy analysis', European Journal of Operational Research, 2/128/300-310.
25. DeTombe, D. J. (2001), 'Introduction to the field of methodology for handling complex societal problems', European Journal of Operational Research, 2/128/231-232.
26. Roberts, N. (2004), 'Public deliberation in an age of direct citizen participation', The American Review of Public Administration, 4/34/315-353.
27. Bongers, F. J. (2000), Participatory Policy Analysis and Group Support Systems, Tilburg University, Tilburg.
28. Renn, O., T. Webler and P. Wiedemann (1994), A need for discourse on citizen participation: objectives and structure of the book. In O. Renn, T. Webler and P. Wiedemann (eds.), Fair and competent citizen participation: evaluating new models for environmental discourse, pp. 1-15, Kluwer Academic Publishers, Boston.
29. O'Neill, J. (2001), 'Representing people, representing nature, representing the world', Environment and Planning C: Government and Policy, 4/19/483-500.
30. Coleman, S. and J. Gøtze (2001), Bowling together: Online public engagement in policy deliberation, Hansard Society and BT, London.
31. Dryzek, J. S. (2001), 'Legitimacy and economy in deliberative democracy', Political Theory, 5/29/651-669.
32. Lenk, K. (1999), Electronic support of citizen participation in planning processes. In B. N. Hague and B. D. Loader (eds.), Digital democracy: discourse and decision making in the information age, pp. 87-95, Routledge, London.
33. Macintosh, A. (2004). Characterizing E-Participation in Policy-Making. Proceedings of the 37th Hawaii International Conference on System Sciences, (CD/ROM), January 5-7, 2004, Hawaii, USA. Computer Society Press.
34. Leuf, B. and W. Cunningham (2001), The Wiki way: collaboration and sharing on the Internet, Addison-Wesley.
35. Fishkin, J. S. (1995), The voice of the people. Public opinion and democracy, Yale University Press.
36. Macintosh, A. and A. Renton (2004). Argument visualisation to support democratic decision-making. Proceedings of the eChallenges e.2004 Conference, 27-29 October, 2004, Vienna, Austria.
37. Carver, S., A. Evans, R. Kingston and I. Turton (2001), 'Public participation, GIS, and cyberdemocracy: evaluating on-line spatial decision support systems', Environment and Planning B: Planning and Design, 6/28/907-921.
38. Peng, Z.-R. (2001), 'Internet GIS for public participation', Environment and Planning B: Planning and Design, 6/28/889-905.
39. Moreno-Jiménez, J. M. and W. Polasek (2003), 'e-Democracy and knowledge. A multicriteria framework for the new democratic era', Journal of Multi-Criteria Decision Analysis, 2-3/12/163-176.
40. Ríos-Insua, D., J. Holgado and R. Moreno (2003), 'Multicriteria e-negotiation systems for e-democracy', Journal of Multi-Criteria Decision Analysis, 2-3/12/213-218.
41. Hämäläinen, R. P. (2003), 'Decisionarium - Aiding decisions, negotiating and collecting opinions on the Web', Journal of Multi-Criteria Decision Analysis, 2-3/12/101-110.
42. Mustajoki, J., R. P. Hämäläinen and M. Marttunen (2004), 'Participatory multicriteria decision analysis with Web-HIPRE: a case of lake regulation policy', Environmental Modelling & Software, 6/19/537-547.

43. Turoff, M., S. R. Hiltz, H.-K. Cho, Z. Li and Y. Wang (2002). Social Decision Support Systems (SDSS). Proceedings of the 35th Hawaii International Conference on System Sciences, (CD/ROM), January 6-9, 2002, Hawaii, USA. Computer Society Press.

44. Abelson, J., P.-G. Forest, J. Eyles, P. Smith, E. Martin and F.-P. Gauvin (2003), 'Deliberations about deliberative methods: issues in the design and evaluation of public participation processes', Social Science & Medicine, 2/57/239-251.

45. Lourenço, R. and J. P. Costa (2006). Discursive e-Democracy support. Proceedings of the 39th Hawaii International Conference on System Sciences, January 4-7, 2006, Hawaii, USA. Computer Society Press.

46. Rousseau, J.-J. (1913), The Social Contract and Discourses by Jean-Jacques Rousseau, J.M. Dent and Sons, London and Toronto.

47. Gordon, T. F. and G. Richter (2002), Discourse support systems for deliberative democracy. In R. Traunmuller and K. Lenk (eds.), E-government: State-of-the-Art and Perspectives (EGOV), pp. 248-255, Springer-Verlag, Aix-en-Provence.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e•Centre for Infonomics

i·Society 2007

Towards advanced e/m-Government platforms

Vassilis Meneklis, Spyros Papastergiou, Christos Douligeris, Despina Polemi

University of Piraeus, Department of Informatics, Piraeus, Greece.
{bmenekl, paps, cdoulig, dpolemi}@unipi.gr

Abstract In the passage of time governmental platforms have changed significantly in terms of offered functionality, number and complexity of the provided services, as well as creation of secure and trusted frameworks within which citizens, enterprises and governmental organizations will operate. The available technological capabilities play a major role towards the transition to more advanced e/m-Government platforms that provide the opportunity to governmental organizations to meet the challenges of a continually developing world. In this context, the main scope of the current paper is to present the major requirements that the next generation advanced e/m-government platforms should satisfy and to propose a series of steps in implementing such a platform. These steps are presented as enhancements to an existing e-Government platform (eMayor). The parts of the architecture that need to be updated or even replaced are identified and the major building blocks of the advanced e/m-Government platform (SWEB) are provided.

Keywords e/m-government platform, modelling enhancements, mobility, Web Services, RM-ODP

1 INTRODUCTION

If one was to categorise the evolution of e-Government technologies until now based on functional criteria, one would identify three main periods. In the first period e-Government was considered a new innovative paradigm for public administration. Platforms were simple internet sites providing information about the organisation and its activities. Governmental employees and citizens had little to none experience with such platforms. During the second period, citizens and governments became more accustomed to the new terms and technologies as well as the opportunities that they presented. Platforms became more functional, offering generic electronic services such as requests for certificates which were then processed conventionally and delivered to the citizens in printed form as before. In the third period, complex security mechanisms were implemented affecting the access to the platform, workflow processes were added and more importantly the offered services became much more extended and wholly electronic. Communication mechanisms between platforms were designed in order to achieve interoperability across organisational and national domains. The e-Government platforms of the third period are usually focused on addressing specific issues and achieve to satisfy only a subset of the corresponding requirements [17], [18]. Representative platforms of this period along with their objectives are presented below.

The Electronic and Secure Municipal Administration for European Citizens (eMayor) [11] project aimed to provide secure, interoperable and affordable web services for small and medium sized government organizations (SMGOs) across Europe, emphasizing especially on the satisfaction of specific security and interoperability issues. The Government User Identity for Europe (GUIDE) [12] program created a European conceptual framework for electronic identity management for e-Government. Its main objectives are the creation of trusted framework within which the transactions will be accomplished in a secure manner.

The Intelligent Natural Language Based Hub for the Deployment of Advanced Semantically Enriched Multi-channel Mass-scale Online Public Services (HOPS) [13] focused on the deployment of advanced semantically ICT voice-enabled front-end public platforms in Europe that provides interactive access to public services information through voice technologies. The Intelligent Cities (IntelCities) [14] project focused on the development of an interoperable platform for cross-border public services to exchange data enabling comparison and analysis. The Impact of e-Government on Territorial Government Services (TERREGOV) [15] program addresses the issue of interoperability of e-Government services for local and regional governments.

The usability-driven open platform for Mobile Government (USE-ME.GOV) [16] project aimed at satisfying specific mobile issues. Such issues are the development of common interface software between administrations and mobile service providers, the investigation on new ways of user interaction with mobile devices, proposing business models for mobile e-Government services, and extending knowledge transfer to the mobile arena. Most of these issues are fulfilled at a significant level taking into account the technological capabilities and the limitation [19] of the mobile devices of the specific period. The results of this project can be considered as the cornerstone for the next generation platforms in order to develop a more advanced, secure, interoperable, and trusted mobile government framework.

It is our belief that we are currently experiencing a transition between the third and the fourth periods of e-Government technologies. Citizens as well as enterprises constitute the main users of e-Government platforms. These users have come to rely on certain enterprise services and demand extensions to the underlying platforms. Effective service provision and expansion, and extended availability of enterprise services bring benefits to all involved stakeholders and strengthen their relations. Therefore, e-Government practitioners must satisfy the needs to increase the quality of the provided services and to offer new ones taking into consideration the available technological capabilities. On the other hand the penetration that mobile devices and the new technologies have achieved both at the professional and the personal level to our lives as well as the impressive technology developments in the mobility sector give birth to a new way of communicating interpersonally, between citizens and organisations, and between organisations themselves. In that respect, the newly attributed electronic Government term has almost become obsolete, steadily being replaced by the term mobile Government, which we believe marks the beginning of this fourth period.

Although mobile access is the trademark of this transition, it is not the only change that is experienced. Platforms of the next period are also identified by their implementations of advanced security features [9] (such as time stamping or revocation techniques of used certificates) and the introduction of basic privacy aspects [10].

Our contribution in this paper lies on two areas, the identification of the requirements brought to the foreground by the transition from the third generation of e-Government to the fourth generation of m-Government and the development of a prototype for enhancing a contemporary e-Government platform with advanced mobile, security and privacy capabilities.

The rest of the paper is structured as follows; Section 2 discusses the need of adopting advanced e/m-government platforms and the requirements that they impose. Section 3 presents a case study of an existing third period platform and proposes a real time holistic enhancement towards an advanced e-m-Government platform of the fourth period. Finally Section 4 concludes and identifies future research directions.

2 NEEDS AND REQUIREMENTS IN ADVANCED E/M-GOVERNMENT PLATFORMS

The continued development of modern society has created new challenges for governmental organizations. The available technologies and standards provide essential assistance to such organizations. The following sections illustrate the main reasons that the governmental organizations are driven to the adoption of an advanced e/m-government platform, and present several vital requirements that can be considered as prerequisites for platform success.

2.1 The need for an advanced e/m-Government platform

Mobility constitutes one of the major challenges of the modern era. Nowadays, citizens travel continuously all over the world for a multitude of reasons, expecting to receive high quality services and ubiquitous access to public administrations both in the countries they are visiting and the countries they are coming from. Consequently, the need for developing next generation platforms or reconfiguring the existing ones in order to support emerging service capabilities is very current.

Governmental organizations should be able to take advantage of the spread and success of mobile/wireless applications and services, the growth and the continued evolution of Internet, the domination of XML and XML Schemas and the development of the corresponding standards. This will give them the possibility to exploit these new developments and enhance electronic government implementations by moving to the mobile government dimension, allowing at the same time innovative services to be created and deployed in short time addressing citizens' needs.

Therefore, the entrance to the mobile environment is vital for governmental organizations. Enabling these organizations with the ability to provide secure m-services to other public organizations and citizens holds a lot of future prospects by strengthening the fundamental structure of these organizations and by enhancing the collaboration of the public sector. M-Services will contribute towards the solution of difficult problems that the governmental sector faces in a more easy and flexible way.

2.2 Requirements

The design and development of a flexible, accessible and efficient advanced e/m-government platform that supports innovative services and applications in order to be widely accepted by the citizens and organizations that interact with the governmental organizations, have to address a set of requirements. In this section we present these requirements.

- *Mobility*: The citizens have to receive high quality services and ubiquitous access to public administrations regardless of their geographical position. The mobile devices due to the adequate offered capabilities such as in memory, battery and computing power are capable to be used by the citizens in order to access the provided mobile enterprises services. Of course, the limitations of the mobile devices still remain and further investigation on this area has to be performed.
- *Interoperability*: The interconnection with many different infrastructures is a difficult task, requiring easily identifiable and publishable mobile enterprises services, as well as clear mobile interfaces for the establishment of secure and reliable connection points. For this reason, new technologies that promote the interoperability (i.e. Web Services) and light protocols in the messaging exchange have to be adopted. Components that are able to handle the exchanged messages must

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

be implemented both in the platform and mobile devices.

- *Security and Trust*: The various provided e/m-services have to be secure in all aspects (confidentiality, integrity, authenticity, long lasting integrity, authorization, non-repudiation, privacy), so that all users come to trust the system and feel confident in using it. The adoption of the proper advanced security and privacy mechanisms that will undertake the responsibility to secure the exchanged messages creating a trusted framework is considered crucial for an advanced e/m-government platform.
- *User Friendliness and Accessibility*: The wireless environment needs to be easily accessible, with user-friendly interfaces covering the specific needs of various types of mobile users and services. Therefore, new mobile applications have to be designed and developed properly taking into consideration the constraints that come from the size and the capabilities of the mobile devices. These applications should offer basic functionality while the complex operations must be completely transparent to the user.
- *Support of cross-border interconnection*: Exchange of information, referring to data or documents between citizens and public administration in an international context and across administrative boundaries, is a strong demand. The design and development of new enterprise services that cover this possibility without aggravating the performance of the mobile devices is a fact that should not be ignored.
- *Reduced organizational and technical complexity*: The adoption of advanced e/m-Government platforms and all provided services should add a small amount of organizational and technical complexity, concerning financial and temporal parameters. This is achieved by introducing standard solutions, applicable in different governmental organizations that are quick and easily customizable to the organization's requirements, trying to diminish the need for maintenance during operation.
- *Scalability and extensibility*: The platforms have to be simple, open, reconfigurable, scalable and easily extensible. Capable to serve a large number of citizens with acceptable levels of quality of service. The increased usage of mobile devices in everyday life reveals the existence of a huge potential group of users that make the utilization of widely used technologies and standards a one-way road for the creation of a scalable and extensible platform.
- *Limited Connectivity and Processing Capabilities of the mobile Clients*: The limited capabilities (e.g. battery, memory, processing power, screen, bandwidth capabilities) of the mobile devices must be taken into account during the design and development of the platforms. The constraints that come from these limitations must play a major role on the adoption of the most suitable solution for the provided enterprise services.
- *Compliance with European Legal & Policy frameworks*: Compliance with the underlying legal and policy framework as dictated by the laws and directives of the states where an e/m-service will be deployed and operated.

The satisfaction of the aforementioned requirements constitutes the basic step for the development of a secure, interoperable, open, affordable platform upon which innovative, and scalable enterprise services may be built respecting the will of the citizens.

3 CASE STUDY

In this section we present an e-Government platform that has all the characteristics of the third period. Through projecting all the requirements, that were discussed in the previous section, on this platform, we identify the parts of the architecture that need to be updated or replaced. We continue by presenting a real life enhancement to the platform considering all the aforementioned problems.

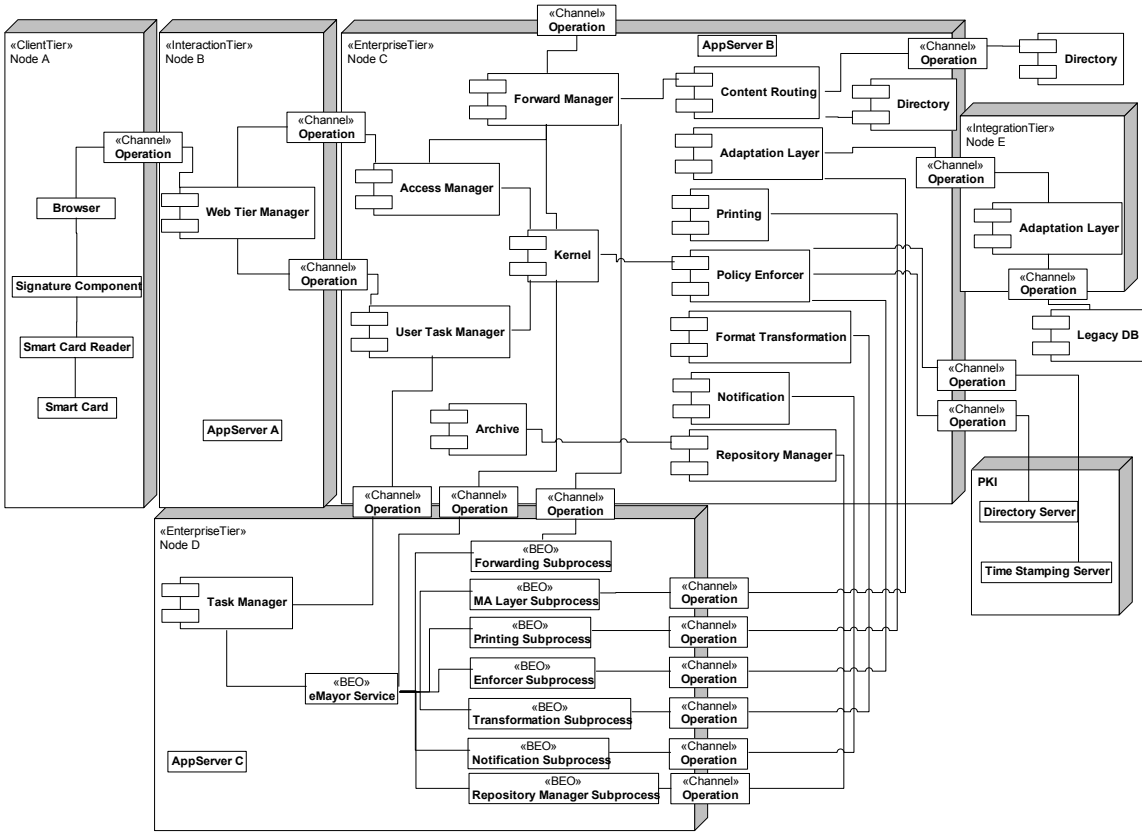
3.1 A contemporary e-Government platform

Our case study was developed as a result of the European project eMayor [11], a successful e-Government security project in which a Web Services based platform was built as a holistic service framework for the deployment and delivery of e-Government enterprise services for European Small to Medium Governmental Organizations (SMGOs). In the eMayor project five European municipalities decided and tested two secure and cross-border e-Government enterprise services. These two enterprise services are the issuance of online residence certification e-documents and the management of taxes payment by direct debit. Both services need a citizen to be authenticated to the platform before requesting service. The workflow process then starts with the citizen's electronic request which is properly stored and processed by the platform. Although the specific workflow sub-processes differ for each of the two enterprise services, there are some common characteristics that allow for a basic grouping of these two enterprise services. Notifications are sent to the requesting citizen via email at important milestones of the service execution (at request submission, at service proper completion or at service improper completion). Both enterprise services generate an electronic document (on the first occasion the online residence certification e-document and on the second an electronic confirmation document concerning taxes payment) which is digitally signed and provided to the citizen.

Access to the platform is accomplished through a Java applet that runs on the computer where the user connects from. The applet handles all the interaction processes with the platform, like sending authentication information about the user to the platform, receiving data from the platform parsing it and presenting it to the user, and handling digital signatures of the user requests. Even though the applet solution is highly customisable from a programming point of view, it poses some practical difficulties and constraints to the easiness of access to the platform from the user's point; the need for installation at the computer that the user connects from

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 1. eMayor Engineering Viewpoint



and the consumption of a fair amount of resources being the two most important. These constraints led us to reconsider the implemented access to the platform scheme and to propose a more flexible based on mobile technologies. Our proposition is described in more detail in Section 3.2.

In order to design and implement the eMayor platform, the ISO reference model for open distributed processing systems was used (RM-ODP) [1]. For detailed information and descriptions about the use of RM-ODP in e-Government please refer to [2], [3]. In this paper we will focus on the two viewpoints that most descriptively present the enhancements that we propose as well as the impact they will have on the architecture and the utilisation of the platform; namely the engineering and the information viewpoints.

An engineering viewpoint describes the mechanisms and components that are required to support distributed interaction between objects in the system [1]. As shown in Figure 1, the eMayor engineering viewpoint specification is divided into five nodes (using RM-ODP terminology) each of whom represents one tier of the architecture. The types of tiers are represented as UML stereotypes. These are:

- Node A: The Client Tier, comprising the browser components.
- Node B: The Interaction Tier, comprising the web server and its components.
- Node C: The main Enterprise Tier, comprising the main application server and the platform main components that run as Web Services and Enterprise Java Beans.
- Node D: The secondary Enterprise Tier, comprising another application server managing the choreography of the main platform services in order to implement the business logic of eMayor.
- Node E: The Integration Tier, comprising the Adaptation Layer components that sit "on-top" of any municipal existing / legacy system.

For a detailed specification of the eMayor engineering viewpoint, one should refer to [4].

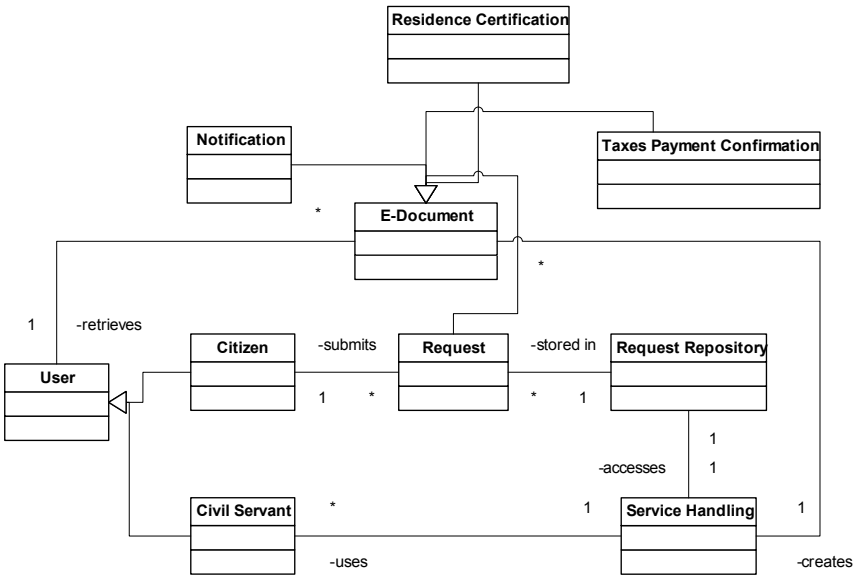
Figure 2 presents an overall information viewpoint of the eMayor platform. The information viewpoint describes the platform and its environment focusing on the semantics of information and the semantics of information processing [1]. It identifies information objects that are communicated and processed during the system's operation.

The User information object is aggregated to the Citizen and Civil Servant objects. A Citizen can submit an enterprise service Request which is stored in a Request Repository for future processing. When the Request is stored and assigned a pending status, the Citizen is informed via email (Notification information object). The Civil Servants use the Service Handling module (actually they use a set of modules, but for the scope of this viewpoint this is irrelevant) to access the Request Repository and process all the pending Requests. When a Request is processed an e-Document is created (either a residence certification e-document or a taxes payment confirmation) and stored temporarily waiting for Citizen Retrieval.

3.2 An advanced e/m-Government platform

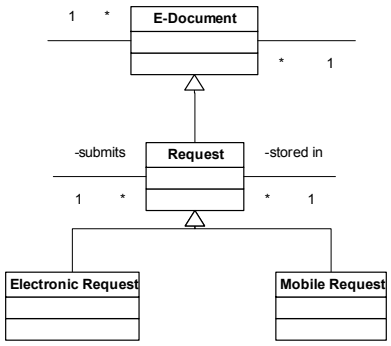
Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. eMayor Information Viewpoint



In view of the requirements for advanced e/m-Government platforms that were presented in section 2, the basic updates that must be made to the eMayor information viewpoint specification concern the information objects Request and e-Document. These updates and the corresponding tasks that implement them have as result the major building blocks of the SWEB platform [37]. Figure 3 depicts the extensions that must be made to these information objects.

Figure 3. Request Information Object Extension



The Request information object has to be extended with two aggregates, the Electronic Request which will be the type of Request that is already implemented in eMayor (and which is until now was referred to simply Request) and the Mobile Request information object which represents the new types of requests that will be developed for mobile access.

The updates and extensions that will be made to the information objects of the platform will reflect on the computational procedures as well. The overall engineering updates that will be implemented to the eMayor platform are depicted in Figure 4:

One can identify several tasks resulting from the updates in Figure 4. We present these tasks below.

1. Task A (Client Tier)

New stand alone Web Service-based mobile applications should be implemented. Citizens would be able to download them to their mobile devices (i.e. PDAs) and install them in order to access the provided mobile enterprise services. The applications must have further consideration of predefined fields, limited values for the used fields and administrative operations that can be performed.

The mobile devices will communicate with the Interaction Tier of the platform. Unfortunately, the limited resources and processing capabilities of the mobile devices require optimized implementations of mobile Web Services messaging, which comprises the messages exchange between mobile client devices with Web Services running on the platform and light protocols used in the messaging exchange, including SOAP message handling, message dispatching, and XML serialization / de-serialization.

Finally, the existing browser components for electronic access to the platform have to be reconfigured and parameterized in order to support the extended functionality of the e/m-platform and meet the new technological challenges.

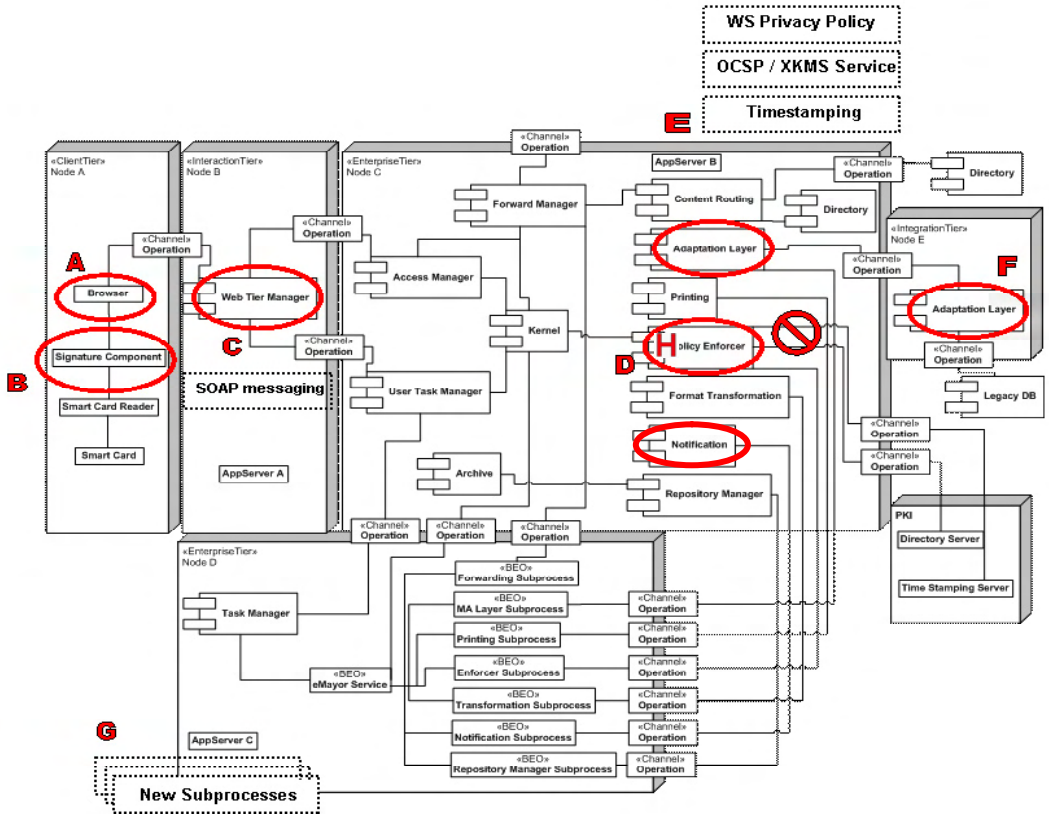
2. Task B (Client Tier)

The mobile applications needs must provide cryptographic functionality in order to satisfy all the security requirements as these were presented in section 2.2. Therefore, the technologies that have to be adopted are the following:

- Digital Signatures based on the W3C XML Signature Recommendation (XML-DIG) [24].
- W3C XML Encryption [25].
- OASIS WSS (WS-Security) [28] in order to add encryption, digital signatures and authorization token support to SOAP messages for web services.
- W3C XML Key Management Protocol (XKMS) [22], [23] in using web services to simplify several PKI [29] protocols.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. Advanced updates to eMayor e-Government platform



- XML Advanced Electronic Signature (XAdES) [26] Standard.
- The use of carious types of tokens, such as X.509 certificates and SIM cards.
- Communication with various Trust Third Parties (TTP) such as OCSP server, Timestamp server and a Secure Token Service (STS) [34] using WS-Trust [33] and Security Assertion Markup Language (SAML) [21] technologies.
- Pprogramming language capabilities and mobile Government policy issues must seriously thought of.

3. Task C (Interaction Tier)

The interaction tier of an advanced e/m-platform undertakes the responsibility to operate and communicate with the neighbouring components which are the client tier (including the mobile devices) and on the other side with the enterprise tier that performs the business logic of the provided services. The upgraded interaction tier has to contain a SOAP Proxy ("SOAP messaging" component in Figure 4) that will enable support for SOAP message exchange with the Client Tier. Every message that is directed to a mobile device from the platform will be enveloped in a SOAP message. On the opposite direction every message that originates from mobile clients directed to the platform will be in SOAP format and parsed by the SOAP proxy.

This approach is in line with the proposal of Kim et al. [5], which opts for a transcoding proxy server between the mobile client and the platform. Another option for the implementation on the Interaction Tier is also that of Mobile Agents, as discussed by Cheng et al. [6], and Bellavista et al. [7]. The choice in favour of the SOAP Proxy can be considered that is preferable due to the performance evaluation of a proxy based scenario that presented very promising results, as discussed by Boutrous-Saab et al. in [8].

4. Task D (Enterprise Tier)

The Policy Enforcer, in the current implementation plays the role of an Access Control Manager combined with a Digital Certificate Manager. On one hand it receives the authentication information form the Access Manager and enforces the corresponding security policies for each user. On the other hand, it communicates with the PKI in order to validate the authentication information (credentials) it receives. The distinction of these two different functionalities is considered necessary and will result in a greater level of flexibility. Therefore, the Policy Enforcer has to be renamed to Access Control Manager and be responsible only for authentication and authorization tasks. Furthermore, it must not be connected to the PKI which is an external component of the platform.

5. Task E (Enterprise Tier)

An advanced e/m-platform is essential to provide advanced security and privacy services. The support of such services distinguishes a platform belonging to the third from one of the fourth period. The components that will support these are described below.

Time-stamping is one of the most important services that must be supported in the upgraded and more functional advanced e/m-platform. The existence of such a service it

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

will actually enable the platform to embed secure time stamp tokens, communicating with a Third Trusted Party (TTP), using XML-based messages. This functionality provides the opportunity to an entity which trusts the TTP that timestamps to verify that a document was not created after the date that the trusted entity vouches.

For the communication with the Timestamp servers the IAIK [30] or Bouncy Castle [31] implementations can be used which are conformant to the standard Timestamp specification (RFC 3161) [35]. While the proper operation of the service can be tested using online Demo Servers such as the IAIK Demo Time stamping Authority, the Certum Time stamping Authority, the EdelWeb experimental Time stamping service [27].

The second component undertakes the responsibility to check the revocation status of X.509 digital certificates that will be used in order to accomplish the cryptographic functions that are performed. The use of the Online Certificate Status Protocol (OCSP), as described in RFC 2560 [36], could be an option since, in contrast to Certificate Revocation Lists (CRLs), provides the following:

- It can provide more timely information regarding the revocation status of a certificate
- It's operation doesn't need for clients to retrieve the CRLs themselves, leading to less network traffic, better bandwidth management and less processing power, saving client-side complexity.

However, an advanced platform that is based on XML technologies, utilizing the Web Services framework, must also support a solution based on an appropriate XML technology. The use of the XML Key Management Specification (XKMS), developed by W3C, is an option taken into account.

Finally, a Web Service Privacy Policy (WSPP) component has also an important role in order to handle the WS-Privacy Policies that need to be enforced. The major objective of the defined Web Service Privacy Policy is to convey conditions on an interaction between two web service endpoints. All the information, which will be provided in the Privacy Policy, is aimed to describe the capabilities and requirements of the web service entities (i.e. the functional attributes of the offered service, such as confidentiality mechanisms and authentication characteristics). The WSPP component has as main responsibility to decide if the participating web service entities are able to accomplish the document exchange according to the assertions that have been released in their Privacy Policies.

6. Task F (Enterprise and Integration Tiers)

This task is not strongly tied to the enhancement of the platform with advanced and mobile technologies. In order to present this task we considered the strong scalability requirements for advanced e/m-Government platforms. As enterprise services provided by such platforms evolve and address a continually increasing part of the population, these platforms will be installed and operated to many governmental organizations. In that respect there is a requirement for scalability. The Adaptation Layer of the Integration Tier is responsible for the interaction of the platform and the legacy systems of the governmental organizations. If it is to satisfy the above requirement, it must be able to support data processing and exchange with every legacy system that will be used at the new governmental organizations.

7. Task G (Enterprise Tier)

In line with the previous task, task G is about scalability, too. The Enterprise Tier utilizes Business Process Execution Language (BPEL) [32] sub-processes in order to orchestrate and coordinate the execution of the actual platform services. This coordination results in forming the overall enterprise services (business processes) that are offered to the citizens (i.e. the business process of submitting a certificate request). The continual growth of offered services - both in complexity and numbers - brings the need to extend current BPEL sub-processes and even add some new ones to the architecture. These new sub-processes are developed and integrated in the Enterprise Tier of the SWEB platform.

8. Task H (Enterprise Tier)

Governmental organizations must be able to provide to their users advanced notification services informing them about the state of processing their requests are at. A Notification should be either Wired or Wireless taking into account decision of the requesters. Currently, the existing service is able to offer only email notification. This functionality can not completely cover the need of retrieving information's messages regardless of user's position.

Hence, the notification services have to be upgraded providing further functionality such as the capability to send SMS messages or even Voice Notification messages. In this way, the relation of the stakeholders is strengthened while at the same time the reliability of the governmental organizations is increased significantly.

4 CONCLUSIONS AND FUTURE WORK

In this paper we presented a short overview of the evolution of e-Government to date, identifying three main periods and a current transition to the fourth. In our perspective, this transition is guided by the needs of the users of e-Government platforms (user pull) and the technological achievements in the mobile, security and privacy sectors (technology push). We then discussed the current requirements for an efficient advanced e/m-Government platform. Based on our observations concerning the evolution of e-Government and the contemporary platform requirements, we proposed a set of enhancements to a real life e-Government platform (eMayor), which will result in an advanced e/m-Government platform (SWEB).

In this paper we have mainly focused on the mobility aspects of next generation platforms. Future work in this area can focus on a large array of topics. Some of them include

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

(but are not limited to) research in privacy aspects in order to develop innovative services, composition mechanisms of implemented services in order to achieve higher level of automation, standardisation of modelling techniques for e/m-Government platforms, introduction of modelling concepts in order to support efficient and precise design of such platforms which is an essential step of their development process.

With our work we provided a first approach of an analytic framework comprising four phases of e-Government evolution based on technological achievements and organizational requirements. Through our analysis we contributed to the existing literature concerning e-Government requirements. Finally, our case study describes a set of real life enhancements that can be practically applied to existing platforms in order to move on to the next generation of m-Government.

ACKNOWLEDGEMENTS

This work has been supported by the GSRT (PENED) programme and the IST project SWEB (IST-2006-2.6.5). The authors would like to thank all the participants for valuable discussions and the European Union for funding the SWEB project.

REFERENCES

1. ISO Reference Model for Open Distributed Processing — Part 1: Overview, Part 2: Foundations, Part 3: Architecture, Part 4: Architectural Semantics. (1998), ITU-T Rec. X.901 | ISO/IEC 10746 - 1, 2, 3, 4.
2. Meneklis, B., Kaliontzoglou, A., Polemi, D., and Douligeris, C. (2005, March 2-4), "Applying the ISO RM-ODP standard in egovernment". E -government: Towards Electronic Democracy: International Conference, TCGOV 2005, Proceedings, Bolzano, Italy (LNCS 3416/2005, pp. 213). Springer-Verlag GmbH.
3. Kaliontzoglou, A., Meneklis, B., Polemi, D., and Douligeris, C. (2007), "A formalized design method for building e-government architectures", Secure E-Government Web Services, IGI Global.
4. Meneklis, B., Kaliontzoglou, A., Douligeris, C., and Polemi, D. (2005), "Engineering and technology aspects of an e-government architecture based on Web services", ECOWS 2005. Third IEEE European Conference on Web Services, Vajxo, Sweden. On page(s): 12 pp.-
5. Kim, K., Lee, H., and Chung, K. (2001), "A distributed proxy server system for wireless mobile web service", 15th Intl. Conference on Information Networking 2001, pp. 749-754
6. Cheng S., Liu J., Kao J., and Chen C., (2002), "A New Framework for Mobile Web Services". In Proc. 2002 Symposium on Applications and the Internet (SAINT) Workshops, pages 218-222, January 28-February 1
7. Bellavista, P., Corradi, A., and Monti, S. (2005), "Integrating Web Services and mobile agent systems", Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'05)
8. Boutrous-Saab, C., Meliti, T., and Mokdad, L. (2006), "Performance evaluation for mobile access to composite Web Services" Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services (AICT/ICIW 2006), Guadeloupe, French Caribbean
9. Lee Y., Lee J., Song J., (2007). "Design and implementation of wireless PKI technology suitable for mobile phone in mobile-commerce". Computer Communications, Volume 30, Issue 4 (February 2007), Pages: 893-903, Year of Publication: 2007, ISSN:0140-3664
10. Papastergiou S., Karantjias A., Polemi D. (2007), "A federated privacy-enhancing identity management system (FPE-IMS)" PIMRC 2007, Athens, Geece.
11. E.C 6th Framework Programme, (2007), "Electronic and secure municipal administration for European citizens – eMayor", IST-2004-507217, 2004, Available at www.emayor.org (October 2007).
12. Government User Identity for Europe (GUIDE). Available at http://istrg.som.surrey.ac.uk/projects/guide/ (October 2007).
13. Intelligent Natural Language Based Hub for the Deployment of Advanced Semantically Enriched Multi-channel Mass-scale Online Public Services (HOPS). Available at http://www.bcn.es/hops/ (October 2007).
14. Intelligent Cities (IntelCities). Available at http://www.intelcities-project.com/wcm-site/jsps/index.jsp?type=page&cid=5026&cidName=HOME&isAnonymous=true (October 2007).
15. Impact of e-Government on Territorial Government Services (TERREGOV). Available at http://www.terregov.eupm.net/my_spip/index.php (October 2007).
16. Usability-driven open platform for Mobile Government (USEME.GOV). Available at http://www.usemegov.org/ (October 2007).
17. Rose J., Rossum V. M. (2005) "A Roadmap for European Research in Learning and Knowledge Creation in eGovernment". European Conference on Electronic Government, Antwerp, 2005.
18. "EGovernment: Work Programme 2005-06 IST Framework Programme VI Consultation workshop on 5th May 2004". Available at http://www.fomi.hu/hunagi/pdf/2004/recommended/mits/eGov-June5.pdf (October 2007).
19. Archer P., Mitukiewicz E. (Editors). (2005). "Scope of Mobile Web Best Practices". W3C Working Draft 1 September 2005. Available at http://www.w3.org/TR/mobile-bp-scope/ (October 2007).
20. Hartman B. et al. (2003). Mastering Web Services Security, Wiley Publishing.
21. Cahill C. P. et al, "Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0", Available at http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf (October 2007).
22. Ford W. et al, "XML Key Management Specification XKMS", W3C note, March 2001, www.w3.org/TR/xkms (October 2007).
23. Park N., Moon K., Sohn S., "Certificate validation service using XKMS for computational grid", Proceedings of the 2003 ACM workshop on XML security, pp.112-120, Fairfax, Virginia, 2003.
24. XML Signature Recommendation – XML-DSIG, http://www.w3.org/TR/xmldsig-core/ (October 2007).
25. XML Encryption, http://www.w3.org/Encryption/2001/ (October 2007).
26. Cruellas J. et al. (Editors). (2003). XML Advanced Electronic Signatures (XAdES), W3C Note 20 February 2003, www.w3.org/TR/XAdES/ (October 2007).
27. IAIK demo time stamping authority. (2006). http://tsp.iaik.at/ (October 2007).
28. Nadalin A., 2004. Web Services Security: SOAP Message Security 1.0 (WS-Security 2004), OASIS Standard, docs.oasis-open.org/wss/2004/01/oasis-200401-wss-soap-message-security-1.0.pdf (October 2007).
29. Adams C., Lloyd S., Understanding Public-Key Infrastructure – Concepts, Standards and Deployment Considerations, 1st Edition, Macmillan Technical Publishing, 1999.
30. IAIK XAdES library. (2006). Available at http://jce.iaik.tugraz.at/sic/products/xml_security/xades (October 2007).
31. Bouncy Castle cryptographic libraries. (2006). Available at http://www.bouncycastle.org/latest_releases.html. (October 2007).
32. Andrews T., et al, "Business Process Execution Language for Web Services version 1.1". Available at http://download.boulder.ibm.com/ibmdl/pub/software/dw/specs/ws-bpel/ws-bpel.pdf (October 2007).

33. Anderson S., et al, (2005). "Web Services Trust Language (WS-Trust)". Available at http://schemas.xmlsoap.org/ws/2005/02/trust/ (October 2007).

34. Implementing Message Layer Security with a Security Token Service (STS) in WSE 3.0. Microsoft. Available at http://www.willydev.net/descargas/PartnerAndPractices/WillyDev_ImplementingMessageLayerSecuritywithaSTSinWSE3.0.pdf (October 2007).

35. Adams C., Cain P., Pinkas D., Integris R., (2001). "IETF RFC 3161 Time-Stamp Protocol (TSP)". Available at http://www.ietf.org/rfc/rfc3161.txt (October 2007).

36. Myers M., Ankney R., Malpani A., Galperin S., Adams C.. (1999). "X.509 Internet Public Key Infrastructure Online Certificate Status Protocol – OCSP". Available at http://www.ietf.org/rfc/rfc2560.txt (October 2007).

37. Sixth Framework Programme, Priority 2, Information Society Technologies, (2007), "Secure, interoperable, cross border m-services contributing towards a trustful European cooperation with the non-EU member Western Balkan countries – SWEB", IST-2006-2.6.5

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

The diffusion of innovation beyond the tipping point: the case of the regional cancer program formulary software

Michelle Marie Goulbourne

Department of Health Policy, Management and Evaluation
Faculty of Medicine, University of Toronto
Health Sciences Building
155 College Street, Suite 425
Toronto, ON M5T 3M6
michelle.goulbourne@utoronto.ca

There is no guarantee that if you build an electronic knowledge dissemination tool people will use it. This poster examines the development, deployment and evaluation of the Regional Cancer Program Formulary Software (RECAP-FS) in order to shed light on important factors that impact on its rate of adoption and sustained clinical use over time.

In an environment where less than half of technological innovations achieve their goals, the Regional Cancer Program Formulary Software (RECAP-FS) project has experienced some success. In 2004 three staff members from the Juravinski Cancer Centre developed RECAP-FS for use at Hamilton Health Sciences and associated community oncology clinics. Within a year of its formal deployment, RECAP-FS has been adopted by another Regional Authority and is developing a global network of oncology partners.

This poster will:

1. Describe the development and deployment of RECAP-FS.
2. Summarize results of interim surveys, confirmatory evaluation and web statistics
3. Discuss the relative impact of each the following variables on the rate of RECAP-FS adoption: (i) characteristics of the innovation, (ii) the number of people involved in the innovation decision, (iii) communication network structure (iv) cultural context and (v) promotion efforts (Rogers 2003).
4. Assess the importance of key 'agents of change' in taking RECAP-FS beyond the tipping point to the level of cultural change.

Transitioning from novelty to sustained changes in day-to-day clinical practice is an important one. The final section of the poster will discuss key lessons learned regarding how local level innovations such as RECAP-FS can have a positive impact on the quality of patient care within the local and global oncology community.

REFERENCES

Everett Rogers (2003), Diffusion of Innovation, Fifth Ed., New York, The Free Press.

i·Society 2007

Developing a web-based intelligent decision support system for personalized healthcare

Chien-Chih Yu, Wen-Liang Kung, Hsiao-ping Chang

National ChengChi University, Taipei, Taiwan

Abstract Personalized healthcare delivers healthcare–related decision services to individual users for assisting them in making a more coherent and focused healthcare treatment plan based on personal health conditions, medical histories, risk factors, as well as preferences and needs. Although personalized health assessment and management should take into account individual health conditions and needs, most health exam and medical centers only offer uniform health exam packages with general medical advices to their customers/patients. Since no person-centric integrated recommendation services for developing personalized health check-ups, wellness maintenance, as well as illness management plans has been provided in current e-healthcare environment, the goal of this paper is to propose a web-based intelligent decision support system for personalized healthcare to meet the demand. A design framework and a system prototype that encompass personalized health check-ups recommendation, medical interpretation and advisory, as well as clinical care recommendation processes are presented to show the feasibility and effectiveness of the personalized healthcare recommendation and decision support services approach.

Keywords Intelligent decision support system, personalized healthcare, healthcare recommendation, web services.

1 INTRODUCTION

According to the World Health Organization (WHO), healthcare embraces all the goods and services designed to promote health, including preventive, curative and palliative interventions, whether directed to individuals or to populations. In the Internet age, e-healthcare gains its currency and emphasizes on creating a web-based knowledge system to deliver personalized healthcare services for yielding improvements in patient outcomes as well as cost reductions [1]. Targeting on life-long person-centric healthcare delivery, personalized healthcare enables individuals to self-manage their own health information as well as to make intelligent decisions based on their health conditions, risk predictions, preferences and needs for obtaining a more coherent and focused healthcare services [2, 3]. In addition, personalized healthcare is also noted as a consumer-centric system in which customized diagnostic, treatment, and management plans are generated and delivered to healthcare consumers by examining a variety of factors including personal behaviours and preferences, family medical histories, and their unique genetic makeup. There is no doubt that personalized health assessment, wellness management, as well as intelligent decision support for monitoring, interpreting, and managing individual healthcare needs have emerged as major issues in the healthcare domain. However, comprehensive discussions on conceptual and implementation issues regarding the development of a personalized healthcare system are still very rare in the literature. On the other hand, by reviewing previous works, we note that (1) health check-ups, medical treatment, dietary, and exercise/fitness recommendations are personalized needs for healthcare, (2) risk prediction models for chronic diseases and cancers are essential for performing health assessment, (3) specific health domain information, evidence, models, and rules are needed to build the intelligent support system for personalized healthcare, and (4) web-based and service-oriented technologies are helpful for facilitating effective system development. In order to fully support all phases and functions of personalized healthcare, an appropriate person-centric intelligent decision support system (IDSS) architecture encompassing desired application functions and processes, as well as data, model, and knowledge management mechanisms is needed. Therefore, the goal of this paper is to propose a web-based intelligent decision support system for personalized healthcare (PH-IDSS) in which personalized health exam recommendation, medical interpretation and advisory, as well as medical/clinical care recommendation services are provided to improve healthcare qualities. In the following sections, a brief literature review is given in section 2. The PH-IDSS framework for personalized healthcare and related functions, processes, models, and rules are illustrated in section 3. In section 4, a prototype system with a scenario and example views are presented. The final section contains a conclusion.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007

2 LITERATURE REVIEW

In this section, reviews of healthcare information systems and decision support systems, as well as healthcare models and rules related to personalized healthcare are described below.

2.1 Healthcare IS and DSS

Viewing from the functional perspective, Morpurgo and Mussi [4] introduce an Intelligent Diagnostic Support System (IDSS) that is a personal diagnosis oriented tool. Its knowledge base, consisting of a database and a rule base, encompasses both the knowledge of the concerned domain and the knowledge about the entity to be diagnosed. It also provides a complete range of the possible choices for making the best healthcare decision. Abidi [2] presents a Tele-Healthcare Information and Diagnostic Environment (TIDE) that aims to ensure lifelong coverage of person-specific health maintenance decision support services. At the heart of TIDE are two intelligent subsystems, namely the Automated Health Monitoring System (AiMS) and the Illness Diagnostic and Advisory System (IDEAS). Gomez et al. [5] develop a webbased self-monitoring system for people living with HIV/AIDS. The system comprises three modules: a patient selfmonitoring personal diary for creating follow-up patient records, a data analysis and visualisation tool, and a facility allowing patients to ask for remote advice and doctor support. The self-monitoring process implies the registration of a set of health status data including personal data, clinical data, life style data, and treatment data. Saade et al. [6] present a webbased decision support system prototype to assess patients with osteoporosis. The system is embedded into a clinical workflow environment that entails three work stages, the integration, the system, and the expert review stages. Key functional features include acquiring high quality data directly from patients or general practitioners, analyzing collected data, and presenting analysis result to the specialists prior to meeting with the patients. A knowledge base is contained in the system. Furthermore, German and Watzke [7] argue that the key factor to personalizing foods for metabolic health is to understand consumer differences and apply the metabolic knowledge in order to match individual differences in metabolism with specific foods and diets. Haux [8], while indicating that most decision support systems in healthcare are designed for healthcare professionals such as physicians, nurses, and hospital managing staffs, circles the need of an information system architecture to support patient-centered shared care from networking care facilities in health regions to home care. On the other hand, taking the viewpoint from the process perspective, German and Watzke [7] focus on personalizing foods for health and delight, and Phillips [3] proposes a personalized medicine for colorectal and breast cancer. Both of these efforts contribute to service deliveries in the preventive phase. In the prototyped decision support system developed by Saade et al. [6] for assessing osteoporosis as aforementioned, an expert consultation is also generated to provide service in the diagnostic phase. In addition, Kypri and McAnally [9] suggest the use of web-based assessment and personalized feedback as routine intervention in the primary care stage.

2.2 Healthcare Models and Rules

For risk prediction, several websites maintained by major cancer research centers in the USA have implemented various assessment models for predicting cancer risks. Among them, there are National Cancer Institute (NCI) (www.cancer.gov), Memorial Sloan-Kettering Cancer Center (MSKCC) (www.mskcc.org), University of Maryland Medical Center (UMMC) (www.healthcalculators.org), and Harvard Center for Cancer Prevention (HCCP) (www.yourdiseaserisk.harvard.edu), etc. In NCI's online cancer query system, a pre-calculated lifetime risk probability of developing cancers or dying by cancers is provided. NCI is devoted to the statistical research and applications of cancer prediction models, and gain reputation for its dedication to this field. MSKCC is the world's oldest and largest private cancer center devoted to patient cares as well as innovative researches. It offers online cancer risk prediction tools to assess risks of, for instance, breast cancer and lung cancer. Aiming at predicting cancer risks, HCCP has developed the Harvard Cancer Risk Index. Its website generates interactive questionnaires to collect required personal data for estimating the risk of having cancer, and additionally, it also provides personalized tips for cancer prevention. As one of the USA's oldest academic medical center, UMMC casts a website that offers twenty-four health calculators including depression, stress, HIV risk, heart disease risk, diabetes, etc., for users to learn more about health status in various aspects Although these existing websites provide friendly user interfaces and various risk models and health calculators, they do not offer an integrated personalized healthcare process that equipped with required analysis/decision models for measuring personal health conditions, recommending necessary health checkups, and suggesting proper medical/clinical treatments. For predicting health risks, in addition to cancer assessment models for lung cancers [10] and breast cancers [11], other risk models for major causes of death focus on assessing risks of coronary heart diseases [12], cardiovascular & cerebrovascular diseases [13], as well as hepatocellular carcinomas [14], etc.

For health check-up planning, the selection of check items is based on recommendation rules related to chronic diseases/cancers exams, as well as periodic health exams. There are two major sources of the periodic health exam recommendation rules: the Canadian Task Force on the Periodic Health Care (www.ctfphc.org) and the United States Preventive Services Task Force (www.ahrq.gov). To make these rules easier for understanding and operating, Dubey et al. [15] develop a periodic health exam checklist form out of these two rule sets. Other healthcare related rules include guideline-driven preventive screening programs [16]. Up to this point, what remains to be explored is the need to integrate appropriate risk prediction models and associated recommendation rules for seamlessly implementing and linking health risk assessment and health exam recommendation processes. Furthermore, the demand to integrate models and rules for facili-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

tating the diagnostic as well as treatment recommendation processes is equally significant.

3 THE PH-IDSS SYSTEM AND PROCESSES

An ideal intelligent decision support system for personalized healthcare should be able to support all healthcare functions and processes that meet personalized needs including health exam recommendation, medical interpretation and advisory, as well as medical care recommendation. By integrating the context-aware and service-oriented concepts and methodologies for personalized management [17, 18], our PH-IDSS architecture contains three layers, namely the personalized user interfaces, the application functions, and the base management subsystems. In the following subsections, the system architecture, system application and functional processes, as well as integrated data, models and rules for specific health exam recommendation process are described in more detail.

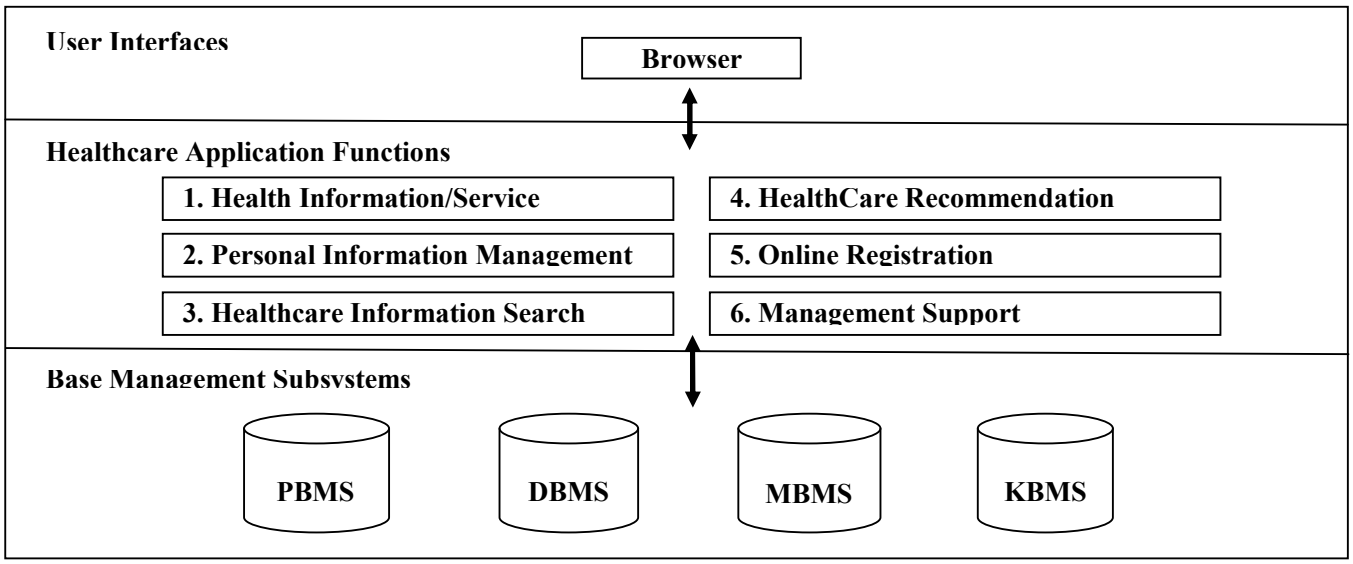
3.1 The System Architecture

The system architecture as depicted in Figure 1 contains three functional levels –the personalized user interfaces, the application functions, and the base management subsystems. Users from the client sites can access the PH-IDSS application server and create their own user interfaces to arrange the layouts of preferred application functions.

To support the personalized healthcare needs, the actionable application functions offered in the application server include health information and services navigation, personal information management, information search, healthcare recommendation, online registration, and management support. Figure 2 shows the contents and services associated with these healthcare related functions. By using the health information and services navigation function, users may navigate through a directory or guide-tour to browse healthcare information, knowledge resources, as well as services and providers. Since health exam is included as a major healthcare service, health exam item descriptions and health exam package plans can also be easily checked and viewed. The personal information management function allows users to input and update personal basic data, medical records, and health history. The healthcare information search function enables users to search and browse medical reports and databases, as well as the hospital and clinical information. By accessing the healthcare recommendation function, users can activate the personalized healthcare recommendation process that retrieves necessary data, triggers and executes suitable decision models and rules for predicting health risks, recommending health exam plans, as well as suggesting nutrition/exercise and clinical care treatments. Users are also allowed to make changes on the health check-up items and to finalize the personalized health exam plan for future implementation. These changes are recorded as feedback to the system for updating the user preferences. Users may then use the online registration function to get health exam appointment and to register for clinical cares based on their needs and preferences regarding hospitals, physicians, and time/location conveniences. The management support function provides a channel for healthcare service providers to transfer and maintain laboratory results, health exam and medical reports, as well as to provide medical interpretation of laboratory result and medical care advices.

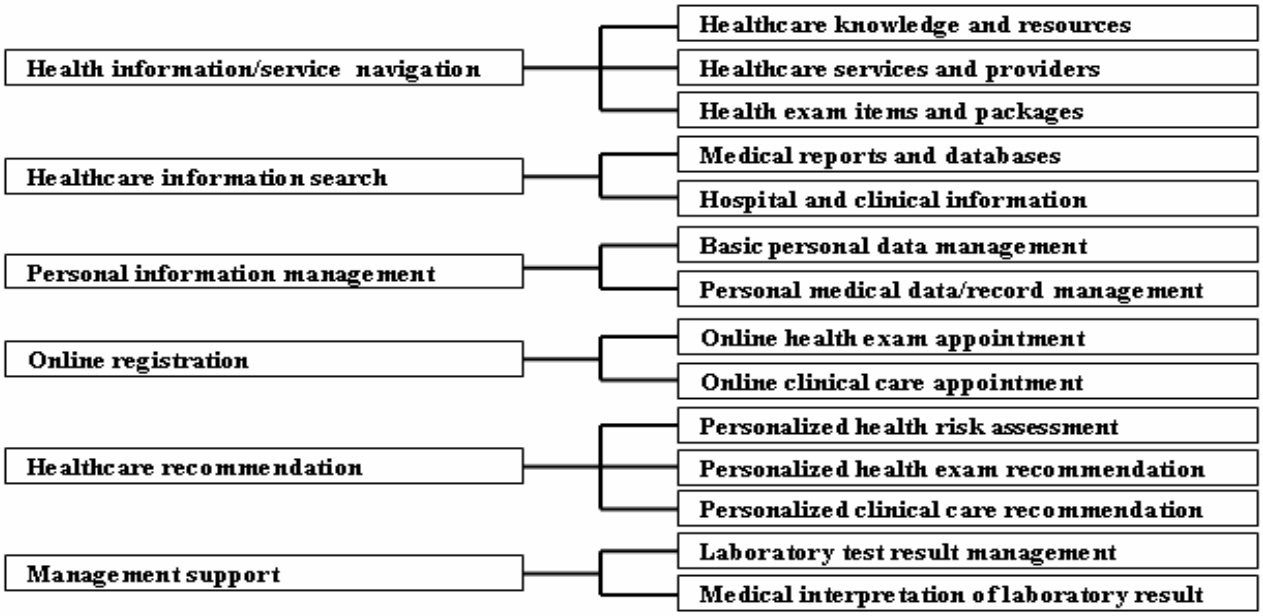
To ensure the efficiency and effectiveness of system management and intelligent decision support for personalized healthcare, base management subsystems in the PH-IDSS backend server environment include the process base management subsystem (PBMS), the database management subsystem (DBMS), the model base management subsystem (MBMS), and the knowledge base management subsystem (KBMS). The PBMS is a management system for efficiently creating, maintaining, and executing healthcare related application processes such as the health risk assessment and the health exam recommendation processes. Developed based on the user's perspective, an application process is designed as a complete operating process for solving a specific healthcare decision problem. In a specific application process, data, models, and knowledge are arranged and integrated accord-

Figure 1. The PH-IDSS architecture



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. PH-IDSS application functions



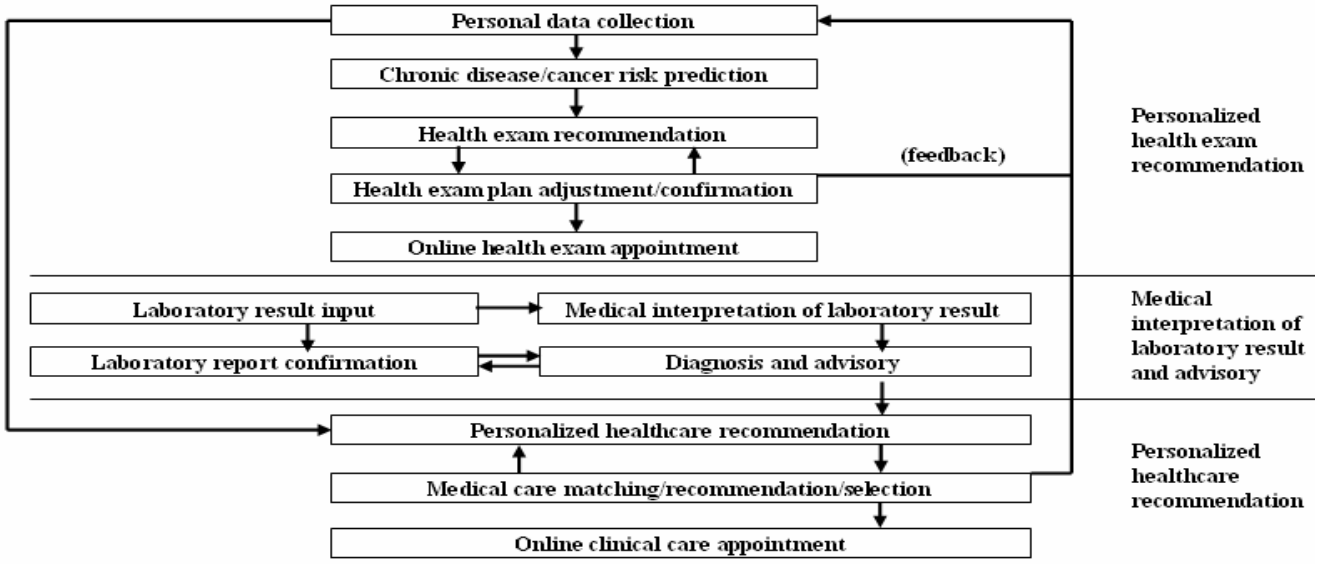
ing to their interoperable relationships for solving the decision problem. The DBMS manages healthcare related databases and web resources with diverse data models and data types. Database entities include document files, relational tables, and multimedia objects generated in the healthcare domain. The MBMS manages the healthcare model base as well as the input, process, and output files for executing specific decision models such as chronic disease and cancer risk prediction models. The KBMS manages all required knowledge to carry out specific healthcare application processes. For example, rule sets used in the risk assessment and health exam recommendation processes. The KBMS creates and manages knowledge base, as well as input, process, and output files for rule inference.

3.2 The System Processes

In Figure 3, three main system processes in the PH-IDSS including the personalized health exam recommendation process, the medical interpretation and advisory process, and the personalized health care recommendation process are illustrated. Descriptions of each system process are as follows.

Before using the personalized health exam recommendation function, users can browse information about health exam packages and health exam item descriptions. They can choose certain health exam package or design their own health exam plans by selecting needed health exam items. If users are not sure about which health exam package to choose and how to form a health exam plan that suits their health conditions, they can activate the personalized health exam recommendation process. Based on the direct input or previously stored personal data, the system performs first the chronic diseas-

Figure 3. PH-IDSS system processes



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

es/cancers risk prediction sub-process and then generates a personalized health exam plan as the responded recommendation. For executing the application process, required personal information includes basic personal data, personal and family medical histories, and personal life styles, etc. Risk prediction models for chronic diseases and cancers cover categories of coronary heart disease, cardiovascular disease, cerebrovascular disease, hepatocellular carcinoma, lung cancer, breast cancer, and others. Three specific rule sets used in the recommendation process include the periodic health exam recommendation rules, the chronic diseases/cancers exam recommendation rules, and the preventive screening recommendation rules. After viewing the system recommended health exam plan, the user may adjust the health exam items to form a final plan and further make online appointments with a selected health exam center based on his needs and preferences. The health exam booking confirmation messages and health exam preparation instructions will be delivered to the user and also displayed on the screen.

After the completion of physical health examination, laboratory results and health exam report are transferred to the system via the management support functions. Based on the exam report, the system performs the medical interpretation and advisory process. First, value of each health check-up item is compared with the standard/normal value including the tolerance range and abnormal range, and the occurrence and meaning of abnormal value, if any, is highlighted. And then, the system integrates related results and performs an overall health evaluation. Based on the medical interpretation and evaluation, medical care advices are proposed and forwarded to the physicians for confirmation, and then presented to the users. As some in-process examples, the hepatitis B blood test result comes from three tests: hepatitis B surface antigen (HBsAg), Hepatitis B surface antibody (HBsAb) and hepatitis B core antibody (HBcAb); and the health evaluation and advisory contents include personal body fitness analysis (body mass index, waist-and-hip ratio, and recommended daily nutrition allowances).

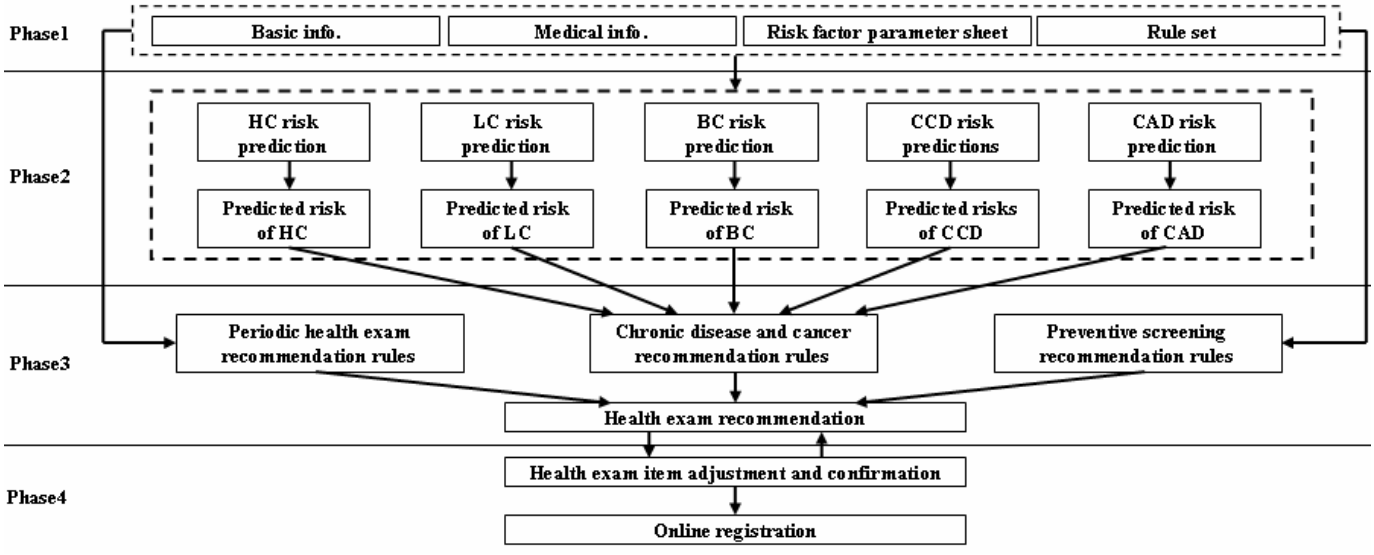
The personalized health care recommendation process uses personal data and health exam reports, as well as associated models and rules to develop healthcare recommendations for wellness maintenance and illness treatment. The healthcare recommendation outputs include personalized nutrition and exercise recommendations and personalized medical care recommendations such as food treatment and medicine treatment. These healthcare recommendations are confirmed by the physicians before they are handed to the users. Users can then use the medical care matching and appointment function to locate suitable hospitals and get clinical appointments based on personal health conditions as well as time and location preferences.

3.3 Data, Models and Rules for Health Exam Recommendation Process

In this section, we will further detail the personalized health exam recommendation process with descriptions and illustrations of inter-related sub-processes, models, and rules. For fully carrying out the personalized health exam recommendation process, associated sub-processes that need to be performed include processes for personal data collection, health risk prediction, health exam recommendation, as well as health exam plan adjustment and confirmation, as shown in Figure 4.

In the personal data collection process, personal data from the users are collected and put into the database for being used as inputs to various risk prediction models as well as health exam recommendation rules. In the health risk prediction process, the main task is to predict users'risks of getting chronic diseases and cancers over ten-year time span. Inputs to the risk prediction process and models include personal data, the parameter value sheet of risk factors, and the rule sets for assigning values to certain parameters. These input data are extracted from the database and rule base and/or collected from the user interface. Among many risk prediction models, six major risk models in the domain of chronic

Figure 4. Personalized health exam recommendation process



Note: HC stands for hepatic cancer. LC stands for lung cancer. BC stands for breast cancer. CCD stands for cardiovascular and cerebrovasular diseases. CAD stands for coronary heart disease.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 1. Personal data elements

Category	Data element
Basic info.	ID, name, sex, age, current marital status, phone, mobile, address, and email.
Basic info. for women	age at menarche (ma), age at first birth (ba), number of children (parity), breast feeding (bm), postmenopausal hormones (ph), and oral estrogen (oe).
Life style	eating habit, sleep habits, sports, smoking (sm), cigarettes per day (cpd), duration of smoking (smyear), duration of smoke-quitting (quit), asbestos exposure (asb), and alcohol drinking (ad).
Previous lab tests	height, weight, waistline, hipline, total cholesterol (tc), HDL-cholesterol (hdl), systolic blood pressure (sbp), diastolic blood pressure (dbp), blood pressure stage (bps), triacylglycerol, AlanineTransaminase (alt), HBsAg serostatus (hhbsag), and previous stool occult blood in last 1 year (stool).
Personal medical history	allergies, diabetes (dm), obesity, osteoporosis, stroke, depression, aspirin, hypertension, dyslipidemia, atrial fib (af), left ventricular hypertrophy (lvh), coronary heart disease, cardiovascular disease, cerebrovascular disease, lung cancer, hepatic cancer, breast cancer (bc), and any other cancer.
Family medical history	First/second degree with high blood pressure, diabetes, high cholesterol, coronary heart disease, cardiovascular disease, cerebrovascular disease, hepatic cancer (fhcc), breast cancer (fhx), lung cancer and any other cancer (bcself).

Table 2. The parameter value sheet for the coronary heart disease risk factor

Risk factor	Male	Female
Sex	c_sex(M)	c_sex(F)
	0.04826	0.33766
Total cholesterol (tc)	c_tc(M, tc)	c_tc(F, tc)
<160	-0.65945	-0.26138
160-199	0.0	0.0
200-239	0.17692	0.20771
240-279	0.50539	0.24385
>=280	0.65713	0.53513
HDL-cholesterol (hdl)	c_hdl(M, hdl)	c_hdl(F, hdl)
<35	0.49744	0.84312
35-44	0.24310	0.37796
45-49	0.0	0.19785
50-59	-0.05107	0.0
>=60	-0.48660	-0.42951
Blood pressure stage (bps)	c_bps(M, bps)	c_bps(F, bps)
Optimal	-0.00226	-0.53363
Normal	0.0	0.0
High normal	0.28320	-0.06773
Stage I	0.52168	0.26288
Stages II-IV	0.61859	0.46573
Diabetes (dm)	c_dm(M, dm)	c_dm(F, dm)
Y	0.42839	0.59626
N	0.0	0.0
Smoking (sm)	c_sm(M, sm)	c_sm(F, sm)
Y	0.52337	0.29246
N	0.0	0.0
Mean value of all parameters (G)	G(M)	G(F)
	3.09755	9.92545
Baseline survival rate over ten years (S_{10})	S_{10}(M)	S_{10}(F)
	0.90015	0.96246

diseases and cancers are the coronary heart disease prediction model, the cardiovascular disease prediction model, the cerebrovascular disease prediction model, the lung cancer risk prediction model, the hepatic cancer risk prediction model, and the breast cancer risk prediction model. Outputs of the health risk prediction process include these six predicted risk levels of getting chronic diseases and cancers over ten years.

In the health exam recommendation process, the core mission is to recommend a personalized health exam plan for the user. Inputs for processing the recommendation task include personal data and risk levels collected and derived from previous processes. Three recommendation rule sets used in this process include rules for the periodic health exam recommendation, the chronic diseases/cancers exam recommendation, and the preventive screening recommendation. The periodic exam rule set is applied to suggest general health exam items based on sex and age. The chronic diseases/cancers exam rule set is for deriving the suggestion of health exam items related to chronic diseases and cancers based on the predicted risk levels. The preventive screening rule set is responsible for recommending the health exam items related to obesity, diabetes mellitus, lipid disorder and hypertension based on personal data and risk levels. Since the obesity-related factors cause the majority of the chronic disease burden, the preventive screening rules trigger the generation of health exam items for people at the risk of obesity-related diseases. Consequently, in the end of the recommendation process, the output is a personalized health exam recommendation plan that is suitable for the user based on his health

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Table 3. The coronary heart disease prediction model

Model name	Coronary heart disease prediction model
Input	age, sex, total cholesterol (tc), HDL-cholesterol (hdl), blood pressure stage (bps), diabetes (dm), and smoking (sm), etc.
Process	$C = c\_sex(sex)*age + c\_tc(sex, tc) + c\_hdl(sex, hdl) + c\_bps(sex, bps) + c\_dm(sex, dm) + c\_sm(sex, sm)$ (1) $A = C - G(sex)$ (2) $B = e^{A}$ (3) $P_{10} = 1-[S_{10}(sex)]^{B}$ (4)
Output	Predicted risk of getting coronary heart disease over ten years (P_{10})

condition and risks. A personalized health exam recommendation plan consists of suggested exam items and associated costs. Finally, the health exam adjustment and confirmation process is to allow users to adjust the system recommended health exam plans by adding or dropping health exam items with further personal concerns. By the time the user confirms his finalized health exam plan, an order is transferred to a system-directed or user-selected health exam center for making a health exam appointment.

As shown in Table 1, there are five categories of personal data including basic information, life style, previous lab test results, personal medical history and family medical history. Examples of the basic information are sex and age. There is some basic information specifically for women to be used as inputs for predicting breast cancer risks. Life style data includes eating habits, exercise habits, smoking habits, and alcohol drinking habits. Examples of the previous lab test results are cholesterol (total cholesterol and HDL-cholesterol) and blood pressure (systolic blood pressure, diastolic blood pressure). Major personal and family medical histories include historical information of high blood pressure, high cholesterol, diabetes, heart diseases, and cancers. Sex and age are inputs to all risk prediction models, the periodic recommendation rules and the preventive screening rules. The smoking habits are inputs to the coronary heart disease prediction model and the lung cancer prediction model. Data of the previous lab test results and personal medical history are inputs to the coronary heart disease, cardiovascular disease, cerebrovascular disease, and hepatic cancer risk prediction models. The family medical history data is used for the breast cancer risk prediction.

In addition to personal data, inputs for processing risk prediction include parameter values of multiple risk factors. Using the coronary heart disease risk prediction as an example [12], Table 2 shows the parameter value sheet of the coronary heart disease risk factors. The first column displays risk factors of the coronary heart diseases and their value ranges. The second and the third columns list associated parameter values for men and women respectively. The risk factors include sex, age, total cholesterol (tc), HDL-cholesterol (hdl), blood pressure stage (bps), diabetes (dm), and smoking (sm). The unit of both total cholesterol and HDL-cholesterol is mg/dL. There are five levels for tc ranging from below 160 mg/dL to equal to or above 280 mg/dL. There are five levels for hdl ranging from lower than 35 mg/dL to equal to or higher than 60 mg/dL. Several bps levels are from optimal to stage IV. The values of dm and sm are either yes or no. Two more derived impact factors are the mean value of all factors (G), and the baseline survival function over ten years (S10). Based on the risk factors and gender, associated parameter values generated by according mapping functions are placed in the second and the third columns. There are totally eight mapping functions for the risk factors of the coronary heart diseases including c_sex(sex), c_tc(sex, tc), c_hdl(sex, hdl), c_bps(sex, bps), c_dm(sex, dm), c_sm(sex, sm), G(sex), and S10(sex). These mapping functions take values of the risk factors and gender as inputs and out parameter values of the risk factors. For example, c_tc(M, 250)= 0.50539 is the risk score for male with 250 of total cholesterol level, and c_sm(M, Y)= 0.52337 is the risk score for male who smoke. The function G(sex) represents the mean risk for certain sex. For instance, G(M)= 3.097547 is the mean value of all risk factor parameters for male. Similarly, S10(M)= 0.90015 shows the survival rate for male over ten years horizon. These parameter values are stored in the database and used as inputs to the coronary heart disease prediction model.

In Table 3, the inputs, processes and outputs of the coronary heart disease prediction model are presented. Inputs for the risk prediction model include age, sex, total cholesterol (tc), HDL-cholesterol (hdl), blood pressure stage (bps), diabetes (dm), and smoking (sm). There are four computing equations in the process for generating the final output P10, that indicates predicted risk level (probability) of getting coronary heart disease over ten years.

To illustrate the coronary heart disease prediction process, the following scenario is used.

Tom, a 55-year-old male smoker, has total cholesterol of 250 mg/dL, HDL-colesterol of 39 mg/dL, blood pressure of 146/88 mm Hg (that falls into stage I hypertension), and no diabetes. The model inputs are age=55, sex=M, tc=250, hdl=39, bps=stage I, dm=N, and sm=Y.

From equation (1),C=55*0.04826+0.50539+0.24310+0.52168+0.0+0.52337=4.4478.

From equation (2), A=4.4478-3.09755=1.35025.

From equation (3), B=e1.35025=3.85839.

From equation (4),
P10=1-0.900153.85839=1-0.666391=0.333609=33%.

Based on the coronary heart disease risk prediction model, Tom has a 33% risk of getting coronary heart diseases over ten years.

Besides the coronary heart disease risk prediction model, the inputs, processes, and outputs of other risk prediction models including the cardiovascular disease prediction model [13], cerebrovascular disease prediction model [13], lung cancer risk prediction model [10], hepatic cancer risk prediction model [14], and breast cancer risk prediction model [11] are stored and managed in the model base. Due to the page constraint, these models are not shown in this paper.

3.4 Personalized Health Exam Recommendation Rules

Adapted from previous works, three types of health exam recommendation rules are integrated: the chronic diseases and cancers recommendation rules, the periodic recommendation rules, and the preventive screening rules [15, 16]. Parts of the rule sets are shown in Figure 5. The inputs for the personalized health exam recommendation rule set include (sex, age) from the personal basic information, (hypertension, dyslipidemia, obesity, etc) from the personal medical history, the first/second degree with breast cancer (fhx) from the family medical history, the predicted risks of getting chronic diseases and cancers (P10, CV10, V10, L, H10, and B), and the average risks of chronic diseases and cancers (Pa, CVa, Va, La, and Ha). Similar as P10 indicating predicted risk level of getting coronary heart disease over ten years, CV10, V10, and H10 represent predicted risk levels of having cardiovascular disease, cerebrovascular disease, and hepatic cancer over ten years respectively. L is the predicted risk level of having lung cancer within one year, and B is the breast cancer screening index. To be extracted from the database, Pa, CVa, Va, and Ha respectively represent the average risks of getting the coronary heart disease, the cardiovascular disease, the cerebrovascular disease, and the hepatic cancer over ten years. And La is average risk of getting the lung cancer within one year. The output of the personalized health exam recommendation rules is the health exam recommendation plan suitable for the specific user, including recommended health exam items, item prices, as well as the total price of the health exam plan. Suppose there are totally n health exam items, denoted as Ei for i=1 to n, in m categories, data elements related to Ei include Ei.category, Ei.name, Ei.price, and Ei.check. Ei.category is the category name of the exam item. Ei.name is the name of the health exam item and Ei.price is the associated item price. Ei.check is a flag having values 0 or 1 to indicate whether the health exam item Ei is selected or not. L stands for the total list of all recommended health exam items, and P is the add-up total price of all recommended health exam items.

Considering the Tom's example, if the inputs for the personalized health exam recommendations rules are sex=M, age=55, hyptertension=Y, dm=N, sm=Y, P10=33%, Pa=16%. Then according to the chronic diseases and cancers recommendation rules, P10 is higher than Pa that results in the recommendation of health exam items: fasting lipid profile, chest x-ray, and ECG. Furthermore, by the fact that Tom is 55 years old and the periodic recommendation rules, ear nose throat exam, fecal occult blood, and UGI endoscopy are added to the recommended health exam list. Finally, accord-

Figure 5. Personalized health exam recommendation rules (partial rules)

```
// chronic diseases and cancers recommendation rules
If (P10 > Pa) then
    fasting_lipid_profile.check=1; chest_x-ray.check =1; ECG.check =1;
If (CV10 > CVa) then
    MRI.check=1; duplex_extracranial.check =1; duplex_Intracranial.check =1; fasting_lipid_profile.check =1;
If (V10 > Va) then
    fasting_lipid_profile.check =1; serology_test.check =1; ECG.check =1; 64_slices_Volume_CT_Scan.check =1;
    EEG.check =1
If (L > La) then
    chest_x-ray.check =1; CT_scan.check =1; Bronchoscopy_test.check =1; Screening_Spirometry.check =1;
If (H10 > Ha) then
    Angiography_of_abdomen.check =1; CT_scan.check =1; abdominal_ultrasound.check =1;
    radio_isotope_scan.check =1; liver.check = 1; bile_duct.check =1; pancreas.check =1;
If (B > 0 and fhx=Y) then
    Mammography.check =1;
//periodic recommendation rules
If (age>35) Then
    ear nose throat exam.check =1
If (age>=40) Then
    fecal occult blood.check =1;
If (sex=M and age>=45) then
    UGI endoscopy.check =1;
//preventive screening rules
If (((sex=M and age>35) or (sex=F and age>45) or (P10>20%)) and hypertension=N) then
    TC_test.check =1; HDL_test.check =1; triacylglycerol_test.check =1;
if ((hypertension=Y or dyslipidemia=Y or obesity=Y) and dm=N) then
    OGTT.check =1;
    fasting_blood_glucose.check =1;
if (BMI>24) then
    TC_test.check =1; HDL_test.check =1; triacylglycerol_test.check =1; OGTT.check =1;
    fasting_blood_glucose.check =1;
// outputs
For i=1 to n
If Ei.check=1 then
SET L = L + Ei.name, and P = P + Ei.price.
```

ing to the preventive screening rules, OGTT is also added to the list of health exam recommendation.

4 IMPLEMENTATION RESULTS

To develop the PH-IDSS prototype, object-oriented system development approach is adopted with the use of Unified Modelling Language (UML) as a modelling tool. Software and hardware environment for the prototype system include Windows 2000 professional as the operating system, Microsoft SQL server and Access as the DBMSs. The application functions and web pages are programmed using Java and Java Server Page (JSP) and are put in the Apache Tomcat servlet container. To access the system and functions, users only need a web browser.

Using the same Tom's case as an example, Figure 6 shows the risk factors and predicted risk level of getting coronary heart diseases in a table format. He can click on the button of "how to lower your risk"to get health tips or click the "show the bar chart"button to view the comparisons of his risk level with the ideal and average risk levels in a bar chart. In Figure 7, Tom's personalized health exam recommendation plan is presented with recommended exam items: fasting lipid profile, chest x-ray, ECG, fecal occult blood, UGI endoscopy, ear nose throat, and OGTT, and associated prices.

Figure 6. Risk prediction results

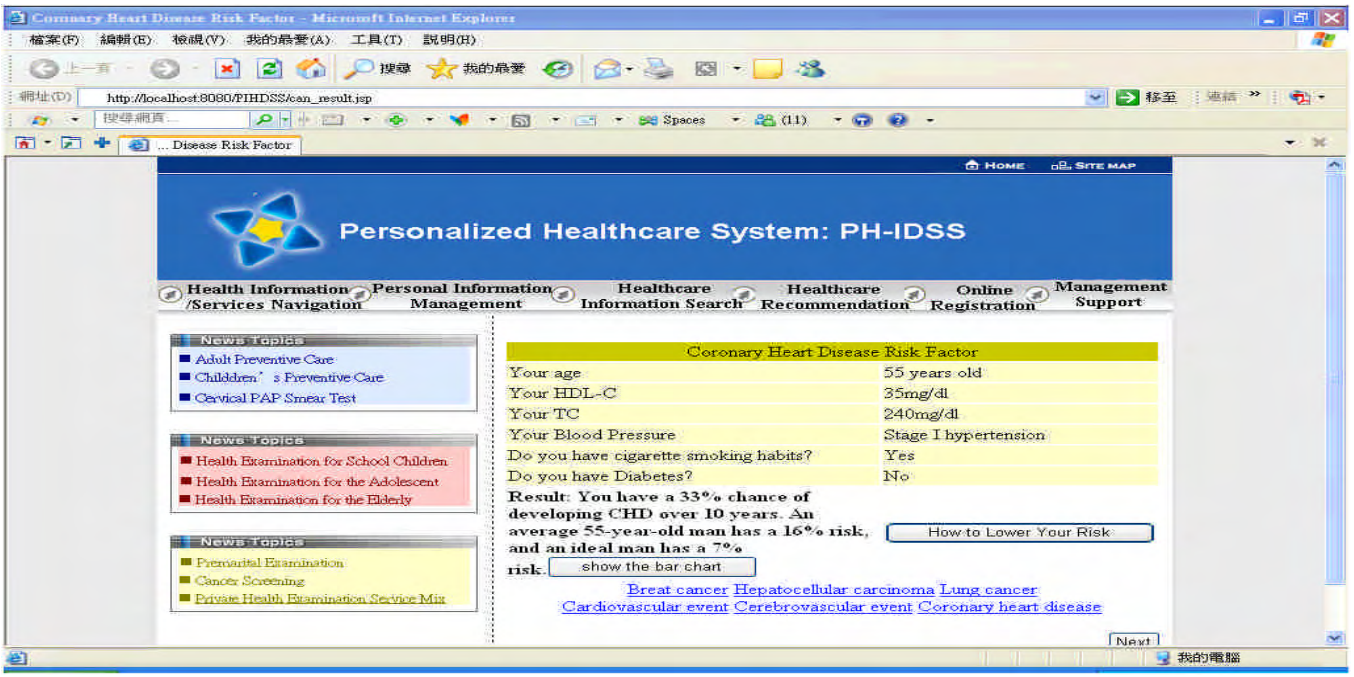
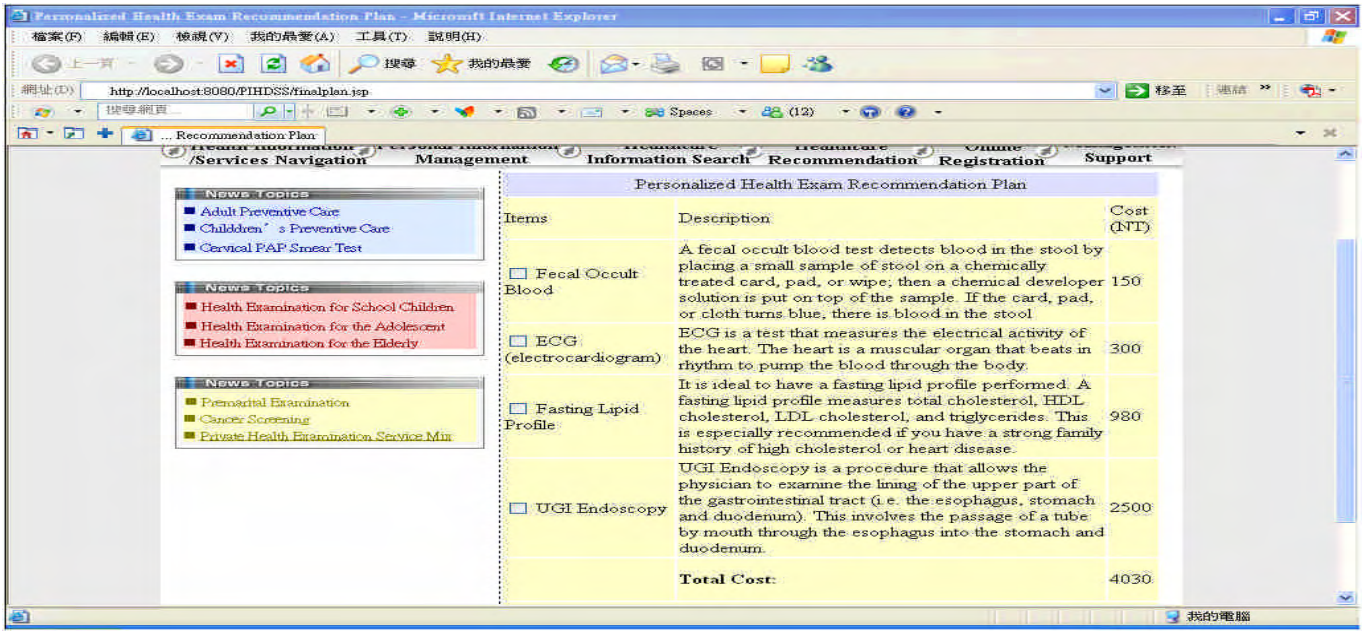


Figure 7. The health exam recommendation plan



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

5 CONCLUSIONS

In this paper, we propose the architecture and process of a web-based intelligent decision support system for personalized healthcare, and also develop a system prototype with illustrated examples. The PH-IDSS design framework, operating processes, and system prototype encompass personalized health check-ups recommendation, medical interpretation and advisory, as well as clinical care recommendation functions. Integrated data, models and rules for performing the personalized health exam recommendation process are described in detail with a scenario to show the feasibility and effectiveness of the proposed personalized healthcare recommendation and decision support approach. The main contribution of this paper is to provide an innovative and unified approach for linking all processes, data, models, and rules to support intelligent personalized healthcare. Future research works will focus on the validation of this personalized healthcare approach from the users and institutions perspectives, as well as on the evaluation of feasibility, importance, and cost/benefit effectiveness when adopting and implementing the PH-IDSS method.

REFERENCES

1. Hesse, B. W. and Shneiderman, B. (2007), 'eHealth research from the user's perspective', American Journal of Preventive Medicine, 32(5), 97-103.
2. Abidi, S. S. R. (2001), 'An intelligent tele-healthcare environment offering person-centric and wellness-maintenance services', Journal of Medical Systems, 25(3), 147-165.
3. Phillips, K. A. (2006), 'Personalized medicine for colorectal & breast cancer', Retrieved from the Pharmacogenetics and Pharmacogenomics KB website, http://www.pharmgkb.org/views/project.jsp?pId=52 (June 1, 2007).
4. Morpurgo, R., and Mussi, S. (2001), 'I-DSS: An intelligent diagnostic support system', Expert Systems, 18(1), 43-58.
5. Gómez, E. J., Cáceres, C., López, D., and Pozo, F. D. (2002), 'A web-based self-monitoring system for people living with HIV/AIDS', Computer Methods and Programs in Biomedicine, 69(1), 75-86.
6. Saadé, R. G., Tsoukas, A., and Tsoukas, G. (2004), 'Prototyping a decision support system in the clinical environment: Assessment of patients with osteoporosis OSTEODSS', Expert Systems with Applications, 27(3), 427-438.
7. German, J. B., and Watzke, H. J. (2004), 'Personalizing foods for health and delight', Comprehensive Reviews in Food Science and Food Safety, 3(4), 145-151.
8. Haux, R. (2006), 'Health information systems –past, present, future', International Journal of Medical Informatics, 75(3-4), 268-281.
9. Kypri, K., and McAnally, H. M. (2006), 'Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors', Preventive Medicine, 41(3-4), 761-766.
10. Bach, P. B., Kattan, M. W., Thornquist, M. D., Kris, M. G., Tate, R. C., Barnett, M. J., Hsieh, L. J., et al. (2003), 'Variations in lung cancer risk among smokers', Journal of the National Cancer Institute, 95(6), 470-478.
11. Rockhill, B., Byrne, C., Rosner, B., Louie, M. M., and Colditz, G. (2003), 'Breast cancer risk prediction with a log-incidence model: Evaluation of accuracy', Journal of Clinical Epidemiology, 56(9), 856-861.
12. Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., and Kannel, W. B. (1998), 'Prediction of coronary heart disease using risk factor categories', Circulation, 97(18), 1837-1847.
13. McCormack, J. P., Levine, M., and Rangno, R. E. (1997), 'Primary prevention of heart disease and stroke: A simplified approach to estimating risk of events and making drug treatment decisions', Canadian Medical Association Journal, 157(4), 422-428.
14. Yang, H. I., Lu, S. N., Liaw, Y. F., You, S. L., Sun, C. A., and Wang, L. Y., et al. (2002), 'Hepatitis B e antigen and the risk of hepatocellular carcinoma', The New England Journal of Medicine, 347(3), 168-174.
15. Dubey, V., Mathew, R., Iglar, K., Moineddin, R., and Glazier, R. (2006), 'Improving preventive service delivery at adult complete health check-ups: The preventive health evidence-based recommendation form (PERFORM) cluster randomized controlled trial', BMC Family Practice, 7, 44.
16. Lin, J. W., Chu, P. L., Liou, J. M., and Hwang, J. J. (2007), 'Applying a multiple screening program aided by a guideline-driven computerized decision support system - a pilot experience in Yun-lin, Taiwan', Journal of the Formosan Medical Association, 106(1), 58-68.
17. Ardissono, L., Leva, A. D., Petrone, G., Segnan, M., and Sonnessa, M. (2006), 'Adaptive medical workflow management for a context-dependent home healthcare assistance service', Electronic Notes in Theoretical Computer Science, 146(1), 59-68.
18. Yu, C. C. (2004), 'A web-based consumer-oriented intelligent decision support system for personalized e-services', Proceedings of the 6th International Conference on Electronic Commerce, ACM, 429-437.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

i·Society 2007

Towards collaborative user-centric healthcare services

Paul Zernicke
Carsten Wirth
Sahin Albayrak

DAI Labor of the Technical University of Berlin,
Secretary TEL14, Ernst-Reuter-Platz 7, D-10587 Berlin, Germany
sahin.albayrak@dai-labor.de

Abstract While being a very important field both from a social and economic perspective, the healthcare domain lacks supporting IT systems and services that really fit the process schemes used in this domain, with one very important one being collaboration. This paper describes our approach to healthcare services utilizing software agents which in turn employ user-centric collaboration. This approach is grounded on the properties of the healthcare domain as a whole and key prerequisites defined by related research. We introduce the use of software agents in our healthcare system SHA as to implement the domain goals by decomposition and collaborative recomposition. Using two scenarios we show that collaborative user-centric healthcare services fit the structure of the underlying domain naturally, allowing a comprehensive, integrated solution, increasing service quality and adaptability and providing collaboration support for human users, especially medical experts.

Keywords Collaboration, Service, Health, Healthcare, eHealth, User-Centric, Agent

1 INTRODUCTION

Society realises more and more that lack of exercise and malnutrition - especially in industrialized countries - are serious social problems which need to be antagonized. Increasing fees in the health insurance systems, demographic health care aspects such as typical diseases of an affluent society like diabetes or obesity and the need to relieve the growing occupational stress leads to a higher demand for solutions in the training and nutrition domain [1], [2].

The market adapted to this trend – apparel and equipment manufacturers for instance - by bringing up new innovative products addressing the aforementioned problems. One good example for this development is the combination of a wearable MP3-player with trainers, which have integrated sensors for collecting data while running to assist the exercising person with acoustic training advice [3].

This movement is opening new perspectives in computer science and software industry concerning the demand for new software systems and services which integrates and emphasizes these new upcoming ideas, concepts and products in the health domain accordingly.

Based on this background we wanted to develop a software system which provides services which cover most aspects of the health domain for an average user thus assisting him in his goal to live a healthier life like a personal trainer would do.

Using existing work done by Nealon and Moreno we derived some critical key requirements we think that ideally every system related to healthcare should provide, which are mutual syntactic and semantic understanding, user acceptance, privacy and openness. Additionally, for computer science applied to another domain with – like in our case - eHealth being a pretty good example, it is of utmost importance to always take the inner workings of the target domain into account.

Doing so, we discovered another important aspect for healthcare services: In the healthcare domain human experts have a dominant role in decision processes. They are needed, especially since they provide soft skills that are not easy to implement in computer systems, with knowledge as a prime example, but also things like experienced intuition are phenomena deeply rooted into the healthcare system, often on a psychological level.

Our approach is to use software agents as building blocks to develop electronic healthcare services that sport proactive, collaborative behaviour. Agents can support experts in making the right decisions. Taking into account that experts do not work alone, but also collaborate in many diverse fashions, and that we believe eHealth is user-centric, user collaboration support is just the next logical step – agents collaborating to support users in doing the same. Information filtering approaches like collaborative feature based information filtering can be naturally integrated into this scenario, with the SHA (Smart Health Assistant) collaboration approach doing exactly so.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

We will identify key requirements for this approach, present the state of the art in both the eHealth and the underlying healthcare domain and finally describe how we modelled and in some instances also implemented these concepts in our eHealth system called SHA. We will conclude with analyzing how our approach fares in respect to the requirements posted, a summary of our work and a brief outlook.

2 PROBLEM DESCRIPTION

As proposed in the introduction the diseases of an affluent society and the growing occupational stress is a problem which needs to be solved. Computer science can help to engage this problem if sufficient solutions can be developed. Due to complexity and interdependency of the healthcare domain a comprehensive and integrated solution is needed. Sadly, there is no such solution today, so our aim is to present a solution in the form of collaborative user-centric healthcare services.

We will first present generic key prerequisites for healthcare services as a guideline what requirements our services should fulfil. Nealon and Moreno researched the health domain from a rather generic perspective [4] and identified four aspects most relevant for electronic health services:

- The need for open standards for the representation and communication of medical knowledge
- Privacy and Security issues because of the very personal nature of medical data
- Acceptance by users and experts alike
- Interoperability of Platforms and servicesAlso, we define collaboration as an essential means to implement healthcare services that fulfil the aforementioned requirements. As the goals of collaboration in healthcare services are centered on the human users of the system and the collaboration processes evolve around – mostly medical – user data and preferences, we speak of user-centric collaboration. This decision will be motivated in chapter 3.

To evaluate our efforts, we will define two service usage scenarios in the healthcare domain that we consequently apply our approach of user-centric collaboration to:

The wellness scenario is about a user that wants to maintain his fitness by means of a healthy nutrition and training schedule. To do so, he needs instructions from an expert which are based on this expert's knowledge and skills.

The diagnostic scenario is about a user with health problems. These problems need to be recognized first, then communicated to the respective experts and finally be treated adequately.

3 STATE OF THE ART

In this chapter, we give an overview about the healthcare domain as a whole and related work on healthcare services.

3.1 Understanding the underlying domain

For the proposed electronic healthcare services to be usable in the broader healthcare domain, they have to fit the process patterns that are used within this underlying domain. To do that, we will approach the healthcare domain as a whole and postulate what we call the "prime principle" of healthcare. Then, we will identify its core process patterns and partaking entities.

So what about this all-encompassing task of healthcare? Quite straightforward, it is maintaining the human being in a physical and psychological state deemed normal by a common standard. Or to put it in simpler terms, keep people healthy [i]. This principle must be considered in every healthcare process, be the participants of human or technical nature – doctors even swear on it within the Hippocratic Oath. On a side note, the focus of this principle, the human health, also leads to consequences that are in their direct nature specific to the healthcare domain: Failing this principle, a healthcare process may impair the patient's health, which in turn is often neither substitutable with other goods nor reversible. A patient's death is the ultimate culmination of this fact, as it is very difficult to compensate and impossible to reverse on a general level. This may explain why there are especially strict considerations for healthcare processes in different domains, may they be juristic or regard user acceptance of said processes. One could argue that other sciences else than the medical one also strive to fulfil the mentioned "prime principle". Though this is correct, the critical distinction to other sciences is that every healthcare product is expected to have this driver as the product core. It is not a feature, it is the feature. These considerations also motivate the notion of user-centric collaboration, as both the goals and the data collaborated on are related to human users. The "prime principle" is to be understood as an abstract root of grounding for the goal decomposition hierarchy – it has no concrete properties by itself but is rather defined by the subgoals it is decomposed into.

Having internalized the "prime principle" of healthcare, we delved into the historical grown role of human experts in healthcare and the implications of today's technology for this very role. Historically, medical experts have always had a powerful position in human society. With the upcoming mechanisation of medical science in the 20th century, this position has been somehow impacted, as the technology and modern science used often allow a better verification of the expert's medical decisions. The trends range from medical schools such as evidence based medicine (EBM) [5] which promotes use of diagnosis and treating methods that have been independently checked by empirical studies, up to the vision of virtual software physicians.

This process is taken with some unease in parts of the medical community, with a common point of criticism being that the ultimate goal of it would be the gradual replacement of human experts by machines – remember the intelligent systems – or their incapacitation by scientific frameworks – an often-uttered resentment against the EBM.

Our opinion on this matter follows Norman [6] and others, who propose that every technology has one single primary driver: To be a tool for mankind. Consequently, we see the invention of technology into healthcare that culminated into electronic health using the same basic perspective. Therefore our mission statement for eHealth services is for them to support the medical experts and patients. Granted, there are areas where human participation in a healthcare process actually gets replaced by machines and software systems for good. But these areas must be selected wisely, considering the specific advantages of men and technology.

The second important lesson to learn from the healthcare domain is that collaboration is an imminent building block for healthcare processes. How so? One of the central process pattern encountered in healthcare is collaboration [7]. Collaboration happens between experts, users and healthcare services in any combination. Examples for this are the two service scenarios that we introduced in chapter 2. There also is a strong connection between collaboration and privacy aspects, as [8] suggests.

3.2 Healthcare services

In our knowledge most research and project works in the healthcare domain are either focusing on particular issues like nutrition, training or medical decision making individually or outlining it in a generic consideration. Regarding these individual aspects several different research activities and products can be found related to them such as BodyCap [9] and BodyForm Professional [10] in the nutrition domain, several Personal Trainer applications in the training domain and as examples for the medical domain the AADCARE system [11], the DILEMMA [12] project and as commercial example the BodyMedia [13] products.

In our opinion the healthcare domain lacks an integrated open solution which due to its extensibility is able to comprehensively cover all different aspects, such as Wellness, Fitness, Nutrition, Medical Expert Integration and Diagnosis among others. Our goal is to provide a software system related in the healthcare domain which can be used as fundament for such a high aiming solution.

3.3 Agent related research

The aforementioned Nealon and Moreno also note that "the usual properties of intelligent agents match quite precisely with … needs in [the healthcare] field (basically autonomous, intelligent, proactive, collaborative and distributed)" [4]. Concerning other key features of Multi Agent Systems (MAS[ii]) this software methodology is predestined for a realisation of an electronic healthcare application. This is also proven by the research work of Huang and Jennings [16] with the focus on agent technology in the health domain.

Thus, considering the proposed properties of agents and agent systems, based on the BDI-approach (Belief, Desire and Intention), it became clear to us that using the agent metaphor as the core of our further work is a good choice:

- Having desires, they strive to fulfil these by goal decomposition. Fulfilling a common goal is also a core feature in collaboration. This similarity allows agents to act as natural players in a collaboration process, much like humans do.
- Also, they have a set of Intentions that they can choose the most promising from to reach their desires. They are also able to decompose a given task into smaller work units. Thus, as collaboration depends on communication and coordination, agents can use their intentions to advance the collaboration process one step at a time.
- Due to their mentalistic stance, software agents are able to act pro-actively. Considering the coordinative side of collaboration, they can act on the received need for collaboration on their own, making a superseded control instance less important.Applied to the use of collaborative, proactive agents in healthcare, it becomes clear that they can be used to derive requirements for a collaborative service. For the agents to be able to collaborate, they need to have a common semantic understanding of the aim of their collaboration and about the information that is exchanged during the process.

When considering agents collaborating to reach a user-centric goal, the aforementioned privacy concerns become fully visible, as the collaborating agents need to access and exchange the users' preferences and profile data. The more agents involved, the bigger the threat of personal data emerging to entities that are not meant to know that data becomes. A prominent approach for this problem is to use a need-to-know policy in the collaboration context.

User acceptance is also connected to the collaborative approach: In the case of hard interdependencies between services, coordination becomes a must. The lack thereof leads to services that are not able to fulfil their given functionality, which will in turn lead to strict refusal by the user. Even services that are only loosely correlated to each other can profit from collaboration, as the result can become "more than the sum of its parts", sometimes leading to whole new service innovations. This aspect can be studied in the wide domain of the web 2.0, where service interconnection and also collaboration form one of the foundations for innovations that are being awarded by the users with strong interest.

4 COLLABORATIVE USER-CENTRIC HEALTH SERVICES

We will now describe our approach to agent-based healthcare services that use user-centric collaboration. First, we will give a short introduction into the SHA system that we employed our collaborative services in. Then, we will present the generic aspects of our collaboration model and apply them to the scenarios at hand.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 1. SHA End User Interface (Hi-Res Version) – Home Page

4.1 The SHA System

The SHA system consists of three components – the SHA framework, the SHA assistant services and the eHealth test bed.

The SHA framework is based on the general purpose JIAC (Java Intelligent Agent Componentware) agent framework. JIAC was chosen over other MAS frameworks as it has some distinct advantages: It uses the BDI approach deemed fitting for our purpose and there are many components readily available together with a comfortable toolchain that uses a graphical eclipse plugin for IDE purposes. Also, the JIAC framework was certified by the german BSI (Bundesamt für Sicherheit in der Informationstechnik) according to the CC (Common Criteria) [17] , so the security aspect that is very important in regard to healthcare services is also covered. The JIAC framework was extended with domain-specific functionalities needed for efficient development of eHealth applications. As the whole SHA project was focused on prophylactic services, the SHA framework is also endowed best in this area, but there are more generic functionalities nevertheless.

Apart from standard JIAC framework functionalities like service provisioning, persistence and multiple open interface support (e.g. web services), the SHA framework sports functions like

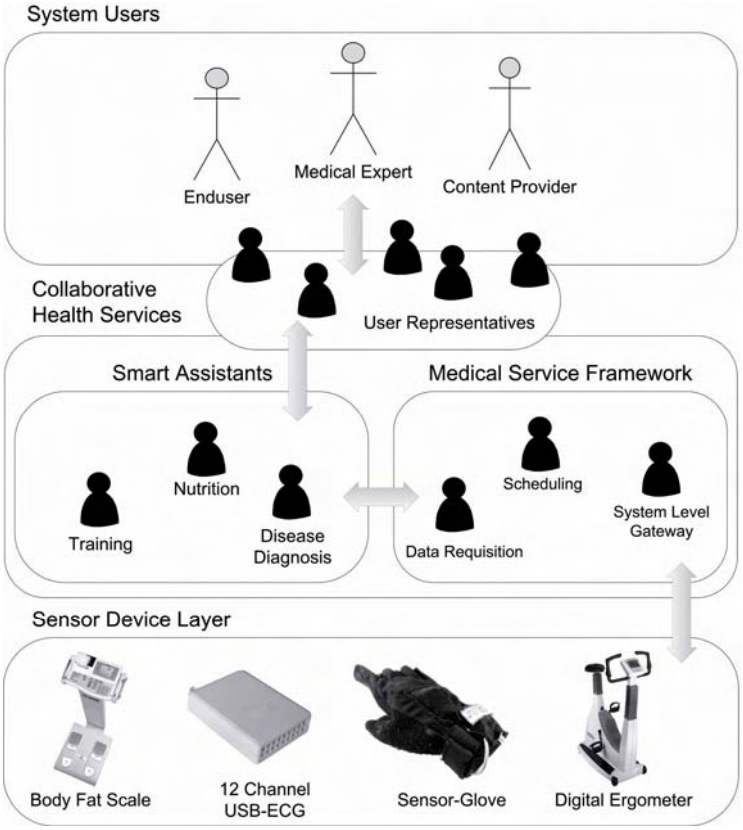
- An user representation agent model
- A powerful, multi-modal user interface system
- A term-scheduling engine based on a constraint server
- An Location-Based-Service (LBS) engine with integrated, dynamic Point-of-Interest routing (POI) functionality
- An abstract hardware interface model with a multi-layer meta-protocol stack
- An abstract medical feature model with an extensible algorithm libraryThe second component of the SHA system is the application services. These are grouped by their respective target audiences, namely end users, medical experts and knowledge providers, content providers and service providers. For each group, there is a portfolio of services tailored to the groups specific needs. In the course of this paper, we will concentrate on the end user services, which also had the development focus in the project SHA.

The core services in the end user domain are the nutrition assistant, the training planning and the training supervision assistants. Both of the planning assistants try to create a weekly schedule for the users' diet and physical exercises, respectively, by taking the users preferences, goals and other contextual data, like his PIM schedule, into account. There also is a disease diagnosis assistant, but due to the wellness-oriented approach we took with the SHA application services, the usable functionality of this assistant is more restricted than the others.

These high-level assistants are supported by a number of auxiliary agents, such as a data requisition agent that is aimed at concealing all data requisition and management functions from the user and other agents. Thus, the data requisition agent features functionalities from device and device capability management, measurement history and other sources, allowing other services, especially the user agent – or even the user himself – to simply request the type of data they need, optionally constrained by conditions, with the data requisition agent handling all necessary further steps like locating appropriate devices and telling the user how to perform the measurement.

Figure 1 gives an impression of the SHA user interface, while figure 2 illustrates the abstract system design.

Figure 2. SHA System model overview



4.2 Generic Collaboration Aspects

During our design of the SHA system and especially its collaborative aspects, we identified multiple patterns for user-centric collaboration schemes. We structured the collaborating agents into three distinct levels, set apart by the corresponding collaboration goals position in the underlying goal decomposition hierarchy. These levels also match the system design very closely, as can be studied in figure 2.

As already mentioned, the main goal of user-centric collaboration in healthcare services is keeping the patient healthy. This goal decomposes into two sub-goals, being comprised of prophylactic and rehabilitative aspects. The two proposed scenarios focus on these two sub-goals respectively.

The decomposition hierarchy continues to broaden further until it terminates with atomic goals. The levels themselves are ordered from high to low level.

User Representation Level – This level is situated directly at the system boundary towards the user. It houses the user representative agents (UA's). They represent the user him- or herself, and thus they have three distinct duties:

The first is to decompose the highest-level goals and to initiate collaboration processes with primary collaboration level agents either as instructed by their user, by other (expert) users or in consultation with the user, where the impetus for a non-user-initiated collaboration may also stem from primary collaboration level assistants.

The latter duties refer to the UA's role as an "intelligent" user profile, which is the data base for virtually every user-centric collaborative process. The data he provides to other agents should describe the user as exactly as possible. While this may be easy to achieve in regard to "hard" facts like physical properties, with "soft" properties like preferences it becomes much more difficult. Due to this reason we equipped the UA's with an information filtering technique that we will describe in Chapter 4.3. Additionally, he has to protect his users' privacy by employing a need-to-know principle when providing data to other agents.

Assistant Collaboration Level – This level comprises of the agents that can handle second-tier goals, such as the two prophylactic sub-goals nutrition and fitness planning. The agents on this level are addressed by the UA's right after they decomposed the superior goals, but they may also propose a collaboration themselves (if they see a specific need for it). The specific collaboration agents utilize services provided on the

Framework Collaboration Level – one example for the agents at this level is the scheduling agent that knows the users schedule and can create schedules containing the units requested by the planning agents. Agents at this level neither pursue the first nor the second tier goals, but more specialized ones like the scheduling agent: To maintain the user's schedule, to plan different activities and to interact with the user, trying to enforce him to keep the schedule.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4.3 Feature-Guided Collaborative Filtering

The initial motivation for the use of Feature-Guided Collaborative Filtering (FGCF) was to answer the question how a health system can be personalized such that the suggestions it offers are matching the users preferences. In many systems personalization is accomplished by applying information filtering techniques. Therefore we evaluated different information filtering mechanisms to find an appropriate for the health domain. As result we considered FGCF to be suitable for our system, which is based on a passive or active collaboration of the system users and the characteristics of the content objects within the domain [18]. Although the characteristics of content objects like a training unit or a meal can not be defined in a trivial way the existent limitations can be described as applicable attributes. A description of a meal by listing all of its ingredients will still lack of information about the taste for example but it enables a classification concerning allergies thus making it possible to be approved as valid for a certain user by a computer system with respective allergies information about the user. For training units this can be shown likewise considering impacts on body parts and illnesses.

Since the Collaborative Filtering is performed on user profiles by matching them to find preferences or generate recommendations we included agents in the SHA system adopting the user representative role. These agents not only encapsulate the user profile but provide several functionalities needed for a comprehensive health system.

To return to the before shown scenario of the generation of a training and nutrition plan the assistants can request recommendations from the user representative agents to generate their plans by not only regarding the calculation facts but also taking the likes and dislikes of the user into account.

Another important feature regarding the collaboration aspect is that they allow the users to communicate with each other with the help of a messaging service thus enabling an active collaboration between them by exchanging experiences with different training modules and meals. But unlike a simple message system the user representative agents are able to extend the messages with information important for the receiver as far as the message format is designed in a way allowing the agents to understand the context. To point out this approach we want to show another example:

User A is enthusiastic about a training module T and wants to share his experience with User B. To do so A sends a message with a reference to the module to B suggesting him to give it a try. The user representative agent of B receives the messages from user representative agent of A, extracts the information about the train module and can compare it with the user profile of B. If T is not suitable for B due to the fact that B can not perform this unit because of a specific illness, user representative agent of B can enrich the message with a warning.

4.4 Implementing the Wellness scenario

As mentioned in the Problem description, the wellness scenario takes a prophylactic approach.

The following parties partake in the wellness scenario:

User and user representative agent levels – The end user himself (EN-U) and his personal trainer (TX-U), both being represented by their respective user agents, EN-A and TX-A.

Assistant collaboration level – The nutrition and training planning assistants (NU-A, TR-A).

Framework collaboration level – The scheduling agent (SC-A) and the data requisition agent (DR-A).

In his constant pursuit of fulfilling the "prime directive" for the user, EN-A constantly checks for other agents that can help him decompose this goal. Collaborating with TX-A they realize that the user is overweight. After rechecking with their respective users and them allowing further action, they derive a goal decomposition into two sub-goals: Healthy diet and training schedule. TX-A calculates an adequate calorie delta between input and consumption. Looking for agents that can achieve the two sub-goals, they find NU-A and TR-A and provide them with the goals and the delta constraint. NU-A and TR-A in turn decompose the goal further, delegating simpler goals to framework agents such as SC-A or DR-A. The framework agents initiate collaboration processes by themselves as needed. All access to user specific data is performed through EN-U, who keeps control over the data. TR-A, for example, needs information about the user preferences in terms of training modules. This information is gathered on-the-fly by EN-U, who initiates a FCGF collaboration in the background. The finished schedules are proposed to the users, who may request changes, either by preference (EN-U) or by expert knowledge (TX-U).

During this process, TX-U discovers that TR-A has planned too much training units for the user to be effective. TX-A rechecks with EN-A and then requests TR-A to revise the schedule. TR-A does so but now, he realizes that the resulting schedule will not fulfil the calorie delta goal. Thus, he requests NU-A to plan a slightly lower calorie consumption. NU-A does that, and the calorie delta goal is met again. The current state of the schedules is presented to EN-A, who thinks that his user will be satisfied with the result (as calculated by the known preferences of EN-A). Before he presents the result to the user for final confirmation, he lets TX-A examine the schedule, who in turn rechecks with TX-U. All users and agents are satisfied with the result, and thus it is used as a prophylactic means of fulfilling the "prime directive" until changes occur. To keep the schedule consistent with reality, events that might possibly make the schedule sub-optimal trigger another evaluation round.

Implementing the Diagnostic Scenario - Bringing Collaboration support into play

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Having discussed the wellness scenario, we will now deal with the diagnostic chain scenario, which is based on the task of deriving abnormal medical states of the patient (EN-U) by using knowledge and skills scattered over a group of human experts and agents and subsequently finding appropriate means of handling the illness. Sub-goals are

- realizing the abnormal state
- efficient knowledge and information transmission between users and
- the optimal choice of fitting experts for each individual case.Our expert group is populated by the general practitioner GP-U, the radiology expert RA-U and the cardiology expert CA-U. They are represented by their respective user agents GP-A, RA-A and CA-A, where the EN-U is represented once again by EN-A.

The assistant collaboration level is populated by the disease diagnosis assistant (DD-A), while the framework collaboration level comprises of the knowledge database agent KD-A, the data requisition agent DR-A and the scheduling agent SC-A.

The KD-A facilitates information transmission between the user agents (and thus the users) by

- Attributing interactions with context-based information
- Translating information contained within the interactions to a form understandable by the target user
- Providing an extensive case databaseThe core collaboration process in this scenario is between users, not between users and agents nor between agents. After being notified by the DD-A about a abnormal rise in body temperature paired with circulation problems indicated by a low pulse, who in turn used the DR-A to get the data he deduced this information from, the expert group is trying to derive a correct diagnosis for the patient.

The collaboration on the assistant level is concerned with finding the matching expert for the users problem. The user is sent to his GP-U first, who then sends the user to the RA-U, who does an x-ray of the users torso. When the GP-U could not find any concluding evidence in the image, he sent the user to the CA-U, who in turn diagnosed him with a slight cardiac infection. Utilizing the help of KD-A, CA-U proposes a medical treatment after having diagnosed EN-U in person, which is in turn checked by CA-A for incompatibilities using profile data provided by EN-A.

The first sub-goal to fulfil is the selection of further diagnosis. As the GP-U decides that the user should get an x-ray (supported by the GP-A, who in turn collaborated with KD-A in this matter), virtually any radiologist will be able to do so. The next step becomes more difficult, as to decide for the GP-U where to send the user next can be quite a vital choice for successful diagnosis. At this point, he would profit from any information provided to him by the RA-U, apart from the radiograph itself.

The GP-A will try to locate matching experts (being radiologists in this example) and will provide this list to EN-A, who in turn will contact the RA-A to set up a meeting with both agents being supported by the SC-A.

As soon as the actual x-ray has been taken, the radiologist will check the image and comment on his findings. The information is passed back to EN-A, who notifies GP-A (and in turn, the GP-U himself). GP-U checks the image and also the radiologists' comments. This information has been translated before into the GP-U's knowledge domain. The translation process can be as simple as to add generic medical definitions of notions specific to radiology. If the EN-U himself is interested in RA-U's findings, the same translation logic can be applied by EN-A. The translation itself is performed by KD-A.

As the GP-U is not really sure what to do next, he can use the agent services once again, being the KD-A case database.

The important thing to notice at this point is that the system is not aiming at patronizing the GP-U, but to support his medical decision. So, if we consider the scenario, where GP-U has a 'soft' intuition about the problem to be of cardiac nature, the KD-A would assist him with evaluating this, by providing him with similar cases. Additionally, the users' current state of diagnosis is transferred to CA-A (and in turn to CA-U), so CU is able to assert on the case at hand. As the information is not sufficient for CA-U to derive an ultimate decision, he instructs CA-A to set up a meeting with EN-U, with this process being facilitated by EN-A and SC-A again.

4.6 Model Implementation by SHA

While the current SHA implementation does not yet provide the functionality to implement the aforementioned scenarios fully (especially the diagnostic chain scenario, since the SHA focus lies within the wellness domain), the basic functionality is still fully working. SHA as a system facilitates collaboration by using a powerful agent framework together with an abstract knowledge representation and evaluation model. It also features user agents that employ FGCF, the constraint-server based scheduling engine and fully operable nutrition and training planning and supervision assistants.

4.7 Evaluating the Solution

Having presented our agent-based user-centric collaboration model, both in generic nature and applied to the two scenarios at hand, we already showed how the involved users and experts would profit from the collaborating user-centric agents, enabling both increased service quality (wellness scenario) and the support of healthcare collaboration processes (diagnostic chain scenario).

We will conclude this chapter with an evaluation if services employing this model will be able to fulfil the four key requirements stated in the problem description.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Open Standards for knowledge representation and communication

Agents commonly use ontologies as grounding for both knowledge representation and communicating with each other. In medical sciences, serious efforts are being invested into collecting the medical knowledge in ontologies, one very prominent example being the US National Library of Medicine (US-NLM) "Unified Medical Language System" (UMLS) [19], consisting of the Semantic Network ,the Meta-Thesaurus and the SPECIALIST Lexicon. With JIAC agents being FIPA (Foundation for Intelligent Physical Agents) conformant, they allow direct and indirect support for a variety of ontology languages. Thus, the first requirement is met by open standard agent systems in general and also by the collaborating SHA agents.

Privacy and Security issues because of the very personal nature of medical data

The access to all personal data is routed through the user representative agents in our approach. Thus, user representatives employing a need-to-know approach when providing data to other agents will be able to ensure the users privacy. To prevent security issues, the agent system itself must feature a sturdy design, a requirement met by the JIAC agent platform.

Acceptance by users and experts alike

User acceptance is governed by many aspects, such as usability, service quality and others. Collaborative healthcare services can bring in their fair share of acceptance by users and experts, as they can provide adaptive behaviour. The specific role of expert and user collaboration in healthcare also indicates that collaborative services that try to support users and experts in what they are doing will be much more acceptable than solutions taking a rather exclusive, substitutive approach.

Interoperability of Platforms and services

The interoperability is a rather technical requirement, which is a natural requirement for collaboration also. Thus, the SHA system design and implementation heavily rely on interoperability as it insures encompassing collaboration processes. As mentioned in chapter 4.1, SHA features multiple open standard interfaces and gateways, such as web services.

5 SUMMARY

In our paper, we presented an approach to healthcare services that is able to solve the problems within the healthcare domain posted before. We showed that agent-based collaborative user-centric healthcare services fulfil the key requirements for healthcare services postulated by Nealon and Moreno. Additionally, the collaborative nature of such services allows them to support the healthcare processes in a natural way, as they are user-centric and collaborative too. When implementing healthcare services, the abstract "prime principle" of healthcare needs to be decomposed into more concrete goals, which is in fact one of the key features of intelligent agents. To re-merge the individual solutions, collaboration is needed and thus agent-based collaboration provides an adequate model for this process pattern and also a means of maintaining this very principle in reality. The system model and its application to the selected scenarios illustrate the strengths of this approach.

Both users and experts profit from this approach, as users are provided with integrated, intelligent services that will support their need for both prophylactic and rehabilitative, personalized solutions for healthcare, while still keeping their personal profiles as private as possible. Experts on the other hand need services that efficiently improve the quality of their own services provided to patients, without being patronized by technical systems.

These considerations culminate in the idea of collaborating agents supporting collaborating experts and users. This very model shows the true potential of our approach: To create a service environment where humans and machines both provide their individual skills and capabilities in harmonic co-existence, united under one common goal: To keep the people healthy.

REFERENCES

1. FOCUS (2005), 'Der Markt für Fitness und Wellness - Daten, Fakten, Trends' , IHK web site, German Industrie- und Handelskammer; http://www.dienstleister-info.ihk.de/branchen/Fitnesswirtschaft/Merkblaetter/05fitness.pdf (31/07/07)
2. International Obesity Task Force (2005), 'EU platform on diet, physical activity and health - EU platform briefing paper', European Communities web site, European Communities; http://ec.europa.eu/health/ph_determinants/life_style/nutrition/documents/iotf_en.pdf (31/07/07)
3. Apple Computers and Nike (2007), 'Nike + Ipod', german Apple homepage, Apple Computers and Nike, http://www.apple.com/de/ipod/nike/ (31/07/07)
4. J. Nealon and A. Moreno (2002), 'The application of agent technology to healthcare', in Proc. 1st International Joint Conference on Autonomous Agents and Multiagent Systems, Bologna, July 2002
5. D. Sackett, W. Rosenberg, J. Gray, R. Haynes, W. Richardson (1996), 'Evidence based medicine: what it is and what it isn't', BMJ, 312(7023):71-2.
6. D. Norman (1994), 'Things that make us smart ,1st ed.', Addison Wesley Publishing Company,1994
7. V. Patel, K. Cytryn, E. Shortliffe and C Safran (1999), 'The collaborative health care team: the role of individual and group expertise', Teaching and Learning in Medicine, 1999
8. G. Wiederhold (2001), 'Collaboration requirements: a point of failure in protecting information', Computer Graphics and Applications, Vol. 31, No. 4, July 2001, 336-342
9. Trebaxa Company, 'BodyCap extended', Trebaxa BodyCap homepage, Trebaxa Company, http://www.bodycap.de/ (31/07/07)
10. Alexander Ruhmann (2007), ,' BodyForm professional', Body-Soft homepage, Alexander Ruhmann, http://www.body-soft.de/
11. J. Huang , N. Jennings and J. Fox (1995), 'Agent based approach to health care management', Applied Artificial Intelligence, 9, 401- 420.
12. EHTO Enterprise (1997), 'DILEMMA', EHTO homepage, EHTO Enterprise, http://www.ehto.org/aim/volume2/dilemma.html (31/07/07)

13. Bodymedia Inc. (2007), 'BodyMedia – solutions for metabolic disorders', Bodymedia homepage, Bodymedia Inc., http://www.bodymedia.com (31/07/07)

14. S. Albayrak (1998), 'Introduction to agent oriented technology for telecommunications', Intelligent Agents for Telecommunications Applications, pages 1–18, Basel

15. R. Cissée, A. Rieger, N. Braun, and S. Albayrak. 'An agent-based framework for secure and privacy-preserving personalized information service', Proceedings of 1st International Symposium on Wireless Communication Systems, ISWCS 2004, Mauritius, 20-22

16. J. Huang, N. Jennings and J. Fox (1994), 'Cooperation in distributed medical care', in Proc. 2nd International Conference on Cooperative Information Systems, Toronto University Press, Toronto, 255-263

17. Tim Geissler, Olaf Kroll-Peters (2004), 'Applying security standards to multi agent, Proceedings of the First International Workshop on Safety and Security in Multiagent Systems (Sasamas'04 at AAMAS'04), New York

18. M. Runte (2000), 'Personalisierung im Internet - Individualisierte Angebote mit Collaborative Filtering', personal PhD thesis homepage , M. Runte, http://www.runte.de/matthias/publications/personalisierung_im_internet.pdf (31/07/07)

19. United States National Library of Medicine (2007), 'Unified Medical Language System ', US-NLM homepage, United States National Library of Medicine, http://www.nlm.nih.gov/research/umls/umlsmain.html

ENDNOTES

i As mentioned in the introduction, we restrict our notion of healthcare to human medicine

ii See [14] for a introduction to MAS and [15] for the features of a MAS System like i.e. Scalability, Extensibility

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e•Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Multimedia evaluation: understanding the user-needs gap

Olatubosun Olubusuyi Ojo

1 INTRODUCTION

This essay focuses on the elements / factors that need to be considered when evaluating the 'fitness for purpose' of a multimedia application with emphasis on user requirements, business objectives, metric selection and user-needs gap.

Multimedia as the name implies is the combination or blending of different forms of media as against using a single component of media.

Tony Feldman, a multimedia consultant defined multimedia as "the seamless integration of text, sound, images of all kinds and control software within a single digital information environment" England, E., & Finney, D. (1999).

The term 'fitness for purpose' of a multimedia application can be thought of as synonymous with the word usability but in actual fact it is a much broader term as it encompasses other factors such as functionality, navigability, reliability and efficiency.

Let it be stated clearly from the onset that what is considered simple and exciting to one user may be regarded as complex and clumsy for another user, therefore for obvious reasons when planning a multimedia application it is essential to have the needs of the potential users of the product in mind.

Therefore in this essay I will focus on user requirements, business objectives, metrics selection and other considerations such as user-needs gap which is responsible for certain multimedia applications being attractive to some users while repelling other users.

Knowledge of this gap will be useful to application developers if success in a project is to be achieved.

1.1 MULTIMEDIA APPLICATION EVALUATION

The main purpose of Multimedia application evaluation is to match a prescribed set of quantifiable criteria against performance to find errors or faults and to seek ways of improving the design during development or post-development of the application.

This evaluation can be thought of in simplistic terms as a form of quality assurance since its aim is to generate confidence and trust among its users (clients) and other stakeholders that the multimedia application meets needs, expectations and other requirements.

There has been a rich combination of disciplines coming together in multimedia which implies that various methods can be adopted in evaluation also the "multimedia interface depends on usability factors and aesthetic judgement to provide the application with an interface that is well suited for its technical and business objectives" (Martin, S., Bolissian J., & Pimendis, E. 2003), all these implies that for a robust evaluation there is need to include both quantitative and qualitative analysis.

Important factors to be considered when evaluating the fitness for purpose or quality of a multimedia application are usability, functionality, navigability, reliability, efficiency or consistency. The list is not exhaustive because it depends on the background and specialisation of the evaluator, for a comprehensive and robust evaluation it is good practise to involve multiple evaluators and then find the aggregate of their various evaluations.

2 MULTIMEDIA EVALUATION CRITERIA & METRICS

Multimedia applications differ in what they are expected to achieve therefore evaluation of each application requires a knowledge of what the application is expected to deliver.

The evaluation carried out against a prescribed set of quantifiable criteria is objective evaluation while subjective evaluation is based on observation and analysis of non-quantifiable factors and is influenced by the experience and preference of the evaluator.

The final decision on which method to use will have to be made against the background of which one will well serve the project's quality constraints, time, cost and requirements.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Generally some factors have been recognised as recurring on most quantitative evaluations carried out by WebQEM and CIDOC Multimedia Working Group and they are:

- Efficiency
- Functionality
- Navigability
- Reliability
- Usability

Metrics Selection: For each of the factors listed to be evaluated a metric system is selected for example it can be graded on a scale of 1 to 4, where 1 is the lowest performance rating while 4 is the highest performance rating.

1	2	3	4
Unsatisfactory	Average	Very Good	Excellent

These are all added up depending on which formula is deemed suitable to get an idea of the fitness for purpose of the multimedia application. Using only this quantitative approach is not sufficient to obtain a robust picture I will suggest also applying qualitative analysis in addition irrespective of whatever additional costs it may incur.

2.1 Efficiency

When evaluating the fitness for purpose of multimedia products there are some quantifiable qualities that will need to be working properly for the product to be considered efficient. This can be assessed by answering some basic questions about the product, the same standard format must be used in evaluating the other factors listed above to create harmony and consistency in the evaluation.

Questions:

2.1.1 Will the application work across platforms?
2.1.1.1 Will it work with different browsers?
2.1.1.2 Which platform provides the best quality?
2.1.1.3 Which platform gives the worst quality and why?
2.1.2 Is the content migratable?
2.1.2.1 Can it be archived and re-used?
2.1.2.2 Is the system process independent of machine?
2.1.3 How easy is installation?
2.1.3.1 Does it alter existing system parameters during installation?
2.1.3.2 Is it clear about parameters altered during installation?
2.1.3.3 Is there an uninstall program?
2.1.4 Can you extend the architecture?
2.1.4.1 Is the meta-design explicit?
2.1.4.2 Can you upgrade the application?
2.1.5 Do you need a special environment to run the application?
2.1.5.1 Does it require special configurations?
2.1.5.2 Does it require special Plug-ins to work?

For each question a grade (from 1-4) is awarded depending on performance.

2.2 Functionality:

Questions:

2.2.1 Does the system incorporate tools or methods which enable interaction with contents?
2.2.1.1 How many of such tools are provided?
2.2.1.2 What kinds of tools are they?
2.2.1.3 Are the tools relevant to the nature of the content?
2.2.1.4 Are the tools appropriate and engaging?
2.2.2 Is a user search engine provided?
2.2.2.1 Do the search criteria match the user criteria?
2.2.2.2 Does it give accurate results even when some values are missing?
2.2.2.3 Does it give a wide range of results?
2.2.3 Any descriptions of the systems functions?
2.2.3.1 Are the descriptions clear and consistent?
2.2.3.2 Are all the functions described?
2.2.4 Is the system Interactive?
2.2.4.1 Does it respond to user input and give appropriate output?
2.2.4.2 Does it respond by giving only pre-defined choices?
2.2.4.3 Does it provide facilities for users to save and build upon?
2.2.5 Does the application provide a print facility?
2.2.6 Is there a feedback facility for users to comment?
2.2.7 Is there a metrics facility to record user activity?
2.2.8 Is it possible to update its content?
2.2.9 Is the multimedia application easy to access?
2.2.9.1 Is it available in the indexes of popular search engines?
2.2.9.2 How is the fixed format multimedia distributed?

For each question a grade (from 1-4) is awarded depending on performance.

2.3 Navigability

Questions:

2.3.1 Are the navigation paths easy to use?
2.3.1.1 Are they structured consistently?
2.3.1.2 Do they respond to user requests appropriately?
2.3.2 Is there a navigation guide?
2.3.2.1 Does it guide the user to desired content?
2.3.2.2 Is it accurate and sufficient?
2.3.3 Are the visual symbols obvious?
2.3.3.1 Are the icons easy to understand?
2.3.3.2 Is there an exit icon?
2.3.4 Is the structure of the system obvious?
2.3.4.1 Is it consistent?
2.3.4.2 Is it intuitive and relaxing?
2.3.5 Is the complete design appropriate for use?
2.3.5.1 Are the colours appropriate?
2.3.5.2 Are the fonts appropriate?
2.3.5.3 Are the screen shape and outline appropriate?
2.3.6 Can the user control the multimedia presentation?
2.3.6.1 Can the user alter the size of the window?
2.3.6.2 Can the user alter the layout of the screen?
2.3.6.3 Can users move around the presentation?
2.3.7 Can users resume from where they stopped?
2.3.7.1 After exiting can users resume from exit point?

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

For each question a grade (from 1-4) is awarded depending on performance.

2.4 Reliability

2.4.1 Does the application display properly?
2.4.1.1 Are there broken links?
2.4.1.2 Are there dead-ends?
2.4.2 Is the application engaging?
2.4.2.1 Does it provide a unique and rewarding experience?
2.4.2.2 Does it fulfil its original purpose?
2.4.2.3 Is it easily available?
2.4.2.4 Is contact information provided on the CD-ROM for inquiries?
2.4.3 Is the application appropriately priced?
2.4.3.1 Compared with competitors is it a fairly priced?
2.4.3.2 Is there anything that can be considered as added value?

For each question a grade (from 1-4) is awarded depending on performance.

2.5 Usability

Usability as a factor of fitness for purpose has evolved into a growing discipline. Usability principles are also known as heuristics and heuristics evaluation involves making sure that the user interface is compliant to predetermined heuristics.

It is good practice to involve multiple evaluators because a single individual cannot find all the usability problems in an interface.

It is possible to perform heuristics evaluation of user interfaces that exist at the planning stage before reaching implementation which is an advantage.

Nielsen & Molich (1990) described nine basic heuristics that can be used to evaluate the usability of a multimedia application and they are:
2.5.1 The use of simple and natural language
2.5.1.1 Is the language simple and natural?
2.5.2 Speak the user's language
2.5.2.1 Are there various language versions for diverse users?
2.5.3 Minimize user memory load
2.5.3.1 What is the format used for presentation?
2.5.4 Consistency
2.5.4.1 Is there consistency in the application?
2.5.5 Provide feedback
2.5.5.1 Is there a feedback facility?
2.5.6 Provide clearly marked exits
2.5.6.1 Are there clearly marked exits?
2.5.7 Provide shortcuts
2.5.7.1 Are there shortcut links?
2.5.8 Provide good error messages.
2.5.9 Prevent errors.

All these are performance measures and do not constitute the only means by which usability can be measured. Usability can also be measured through subjective means. Subjective means refers to people's perception, opinions and judgements. The bulk of research on usability has tilted heavily on performance measures to the detriment of subjective analysis; this may be an explanation for the success of certain multimedia applications which continue to attract users in the millions while disobeying all the rules of conventional usability theories.

For this reason I believe there a need to merge the two types of usability evaluation (performance and subjective) and also to include some elements of fitts' law into determining the fitness for purpose of multimedia applications and finding ways to seek improvements.

2.6 Usability Analysis

Usability analysis of a multimedia application will be an exercise in futility if it is not user-centred. It is necessary to analyse taking into consideration the needs of users with or without disabilities. "The type of data collected can be either quantitative where specific performance is measured while the second is qualitative analysis which is based on non-quantifiable personal factors". England, E., & Finney, D. (1999)

Experts in usability argue that the choice of which method to use should be based on business objectives and budget but I think otherwise and feel that the two should be incorporated together to get a more robust result.

2.6.1 Performance Measures: These are also called objective measures and are all a prescribed set of quantifiable criteria such as time taken to finish a task, time taken in navigating menus or time taken to recover from errors.

They also include any signs of frustration or expressions of satisfaction which must be documented within a time frame for analysis.

2.6.2 Subjective Means: Subjective means refers to people's perception, opinions and judgements, I consider this to be as important as performance measures because of the promotional slogan "The customer is king" used in many adverts worldwide.

No matter how usable an application is if it is not specific user-centred with a direct appeal to its preferred target audience then it can be considered as having failed in its primary objective.

Dumas, J.S., & Redish, J.C. (1999) listed examples of subjective measures in usability tests as in table below:

To dismiss the qualitative analysis as an unserious component of multimedia application evaluation will amount to a serious mistake as research and reality has shown that it is the user that dictates the pace in multimedia success.

Although most applications are designed with clear business objectives it is very important to incorporate user require-

Ratings	Reasons for Preferences	Predictions of Behaviour & Reasons	Spontaneous comments
Ease of learning the application	Over a previous version	Would you buy this application?	"I'm totally lost here"
Ease of using the application	Over a rival application	Would you pay extra for the manual?	"That was easy"
Ease of doing a task	Over the way tasks are done now	How much would you pay for the application?	"I'd call tech support now"
Ease of Installation			"I don't understand this message"
Relevance of the online help			"Whoa very nice"
Ease of finding information in the manual			
Ease of understanding the information			

ments into planning and design even though balancing these goals may be very challenging.

One way of meeting this challenge is through branding whereby users already have a positive impression about an application or product.

Branding makes businesses stand out in the crowd and it not limited to big organisations, small and medium enterprises can also build and manage brands as it is all about focusing on end users and making sure that their requirements are met.

Jacob Nielsen a usability expert in a recent interview "usability makes business sense" admitted that there are some multimedia websites that break all the rules of usability yet remain successful in attracting users in the millions why?

I believe the answer lies in understanding the user-needs gap.

3 USER-NEEDS GAP

We cannot afford to ignore the importance of the study of human behaviour and its interaction with machines and applications in general which includes multimedia applications.

Knowledge of human-computer interaction (HCI) is imperative if we are to understand the dynamics of multimedia applications and how it can be developed and evaluated to meet the specific needs of users.

Current research based upon the fundamental work of Paul Fitts in 1954 and 1964 exists which allows us to predict the time required for computer users to move from a starting or rest position to a final target area.

In summary, Fitts law predicts the time it takes to point at a target based on the size and distance of the target object and is very useful for accurate design of time dependent applications. Its major limitation is that it predicts the time for movement in one direction only.

Hick's law is a major improvement and refinement of Fitt's law as it enables the prediction of time taken for people to make a decision while on a user interface for example clicking a tool bar, choosing from a menu list etc.

Hick's law states the time taken (T) to make a decision is:

$$T = gH \,,$$

Where g is a constant (approximately 150milliseconds)

H is the information-theoretic entropy of a decision,

$$H = \sum_{i}^{n} p_i \log_2 \left(1/p_i + 1\right)$$

n – number of alternatives

p_i – probability of alternative i for n alternatives of unequal probability.

All these tools allow multimedia applications to be designed to target users based on how much time we predict they can afford to spend without get bored or agitated.

This particular field is one of the most ignored areas in multimedia evaluation considerations yet its value in improving usability has been tested and found to be reliable and accurate, why it remains largely ignored as an evaluation tool seems curious to me.

3.1 SUMMARY

Understanding user-needs gap is important if we are ever going to be able to resolve the issues behind why certain 'unusable' multimedia applications succeed in the real world while a few usable applications fail in achieving its business objectives.

"If you're making a choice between user and business goals, you're making the wrong choice. Period" Robinson, D.K (2004)

The users are the main focus and without them there cannot even be a business plan therefore a comprehensive psy-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

choanalysis of potential users with the aid of mathematical models of expected fine motor control is necessary if we are to develop an application that will be fit for purpose while also achieving business objectives.

4 CONCLUSION

Bringing to bear the importance of user-centred design in multimedia applications this paper has focused on user requirements, business objectives to a lesser extent, metrics selection and user-needs gap.

Multimedia evaluation is aimed at improving the quality of the product thereby building confidence among clients and other stakeholders.

The need to integrate both quantitative and qualitative analysis as a hybrid system for a more robust evaluation instead of using just one method has been argued.

The elements / factors selected for consideration when evaluating the fitness for purpose of a multimedia application are:

- Efficiency
- Functionality
- Navigability
- Reliability
- Usability

The metric system advocated in this paper is a system of grading from 1 to 4 which translates to unsatisfactory (1), average (2), very-good (3) and excellent (4).

Nine basic usability heuristics as postulated by Nielsen & Molich (1990) was reaffirmed as still relevant while the paper ends with a need to attach more importance to the research on the user-needs gap because of its accuracy and the potential of removing evaluation by trial and error.

REFERENCES

CIDOC Multimedia Working Group, Multimedia Evaluation Criteria (1997)

http://www.archimuse.com/papers/cidoc Revised Draft, Nuremberg, Germany September 8-10, 1997 accessed 16/04/07

Dumas, J.S., & Redish, J.C. (1999), 'A Practical Guide to Usability Testing' Revised Edition, Intellect. ISBN 1841500208

England, E., & Finney, D. (1999), 'Managing Multimedia: Project management for Interactive Media' 2nd Ed, Addison-Wesley. ISBN 0201360586

Foraker Design (2002-2006), "Introduction to Usability" http://www.usabilityfirst.com accessed 30/04/07

Fosythe, C., Grose, E., & Ratner, J. (1998) 'Human Factors and Web Development' Lawrence Earlbaum, Mahwah, NJ

Hartwig, R., Darolti, C., Herczeg, M. (2003) "Lightweight Usability Engineering Scaling Usability-Evaluation to a Minimum?" http://www.imis.uni-luebeck.de accessed 02/05/07

I. Scott Mackenzie (1992). "Fitts' law as a research and design tool in human-computer interaction" Human-Computer Interaction, volume 7, 1992, pp.91-139

Krug, S. (2000), 'Don't Make Me Think: A Common Sense Approach to Web Usability' QUE: Indianapolis, IN, ISBN: 0789723107

Martin, S., Bolissian, J., & Pimenidis, E. (2003), 'PURE and SIMPLE: a framework for the evaluation of business multimedia products' 10th European Conference on IT Evaluation, Madrid, September 2003

McKerrow, P.J., (2005) "Teaching Content Creation with Programming", IEEE Multimedia 2005, pp36-38

Mendes, E., Mosley, N., & Counsell, S. (2003), "Web Metrics: Estimating Design and Authoring Effort", IEEE Multimedia 2001, pp50-54

M.Hurst, "Dot Com Survival Guide", 2000,

http://www.CreativeGood.com/Survival accessed 28/04/07

Mich, L., Franch, M., & Gaio, L. (2003), "Evaluating and Designing Website Quality", IEEE Multimedia, Vol10, Jan-March, pp34-40

Olsina, L., Rossi, G. (2002) "Measuring Web Application Quality with WebQEM", IEEE Multimedia, Vol.9, Oct-Dec, pp20-25

Pearrow, M. (2000) 'Web Site Usability Handbook' Charles River Media: Rockland, MA

Robinson, D.K (2004) Finding the Sweet Spot

Bridging the Gap between User and Business Goals

http://www.digital-web.com/articles/finding_the_sweet_spot/

Accessed 15/04/07

BIBLIOGRAPHY

Jordan, P. (1998) 'An Introduction to Usability' Taylor & Francis, ISBN 0748407626. A book for beginners in the study of usability covering usability requirements, general principles of design, requirements gathering methods and procedures for conducting usability evaluations.

Nielsen, J. (2005). www.useit.com (accessed 27/04/07)

Jacob Nielsen is a renowned expert on usability and this site includes some articles, information and interviews that can broaden a researcher's knowledge in the field.

Perfetti, C. (2004) "Product Usability: Survival Techniques" User Interface Engineering, User Interface Conference October 11-14, 2004

Booklet providing techniques for developers trying to create usable products with small budgets or tight schedules.

Rivers, H. (2004) "Deconstructing Web Applications: Learning from the Best Designs" Two Rivers Consulting, User Interface Conference October 11-14, 2004. A comprehensive analysis and deconstruction of well designed and highly usable websites like Amazon, Wal-mart, eBay, Hotmail, Match.com, Fidelity.com and ways of improving usability. Chapter 7 of the book deals with Flash and some deconstructions of applications produced with it.

Schaffer, E. (2004) "A Guide to the Institutionalization of Usability" Human Factors International, User Interface Conference October 11-14, 2004

A manual or guide that provides an outline for making usability an integral part of a design process in any organisation.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

i·Society 2007

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Intelligent data classification techniques

T. Shatovska, V. Repka, A. Kharchenko

Department of Computer Science,
Kharkiv National University of Radioelectronics, Ukraine

Abstract The intelligent data analysis (IDA) system is based on data mining, machine learning and data visualisation package. The IDA provides comprehensive analysis of row data by partitioning, hierarchical, density-based methods and assembles of classifiers. Furthermore, the IDA consists of modified clustering algorithm Chameleon, that can be used for different shapes with non equal densities, multiple graphical data representation, 2D/3D rotatable plots and dendrograms. The paper focuses on the comparison of accuracy and finding methods for efficient and effective cluster analysis for complex shapes.

Keywords partitioning, hierarchical, density-based clustering, classification, Chameleon algorithm, shapes, densities.

1 INTRODUCTION

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group in many applications (see Figures 1 and 2). The cluster analysis has been widely used in numerous applications, including pattern recognition, data analysis, image processing, and market research [1]. By clustering, one can identify dense and sparse regions and, therefore, discover overall distribution patterns and interesting correlations among data attributes. Data clustering is under vigorous development. As a branch of statistics, cluster analysis has been studied extensively for many years, focusing mainly on distance-based cluster analysis.

2 IDA PACKAGE

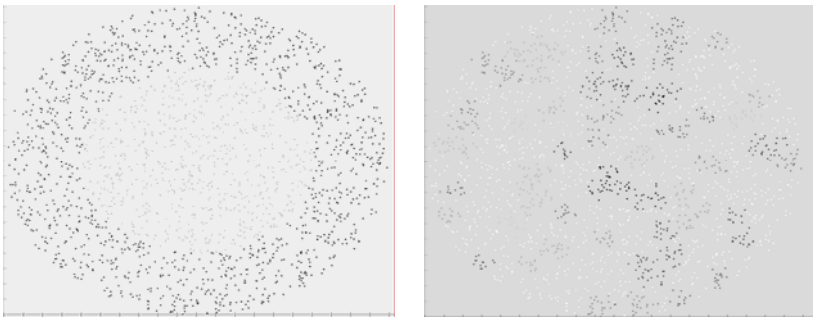
The IDA system provides facilities for manual editing or manipulation of data and analytical techniques. The classification of existing or new data is performed on the basis of class models selected by the user. The IDA system is able to learn from such classification results and update class memberships accordingly. The full range of classification approaches provided by IDA are partitioning methods (k-means, k-medoids); hierarchical clustering (agglomerative hierarchical clustering, balanced iterative reducing and clustering using hierarchies (BIRCH), clustering using representatives, ROCK), COBWEB algorithm, a hierarchical clustering algorithm using dynamic modeling (Chameleon) and modified Chameleon algorithm; a density-based methods (density-based spatial clustering of applications with noise (DBSCAN), ordering Points To Identify the Clustering Structure (OPTICS)) [1, 2]; according to machine learning approach includes: K-nearest neighbors classification (K-NN), classifiers ensemble, Native Bayes classification, Classification trees (ID3) (see Fig.1). For each input learning data set it can be chosen two directions – supervised learning and unsupervised learning.

The symmetric edge between two points is the closest neighbor among all existing neighbors, which gives value of k. We compute the weight of an edge connecting two objects in the k-NN graph that is inversely related to their distance (see Figures 1 and 2). The asymmetric is the hypergraphs construction of k-NN graph. But the IDA package is used to construct a set of small hypergraphs which consists of 200 to 1500 hypervertices. Moreover, IDA is used in the coarsing phase of hypergraph reduction two approaches: via (i) heaviest edge matching without limitation on number of vertices in hypervertex and (ii) heaviest edge matching with limitation on number of vertices in hypervertex. The third partitioning of the hypergraph were possible by 2 algorithms: using k-way multilevel paradigm [6] and recursive bisection Kernighan-Lin / Fiduccia - Mattheyses algorithm.

ROCK is an agglomerative hierarchical clustering algorithm used for clustering categorical attributes. It measures the similarity of two clusters by comparing the aggregate interconnectivity of two clusters against a user-specified static interconnectivity model, where the interconnectivity of two clusters is defined by the number of cross links between the two clusters (link is the number of common neighbors between two points) [4].

The Chameleon is a clustering algorithm that explores dynamic modeling in hierarchical clustering [5]. In its clustering process, two clusters are merged if the interconnectivity and closeness (proximity) between two clusters are highly related. The Chameleon is based on the observation of the weakness of two hierarchical clustering algorithms: CURE and ROCK. Because of its complexity and multiphase pow-

Figure 1. DB Scan



Minimum points = 10, Radius = 0.1
Distance measure = Euclidean

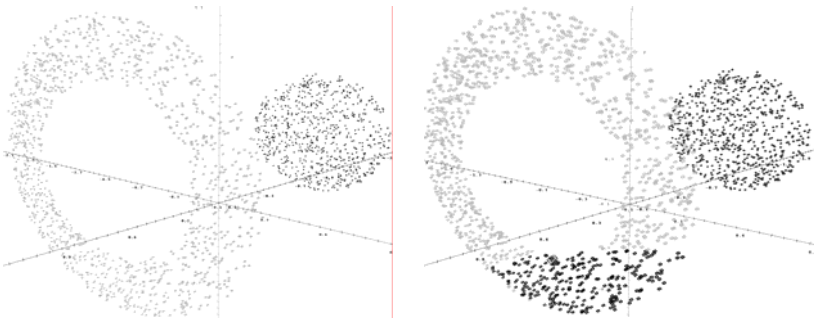
Minimum points = 10, Radius = 0.03
Distance measure = Euclidean

Figure 2. k-means

Number of Clusters = 2
Distance measure = Euclidean

Number of Clusters = 3
Distance measure = Chebychev

Figure 3. k-medoids



Number of Clusters = 2
Distance measure = Euclidean

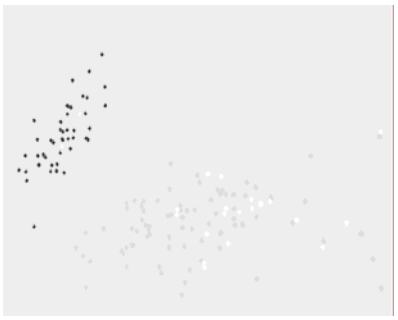
Number of Clusters = 3
Distance measure = Camberra

erful data mining analysis software packages do not adopt the latter approach.

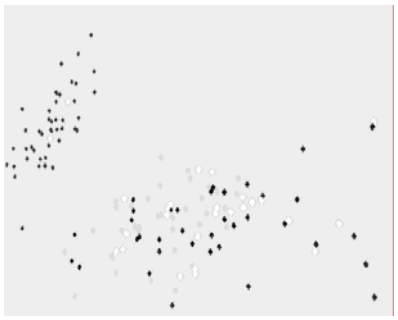
Moreover, IDA consists of two versions of this algorithm: the classic Chameleon [5] and our proposed version. Why we need to use two variants of this approach? The overall computational complexity of the CHAMELEON depends on the amount of time it requires to construct the K – nearest neighbors graph and the amount of time it requires to perform the two phases of the clustering algorithm. The CHAMELEON is not very sensitive to the value of k for computing the k-nearest neighbor graph, of the value of MINSIZE for the phase I of the algorithm, and of scheme for combining relative inter-connectivity and relative closeness and associated parameters, and it was able to discover the correct clusters for all of these combinations of values for k and MINSIZE. Our experimental evaluation of clustering using METIS hypergraph partitioning package for k-way partitioning of hypergraph and for recursive bisection [7] and CLUTO – A Clustering Toolkit Release 2.1.1 [8] – experimented with five different data sets containing points in two dimensions: "disk in disk", t4.8k, t5.8k, t8.8k, t7.10k [9]. We choose the number of neighbors k=5, 15, 40, MINSIZE = 5%. The results of the k-way partitioning of hypergraph by hMETIS package [7] and by CLUTO package [8] with k=5 nearest neighbors. Figure 6 presents cases where the genuine clusters have not been correctly identified. The CLUTO can identify the border between 2 classes only by symmetric k-NN graph, where the weights of edges are the number of common neighbors of two vertices. The data set t8.8k consists eight clusters of different shapes, size and orientation, some of which are inside the space enclosed by other clusters. It also contains random noise (collection of points forming vertical streaks). In the k=5 nearest neighbors hMETIS computes k-way partitioning of hypergraph with mistakes closer to the border of two classes and CLUTO can not effectively merge clusters for such type of dataset using asymmetric k-NN. Thus, the partitioning phase is very sensitive to the value of k for spherical shapes of clusters and to the types of k-NN graph (symmetric and asymmetric). It is very important to choose an optimal value of k, so if k=16 and more, only the symmetric k-NN with

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. OPTICS

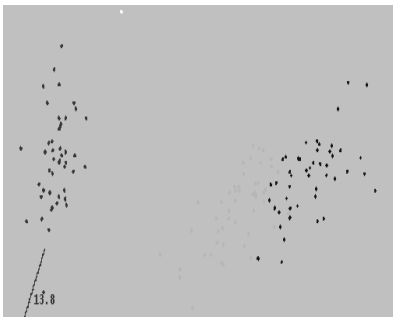


Minimum points = 4, Radius = 0.45
Distance measure = Euclidean

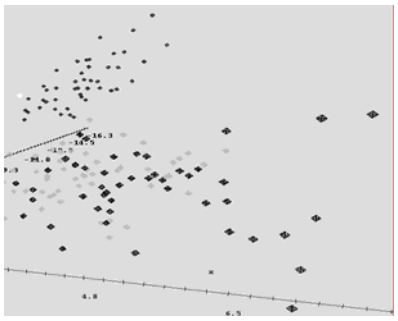


Minimum points = 4, Radius = 0.42
Distance measure = Euclidean

Figure 5. Ensembles of Classifiers



Method = 3-NN, Count = 9,
Method of Manipulation = Bagging,
Method of Voting = Simple



Method = Naïve Bayes, Count = 2,
Method of Manipulation = Cross-Validation, Method of Voting = Weighted

weights of edges equal to the number of common neighbors would be obtained for final clustering with minimum percentages of errors.

Our approach is based on the classic approach, the weight of an edge that was computed has weighted distance between objects; usually the weight of an edge connecting two nodes in a coarsened version of the graph is the number of edges in the original graph that connect the two sets of original nodes collapsed into the two coarse nodes. In our experiments, we computed the weight of the hyperedge as the sum of the weights of all edges that collapse on each other during coarsening step. On the next level of algorithm we produce a set of small hypergraphs using k-way multilevel paradigm. One of the most commonly used objective functions is to minimize the hyperedge-cut of the partitioning; i.e., the total number of hyperedges that span multiple partitions [10]. We used the cut size of hypergraph sum of the weights of edges that span partitions and as the gain value of vertex (value of difference between sum of weighted edges that leave the subgraph and stay within it). We do some modification to the expression of interconnectivity and closeness between two clusters by adding value that estimates the average density of each subgraph using weights of edges. The Figure 7 presents our results of clustering of data sets with different densities and size 2000 pt and 3000 pt (see Figures 6(a) and (b)), for k=5 neighbors, asymmetric k-NN, modified expression of interconnectivity and closeness.

Our result shows an accuracy border between two classes with the total error of clustering which equals or near 3% in the first case and 1%, respectively (see Figure 7).

3 IDA SOFTWARE PACKAGE FEATURES

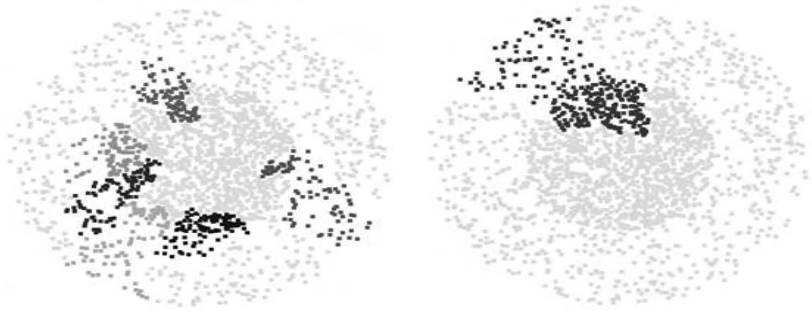
All operations performed using the IDA are carried out within the main window which appears when you start the program. The IDA provides the functionality for processing data starting from importing, analysis, clustering, modeling and classification. There are three possible graphical representations – dendrogramm, 2-3D scatter plot, on which the data is displayed. They all provide functions which allow you to view the data in different ways and generally get to know its inter-relationships and characteristics. They can help to identify data which may exclude from the analysis, or corrupt data; or change. Several weighting and scaling functions are available to normalise the project data across all variables. The weighted values are plotted on the scatter plot, which consists of data from different ASCII text files, Excel and Access. As a result the IDA performs the report window for tables and pictures classification and can be integrated with Word, XML, Excel formats.

REFERENCES

1. Jiawei Han and Micheline Kamber, Cluster Analysis (Chapter 8). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, pp.335-393, 2000.
2. P. Michaud. Clustering techniques. Future Generation Computer systems, 13, 1997.
3. D. Gibson, J. Kleinberg, and P. Raghavan. Clustering categorical data: An approach based on dynamic systems. In Proc. VLDB'98. 1998.
4. S. Guha, R. Rastogi, K. Shim. ROCK: Robust Clustering using linKs, (ICDE'99).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 6. The results of clustering with k=5 nearest neighbors and asymmetric k-NN



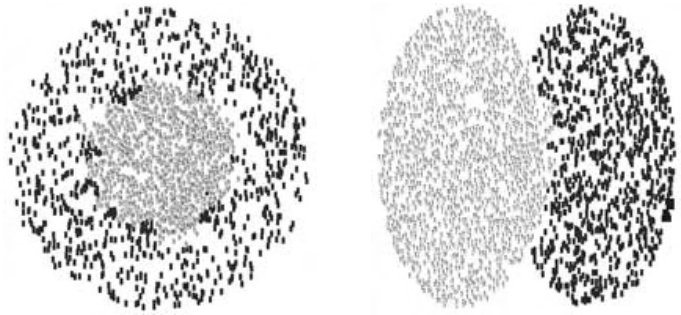
a) k-way partitioning by hMETIS b) final clusters by CLUTO

Data set "disk in disk"

a) k-way partitioning by hMETIS b) final clusters by CLUTO

Data set "t8.8k"

Figure 7. The results of modified hierarchical clustering using dynamic modeling



5. G. Karypis, E.-H. Han, and V. Kumar. CHAMELEON: A Hierarchical Clustering Algorithms Using Dynamic Modeling. IEEE Computer, Vol.32, No.8, 1999. pp. 68-75.
6. G. Karypis and V. Kumar. Multilevel k-way hypergraph partitioning. In Proceedings of the Design and Automation Conference, 1999.
7. G. Karypis, and V. Kumar, hMETIS 1.5.3: A hypergraph partitioning package. Technical report. Department of Computer Science, University of Minnesota, 1998.
8. G. Karypis, CLUTO 2.1.1. A Clustering Toolkit. Technical report. Department of Computer Science, University of Minnesota, 2003
9. KARYPIS LAB, http://www.cs.umn.edu/~karypis
10. G. Karypis, R. Aggarwal, V. Kumar and Sh. Shekhar, Multilevel hypergraph partitioning: Application in VLSI domain. In: Proceedings of the Design and Automation Conference. 1997.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Survey of agent oriented software engineering methodologies

Rawan Abu lail
Philadelphia University, Jordan

Mohd Shkoukani
Applied Science University, Jordan

Saed Ghoul
Philadelphia University, Jordan

Abstract This paper provides a summary of software engineering process and its importance in open system industry. It focuses on orientation of multi agent systems and on some representative agent oriented software engineering methodologies such as Gaia, ROADMAP, Tropos and MaSE with their strengths and weaknesses. Then it describes the agent oriented software engineering development lifecycle it also presents a comparative evaluation of Agent oriented software engineering methodologies, finally it recommends further research and improvements for the existing methodologies. This is an important conclusion in support of agent-oriented methodologies, as it may promote these enhancements and help arriving at industry-grade methodologies.

Keywords Software engineering process, Multi agent systems, Agent oriented software engineering methodologies.

1 INTRODUCTION

Software development is very different from other types of product development.

Software is an intangible product where other engineering disciplines such as civil engineering produce tangible products such as buildings and bridges. The physics behind the structures necessary to support a bridge and the engineering principles used to build a bridge are long proven. Software is a relatively new engineering discipline. Because it is intangible, principles used in other engineering disciplines to produce a product free from failure do not apply. No two software development projects are alike.

Software engineering is an engineering discipline that addresses all aspects of software production [23]. It is concerned with all phases of the development process from defining the requirements and early stages of the system specification through maintenance for all types of systems. So, it can be defined as the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software; that is, the application of engineering to software [12]. It is based on many independent processes involving many interacting stakeholders with conflicting interests and point of view, and is the application of tools, methods and disciplines to produce and maintain an automated solution to a real world problem, called software system [9]

Since the product is intangible, tracking the building process is an integral portion of software engineering. Unlike the building of a bridge, the production manager cannot look at the product to determine its progress. Software engineering is a methodological process of developing software in a repeatable manner on time within budget such that the software has the following attributes [22]: conforms to specifications, maintainable, dependable, efficient, and usable.

The complexity associated with MAS in an open setting involves numerous facets and dimensions. When a large set of agents interact over heterogeneous environment, several problems appear. It makes their coordination and management more difficult and increases the probability of exceptional situations, security holes, and unexpected global effects, and so on Commercial success for open agent-based applications will require software engineering approaches in order to enable effective scalable deployment [3].

Multiple modeling methods for constructing agent-based systems have been suggested, however no of them have been accepted as a standard. A prominent reason for this is the gap that exists between agent oriented methods and the modeling needs of agent based systems [1].Another problem in AOSE methodologies that there is no agreement on how to identify and characterize roles in the analysis phase and

agent types in the design phase [20], that we are trying to solve in our work.

2 MOTIVATION

Software development and management is a smart activity, necessitating high skills of analysis, design, coding, and testing. These activities integrate harmoniously and consistently different paradigms, tools, methods, and methodologies. The enactment of these activities and the coordination between them requires knowledge based reasoning, diagnosing, deciding, adapting, which is conceptually more supported by the agent paradigm than another one.

Effectively, in recent years, researchers and practitioners have recognized the advantages of applying the agent paradigm for system development. Yet the number of deployed commercial agent based applications is quit small, a major reason for this slow technology transfer is the lack of an industry –standard method for agent-based application development [1].

Software is becoming present in every aspect of our lives, pushing us inevitably towards a world of distributed, context aware computing systems. Multi-agent systems (MASs) are prominent technology to model and develop context-aware computing systems, as MAS intrinsically consists of large numbers of cooperating entities that consider their context in performing their task. Context is any information about the circumstances, objects or conditions by which an agent is surrounded that is considered relevant to the interaction between the agent and the computing environment [2].

Advances in networking in the last few years have turned the agent technology into promising paradigm to engineer complex distributed software systems .Nowadays , it has been applied to a wide range of application domains, including e-commerce, human-computer interfaces , telecommunications, and concurrent engineering. Agent technology is now being applied to the development of large open industrial software systems [3]. Since a software agent is an inherently more complex abstraction, the development of multi-agent systems (MAS) poses new challenges to software engineering [3].

Future software systems will be intelligent and adaptive. They will have the ability to seamlessly integrate with smart applications that have not been explicitly designed to work together. Traditional software engineering approaches offer limited support for the development of intelligent systems [4].

The explosive growth of application areas such as electronic commerce, enterprise resource planning and peer-to-peer computing has deeply and irreversibly changed our views on software and software engineering. Software must now be based on open architectures that continuously change and evolve to accommodate new components and meet new requirement. Software must also operate on different platforms, without recompilation, and with minimal assumptions about its operating environment and its users. As well, software must be robust and autonomous, capable of serving end users with a minimum of overhead and interfaces. These new requirements, in turn, call for new concepts, tools and techniques for engineering and managing software.

For these reasons –and more- agent oriented software development is gaining popularity over traditional development techniques, including structured and object-oriented ones.

After all agent based architecture do provide for an open, evolving architecture that can change at run time to exploit the services of new agents, or replace under-performing ones. In addition, software agents can, in principle, cope with unforeseen circumstances because their architecture includes goals along the planning capability for meeting them [10].

This recognized promising area and its open problems encourage investigating it in order to contribute to its enrichment with effective solutions to some open problems.

3 CURRENT RESEARCHES ON AGENTS- BASED SOFTWARE ENGINEERING

3.1 Agent Oriented Software Engineering development Lifecycle

Agent oriented software engineering development lifecycle covers the following stages:

Analysis: this stage relates to the expression of requirement. In the agent domain, there are agent approaches that deal directly with these requirements.

Design: this stage considers how to facilitate a design of multiagent systems (MAS) using agent concepts and technology. In software engineering, design covers the study of how to realize analysis elements into another specification that can be directly implemented

Implementation: is the translation of design concept to programs compliable to executable code or interpretable. To implement MAS, the language may be conventional or agent-oriented.

Testing: enables to identify the existing failures and to check if the code sticks to the specification of the system or at least, if it satisfies the requirements of customers. In this stage, classic software engineering distinguishes between validation and verification. Verification is concerned with checking the internal consistency of specification, and validation is concerned with checking the specifications` consistency with the stockholder's intensions.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

3.2 A survey on Agent-Oriented-Software engineering

Agents and multiagent system (MAS) have emerged as a powerful technology to face the complexity of variety of todays IT scenarios. Several industrial experiences already testify to the advantages of using agents in manufacturing processes [10].

Agent-based computing promotes designing and developing applications in terms of autonomous software entities (agents), situated in an environment, and which can flexibly achieve their objectives by interacting with one another in terms of high-levels protocols and languages. These features are definitely well suited to tackle the intrinsic complexity in developing software in modern scenarios. In fact: 1) the autonomy of the application components reflects the intrinsically decentralized nature of modern distributed systems, and can be considered as the natural extension to the notions of modularity and encapsulation for systems that are owned by different stakeholders; 2) the flexible in which agents operate and interact (both with each other and with the environment) is suited to the dynamic and unpredictable situations in which software is expected to operate today; and 3) the concept of agency provides for the unified view of artificial intelligence results and achievements, which eventually can be used to solve world problems, by making agents and MAS act as sound and manageable repositories of intelligent behaviors [10].

Software development for enterprise systems has been notoriously difficult. Computing architecture have gone from centralized to rigidly distributed to fully open. Open architectures are characterized by the fact that they enable autonomous, hydrogenous components to be added and removed dynamically .Open architectures are becoming increasingly common with the expansion of e-business [10].

Agent concepts are natural to describe intelligent adaptive systems which are able to act rationally to seek optimal solutions for their design objectives [4]; they are simply computer systems that are capable of autonomous in some environment in order to meet their design objectives [8, 11]. An agent, also called a software agent or an intelligent agent, is a piece of autonomous software, the words intelligent and agent describes some of its characteristic features. Intelligent is used because the software can have certain types of behavior ("Intelligent behavior is the selection of actions based on knowledge"), and the term agent tells something about the purpose of the software. An agent is "one who is authorized to act for or in the place of another" [7].

The Agent oriented (AO) approach promises the ability to construct flexible systems with complex and sophisticated behavior by combining highly modular components. The intelligence of these components –the agents – and their capacity for social interaction results in a multi agent systems (MAS) with capabilities beyond those of a simple 'sum' of agent [5]. Agent-Oriented Software Engineering is being described as a new paradigm for the research field of Software Engineering. But in order to become a new paradigm for the software industry, robust and easy-to-use methodologies and tools have to be developed [7].

Agents in environments should be able to acquire and reason about contexts to adapt the way they behave. However, contextual information poses some interesting problems since different agents could have different understanding of the current context. They might use different terms to describe context, and even if they use the same terms they might attach different semantics to these terms. Context awareness must address this problem by insuring that there is no semantic gap between different agents when they exchange contextual information [2].

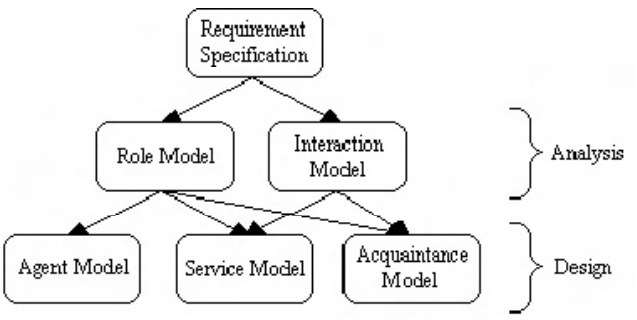
The main purposes of Agent-Oriented Software Engineering are to create methodologies and tools that enable inexpensive development and maintenance of agent-based software. In addition, the software should be flexible, easy-to-use, scalable and of high quality [7]. The development of multi-agent systems (MAS) is not a trivial task. In addition, with the advances in internet technologies, MAS understand a transition from closed to open architectures composed of a huge number of autonomous agents, which operate and move across different environment. In fact, openness introduces additional complexity of the systems modeling, design and implementation it also impacts on most quality attributes of MAS, including scalability, interoperability, reliability and adaptability [3].

There is an urgent need not only for theoretical foundation but also for specific methodologies driving the development of MAS's, and for powerful manageable architecture making multi agent system a viable approach to build context aware software systems. Without adequate development techniques and methods, such systems will not be sufficiently dependable, robust, trustworthy, and extensible [2].

Many well known agent-oriented software methodologies have been proposed such as GAIA, ROADMAP, MaSE and TROPOS methodology.

The Gaia methodology models both the macro (social) aspect and the micro (agent internals) aspect of the multiagent system. Gaia takes the view that a system can be seen as a society or an organization of agents. The methodology is applied after the requirements are gathered and specified, and covers the analysis and design phases. Figure 1 shows the artifacts produced by using Gaia [14].

Figure 1 The Gaia Models



Gaia was designed to handle small-scale, closed systems. Thomas and et al have identified the following weaknesses through their work on the motivating example, in order of encounter [14]:

1. Gaia assumes complete specification of requirements and does not address the requirement-gathering phase. It does not guide developers to take advantage of richer requirements enabled by agent technologies. The methodology does not facilitate regular changes of requirements typical to open systems.
2. Environmental information is implicitly encoded in the permissions and protocols of individual roles. Gaia does not present a holistic model of the execution environment to the developers. This omission renders Gaia inappropriate for engineering applications with dynamic and heterogeneous environments.
3. Domain knowledge in the system is implicitly encoded in the attributes of the individual roles. Gaia does not present a holistic model of the structure of the domain knowledge and the interaction and dependencies of knowledge components in the system. This omission prohibits knowledge in the system to be shared, reused, extended and maintained in a modular fashion.
4. Roles are not hierarchical in Gaia. The role model provides strictly one level of abstraction for the developers to conceptualize the system. The methodology does not facilitate iterative refinement of the system through different levels of abstraction. As a result, Gaia does not scale to handle complex systems.
5. Gaia cannot explicitly model and represent important social aspects of a multi-agent system. Gaia cannot explicitly model the organization structure of the agents in the system, or alternatively, the architecture of the system. It also lacks the ability to explicitly model the social goals, social tasks or social laws within an organization of agents.
6. Gaia offers no mechanisms to model the dynamic reasoning, extension and modification of the above social aspects at runtime.
7. The roles, representing responsibilities and capabilities of agents, are not realized in design or at runtime. The lack of such information at runtime makes peer verification of agent behaviors difficult.
8. At an individual agent level, Gaia offers no mechanisms to model the dynamic reasoning, extension and modification of responsibilities and capabilities of agents at runtime.

The ROADMAP methodology extends Gaia with four improvements - formal models of knowledge and the environment, role hierarchies, explicit representation of social structures and relationships, and incorporation of dynamic changes [14].

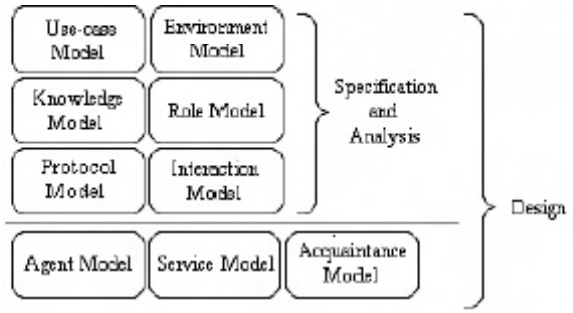
The role hierarchy represents the agent organization and constrains the behavior of member agents [15].

To extend Gaia for open systems, Thomas and et al expect the following features to be included [14]:

1. Support for requirements gathering.
2. Explicit models to describe the domain knowledge and the execution environment.
3. Levels of abstraction during the analysis phase, to allow iterative decomposition of the system.
4. Explicit models and representations of social aspects and individual agent characteristics, from the analysis phase to the final implementation.
5. Runtime reflection, modeling mechanisms to reason and change the social aspects and individual agent characteristics at runtime.

Figure 2 shows the artifacts produced by ROADMAP [14].

Figure 2. Structure of ROADMAP models in two phases of development



One issue of ROADMAP is the support of open systems via formal models of the environment, domain knowledge and the organization in terms of agent roles. On the other hand there is a lack of support for the detailed design stage. ROADMAP must rely on other methodologies for detailed design [15].

Tropos is a novel agent-oriented software development methodology founded on two key features: (i) the notions of agent, goal, plan and other knowledge level concepts are used uniformly throughout the software development process; and (ii) a crucial role is assigned to requirements analysis and specification when the system to be is analyzed with respect to its intended environment [17]. It is based on two key ideas (1) the notion of agents and all related mentalistic notions (for instance: beliefs, goals, actions and plans) are used in all phases of software development, from the early phases of requirements analysis, thus allowing for a deeper understanding of the environment where the software must operate, and (2) of the kind of interactions that should occur between software and human agents [18, 19].

Tropos adopts Eric Yu`s i* model which offers actors (agents, roles, or positions), goals and actors dependencies as primitive concepts for modeling an application during early requirement analysis.

Early requirement analysis it focuses on the intensions of stakeholders. Intensions are modeled as goals. Through some form of goal oriented analysis, these initial goals eventually lead to the functional and non functional requirements of the system-to-be [24].In i*stakeholders are represented as (social actors who depend on each other for goals to be achieved, tasks to be performed, and recourses to be furnished[10]. The i* frame work includes the strategic dependency model

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

for describing the network of relationships among actors, as well as the strategic rational model for describing and supporting the reasoning that each actor goes through concerning its relationships with other actors[10].

A strategic dependency model is a graph, where each node represents an actor, and each link between two actors indicates that one actor depends on the other for something in order that the former may attain some goal. We call the depending actor the depender and the actor who is depended upon the dependee. The object around which the dependency centers is called the dependum. By depending on another actor for a dependum, an actor is able to achieve goals that it is otherwise unable to achieve, or not as easily, or not as well. At the same time, the depender becomes vulnerable. If the dependee fails to deliver the dependum, the depender would be adversely affected in its ability to achieve its goals[25].

The type of the dependency describes the nature of the agreement. Goal dependencies are used to represent delegation of responsibility for fulfilling a goal; softgoal dependencies are similar to goal dependencies, but their fulfillment cannot be defined precisely (for instance, it is a matter of personal feeling, or the fulfillment can occur only to a given extent); task dependencies represent situations where the dependee is required to perform a given activity, while resource dependencies require the dependee to provide a resource to the depender[26].As shown in figure(3) , actors are represented as circles; dependums – goals, softgoals, tasks and recourses – are respectively represented as ovals, clouds, hexagons and rectangles ; and dependencies have the form depender à dependumàdependee [27].

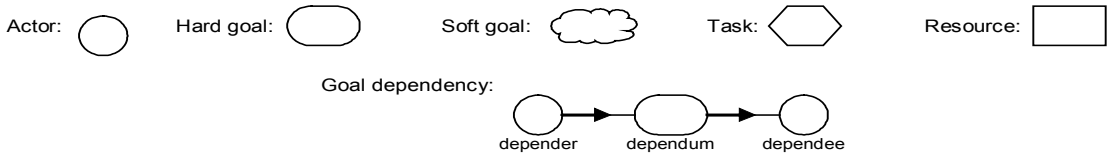
MaSE was originally designed to develop general-purpose multiagent systems and has been used to design systems ranging from computer virus immune systems to cooperative robotics systems [21]. While it provides many advantages for building multiagent systems, it is not perfect. It is based on a strong top-down software engineering mindset, which makes it difficult to use in some application areas [21]:

1. MaSE fails to provide a mechanism for modeling multiagent system interactions with the environment.
2. MaSE also tends to produce multiagent systems with a fixed organization. Agents developed in MaSE tend to play a limited number of roles and have a limited ability to change those roles, regardless of their individual capabilities. As discussed above, a multiagent team should be able to design its own organization at runtime. While MaSE already incorporates many of the required organizational concepts such as goals, roles and the relations between these entities, it cannot currently be used to define a true multiagent organization.
3. MaSE also does not allow the integration of sub-teams into a multiagent system. MaSE multiagent systems are assumed to have only a single layer to which all agents belong. Adding the notion of sub-teams would allow the decomposition of multiagent systems and provide for greater levels of abstraction.
4. The MaSE notion of conversations can also be somewhat bothersome, as it tends to decompose the protocols defined in the analysis phase into small, often extremely simple pieces.

There is a number off issues that are problematic for the previous methodologies [20]:

- There is no agreement on how to identify and characterize roles in the analysis phase and agent types in the design phase.
- The concepts used in the methodologies, like responsibility, permission, goals and tasks do not have a formal semantics or explicit formal properties. This becomes an important issue when these concepts are implemented; implementation constructs do have exact semantics.
- There is a gap between the design models of the methodologies and the existing implementation languages. It is unreasonable to expect a programmer to implement the proposed complex design models. To bridge the gap, a methodology should either introduce refined design models that can be directly implemented in an available programming language, or use a dedicated agent-oriented programming language which provides constructs to implement the high-level design concepts.
- The methodologies that include an implementation phase, such as Tropos, propose an implementation language in which it is not explained how to implement reasoning about beliefs, reasoning about goals and plans, reasoning about planning goals, or reasoning about communication.
- It is widely recognized that an agent may enact several roles. None of the methodologies addresses the implementation of agents that need to represent and reason about playing different roles.
- Open systems are not really supported. The methodologies implicitly suppose that agents are purposely designed to enact roles in a system. But as soon as agents from the outside may enter the analysis, design and implementation needs to treat agents as given entities.
- In the analysis, methodologies do not consider the environmental embedding of a system. The structure of the organization in which a system will be embedded, has a large influence on the type of organizational structure of the system, at least when it interacts with more than one person.

Figure 3 . Examples of Tropos Notation



We can conclude that out of the numerous proposed methods, as every method has its own advantages and disadvantages [16]. Most of these methodologies do not address the characteristics of architectural independence, robustness and scalability adequately. In particular, there has been insufficient coverage on facilitating the specification of dynamic social interactions [14].

Some methodologies offer good software tools and processes for system analysis and design, but do not take into account social norms. Others are mainly focused on analysis, whereas design and implementation phases lack or are redirected to agent-oriented methodologies (which do not cover organizational concepts) [13].

Moreover, most of the proposed methodologies only deal with groups but do not consider other topological designs (ex. hierarchies, matrix, and markets) [13].

The existing methodologies generally do not consider the following all at-once [16]:

- Most approaches analyze the functional and non-functional requirements in a single module.
- Most approaches do not integrate the quality attributes with functional attributes at the requirements gathering and analysis stage and moreover the integration is done at a later stage of the software process where it is difficult to achieve it.
- These methodologies do not take into account the nature of users, user's interpretation and the user's perception of the quality attributes.
- These methodologies do not take into account the crosscutting nature of requirements.

3.3 A comparative evaluation of Agent-Oriented-Methodologies

Here is an evaluation framework that based on a feature analysis technique. That is, the features on each of the examined methodologies are evaluated. The evaluation is performed based on information regarding the examined methodologies available in publications. The framework's three facets are: concepts and properties, notations and modeling techniques, and development process [10].

3.3.1 Metric

To enable ranking of the properties examined in the evaluation process, the framework proposes a scale of 1 to 7 with the following interpretations[10]:

1. Indicates that the methodology does not address the property.
2. Indicates that the methodology refers to the property but no details are provided.
3. Indicates that the methodology address the property to a limited extent. That is, many issues that are related to the specific property are not addressed.
4. Indicates that the methodology address the property, yet some major issues are lacking.
5. Indicates that the methodology addresses the property, however, it lacks one or two major issues related to the specific property.
6. Indicates that the methodology addresses the property with minor deficiencies.
7. Indicates that the methodology fully addresses the property.

3.3.2 Methodologies evaluation summary

Table 1. Methodologies evaluation summary [10]

Framework Criteria	Gaia	Tropos	MaSE
1.Concepts and properties			
1.1 Autonomy	7	7	7
1.2 Reactiveness	7	4	4
1.3 Proactiveness	7	7	7
1.4 Sociality	4	4	4
1.5 Building blocks coverage	4	5	5
2.Notations and modeling techniques			
2.1 Accessibility	5	4	5
2.2 Analyzability	1	5	6
2.3 Complexity Management	1	5	4
2.4 Executability	1	4	4
2.5 Expressiveness	4	4	5
2.6 Modularity	4	7	4
2.7 Preciseness	7	7	6
3.Development process			
3.1 Development context	5	6	5
3.2 Lifecycle coverage	3	6	5
3.3 Stages activities	4	4	7
3.4 Validation and verification	1	5	4
3.5 Quality assurance	1	1	1
3.6 Project management	1	1	1

4 CONCLUSION

In conclusion, the examined agent-oriented methodologies provide an appropriate infrastructure, however there is a need for further research and improvements. This is an important conclusion in support of agent-oriented methodologies, as it may promote these enhancements and help arriving at industry-grade methodologies. Additionally, the evaluation performed here provides researchers and practitioners with a detailed comparison among the leading agent-oriented methodologies. Further, the framework used in this study may be utilized by others to evaluate and compare other methodologies as needed [10].

REFERENCES

1. Arnon Sturm, Dov Dori, Onn Shehory, 2003, Single-model method for specifying multi-agent systems, ACM Press, pp 121-128

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

2. Alessandro Garcia, Ricardo Choren, Carlos Lucena, Alexander Romanovsky, Holger Giese, Danny Weyns, Tom Holvoet, Paolo Giorgini, Jul 2005, Software Engineering for Large-Scale Multi-Agent Systems - SELMAS 2005: workshop report, ACM SIGSOFT Software Engineering Notes, issue 4, vol. 30, pp 1-8
3. Ricardo Choren, Alessandro Garcia, Carlos Lucena, Martin Griss, David Kung, Naftaly Minsky, Alexander Romanovsky , 2004, Software Engineering for Large-Scale Multi-agent Systems SELMAS'04, ACM Press , International Conference on Software Engineering Proceedings of the 26th International Conference on Software Engineering pp 752-753
4. Leon Sterling, Thomas Juan, 2005,The software engineering of agent-based intelligent adaptive systems, ACM Press , International Conference on Software Engineering
Proceedings of the 27th international conference on Software engineering, pp 704-705
5. Philippe Massonet, Yves Deville, Cédric Nève , 2002,From AOSE methodology to agent implementation, ACM Press , International Conference on Autonomous Agents
Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1, pp 27-34
6. Gregor Engels, Wilhelm Schäfer, Robert Balzer, Volker Gruhn , 2001, Process-centered software engineering environments: academic and industrial perspectives, ACM Press , International Conference on Software Engineering Proceedings of the 23rd International Conference on Software Engineering, pp 671-673
7. Tveit A., 2001, A survey of agent-oriented Software Engineering, www.csgsc.org.
8. Holger Knublauch, 2002,Extreme programming of multi-agent systems, ACM Press , International Conference on Autonomous Agents Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2, pp 704-711
9. Claudine Toffolon, Salem Dakhli, 2000, A framework for studying the coordination process in software engineering, ACM Press , Symposium on Applied Computing
Proceedings of the 2000 ACM symposium on Applied computing - Volume 2, pp 851-857
10. Bergenti F., Gleizes M, Zambonelli F. , 2004, Methodologies and software engineering for agent system, Kluwer Academic publishers.
11. Wooldridge M., Jennings N., 1995, Intelligent Agents: Theory and Practice, www.citeseer.ist.psu.edu,
12. R.S. Pressman, 2005, Software Engineering: A Practitioner's Approach, 6th edition,
13. Estefania Argente Villaplana, 2005, A proposal for an organizational MAS methodology, ACM Press , International Conference on Autonomous Agents Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, pp 1370-1370
14. Thomas Juan, Adrian Pearce, Leon Sterling, 2002, ROADMAP: extending the gaia methodology for complex open systems, ACM Press, International Conference on Autonomous Agents Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1, pp 3-10
15. Thomas Juan, Leon Sterling, Maurizio Martelli, Viviana Mascardi, 2003, Customizing AOSE methodologies by reusing AOSE features, ACM Press , International Conference on Autonomous Agents Proceedings of the second international joint conference on Autonomous agents and multiagent systems, pp 113-120
16. Prabhat Ranjan, A. K. Misra, May 2006, A hybrid model for agent based system requirements analysis, ACM Press, ACM SIGSOFT Software Engineering Notes, issue 3, vol. 31, pp 1-7
17. Fausto Giunchiglia, John Mylopoulos, Anna Perini, 2002, The tropos software development methodology: processes, models and diagrams, ACM Press , International Conference on Autonomous Agents Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1, pp 35-36
18. Anna Perini, Angelo Susi, Fausto Giunchiglia, 2002, Coordination specification in multi-agent systems: from requirements to architecture with the Tropos methodology, ACM Press , ACM International Conference Proceeding Series; Vol. 27 Proceedings of the 14th international conference on Software engineering and knowledge engineering, pp 51-54
19. Paolo Bresciani, Anna Perini, Paolo Giorgini, Fausto Giunchiglia, John Mylopoulos, 2001, A knowledge level software engineering methodology for agent oriented programming, ACM Press , International Conference on Autonomous Agents Proceedings of the fifth international conference on Autonomous agents, pp 648-655
20. Mehdi Dastani, Joris Hulstijn, Frank Dignum, John-Jules Ch. Meyer, 2004,Issues in Multiagent System Development, ACM Press , International Conference on Autonomous Agents Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2, pp 922, 929
21. Scott A. DeLoach, 2005, Multiagent systems engineering of organization-based multiagent systems, ACM Press , International Conference on Software Engineering

Proceedings of the fourth international workshop on Software engineering for large-scale multi-agent systems, pp 1-7

22. Barbara Bracken, Dec 2003, Progressing from student to professional: the importance and challenges of teaching software engineering, Journal of Computing Sciences in Colleges, ACM, issue 2, vol. 19, pp 358-368
23. Ian Sommerville, SOFTWARE ENGINEERING 6th edition, Addison-Wesley.
24. Dardenne, A.,lamsweerde, A., and fixckas,S., 1993, Goal directed requirement acquisitions. Science of computer programming, pp 3-50.
25. Brinkkemper, J. and Castro J., June 2000, Tropos: A Framework for Requirements-Driven Software Development , Information Systems Engineering: State of the Art and Research Themes, Lecture Notes in Computer Science, Springer-Verlag.
26. Fuxman A., Pistore M., Mylopoulos J., Traverso P., Model Checking Early Requirements Specifications in Tropos
27. Cervenka R., 2003, Modeling Notation Source Tropos, Foundation for Intelligent Physical Agents.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

A proposed model for agent-oriented software engineering

Mohd Shkoukani
Applied Science University, Jordan

Rawan Abu lail
Philadelphia University, Jordan

Saed Ghoul
Philadelphia University, Jordan

Abstract This proposal provides a summary of software engineering process and its importance in open system industry. It focuses on orientation of multi agent systems and on some representative agent oriented software engineering methodologies such as Gaia, ROADMAP, Tropos and MaSE with their strengths and weaknesses. Finally it proposes the development of a new model that combines the features of two of the existing methodologies which are Gaia and Tropos by concentrating on their strengths and avoiding their weaknesses, helping to formally identify and characterize roles in the analysis phase and determination of agent types which are recognized as open problems in actual active researches.

Keywords Software engineering process, Multi agent systems, Agent oriented software engineering methodologies.

1 INTRODUCTION

Software development is very different from other types of product development. Software is an intangible product where other engineering disciplines such as civil engineering produce tangible products such as buildings and bridges. The physics behind the structures necessary to support a bridge and the engineering principles used to build a bridge are long proven. Software is a relatively new engineering discipline. Because it is intangible, principles used in other engineering disciplines to produce a product free from failure do not apply. No two software development projects are alike.

Software engineering is an engineering discipline that addresses all aspects of software production [23]. It is concerned with all phases of the development process from defining the requirements and early stages of the system specification through maintenance for all types of systems. So, it can be defined as the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software; that is, the application of engineering to software [12]. It is based on many independent processes involving many interacting stakeholders with conflicting interests and point of view, and is the application of tools, methods and disciplines to produce and maintain an automated solution to a real world problem, called software system [9]

Since the product is intangible, tracking the building process is an integral portion of software engineering. Unlike the building of a bridge, the production manager cannot look at the product to determine its progress. Software engineering is a methodological process of developing software in a repeatable manner on time within budget such that the software has the following attributes [22]: conforms to specifications, maintainable, dependable, efficient, and usable.

The complexity associated with MAS in an open setting involves numerous facets and dimensions. When a large set of agents interact over heterogeneous environment, several problems appear. It makes their coordination and management more difficult and increases the probability of exceptional situations, security holes, and unexpected global effects, and so on Commercial success for open agent-based applications will require software engineering approaches in order to enable effective scalable deployment [3].

Multiple modeling methods for constructing agent-based systems have been suggested, however no of them have been accepted as a standard. A prominent reason for this is the gap that exists between agent oriented methods and the modeling needs of agent based systems [1].Another problem in AOSE methodologies that there is no agreement on how to identify and characterize roles in the analysis phase and agent types in the design phase [20], that we are trying to solve in our work.

2 MOTIVATION

Software development and management is a smart activity, necessitating high skills of analysis, design, coding, and testing. These activities integrate harmoniously and consistently different paradigms, tools, methods, and methodologies. The enactment of these activities and the coordination between them requires knowledge based reasoning, diagnosing, deciding, adapting, which is conceptually more supported by the agent paradigm than another one.

Effectively, in recent years, researchers and practitioners have recognized the advantages of applying the agent paradigm for system development. Yet the number of deployed commercial agent based applications is quit small, a major reason for this slow technology transfer is the lack of an industry –standard method for agent-based application development [1].

Software is becoming present in every aspect of our lives, pushing us inevitably towards a world of distributed, context aware computing systems. Multi-agent systems (MASs) are prominent technology to model and develop context-aware computing systems, as MAS intrinsically consists of large numbers of cooperating entities that consider their context in performing their task. Context is any information about the circumstances, objects or conditions by which an agent is surrounded that is considered relevant to the interaction between the agent and the computing environment [2].

Advances in networking in the last few years have turned the agent technology into promising paradigm to engineer complex distributed software systems .Nowadays , it has been applied to a wide range of application domains, including e-commerce, human-computer interfaces , telecommunications, and concurrent engineering. Agent technology is now being applied to the development of large open industrial software systems [3]. Since a software agent is an inherently more complex abstraction, the development of multi-agent systems (MAS) poses new challenges to software engineering [3].

Future software systems will be intelligent and adaptive. They will have the ability to seamlessly integrate with smart applications that have not been explicitly designed to work together. Traditional software engineering approaches offer limited support for the development of intelligent systems [4].

The explosive growth of application areas such as electronic commerce, enterprise resource planning and peer-to-peer computing has deeply and irreversibly changed our views on software and software engineering. Software must now be based on open architectures that continuously change and evolve to accommodate new components and meet new requirement. Software must also operate on different platforms, without recompilation, and with minimal assumptions about its operating environment and its users. As well, software must be robust and autonomous, capable of serving end users with a minimum of overhead and interfaces. These new requirements, in turn, call for new concepts, tools and techniques for engineering and managing software.

For these reasons –and more- agent oriented software development is gaining popularity over traditional development techniques, including structured and object-oriented ones.

After all agent based architecture do provide for an open, evolving architecture that can change at run time to exploit the services of new agents, or replace under-performing ones. In addition, software agents can, in principle, cope with unforeseen circumstances because their architecture includes goals along the planning capability for meeting them [10].

This recognized promising area and its open problems encourage investigating it in order to contribute to its enrichment with effective solutions to some open problems.

In our work we are trying to solve the problem of identifying the agent roles in the analysis phase and the agent type in the design phase, adapting an existing methodology called ROADMAP [14].

3 CURRENT RESEARCHES ON AGENTS- BASED SOFTWARE ENGINEERING

Agents and multiagent system (MAS) have emerged as a powerful technology to face the complexity of variety of today's IT scenarios. Several industrial experiences already testify to the advantages of using agents in manufacturing processes [10].

Agent-based computing promotes designing and developing applications in terms of autonomous software entities (agents), situated in an environment, and which can flexibly achieve their objectives by interacting with one another in terms of high-levels protocols and languages. These features are definitely well suited to tackle the intrinsic complexity in developing software in modern scenarios. In fact: 1) the autonomy of the application components reflects the intrinsically decentralized nature of modern distributed systems, and can be considered as the natural extension to the notions of modularity and encapsulation for systems that are owned by different stakeholders; 2) the flexible in which agents operate and interact (both with each other and with the environment) is suited to the dynamic and unpredictable situations in which software is expected to operate today; and 3) the concept of agency provides for the unified view of artificial intelligence results and achievements, which eventually can be used to solve world problems, by making agents and MAS act as sound and manageable repositories of intelligent behaviors [10].

Software development for enterprise systems has been notoriously difficult. Computing architecture have gone from centralized to rigidly distributed to fully open. Open architectures are characterized by the fact that they enable autonomous, hydrogenous components to be added and removed dynamically .Open architectures are becoming increasingly common with the expansion of e-business [10].

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Agent concepts are natural to describe intelligent adaptive systems which are able to act rationally to seek optimal solutions for their design objectives [4]; they are simply computer systems that are capable of autonomous in some environment in order to meet their design objectives [8, 11]. An agent, also called a software agent or an intelligent agent, is a piece of autonomous software, the words intelligent and agent describes some of its characteristic features. Intelligent is used because the software can have certain types of behavior ("Intelligent behavior is the selection of actions based on knowledge"), and the term agent tells something about the purpose of the software. An agent is "one who is authorized to act for or in the place of another" [7].

The Agent oriented (AO) approach promises the ability to construct flexible systems with complex and sophisticated behavior by combining highly modular components. The intelligence of these components –the agents – and their capacity for social interaction results in a multi agent systems (MAS) with capabilities beyond those of a simple 'sum' of agent [5]. Agent-Oriented Software Engineering is being described as a new paradigm for the research field of Software Engineering. But in order to become a new paradigm for the software industry, robust and easy-to-use methodologies and tools have to be developed [7].

Agents in environments should be able to acquire and reason about contexts to adapt the way they behave. However, contextual information poses some interesting problems since different agents could have different understanding of the current context. They might use different terms to describe context, and even if they use the same terms they might attach different semantics to these terms. Context awareness must address this problem by insuring that there is no semantic gap between different agents when they exchange contextual information [2].

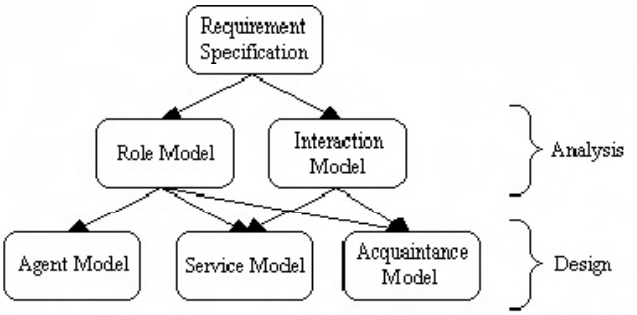
The main purposes of Agent-Oriented Software Engineering are to create methodologies and tools that enable inexpensive development and maintenance of agent-based software. In addition, the software should be flexible, easy-to-use, scalable and of high quality [7]. The development of multi-agent systems (MAS) is not a trivial task. In addition, with the advances in internet technologies, MAS understand a transition from closed to open architectures composed of a huge number of autonomous agents, which operate and move across different environment. In fact, openness introduces additional complexity of the systems modeling, design and implementation it also impacts on most quality attributes of MAS, including scalability, interoperability, reliability and adaptability [3].

There is an urgent need not only for theoretical foundation but also for specific methodologies driving the development of MAS's, and for powerful manageable architecture making multi agent system a viable approach to build context aware software systems. Without adequate development techniques and methods, such systems will not be sufficiently dependable, robust, trustworthy, and extensible [2].

Many well known agent-oriented software methodologies have been proposed such as GAIA, ROADMAP, MaSE and TROPOS methodology.

The Gaia methodology models both the macro (social) aspect and the micro (agent internals) aspect of the multi-agent system. Gaia takes the view that a system can be seen as a society or an organization of agents. The methodology is applied after the requirements are gathered and specified, and covers the analysis and design phases. Figure 1 shows the artifacts produced by using Gaia [14].

Figure 1 The Gaia Models



Gaia was designed to handle small-scale, closed systems. Thomas and et al have identified the following weaknesses through their work on the motivating example, in order of encounter [14]:

1. Gaia assumes complete specification of requirements and does not address the requirement-gathering phase. It does not guide developers to take advantage of richer requirements enabled by agent technologies. The methodology does not facilitate regular changes of requirements typical to open systems.
2. Environmental information is implicitly encoded in the permissions and protocols of individual roles. Gaia does not present a holistic model of the execution environment to the developers. This omission renders Gaia inappropriate for engineering applications with dynamic and heterogeneous environments.
3. Domain knowledge in the system is implicitly encoded in the attributes of the individual roles. Gaia does not present a holistic model of the structure of the domain knowledge and the interaction and dependencies of knowledge components in the system. This omission prohibits knowledge in the system to be shared, reused, extended and maintained in a modular fashion.
4. Roles are not hierarchical in Gaia. The role model provides strictly one level of abstraction for the developers to conceptualize the system. The methodology does not facilitate iterative refinement of the system through different levels of abstraction. As a result, Gaia does not scale to handle complex systems.
5. Gaia cannot explicitly model and represent important social aspects of a multi-agent system. Gaia cannot explicitly model the organization structure of the agents in the system, or alternatively, the architecture of the system. It also lacks the ability to explicitly model the social goals, social tasks or social laws within an organization of agents.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

6. Gaia offers no mechanisms to model the dynamic reasoning, extension and modification of the above social aspects at runtime.
7. The roles, representing responsibilities and capabilities of agents, are not realized in design or at runtime. The lack of such information at runtime makes peer verification of agent behaviors difficult.
8. At an individual agent level, Gaia offers no mechanisms to model the dynamic reasoning, extension and modification of responsibilities and capabilities of agents at runtime.

The ROADMAP methodology extends Gaia with four improvements - formal models of knowledge and the environment, role hierarchies, explicit representation of social structures and relationships, and incorporation of dynamic changes [14].

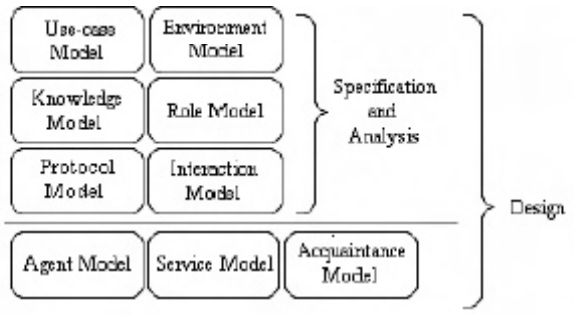
The role hierarchy represents the agent organization and constrains the behavior of member agents [15].

To extend Gaia for open systems, Thomas and et al expect the following features to be included [14]:

1. Support for requirements gathering.
2. Explicit models to describe the domain knowledge and the execution environment.
3. Levels of abstraction during the analysis phase, to allow iterative decomposition of the system.
4. Explicit models and representations of social aspects and individual agent characteristics, from the analysis phase to the final implementation.
5. Runtime reflection, modeling mechanisms to reason and change the social aspects and individual agent characteristics at runtime.

Figure 2 shows the artifacts produced by ROADMAP [14].

Figure 2. Structure of ROADMAP models in two phases of development



One issue of ROADMAP is the support of open systems via formal models of the environment, domain knowledge and the organization in terms of agent roles. On the other hand there is a lack of support for the detailed design stage. ROADMAP must rely on other methodologies for detailed design [15].

Tropos is a novel agent-oriented software development methodology founded on two key features: (i) the notions of agent, goal, plan and other knowledge level concepts are used uniformly throughout the software development process; and (ii) a crucial role is assigned to requirements analysis and specification when the system to be is analyzed with respect to its intended environment [17]. It is based on two key ideas (1) the notion of agents and all related mentalistic notions (for instance: beliefs, goals, actions and plans) are used in all phases of software development, from the early phases of requirements analysis, thus allowing for a deeper understanding of the environment where the software must operate, and (2) of the kind of interactions that should occur between software and human agents [18, 19].

Tropos adopts Eric Yu`s i* model which offers actors (agents, roles, or positions), goals and actors dependencies as primitive concepts for modeling an application during early requirement analysis.

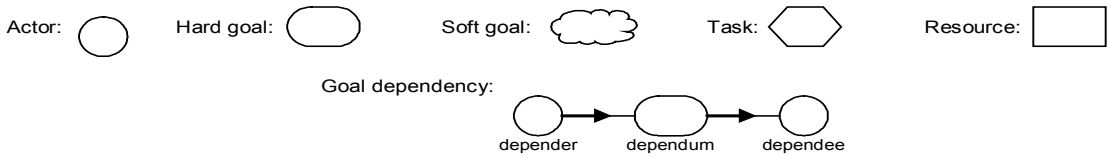
Early requirement analysis it focuses on the intensions of stakeholders. Intensions are modeled as goals. Through some form of goal oriented analysis, these initial goals eventually lead to the functional and non functional requirements of the system-to-be [24].In i*stakeholders are represented as (social actors who depend on each other for goals to be achieved, tasks to be performed, and recourses to be furnished[10]. The i* frame work includes the strategic dependency model for describing the network of relationships among actors, as well as the strategic rational model for describing and supporting the reasoning that each actor goes through concerning its relationships with other actors[10].

A strategic dependency model is a graph, where each node represents an actor, and each link between two actors indicates that one actor depends on the other for something in order that the former may attain some goal. We call the depending actor the depender and the actor who is depended upon the dependee. The object around which the dependency centers is called the dependum. By depending on another actor for a dependum, an actor is able to achieve goals that it is otherwise unable to achieve, or not as easily, or not as well. At the same time, the depender becomes vulnerable. If the dependee fails to deliver the dependum, the depender would be adversely affected in its ability to achieve its goals[25].

The type of the dependency describes the nature of the agreement. Goal dependencies are used to represent delegation of responsibility for fulfilling a goal; softgoal dependencies are similar to goal dependencies, but their fulfillment cannot be defined precisely (for instance, it is a matter of personal feeling, or the fulfillment can occur only to a given extent); task dependencies represent situations where the dependee is required to perform a given activity, while resource dependencies require the dependee to provide a resource to the depender[26].As shown in figure(3) , actors are represented as circles; dependums – goals, softgoals, tasks and recourses – are respectively represented as ovals, clouds, hexagons and rectangles ; and dependencies have the form depender → dependum → dependee [27].

MaSE was originally designed to develop general-purpose multiagent systems and has been used to design systems ranging from computer virus immune systems to cooperative robotics systems [21]. While it provides many advantages

Figure 3 . Examples of Tropos Notation



for building multiagent systems, it is not perfect. It is based on a strong top-down software engineering mindset, which makes it difficult to use in some application areas [21]:

1. MaSE fails to provide a mechanism for modeling multiagent system interactions with the environment.
2. MaSE also tends to produce multiagent systems with a fixed organization. Agents developed in MaSE tend to play a limited number of roles and have a limited ability to change those roles, regardless of their individual capabilities. As discussed above, a multiagent team should be able to design its own organization at runtime. While MaSE already incorporates many of the required organizational concepts such as goals, roles and the relations between these entities, it cannot currently be used to define a true multiagent organization.
3. MaSE also does not allow the integration of sub-teams into a multiagent system. MaSE multiagent systems are assumed to have only a single layer to which all agents belong. Adding the notion of sub-teams would allow the decomposition of multiagent systems and provide for greater levels of abstraction.
4. The MaSE notion of conversations can also be somewhat bothersome, as it tends to decompose the protocols defined in the analysis phase into small, often extremely simple pieces.

There is a number off issues that are problematic for the previous methodologies [20]:

- There is no agreement on how to identify and characterize roles in the analysis phase and agent types in the design phase.
- The concepts used in the methodologies, like responsibility, permission, goals and tasks do not have a formal semantics or explicit formal properties. This becomes an important issue when these concepts are implemented; implementation constructs do have exact semantics.
- There is a gap between the design models of the methodologies and the existing implementation languages. It is unreasonable to expect a programmer to implement the proposed complex design models. To bridge the gap, a methodology should either introduce refined design models that can be directly implemented in an available programming language, or use a dedicated agent-oriented programming language which provides constructs to implement the high-level design concepts.
- The methodologies that include an implementation phase, such as Tropos, propose an implementation language in which it is not explained how to implement reasoning about beliefs, reasoning about goals and plans, reasoning about planning goals, or reasoning about communication.
- It is widely recognized that an agent may enact several roles. None of the methodologies addresses the implementation of agents that need to represent and reason about playing different roles.
- Open systems are not really supported. The methodologies implicitly suppose that agents are purposely designed to enact roles in a system. But as soon as agents from the outside may enter the analysis, design and implementation needs to treat agents as given entities.
- In the analysis, methodologies do not consider the environmental embedding of a system. The structure of the organization in which a system will be embedded, has a large influence on the type of organizational structure of the system, at least when it interacts with more than one person.

We can conclude that out of the numerous proposed methods, as every method has its own advantages and disadvantages [16]. Most of these methodologies do not address the characteristics of architectural independence, robustness and scalability adequately. In particular, there has been insufficient coverage on facilitating the specification of dynamic social interactions [14].

Some methodologies offer good software tools and processes for system analysis and design, but do not take into account social norms. Others are mainly focused on analysis, whereas design and implementation phases lack or are redirected to agent-oriented methodologies (which do not cover organizational concepts) [13].

Moreover, most of the proposed methodologies only deal with groups but do not consider other topological designs (ex. hierarchies, matrix, and markets) [13].

The existing methodologies generally do not consider the following all at-once [16]:

- Most approaches analyze the functional and non-functional requirements in a single module.
- Most approaches do not integrate the quality attributes with functional attributes at the requirements gathering and analysis stage and moreover the integration is done at a later stage of the software process where it is difficult to achieve it.
- These methodologies do not take into account the nature of users, user's interpretation and the user's perception of the quality attributes.
- These methodologies do not take into account the crosscutting nature of requirements.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4 THE PROPOSED MODEL

One of the problems of using AOSE is a formal identification and characterization of agent roles in the analysis phase and a formal determination of the agent type in the design phase. In our work we will combine two of the existing AOSE methodologies, which are Gaia and Tropos , by concentrating on their strengths and avoiding their weaknesses, for developing a new methodology helping in solving the above problems. In our proposed methodology we will combine the early requirement phase from the Tropos methodology with the analysis phase in the Gaia methodology.

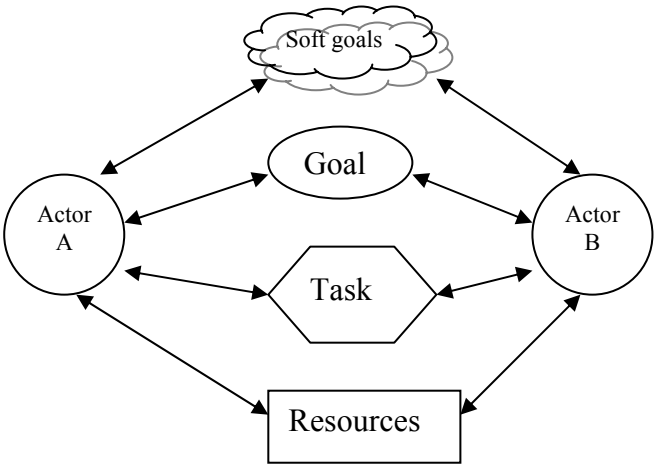
The scope of the methodology includes the analysis and design phases and excludes both collection of specifications and implementation. It is applied after the requirements are gathered and specified [10].

In the analysis stage in Gaia, roles in the system are identified and their interactions are modeled [10].

A role is defined by four attributes: responsibilities, permissions, activities, and protocols. Responsibilities determine functionality and, as such, are perhaps the key attribute associated with a role. Responsibilities are divided into two types: liveness properties and safety properties. Safety properties are properties that the agent acting in the role must always preserve. Liveness properties describe the "lifecycle" or generalized behavior pattern of the role [10].

In order to realize responsibilities, a role has a set of permissions. Permissions are the "rights" associated with a role. The permissions of a role thus identify the resources that are available to that role in order to realize its responsibilities. In the kinds of system that we have typically modeled, permissions tend to be information resources .The activities

Figure 4. i* notation



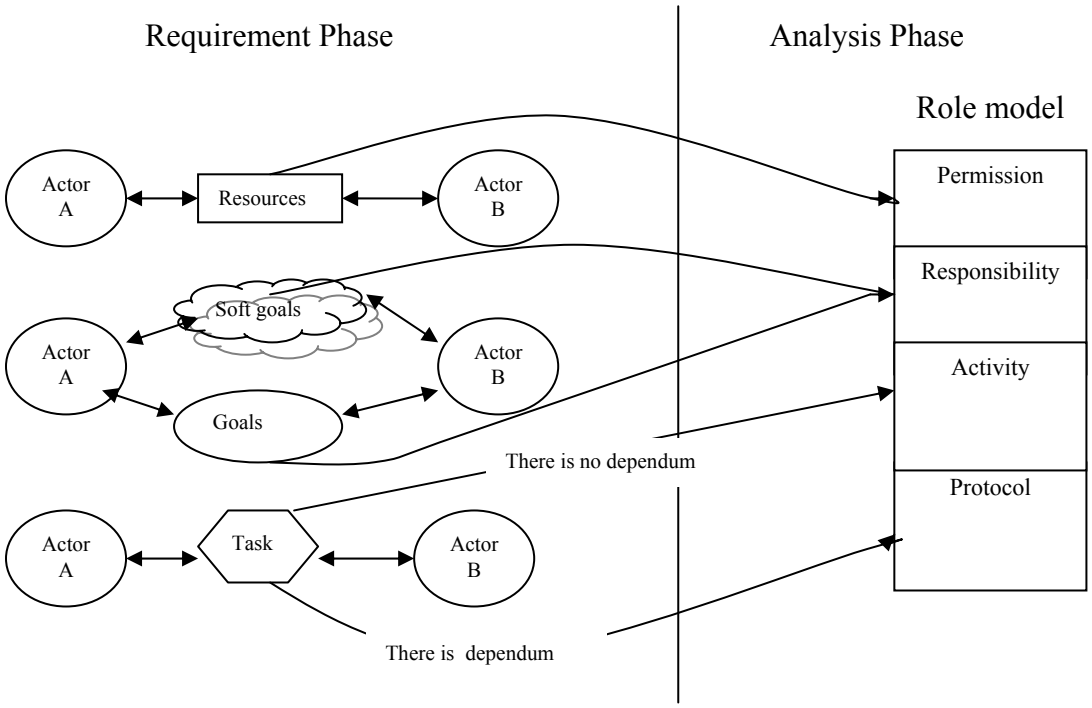
of a role are computations associated with the role that may be carried out by the agent without interacting with other agents. Activities are thus "private" actions, in the sense of. Finally, a role is also identified with a number of protocols, which define the way that it can interact with other roles. [28].

While in Tropos early requirement analysis focuses on the intensions of stakeholders. Intensions are modeled as goals. Through some form of goal oriented analysis, these initial goals eventually lead to the functional and non functional requirements of the system-to-be [24].

In Our proposed methodology we will use the i* notations used in Tropos methodology to analyze the requirements to find the functional and non functional requirements as the first step in the Gaia methodology .

In i* notations actors are represented as circles; dependums – goals, softgoals, tasks and recourses – are respectively

Figure 5. The proposed model



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

represented as ovals, clouds, hexagons and rectangles ; and dependencies have the form depender → dependum → dependee [27].we can summarize the i* notations in the figure(4).

Here is our proposed model figure (5):

In our proposed model we will use the strategic dependency model from tropos as the resource for the requirement phase in Gaia as follows :

1. Resource dependency → permission
2. Softgoals and goals dependencies→responsibilities
3. Task dependencies → Protocol
4. There is no task dependency → activity

REFERENCES

1. Arnon Sturm, Dov Dori, Onn Shehory, 2003, Single-model method for specifying multi-agent systems, ACM Press, pp 121-128
2. Alessandro Garcia, Ricardo Choren, Carlos Lucena, Alexander Romanovsky, Holger Giese, Danny Weyns, Tom Holvoet, Paolo Giorgini, Jul 2005, Software Engineering for Large-Scale Multi-Agent Systems – SELMAS 2005: workshop report, ACM SIGSOFT Software Engineering Notes, issue 4, vol. 30, pp 1-8
3. Ricardo Choren, Alessandro Garcia, Carlos Lucena, Martin Griss, David Kung, Naftaly Minsky, Alexander Romanovsky, 2004,Software Engineering for Large-Scale Multi-agent Systems SELMAS'04, ACM Press , International Conference on Software Engineering Proceedings of the 26th International Conference on Software Engineering, pp 752-753
4. Leon Sterling, Thomas Juan, 2005, The software engineering of agent-based intelligent adaptive systems, ACM Press , International Conference on Software Engineering Proceedings of the 27th international conference on Software engineering, pp 704-705
5. Philippe Massonet, Yves Deville, Cédric Nève, 2002, From AOSE methodology to agent implementation, ACM Press , International Conference on Autonomous Agent Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1, pp 27-34
6. Gregor Engels, Wilhelm Schäfer, Robert Balzer, Volker Gruhn, 2001, Process-centered software engineering environments: academic and industrial perspectives, ACM Press , International Conference on Software Engineering Proceedings of the 23rd International Conference on Software Engineering, pp 671-673
7. Tveit A., 2001 A survey of agent-oriented Software Engineering, www.csgsc.org.
8. Holger Knublauch, 2002,Extreme programming of multi-agent systems, ACM Press , International Conference on Autonomous Agents Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2, pp 704-711
9. Claudine Toffolon, Salem Dakhli, 2000, A framework for studying the coordination process in software engineering, ACM Press , Symposium on Applied Computing Proceedings of the 2000 ACM symposium on Applied computing - Volume 2, , pp 851-857
10. Bergenti F., Gleizes M, Zambonelli F. , 2004, Methodologies and software engineering for agent system, Kluwer Academic publishers.
11. Wooldridge M., Jennings N., 1995, Intelligent Agents: Theory and Practice, www.citeseer.ist.psu.edu.
12. R.S. Pressman, 2005, Software Engineering: A Practitioner's Approach, 6th edition.
13. Estefania Argente Villaplana, 2005, A proposal for an organizational MAS methodology, ACM Press , International Conference on Autonomous Agents Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, pp 1370-1370.
14. Thomas Juan, Adrian Pearce, Leon Sterling, 2002, ROADMAP: extending the gaia methodology for complex open systems, ACM Press, International Conference on Autonomous Agents Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1 , pp 3-10
15. Thomas Juan, Leon Sterling, Maurizio Martelli, Viviana Mascardi, 2003,Customizing AOSE methodologies by reusing AOSE features, ACM Press , International Conference on Autonomous Agents Proceedings of the second international joint conference on Autonomous agents and multiagent systems, pp 113-120
16. Prabhat Ranjan, A. K. Misra, May 2006, A hybrid model for agent based system requirements analysis, ACM Press, ACM SIGSOFT Software Engineering Notes, issue 3, vol. 31, pp 1-7
17. Fausto Giunchiglia, John Mylopoulos, Anna Perini, 2002, The tropos software development methodology: processes, models and diagrams, ACM Press , International Conference on Autonomous Agents Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1, pp 35-36
18. Anna Perini, Angelo Susi, Fausto Giunchiglia, 2002,Coordination specification in multi-agent systems: from requirements to architecture with the Tropos methodology, ACM Press , ACM International Conference Proceeding Series; Vol. 27
Proceedings of the 14th international conference on Software engineering and knowledge engineering, pp 51-54
19. Paolo Bresciani, Anna Perini, Paolo Giorgini, Fausto Giunchiglia, John Mylopoulos, 2001,A knowledge level software engineering methodology for agent oriented programming, ACM Press , International Conference on Autonomous Agents Proceedings of the fifth international conference on Autonomous agents, pp 648-655
20. Mehdi Dastani, Joris Hulstijn, Frank Dignum, John-Jules Ch. Meyer, 2004, Issues in Multiagent System Development, ACM Press , International Conference on Autonomous Agents Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2, pp 922, 929
21. Scott A. DeLoach, 2005, Multiagent systems engineering of organization-based multiagent systems, ACM Press , International Conference on Software Engineering Proceedings of the fourth international workshop on Software engineering for large-scale multi-agent systems, pp 1-7
22. Barbara Bracken, Dec 2003, Progressing from student to professional: the importance and challenges of teaching software engineering, Journal of Computing Sciences in Colleges, ACM, , issue 2,vol. 19, issue 2, pp 358-368
23. Ian Sommerville, SOFTWARE ENGINEERING 6th edition, Addison-Wesley.
24. Dardenne, A.,lamsweerde, A., and fixckas,S., 1993, Goal directed requirement acquisitions. Science of computer programming, pp 3-50.
25. Brinkkemper, J. and Castro J., Tropos: A Framework for Requirements-Driven Software Development , Information Systems Engineering: State of the Art and Research Themes, Lecture Notes in Computer Science, June 2000 Springer-Verlag.
26. Fuxman A., Pistore M., Mylopoulos J., Traverso P., Model Checking Early Requirements Specifications in Tropos
27. Cervenka R., 2003, Modeling Notation Source Tropos, Foundation for Intelligent Physical Agents.
28. Wooldridge J., Jennings N., the Gaia Methodology for Agent-Oriented Analysis and Design

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Controlled semantic tagging – how can topic maps support subject indexing in digital libraries?

Hendrik Thomas, Bernd Markscheffel, Tobias Redmann

Technische Universität Ilmenau,
P.O. Box 100565,
98693 Ilmenau, Germany
+49 (0) 3677 69–3157
[Hendrik.Thomas, Bernd.Markscheffel, Tobias.Redmann]@TU-Ilmenau.de

Abstract Since time immemorial it is the manual indexing which enabled the translation of content subjects in a way which can be sufficiently searched by the reader. This paper deals with the question whether Topic Maps can support the modelling of indexing information as well as the manual indexing process itself. Furthermore we will consider the question how the expressive power of Topic Maps can be used to semantically enhance the subject indexing. In order to answer these questions a Topic Map-based concept for a "controlled semantic tagging" is presented, using a Topic Map based index vocabulary which can be extended collaboratively by users in a controlled process. The objective is to provide indexing information which fulfils high quality requirements regarding the level of detail, correctness and comprehensiveness as well as support the creation of interdisciplinary views on library content.

Keywords Social Software, Digital Libraries, Information Search and Retrieval, Topic Maps

1 INTRODUCTION

Traditional as well as digital libraries are a keystone for the satisfaction of the growing knowledge demands in our society [1]. One of the main challenges in this context is information retrieval providing users with an efficient access to available information resources [2]. Already in 1876, C. A. Cutter demanded that a user has to find documents on subjects and not only via formal metadata [3]. Earlier it was the manual indexing which enabled the translation of content subjects in a way which can be sufficiently searched by the reader [4]. The more accurate, consistent and detailed the indexing process is performed, the better information retrieval can satisfy the user's needs [5].

Lutz Marius Garshol and other authors have shown the potential of Topic Maps[6] in representing, exchanging and merging of index information quite clearly [7,8]. Topic Maps can not only be used to replicate index concepts like thesauri or taxonomies [2,9], it "can go far beyond the possibilities provided by traditional techniques" [7]. However, the discussion in the Topic Map community on these promised enhancements seems to be limited to modeling questions [8,10] as well as to analysis [11] and extraction of informal ontologies from indexing structures [12]. Despite these, it is the indexing process itself, which is a difficult and time-consuming challenge [13]. The involved tasks regarding the identification of the relevant subject of an information resource and the choice of the most suitable index terms have remained constant even in our modern digital information world [5]. As a resulting question, we are asking: Can the Topic Map technology be used to support the indexing process itself?

Furthermore, if we talk about enhancing traditional index techniques by using semantic technologies like Topic Maps, we will have to take a deeper look at the semantic aspects, which are inescapably bound to indexing [4,11]. The semantic interpretation of the relationship between an assigned index term and a document is rather simple. The document contains information on the subject specified by the index term [5]. Other types of relationships and comprehensive context information can not be modeled sufficiently using common indexing techniques [9]. For example in order to index a digitized mechanism model relevant attributes have to be assigned. Thereby, the mechanism does not posses information about his spherical dimension in sense of the traditional index relation, instead it "has the attribute" spherical dimension. This leads to the second question: Can the expressive power of Topic Maps be used to semantically enhance the subject indexing in digital libraries?

Manual indexing is typically performed by trained indexing experts in a controlled process on a high quality level [5]. In the last years the collaborative indexing by users themselves, the so called "social tagging", has become quite successful [14, 15]. A comparison of those two approaches, regarding information retrieval requirements of digital libraries, provides valuable insights for answering the stated research

questions. Based on the conclusions a Topic Map-based concept for a "controlled semantic tagging" (CST) will be proposed as a combination of the features of expert indexing and social tagging. It will be shown how the indexing process can be supported and semantically enriched by applying our concept to the "Digital Mechanism and Gear Library" [16] as an example for application.

2 THE DIGITAL MECHANISM AND GEAR LIBRARY

Before answering the stated research questions a discussion of information retrieval requirements and restrictions, we have to face in digital libraries, is necessary [1]. Take for example the "Digital Mechanism and Gear Library" (DMG-Lib) [16,17]. In this interdisciplinary project of computer experts, information scientists, engineers and librarians, a new digital library is built to preserve and present the knowledge on mechanism and gear science on a new level of quality. The online portal http://www.dmg-lib.org contains a wide range of digitized information resources including full-text and high-resolution scans of many books, journal articles, patents and technical drawings. Additional detailed information about important persons in this field as well as hundreds of digitized mechanism models are available [17].

An appropriate indexing of these heterogeneous resources is obviously challenging [1]. For example, if an engineer searches for a mechanism to open garage doors very fast, he will find relevant information in various sources like a digitized model or a technical drawing on a specific book page. The assignment of a few descriptive index terms to a whole reference book is not sufficient to solve this problem. Garage door opener is a too specific application and it is quite unlikely that an indexer assigns these keywords to a comprehensive textbook. A more detailed indexing of smaller content portions like paragraphs or images is necessary to make them retrievable.

The DMG-Lib is built to satisfy the requirements of different user groups like scientists, engineers or students [16]. The objective is to provide an interdisciplinary and interconnected view on the information resources [18]. An indexing process dominated by the perspective of the engineer experts is not enough [19]. Other science domains like medicine are also interested in this kind of information but use a different domain specific vocabulary and are therefore not able to use index terms intended for engineers.

Furthermore the quality requirements of digital libraries like correctness, consistency and comprehensiveness have to meet the same standards like in traditional libraries [1]. In contrast to the World Wide Web a search is only successful if all relevant documents are found, that contain information on the search subject [13]. For example, it would not be acceptable if the engineer finds a solution for the construction problem in the library, but misses a mechanism which makes the construction less costly. It is even more problematic, if he finds an image of a mechanism, which is indexed with "fast" and "garage door opener". Especially in this domain, images are very limited in showing all features of a complex mechanism. The correctness of index terms therefore is very important. If the feature "fast" was incorrect assigned and the engineer construct the mechanism which turns out to be not very fast, it will cost money.

This brief discussion of the DMG-Lib has shown that the indexing process in a digital library has to fulfil high quality requirements concerning the level of detail, correctness and comprehensiveness. Different interdisciplinary views on the resources have also to be provided to make the available information retrievable for different user groups.

3 MANUAL CONCEPTUAL INDEXING

The main objective of indexing is to construct a surrogate of a content subject by assigning index terms like subject headings to a specified information resource which is being indexed [5]. The indexing process has been performed intellectually by humans for a long time [20]. Recently more and more automated indexing systems have been developed [5], however in this paper we will only focus on manual indexing. According to Lancaster, the subject indexing process involves two main steps: conceptual analysis and translation [4].

Conceptual analysis involves deciding on what a resource is about, mean to identify the relevant content subjects. Note that the result heavily depends on the needs and interests of the person who indexes a resource and therefore quite different aspects of a document subjects can be relevant [5]. The translation is the process of finding an appropriate set of index terms that represents the results of the conceptual analysis. One basis problem involved in translation is the choice of an appropriate index term. The index terms have to represent the subject clearly. Consistency among different indexers is important because users have to predict the assigned index terms in order to be able to search appropriately [5]. Synonyms, spelling problems and homonyms/homographs are common problems in the translation process [21].

3.1 Expert indexing

The indexing process in libraries is typically performed by indexing experts who are familiar with the specific field of knowledge. In order to avoid inconsistent representations of subjects commonly a controlled vocabulary is used for the translation phase [5,20]. Popular examples of controlled vocabularies are the Library of Congress Subject Headings (LCSH) or the Medical Subject Headings (MeSH). A controlled vocabulary can be defined as a systematic and standardized (in terms of usage, spelling and form) selection of preferred index terms, often together with definitions and some semantic structure. As C. A. Cutter stated, the indexer should not choose freely between relevant index terms - he has to choose the index term of the controlled vocabulary, which fits best to represent the content subject [3, 13].

Furthermore the indexer has to adopt "a neutral stand between the reader and the document, giving emphasis to what the author intended to describe rather than to his own view" [21]. The acceptance and prioritization of the author's intent

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

as the way of indexing is one of the basis philosophies of expert indexing, which enables a clear and unambiguous translation of the content subjects. Expert indexing has always been a time-consuming task and requires vast background knowledge to perform the indexing process appropriately [13].

3.2 Social Tagging

In the online world of digital information automated indexing and ranking algorithms have become the basic tools for content analysis. However, in the recent years there is more and more a renaissance of manual subject indexing in terms of the so called social tagging [14,22]. Quite different names are used in literature that refers to this concept like collaborative tagging, social classification, social indexing and folksonomy [23]. It is the basic principle that subject indexing is conducted collaboratively by the end users instead of experts only [12]. The index terms are referred to as tags. The user can choose any tag he likes in order to represent the content subject in natural language [5]. The user "simply creates and applies tags on the fly" [21]. Popular examples are del.icio.us, an online bookmaking system or flickr, an online picture service.

The social aspect of social tagging is one of the most important features. "An author is an authority when it comes to what she intended her work to be about, but not about when [...] it means to others" [21]. During the information retrieval process, the interpretation of the meaning of the information resource for a searcher can be more important than the author's intentions. The different linguistic and cultural backgrounds of the users as well as the valuable individual interpretations of the resources are the inherent strengths of social tagging [22]. The social group can rely on the information and different views of their members enabling different views on the data. In contrast to a single individual expert, who has a limited capacity of information and "blind spots" [24]. However, adding enough of those individual interpretations can lead to inconsistencies, because tags allow contradictions to exist [21]. It is like Friedrich Nietzsche once said "there are no facts, only interpretations" [25].

In contrast to expert indexing, social tagging is unsystematic and might be regarded as not sophisticated from a librarian's point of view. Problems like synonyms or different spellings of tags can lead to inconsistency and ambiguous indexing [21]. Furthermore, the so called "Meta Noise" of inadvertent, irrelevant and inaccurate tagging information is a critical problem. Despite all these facts social tagging provides a flexible method for indexing which has proven to be a great support for users in searching the digital information space [14].

3.3 Discussion

Comparing expert indexing and social tagging, we can note that in both approaches index terms are assigned manually to the documents as a surrogate for the content subject [21]. The common semantic interpretation of an index association is in both accounts equivalent which means that a document contains information about a subject specified by the index term. Other types of relationships can not be modeled flexibly. Note that in social tagging these problems are recognized and as workaround users quite often create tags including information about the association type, e.g. "spherical gear (definition)" which can be interpreted as the document "is a definition" of "spherical gear".

There are, nevertheless, some obvious dissimilarities. First, we can note that indexing experts commonly use a controlled vocabulary in contrast to the natural language indexing in social tagging. The controlled vocabulary enables a consistent and predictable indexing but lacks flexibility [5]. Social tagging enables the users to choose any index term freely which supports the creation of interdisciplinary views [22]. This is especially important for digital libraries like the DMG-Lib which is build by and for engineers [16]. A predefined controlled vocabulary can serve the needs of this user group but it is obviously impossible to predict all possible interpretation of the documents a priori, e. g. for medical or biological scientists.

A second difference relates to the level of detail. Commonly in expert indexing only a few terms are assigned to a whole document to summarize its content [13]. However, based on the intrinsic motivated labor of the user, social tagging can be very detailed [21]. The level of detail is critical for digital libraries. For example in the DMG-Lib over 1000 detailed images and technical drawings in various books are available [16]. They are highly relevant for engineers. Indexing of the whole document is simply not enough. The subject of every content portion has to be retrievable.

Finally, and most important is the difference between the indexing quality of the two approaches. Social tagging has to deal with various language problems which lead to a confusing and inconsistent indexing [21,22]. Commonly these problems are solved due to the amount of users. The assignment of wrong or useless index terms by some users usually has no huge impact, because of the high number of users assigning useful terms which will be more likely recognized by the searcher [24]. In highly specialized digital libraries like the DMG-Lib no such high numbers of users are typically found. Wrong or inconsistent tagging can have a tremendous negative impact on the quality. Summarizing a traditional expert based indexing is more time consuming and in some aspects limited, but using a controlled vocabulary results in consistency and accurate indexing information.

Given the significant disparity between the two approaches, one is tempted to conclude that only a combination of both concepts is suitable for the needs of digital libraries. A flexible and detailed indexing is required to make all available information accessible for the user. Additional appropriate control methods are necessary to ensure the highest possible indexing quality. Overall a digital library has the responsibly to provide at least the same quality as we expected from traditional libraries.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4 CONTROLLED SEMANTIC TAGGING

Based on these conclusion we developed and prototypically implemented a concept for a "controlled semantic tagging" (CST). The key idea of CST is, that the subject indexing process can be enhanced due to the usage of a Topic Map [6] based controlled index vocabulary which can be extended cooperatively by users in a controlled process. Instead of traditional index terms, topic nodes in the semantic network which include information about alternative names and relevant relations are used to represent content subjects unambiguously [7]. Additionally, the user is able to specify the type of relation between the document and the chosen index term. Therefore not only the standard index relation "contains information about" but also other suitable relations like "is attribute of" or "contains references for" can be used. Considering the high quality requirement of digital libraries a pre and post control mechanism is included.

4.1 Control mechanism

As discussed in section III every user can have an individual valid interpretation of a content subject, which is the basis to create multiple and interdisciplinary views [22]. This leads to the first assumption of our concept: Free choice of tags for private usage. Similar to social tagging, users are able to create index terms and association types in natural language easily on the fly. Such individual tags are helpful for the expert, but it does not necessarily mean that it will be valuable for the average user, in terms of clear and consistent indexing [21]. Therefore, our second assumption is: All tags have to be controlled before being published to the community. Such a control has to be performed on two problem fields. First, we have to ensure the correctness and domain relevance of the created tags to avoid the so called "meta noise". On a second level we have to make sure that the documents are indexed correctly. To solve these problems we can consider the trustworthiness of a user who naturally possesses different levels of domain and indexing experiences. For example the indexing of an engineer is more likely to be correct than the indexing of a student. In our concept we propose a weighted tagging approach based on trustworthiness of the user. In the DMG-Lib project we subdivide the users in three groups:

- Novices: any user who has registered to the system. They do not need any special domain knowledge, however they are interested in this field of application. List A-level, without bullet but indented, is used for continuation of A-level lists
- Domain expert: all users who possess above-average knowledge of the domain like a professor of mechanism science or an engineer.
- Indexing expert: this group of persons possesses domain knowledge as well as excellent knowledge about the indexing process itself.The impact of the different knowledge levels on the control process is naturally very domain specific. In our use-case we decided that created tags and index information of the indexing expert group are instantly published. A domain expert possesses excellent domain knowledge, therefore we can trust his judgment in the conceptual analysis and translation phase. However, the creation of new tags has to be controlled. Tests in our library have shown that domain experts tend to use their own specific vocabulary for indexing. For example the name "Koppelgetriebe" is commonly used in East Germany and the synonym "Kurbelgetriebe" for the same subject is typical for West Germany [16]. As a result, all created tags of domain experts have to be reviewed by the indexing expert group, to ensure a consistent indexing and to avoid the creation of tags which represent the same subject. Consequently the novices group has the lowest trustworthiness because their experience and knowledge on the domain are unknown and therefore all information has to be reviewed.

4.2 Creation of a domain specific controlled vocabulary

A domain specific and detailed Topic Map based vocabulary, the so called index topic map, is the basis for our concept. In our concept every index term including all available alternative names are modeled as topic nodes. Additional all relevant relations between the index terms like broader or narrow term relations are modeled, too.

Probably the most suitable approach for creation is to reuse standardized vocabularies, because indexers as well as searcher are already familiar with this set of indexing terms [5]. In the scope of the DMG-Lib we imported the IFToMM dictionary containing over 9000 technical terms in four languages relevant for mechanisms and gears [26].

Despite the manual creation of a vocabulary by domain experts, obviously the back-of-the-book indexes are a relevant source. By definition they are organized collections of relevant terms of a book, enabling the reader to find specific text passages [13]. Every term in the back-of-book index is a potential new index term for the index topic map. However, an automated extraction into topics can not be applied. We have to keep the basic rule of the Topic Map paradigm in mind: every subject is represented by exactly one topic node [18]. Therefore, in a special workflow the experts indexer group has to evaluate and assign every term to the corresponding topic.

4.3 Initial set of indexing information

As you can see on social tagging systems on the Internet, the critical mass of tagging information is one of the most important base for success. The user aspects that for every tag at least a few relevant documents are available, otherwise the systems are not very helpful and do not motivate the user to participate. Bear in mind that user groups of digital libraries are typically relative small which tend to result in fewer tagging information. To solve these problems, the back-of-book indexes can be used to provide the users with an initial set of high quality index information. By definition to every index term several relevant book pages are assigned [13]. These information can be automatically imported as tagging data with the highest trustworthiness, because the author itself had decided which page is relevant.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

4.4 The Controlled Indexing Process

Based on a Topic Map based vocabulary the controlled indexing process can be performed in the following steps:

1. Selection of information resources for indexing

   Based on the requirements for a high level of details whole resources as well as sub items of the resource can be indexed. In our prototype for the DMG-Lib [16] whole books, specific document pages as well as individual images can be tagged. In addition, the available curriculum vitae including important events as well as digitized mechanism models can be indexed.
2. Search for an appropriate index term

   The conceptual analysis is not supported by our concept directly, because it is primarily a cognitive process depending on the individual knowledge and needs of the user [5]. The main strength of our concept lies in the translation phase. During translation the user is supported in choosing an index term which fits best to represent the identified content subject. After typing a keyword, a pre-search in the Topic Map based controlled vocabulary is performed to identify a suitable topic, which represents the subject described by the keyword unambiguously. Synonyms as well as translations are modeled around the topic enabling a precise search. Furthermore, if a known homonym was identified the user can choose between the different meanings based on the semantic information modeled in the index topic map. If no suitable topic is found, a similarity search in the controlled vocabulary will be performed. For example if the user simply misspelled the index term.
3. Creation of new index terms

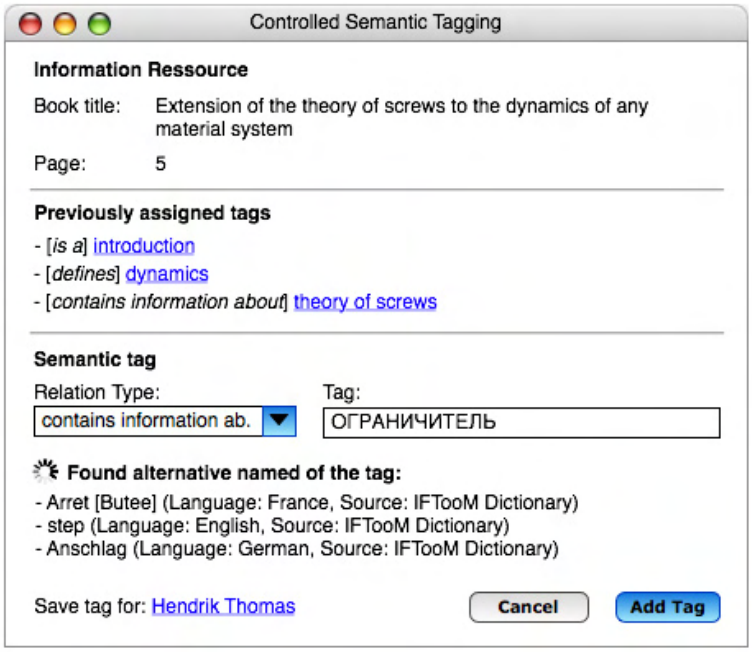
   If no available tags are suitable, the user has the opportunity to create a new index term which means a new topic node with the inserted keyword as basename [6]. Depending on the trustworthiness of the user, the new index term is instantly visible for the community or only for the specific user.
4. Display information about the subject

   If the user has chosen an index term of the controlled vocabulary, all available information about the subject including alternative names and context information as well as related topics are displayed (see figure 1). The objective of this phase is to ensure that the user has chosen the correct and most suitable index term. Displaying related topics will help the user to decide whether a broader or narrower term existed and if probably these are more suitable.
5. Selection of the index relation type

   Additionally, in our concept users can define the association type, in terms of explicit specifying the relationship between the current resource item and the chosen index term. Topic Maps are perfectly suitable for this task, because simply a typed association between the index topic and a topic representing the document item has to be created like "contains information about" or "is attribute of" [18]. If the available association types are not sufficient for the individual needs, the user can create a new type on the fly.

The described prototype of the CST concept is currently internal tested in the DMG-Lib project [16] and will go online in August 2007 during the rollout of the updated web portal.

Figure 1. Controlled Semantic Tagging Interface



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

5 CONCLUSIONS

In this paper we presented a concept which combines aspects of expert indexing and social tagging to provide assistance in the indexing process. However the support is limited on the translation phase where users are supported in identifying the best suitable index term. By using a Topic Map based controlled vocabulary problems like synonyms, homonyms or misspelling can be avoided. The stored semantic information can support the navigation and selection process. A flexible and collaborative expanding of the Topic Map in a controlled process supports the generation of multiple views on the information resources. The high quality requirements result in a high level of control. In our further research we need to develop methods to reduce the control efforts for example by an automated evaluation of tagging information using available semantic information [9, 27]. For answering the second research question, we have shown how Topic Maps can be used to include association types in the index process providing the users with a powerful tool to support the organization of information. Overall only a long term empirical analysis will show whether CST is able to support and improve the manual indexing on a new level of quality.

REFERENCES

1. Tedd L. A. and Large A. J. (2005), 'Digital Libraries - Principals and Practice in a Global Environment', K. G. Saur, New York.
2. Rasmussen E. (2004), 'Information Retrieval Challenges for Digital Libraries', Proceedings of the 7th International Conference on Asian Digital Librariesm (ICADL'04), Springer, Shanghai-New York, p. 93–103.
3. Cutter C. A. (2007), 'Library Systematizer', Libraries Unlimited Inc., U.S.
4. Lancaster F. W. (2003), 'Indexing and Abstracting in Theory and Practise', 3rd ed., facet publishing, Champaign, Illinois.
5. Chowdhury G. G. (1999), 'Introduction to Modern Information Retrieval', Library Association, London.
6. Garshol L. M. and Moore G. (2006), 'ISO/IEC JTC1/SC34, Information Technology – Document Description and Processing Languages, Home of SC34/WG3 Information Association; http://www.isotopicmaps.org/sam/sam-model/ (2007-07-15).
7. Garshol L. M. (2004), 'Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all', Journal of Information Science, 30(4), p. 378-391.
8. Ahmed, K. (2003) 'Beyond PSIs - Topic Map design patterns, Proceedings of the Extreme Markup Languages Conference (EML'03), Montral, Canada; http://www.mulberrytech.com/Extreme/Proceedings/html/2003/Ahmed01/ EML2003Ahmed01.html (2007-07-10).
9. Thomas H., Markscheffel, B. and Brix T. (2007) 'SIREN – a Concept for a Semantic Information Retrieval Environment for Digital Libraries", Proceedings of First international Workshop on Knowledge Media Science - Information Access and Media Technology, Springer, October 2-5 2006, Landsberg Castle, Meiningen, in press.
10. Assem M. v. et. al. (2006), 'A Method to Convert Thesauri in SKOS', Processings of the 3rd European Semantic Web Conference (ESWC 2006), Springer, Budva, Montenegro, p. 95–109.
11. Miller T. and Thomas H., "Indices, Meaning and Topic Maps: Some Observations", Leveraging the Semantics of Topic Maps – Second International Conference on Topic Map Research and Applications (TMRA 2006), Springer, Leipzig, Germany, October 11-12, 2006, p. 130–139.
12. Park J. (2006), 'Tagomizer: Subject Maps Meet Social Bookmarking', Leveraging the Semantics of Topic Maps – Second International Conference on Topic Map Research and Applications (TMRA 2006), Springer, Leipzig, Germany, October 11-12, 2006, p. 200–214.
13. Fugmann R. (1993), 'Subject analysis and indexing: theoretical foundation and practical', advice, Frankfurt am Main.
14. Trant J. and Wyman B. (2006), 'Investigating social tagging and folksonomy in art museums with steve.museum', Proceedings of the WWW 2006 Collaborative Web Tagging Workshop; http://www.archimuse.com/research/

www2006-tagging-steve.pdf (2007-07-10).

15. Szomszor M. et. al. (2007), Integrating Folksonomies with the Semantic Web', Proceedings of 4th European Semantic Web Conference, Bridging the Gap between Semantic Web and Web 2.0, Springer, Innsbruck, Austria, June 2-7, p. 624-639.
16. Brix T. et. al. (2006), 'The Digital Mechanism and Gear Library: a Modern Knowledge Space', Knowledge Media Technologies – Proceedings of the First International Core-to-Core Workshop, Castle Dagstuhl, Germany, Diskussionsbeiträge des IfMK, p. 45-52.
17. Brix T., Döring U., Trott S. (2005), 'DMGLib - ein moderner Wissensraum für die Getriebetechnik", in: Proceedings of the Knowledge eXtended conference (KX'05), Verlag Jülich, Jülich, Germany, p. 251–262.
18. Pepper S. (2002), The TAO of Topic Maps; http://www.ontopia.net/topicmaps/ materials/tao.html (2007-06-12).
19. Weinberger D. (2005), Tagging and Why It Matters, Harvard Berkman Center for the Internet and Society, http://cyber.law.harvard.edu/home/uploads/507/07-WhyTaggingMatters.pdf (2007-07-10).
20. Salton G. and McGill M. J. (1983), 'Introduction to Modern Information Retrieval', McGraw-Hill, New York.
21. Peterson E., (2006) 'Beneath the Metadata – Some Philosophical Problems with Folksonomy', D-Lib Magazine, 12(11), http://www.dlib.org/dlib/ november06/peterson/11peterson.html (2007-07-10).
22. Voß J. (2007), 'Tagging, Folksonomy & Co - Renaissance of Manual Indexing?' http://www.citebase.org/abstract?

id=oai:arXiv.org:cs/0701072 (2007-07-10).

23. Hotho A. et. al., 'Information Retrieval in Folksonomies: Search and Ranking', Processings of the 3rd European Semantic Web Conference (ESWC 2006), Springer, Budva, Montenegro, p. 411–426.
24. Surowiecki J. (2005), 'The Wisdom of Crowds', Anchor Books, New York.
25. Eco U. (1997), 'Kant and the platypus, essay on language and cognition', Harcout Brace, New York.
26. (2007), IFToMM Dictionary, http://www.ocp.tudelft.nl/tt/cadom/IFToMM/web/ online/index.html (2007-07-10).
27. Guy, M. and Tonkin, E. (2006), 'Folksonomies: Tidying up tags?', D-Lib Magazine, 12; http://www.dlib.org/dlib/january06/guy/01guy.html (2007-06-10).

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

The wiki way of knowledge management with topic maps

Tobias Redmann, Hendrik Thomas

Technische Universität Ilmenau,
P.O. Box 100565,
98693 Ilmenau, Germany
+49 (0) 3677 69–3157
[Tobias.Redmann, Hendrik.Thomas]@TU-Ilmenau.de

Abstract Wikis have proven how fast and easy knowledge management can be. In the following paper, we will discuss similarities between Topic Maps and the wiki concept. Based on the conclusions we will present design principles for a Topic Maps based Wiki. The objective is to provide a "wiki way" of Topic Map creation in order to enhance semantic knowledge management.

Keywords Knowledge Management, Social Software, Information Search and Retrieval, Collaborative research, Internet Technologies

1 INTRODUCTION

The Hawaiian word "wiki'" means quick and represents, like no other buzzword, the basic needs of our society for a fast and easy to use capturing and managing of knowledge [1]. Wikipedia, the most comprehensive encyclopedia of the world [2] proved that wiki systems can support knowledge management on new level of quality and quantity.

Looking at the Topic Maps community this simplicity is missing. Standards and concept have been successful developed, but a free and easy to use editor for a collaborative modeling of knowledge in Topic Maps is still missing. Creating and maintenance of topic maps is still insufficient supported. Consequently only very few useful topic maps exists, compared to the high amount of ontology modeling projects using RDF, for example SKOS [3] or DMOZ [4].

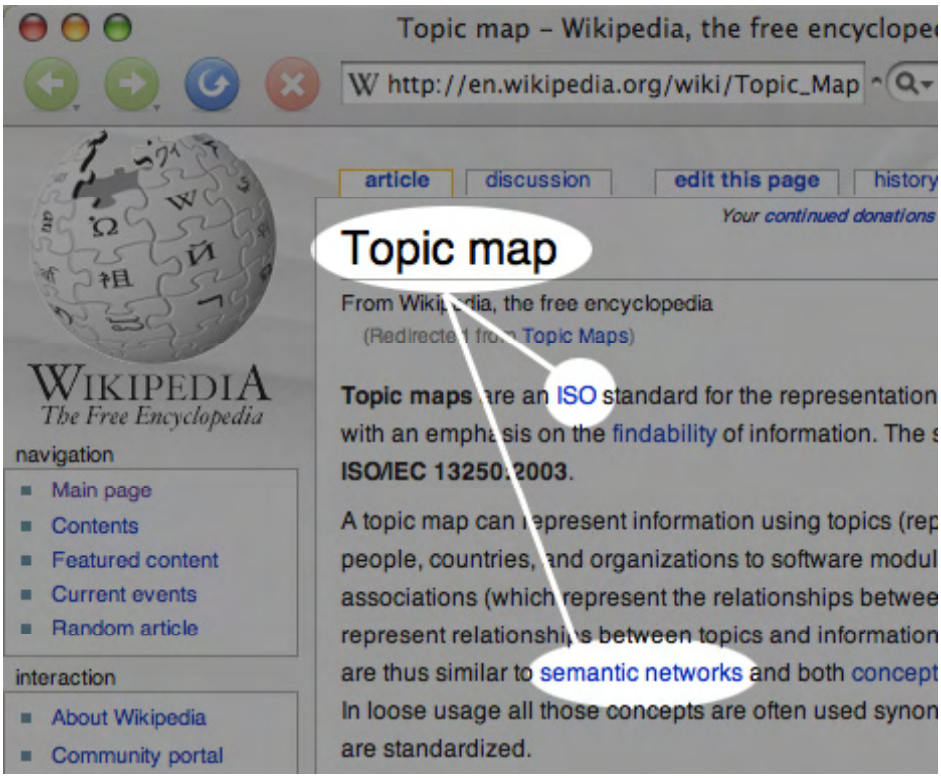
As result, we place ourselves the quesition whether the wiki concept of free and easy modeling of information can be used to support the collaborative modeling process in Topic Maps. In the following section we will discuss similarities between wikis and Topic Maps. Based on the conclusion we will present an approach, how to use wikis for topic map editing. The paper concludes with a summary and an outlook on open questions.

2 WIKIS VERSUS TOPIC MAPS

"A wiki is a web-based software that allows all viewers of a page to change the content by editing the page online in a browser. This makes the wiki a simple and easy-to-use platform for cooperative work on texts and hypertexts." [5] Therefore a wiki provides the following basic features [1]:

- Every wiki user can add or change any wiki page. There are no restrictions or rules. Note, that some wikis allow the locking of specific pages or other access restrictions. However, this violates the basic philosophy on the wiki concept of free editing.
- Changes are stored to prevent malpractice and vandalism - every state of the wiki text can be restored.
- A simple syntax (wiki markup) for text formatting (e. g. for headers, lists and links) is used to support fast and easy knowledge capturing. Currently, there is a standardization process on the go with the objective to establish a standard wiki markup [6]. However, in practice the markup of MediaWiki (http://www.mediawiki.org), which is used for Wikipedia, emerged as the informal standard.
- Wiki links are used to link wiki pages and associate relevant subjects. Wiki links can also be used to link external resources with relevant information about the current subject like websites, images or documents. Any wiki page possesses exactly one unique name, which unambiguously identify the page and therefore the subject of the wiki page. If synonyms of a subject exist, simply for any name a wiki page is created and forwarded to the wiki page with the default name. Homonyms are handled using a meta wiki page for

Figure 1. Wikipedia article on Topic Maps shown as a topic map



disambiguation, on which every known meaning of the word including a textual explanation is listed.

With regard to this, any wiki page captures information about a specific subject. Links between different subjects point out a relevance between two subjects. Links to external resources, like websites or images, mark relevant information resources for the subject.

These concepts can also be modeled with Topic Maps. Topic Maps are a knowledge representing device, which allows the structural modeling of domain knowledge as well as the organization of information resources [7].

Regarding to the Topic Maps Data Model (TMDM) [8], in a topic map one topic is a proxy for one subject. Topics enable the creating of statements about the subject and association for modeling relations between subjects. Occurrences are used to link subject-relevant information resources, like websites or images. Additional associations, occurrences and names can have a type and scope.

Comparing Topic Maps to the wiki concept, we can note that the objectives of both accounts are similar - explicit capturing of knowledge. Furthermore there are other similarities between the two concepts:

- Subject-centric information capturing: wiki page vs. topic.
- Link relevant information resources: external links vs. occurrences.
- Links between relevant subjects: wiki links vs. associations.For example the Wikipedia article on "Topic Maps" contain internal wiki links to other subjects, e.g. "ISO" or "semantic networks". These subjects are in some way relevant for the subject "Topic Maps". There are also "external links" to relevant websites available, e.g. the "XML Topic Maps 1.0 Specification".

Seen from the Topic Maps point of view, we can speak about associated subjects and occurrences. The three subjects are "Topic Maps", "ISO" and "semantic networks" - each represented by a topic. There is an association between "Topic Maps" and "ISO" as well as between "Topic Maps" and "semantic network". Additionally the topic "Topic Maps" have occurrences, e.g. a link to the "XML Topic Maps 1.0 Specification".

Summarizing links to other wiki pages can be interpreted as an association between two topics and external links as occurrences of the current topic. These obviously similarities leads to the following approach.

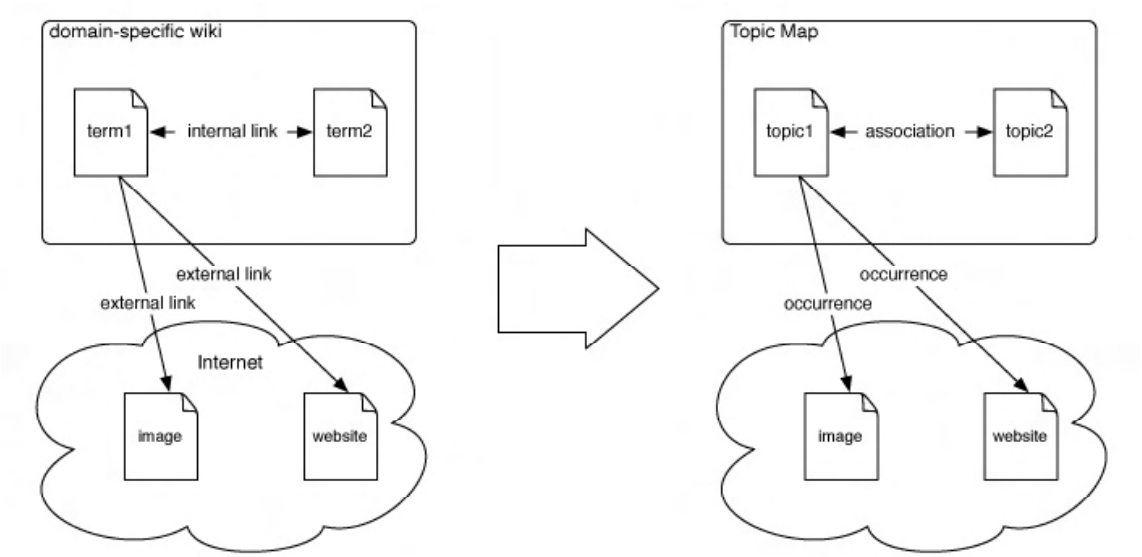
3 MAPPING THE WIKI CONCEPT TO TOPIC MAPS

First lets figure out some predefinations regarding a common sense of Topic Maps and the wiki concept. Each wiki page represent one subject. This agrees with the basic rule "one topic per subject" [7,9]. Adding information to the wiki page means to make statements about the corresponding subject [10].

Every wiki page can have one or more names and they are used to identify the wiki page. One name will be marked as default name. The others indeed exist, but only as a proxy for the default name. The names identifies the topic, there can't be two topics with the same name - this corresponding with the TMDM [8]. Topics with the same name in the same scope will be merged.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. Wiki to Topic Maps transformation



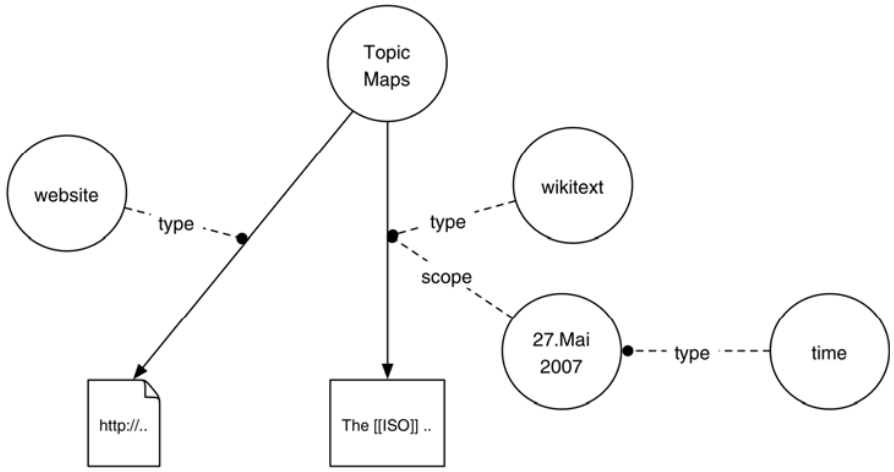
The main challenge is, how to model a whole wiki system with a simple topic map and to transform wiki markup to the Topic Maps concepts. Keep in mind some of the main functions: a permanent storage of all changes and easy information capturing using a simple wiki markup [1]. The complete transformation is shown in figure 2 and will be explained in the next sections.

3.1 Using Topic Maps to store a Wiki System

According to the predefinition from above, every wiki page represent one subject as a topic in a topic map. Information about this subject will be captured using the concept of inline occurrences. This special occurrences save the wiki text in wiki markup. To store the changes, the concept of scopes will be used. Therefore, with every change a new topic representing the time (as timestamp) will be added to the map and used as scope for the occurrence. Every change will result in a new occurrence and a new (time) topic. This method allows to keep track of every change - only the wiki text with the latest time (scope) will be used for presentation. Figure 3 shows the model of this method.

The circles represent topics, the solid arrows are occurrences and the dashed lines connect topics with corresponding types or scopes. The subject "Topic Maps" has two occurrences. The first is an occurrence to an external website (typed as "website"). The other occurrences is used to store the wiki text. This occurrence is scoped (with "time") and typed as "wikitext". How to get the occurrences to external sites will be explained in the following section.

Figure 3. The model for a Topic Maps Wiki



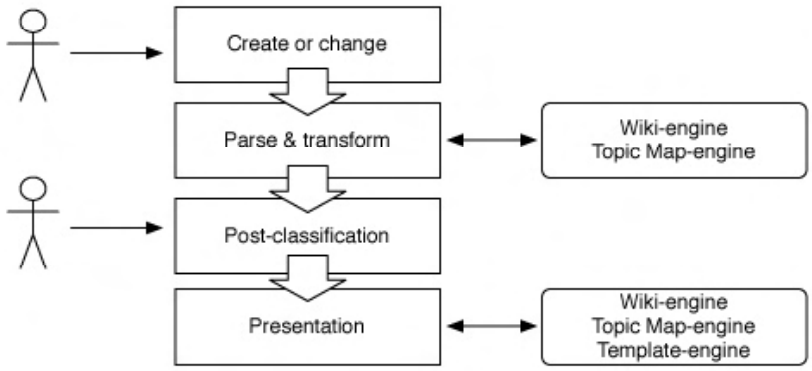
3.2 Wiki Markup to Topic Maps Transformation

The most difficult part is the transformation from wiki markup to Topic Maps concepts. Currently there many wiki systems with individual markups available [11]. The Creole Project [6] tries to define a common and standardized wiki markup. Regarding to the success of Wikipedia, the MediaWiki markup is wide spread. Various wiki engines can be used to parse and render wiki markup to HTML, for example plog4u (http://www.plog4u.org) and Radeox (http://www.radeox.org). These engines can be used to extract internal and external wiki links from the wiki text. These links and their corresponding names can used to create topics, occurrences or associations between topics.

Extracted external links will be used to create occurrences for the topic - extracted internal links will create an association between the current and the extracted topic. If the linked

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. Knowledge Capturing Process



topic, identified by its name, doesn't exist it will be created with the extracted link name.

Regarding to this, adding external links to the wiki text, mean to add occurrences to a topic. Removing the link means removing the occurrence. By adding a link to another topic, an association between both nodes will be created. These associations may, but not have, to be linked by one of both topics. This means, that only in one wiki text this association is emphasized. Removing the association in one topic's wiki text, doesn't indicate to remove the association permanently. Only if both topics doesn't link each other, the assocation have to be removed.

3.3 Topics as Metadata and Topics as Subject Proxies

Another challenge are the topics themself. Some topics describe domain specific subjects and some others are just used to describe "metadata". For example there are topics needed to describe scopes and occurrence types. These topics are not important or needed for the domain or the subjects. The challenge is to separate these topics types and register only relevant topics. This means, that only domain specific subject representing topics have to use for knowledge encoding. The wiki will use some internal topics, identified by PSIs, to describe scopes and types.

Regarding to the example above. Not each timestamps, used to indicate the time when texts were changed, is relevant for the ontology. To get a compact and only domain specific topic map, we use a "working map" to store all changes and metadata as well as a "cleaned map", that only contains the encoded knowledge. This means the "cleaned map" only contains topics, associations and occurrences.

Now let's take a detailed view on the resulting knowledge capturing process.

4 KNOWLEDGE CAPTURING PROCESS

Regarding to the wiki concept, the knowledge capturing process have to be simple and easy to use [1]. New users as well as experts have to be able to use the same system efficiently. Figure 4 shows the whole knowledge capturing process.

4.1 Knowledge Creation

The first process step describes the human driven task of writing wiki text, linking to relevant subjects and adding information resources to the current subject. A user needs only a standard webbrowser to explore the wiki. All wiki pages are generated as HTML. To add a new wiki page an interactive dialog will be provided. This dialog contains an interface for adding a name and wiki text. Saving the page will lead to the next step.

Additional an user can edit an already existing wiki pages. By using the "Edit this page" link an interactive dialog will be provided. Saving the page will also lead to the next step. A new wiki page can be added to system by creating a new page or by adding a link to an other non-existing subject which will be generated automatically by the system.

4.2 Wiki Markup to Topic Maps Transformation

The user generated content from step 1 will be processed and relevant wiki link constructs will be extracted. The used wiki engine provides functions to get internal links with names and external links. These constructs will be transformed in topic map elements according to the process described in section 3.2.

4.3 Post Classification

To model the complex semantic information in Topic Maps, like scopes, association types and roles, basically two different approaches are suitable for this task.

First, we can extent the available wiki markup with specialized constructs. For example the standard construct for associations in Mediawiki could be altered, allowing users to add a association type like "definition of:Topic Maps". However, this approach has one major disadvantage. It forces the user to learn new construct. The wiki markup have always been one of the most important entrance barrier to the usage of wikis. Any change make the input process more difficult for the user.

Therefore we decide to stick to the standardized wiki markup to modeling the basic topic map information, like a association "[[Topic Maps]]". A default type and roles are assigned

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 5. Wiki to Topic Maps transformation

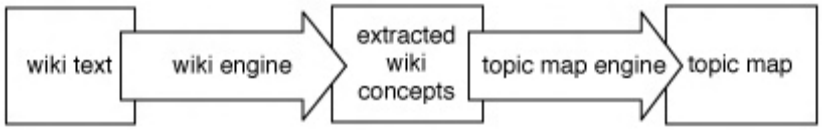
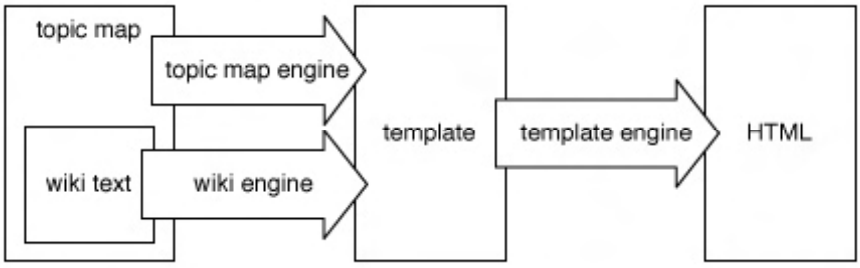


Figure 6. Knowledge Presentation



automatically. In a second step, a post classification is performed, which allows the user to model additional semantic information in a dialog interface. For example the user have the opportunity to select another association types in a menu. He can model as much information as he need for his individual needs - but don't have to. For interoperability [12,13] of the modeled knowledge, the user can also add one or more PSIs to the current topic to indicate the subject more explicit. The post classification can also be skipped by a novice user.

This process step is necessary because we want to reuse the standardized wiki vocabulary without adjustment. In further developments the Topic Maps Constraint Language (TMCL) [14] can be included - to systematic predefine valid roles, and scopes for specific association types.

4.4 Knowledge Presentation

The encoded knowledge and captured information have to presented to the user. Thus the wiki engine renders the wiki text to HTML, but there are more information contained in the topic map.

Information about related subjects, not captured in the wiki texts, can also be presented - see section 3.2 These information have to be generated by the topic map engine. All data will be submitted to the template engine (like Velocity - http://velocity.apache.org) which supports an template language for easy changing the layout of the whole wiki system. All generated information will be merged and rendered as HTML.

5 CONCLUSIONS AND FUTURE WORK

Wikis proved how fast and easy knowledge management can be. In this paper we discussed similarities between Topic Maps and the wiki concept. Based on the conclusions we presented design principles for a Topic Map based Wiki. A combination of both concepts, provides a fast and easy knowledge capturing process in a "wiki way". Experts and novices can easily and in a pragmatic way create topic maps. This can possibly be act as a catalyst for the Topic Maps technology.

There are a lot more concepts in current wiki markups. For example, some systems also provide categories or group concepts. Using these additional features can enhance the expression of the captured subjects and their associations, e.g. super- and subclass relations.

REFERENCES

1. Leuf, B. and Cunningham, W. (2001), 'The Wiki Way: Quick Collaboration on the Web', Addison-Wesley
2. Möller, E. (2005), 'Die heimliche Medienrevolution: Wie Weblogs, Wikis und freie Software die Welt verändern', Heise
3. Assem, M. v., et. al. (2006), 'A Method to Convert Thesauri in SKOS', in: Processings of the 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, Springer, 95–109
4. DMOZ – Open Directory Project (2007), http://dmoz.com/
5. Ebersbach, A. and Glaser, M. and Heigl, R. (2005), 'Wiki: Web Collaboration', Springer
6. Smith, C. (2007), 'Wiki Creole Press Release', http://www.wikicreole.org/wiki/WikiCreolePressRelease, Accessed: 4th July 2007
7. Pepper, S. (2000), 'The TAO of Topic Maps', Proceedings of XML Europe
8. (2006), 'ISO 13250-2 Topic Maps Data Model'
9. Park, J., Hunting, S. (2003), 'XML Topic Maps: Creating and using topic maps for the web', Pearson Education Inc., USA
10. T. Miller and H. Thomas (2007), 'Indices, Meaning and Topic Maps: Some Observations', in: L. Maicher, A. Siegel and L. M. Garshol, Leveraging the Semantics of Topic Maps – Second International Conference on Topic Map Research and Applications, TMRA 2006, 130–139.
11. Wikimatrix - compare them all (2007), http://www.wikimatrix.org
12. Pepper, S. (2002), 'Ten Theses on Topic Maps and RDF', Ontopia, http://www.ontopia.net/topicmaps/materials/rdf.html
13. Pepper, S. (2007), 'The Case for Published Subjects', Ontopia, http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf
14. (2005) 'Topic Maps Constraint Language', http://isotopicmaps.org/tmcl/tmcl-2005-02-12.html

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

i·Society 2007

Data attribute selection with genetic programming

Joel Hickman

Gina Hope

Taehyung (George) Wang

Department of Computer Science
California State University Northridge
joel.hickman@csun.edu
gina.marchetti@csun.edu
twang@csun.edu

Abstract Over the last several years, much work has been done to improve the data mining process. For all the gains made, there are still many challenges to overcome. This paper focuses on the data selection process. Data warehouse design is a complicated and labor-intensive endeavor, and at its heart is determining which data attributes are best suited for fact table dimensions. The inclusion of too many data attributes can slow down the data mining process, and the inclusion of too few data attributes can limit the usefulness of the results. This work examines several proposed solutions to this problem and proposes a solution to this problem using a genetic program for data attribute selection that will aid domain experts and data warehouse architects. Initial tests indicate the use of this solution results in a set of data attributes of statistical significance. We describe the design and implementation of a prototype system as a proof of concept of our proposed solution.

Keywords data mining, genetic programming, semantic computing, data attribute selection, data warehouse

1 INTRODUCTION

There are many challenges in the field of data mining. There is no standard data mining query language. The appropriate algorithm for a data-mining problem is not always obvious, but in many cases a well designed data warehouse is a real asset. Designing a data warehouse however is a difficult undertaking with long lasting effects. It is to this last problem that we seek to provide a solution.

The data warehouse itself is only a means to an end; its architecture serves as the foundation of future data mining operations. If too much data is included, the data mining efficiency is detrimentally affected. The inclusion of too little data may make the desired knowledge discovery impossible. To further complicate matters, a data warehouse should make possible the discovery and use of data relationships that are not yet understood. This forces the data architect and domain expert must to make important decisions. Together they will decide what data to import from the relational databases into the data warehouse.

Many of these decisions will be based upon the domain expert's knowledge and experience. While the importance of the domain expert's knowledge cannot be overstated, it may be insufficient. The cultural and professional biases of the domain expert may cause important data to be overlooked or irrelevant data to be included. In addition, the domain expert may have knowledge gaps, which force the experts to guess at what data to include. The inclusion of irrelevant data may degrade the performance of data mining algorithms such as the C4.5 algorithm [6].

In this paper, we describe an approach for selecting data attributes using a genetic program, and a tool for facilitating data attribute selection. This tool is called Genetic Data Attribute Selection System (GDASS). A working prototype for GDASS has already been implemented. The results of several experiments show that the chi-squared values greatly exceed the critical .005 values for statistical significance for both the testing and the verification data.

In Section 2, we introduce the concepts related to data attribute selection and genetic algorithms, and provide a summary of related works. Section 3 describes the genetic programming solution, as used in GDASS, with descriptions of its design and implementation. In Section 4, we present the results of several experiments conducted with GDASS. Section 5 summarizes the work and recommends future improvements.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

2 BACKGROUND AND RELATED WORK

Data selection significantly affects both the efficiency and the quality of knowledge extraction, so it is extremely important for the data selection method to quickly identify and remove redundant and irrelevant attributes. This reduces the size of the data set and allows the data mining algorithms to operate faster.

Most of the algorithms for data selection fall into two categories, filter and wrapper models. Figure 1 shows the flows for both models. Both perform data selection, but the wrapper model is dependent on the induction algorithm that will later use the selected data. Wrapper models often achieve better results because of the interaction between the training data and induction algorithm. In addition, the wrapper model usually produces smaller sets of data. This is not without drawbacks. It does requires that the user know in advance which induction method will be used for data mining and it is slower that algorithms using the filter model.

There are many existing solutions using the wrapper model with a search algorithm for data selection, but these solutions are limited by the problems generated from noise and dependencies between features. John et al. recommended using the wrapper model with a search algorithm such as forward stepwise selection and backwards stepwise selection [6]. Their results with this approach did not result in better accuracy, but the data sets generated when used in conjunction with ID3 and C4.5 were smaller which ultimately allows the data mining algorithm to run faster.

Another approach is to use the wrapper model with a genetic algorithm. Genetic algorithms are modeled after biological evolution [5, 11, 16]. They contain a population of individuals, and the algorithm evaluates each individual for its overall fitness with respect to the application domain. New individuals are produced by selecting "high performing" individuals to produce offspring that retain the best characteristics of their parents. The result is a population of individuals that has a maximized fitness with respect to the application domain.

In one approach to data selection using genetic algorithms, an individual represents a feature subset [17]. The individual fitness is determined by criteria of interest such as accuracy or cost. The authors employ the wrapper method by combining this algorithm with neural networks to perform the data selection. The data used by the genetic algorithm is not partitioned into training and test sets as described above. The authors claim this avoids the problem caused by greedy search algorithms. They claim the data selected by the search algorithm with the initial partition will perform poorly on random partitions of the data in to test and trainings sets.

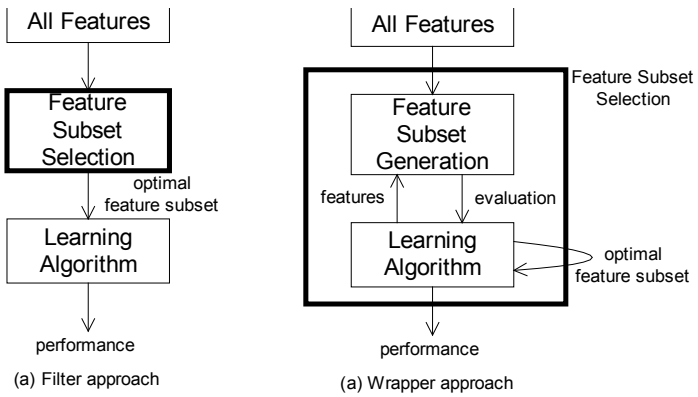
In [12], the algorithm proposed applies directly to data preprocessing to select the user features for knowledge mining. This allows users to find the variables that represent the data in a more precise way to eliminate the later problems that arise in knowledge extraction using the filter model. However, they did not validate the accuracy of their approach against actual data but in mathematical form. They use standard tests proposed by others in the research community, but the input for those tests are introduced artificially. The algorithm needs validation against actual data sets.

3 OUR APPROACH

Within the scope of this paper, it is assumed that the data exist within a relational database, and that a good data warehouse design may be derived by understanding both the data relationships and significance of these relationships within this database. A possible scenario would involve designing a data warehouse that will contain some subset of the data contained within a business's transactional database. Because the desired result of this research is a design tool and not one intended to routinely mine large quantities of data, the ability to find and verify complex relationships is emphasized over data mining efficiency. This emphasis led to wrapper approach.

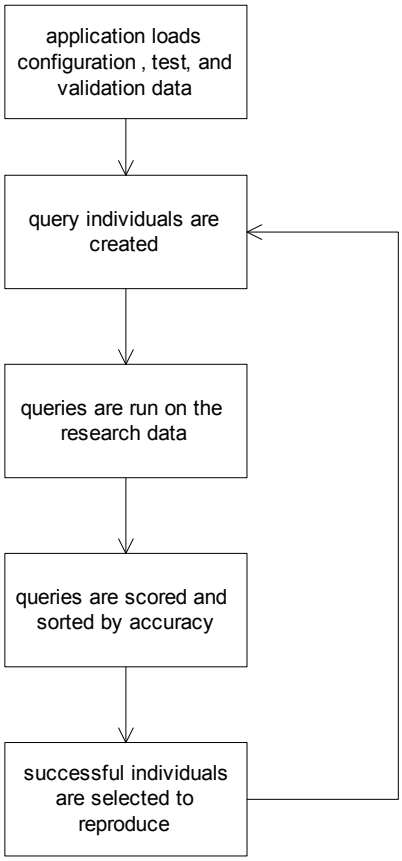
We address the problem of data selection using genetic programming. With a genetic algorithm, the individual must embody the solution. In the case of genetic program, the individual must provide the means to create a solution. In our approach, the individual takes the form of a SQL query, and the selection criteria will be the closeness of the match between the query result and the training data. The recommended data attributes will be embedded in the where clause of successful and partially successful individuals [2, 7, 15].

Figure 1. Subset selection models [17]



In a program designed to search out data relationships in a database environment, the choice to represent an individual as a SQL query seemed advantageous. This query language has been standardized over most relational databases, and it was designed to retrieve data from a database by specifying the relationships that it must satisfy. From a design standpoint, the text composition of SQL helped to keep the complexity of the program within manageable limits. In addition, the portability of SQL enhances the flexibility of the architecture. With the use of standard interfaces such as JDBC, it is easy to achieve a distributed and parallelized architecture. Far more important than these technical considerations, however is that unlike other approaches such as representing individuals as neural networks, humans may easily understand SQL queries. The domain expert can use the resulting queries as a springboard for further research. Attribute restrictions may be modified and retested, and results from different generations or data runs may be contrasted, compared, or even combined. In short, this format is as accessible to the researcher as it is to the application. Figure 2 illustrates the application flow.

Figure 2. Island application flow



We tested this approach on actual data examining a problem of some interest. The problem in question concerns the prediction of graduation for first time university applicants. Students entering a university with little or no college experience are known as first time freshmen (FTF), and the desired outcome is for these first time freshmen to earn their bachelor's degree within six years. Unfortunately, this is often not the case; the chance of a typical FTF graduating within six years at the university examined is likely to be less than 40% [18]. Little is known about FTF when they arrive; there may only be their application, the high school transcript, and SAT results. We attempt to determine which attributes about the freshman will serve to predict whether a student is likely to graduate within six years.

The training data for our approach comes from a single university. We will use this data to help with the attribute selection. Typically, colleges and universities measure the rate at which their students graduate. To do this, groups of students entering the institution in fall semesters, known as cohorts, are followed throughout their careers, and a graduation rate for these students is calculated at the end of spring semesters. Graduation rates, especially the six year graduation rates, contribute to planning within the university as well as comparison with other institutions.

For attribute selection, we evaluate thirty-eight data attributes known by the university when it admits its students. Chosen on the basis data completeness, support, and applicability over time, these attributes include high school general education course completion, residence area, age, gender, SAT verbal, and high school GPA. We rejected attributes from the study if they did not affect a significant number of students or if the nature changed significantly over time. An example of the latter case is the application type. For many years, application type was paper, but on-line applications now dominate. The rapidly changing nature of these kinds of attributes prevents them from serving as predictors.

Utilizing a supervised learning model, we used the applications, test results, and high school transcripts from the fall 1999 FTF cohort as a basis for prediction. The predictions, in the form of list of student ID numbers generated by the query individuals, were tested against the list of all students from the fall 1999 FTF cohort that graduated by end of spring 2005. The scoring rewards individuals for each student in their result list that graduates and punishes them for each student that does not. The individuals are validated by modifying them to make predictions about the fall 2000 FTF cohort about scoring them with that cohort's graduation data.

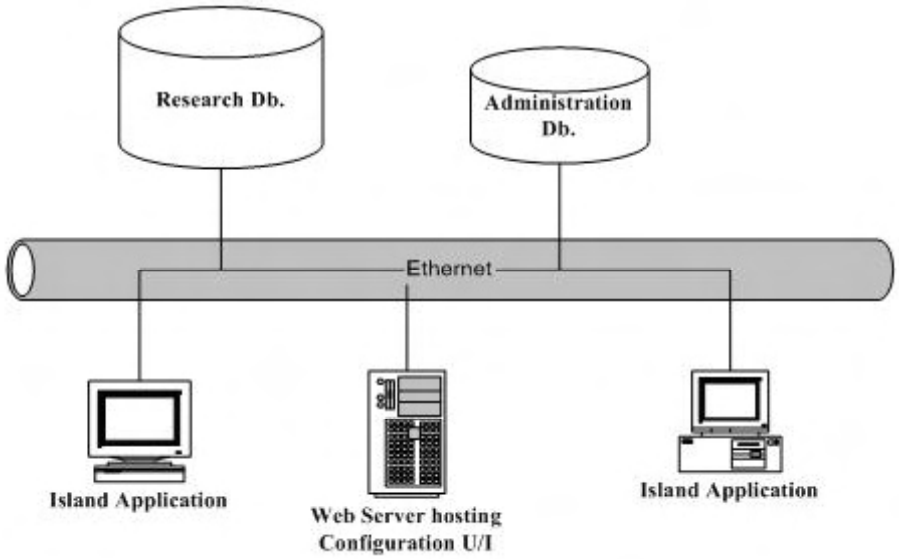
3.1 Architecture

The GDASS architecture is illustrated in Figure 3. The design and implementation for our application built upon an earlier prototype [4, 9, 14]. The application consists of several major components including the island applications, research database, and administration database. The large amounts of data required that we balance the data across the application. This required partitioning the data into that which spanned the all individuals in the population and data unique to each individual. Population wide data is loaded upon initialization and accessed locally. Individuals access unique information from a central location. This strategy tends centralize administration while limiting both the client and network footprints. If client (island) data are accessed and stored locally, not only can its footprint increase to the point were population size is limited, but administration may become a burden. However, centralizing too much data may result in untenable network, server, and database loads.

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. GDASS Architecture



Each island (sub-population) [13] is a Java console application running on a separate server. Upon starting up, the island's identification number and the identification number of the island to which it will send messages is provided. Each island loads additional configuration data from the administrative database. This island maintains a separate population where it orchestrates population generation, crossover, mutation and fitness testing. Periodically, it sends part of its population to its recipient island and picks up immigrants via the administrative database. In addition, the island population, with its inherent distributed parallel approach, enhances both scalability and flexibility.

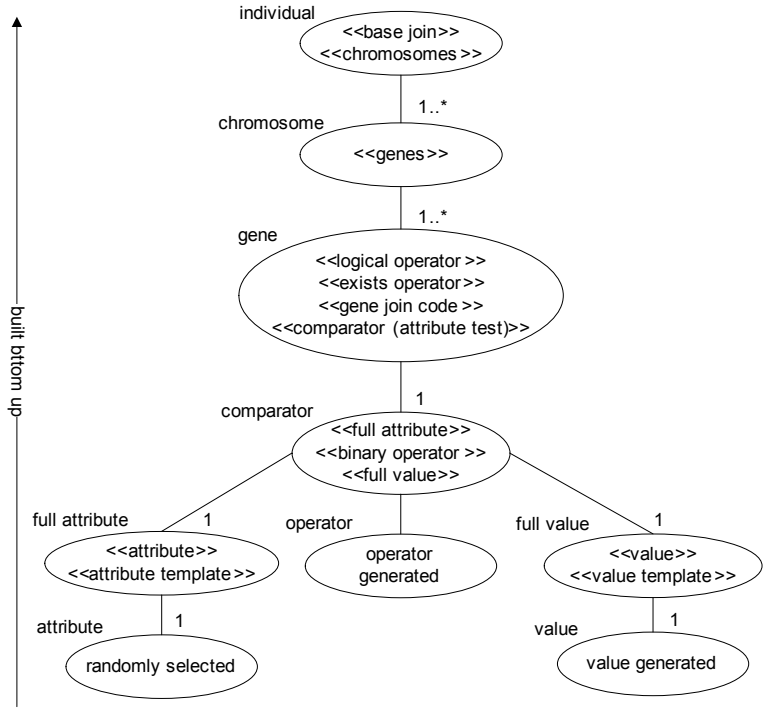
We use two databases in our application. The administrative database coordinates communication between islands, promulgates user orders, records both evolutionary history and outstanding individuals, and maintains research database metadata. Commands are issued to the administrative database via Oracle's SQL Plus. The research database contains the training and validation data used by the application. This database may also contain the data used to populate the data warehouse.

3.2 Individual Representation

Individuals are generated from the bottom up as depicted in Figure 4. Base values are fit into templates are fit into more general templates until first the genes, chromosomes, and finally the complete query is built. The logical operator for the outer gene that connects the gene to the rest of the query is constrained to be an AND. This makes the chromosome order within the individual unimportant. For example, (1 and 0 and 1) = (1 and 1 and 0).

All individuals start with a base join loaded during initialization; this query selects students from the cohort being

Figure 4. Individual Query



studied. In this case the queries are derived from two tables; stu_admission contains information entered into the college application by the applicant as well as administrative data entered by the university in the application process. While there is some overlap between stu_admission and the ERSA table, this second table also contains standardized test and transcript information. Together, these tables provide a summary of the university's knowledge about an incoming freshman.

Figure 5 is a sample of the base join. The base join restricts the student population to those admitted (ADM_PROG_STATUS like AD%) in a specific year and term (strm = 0997 for fall 1999). Additionally, the population is restricted to first-time applicants (admit_type = 5) with an academic level of undergraduate (academic_level <= 40).

Figure 5. Base Join

```
select stu_id
from stu_admission b
where(b.strm = '0997') and //year and term
(b.Academic_level <= '40') and //undergrad
(b.Admit_Type  = '5') //first time
(b.ADM_PROG_STATUS like 'AD%') //admitted
```

Genes consist of sub-queries that each test against an attribute. One or more of these genes make up a chromosome. Simple chromosomes contain only one gene. Figure 6 is an example of a simple sub-query. In this example, the sub-query is joined to table stu_admission by its primary key, and student with mothers with only a high school degree or less are excluded (not (c.ED_LEVEL_MOTHER in ('1','2','3', '8')). In Figures 6 and 7, randomly generated portions are in shown in bold font. In Figure 6, the AND operator is bold and joins the chromosome to the query. It is generated from a set containing only AND.

Figure 6. Simple Chromosome

```
and(not exists (select * from
stu_admission c where
(c.STU_ID = b.STU_ID) and
(c.ACAD_CAREER = b.ACAD_CAREER) and
(c.STDNT_CAR_NBR = b.STDNT_CAR_NBR) and
(c.ADM_APPL_NBR = b.ADM_APPL_NBR) and
(c.ED_LEVEL_MOTHER in
('1','2','3','8'))))
```

If there is more than one gene, the chromosome is complex as is illustrated in Figure 7. The first (outer) query is connected to the rest of the individual with an AND operator so that the results are limited to the cohort, but later (inner) genes within a chromosome may be connected by other operators. The first sub-query is same as used in figure 6; the second sub-query joins to the ERSA table by its primary key and restricts the prediction to those students with a high school grade point average higher than 3.21 out of 4 or 5 (HS_GPA > 3.2). The net effect of the chromosome in figure 7 would be to restrict predictions to only those students with a high school grade point average greater than 3.21 that have a mother with at least some college education.

Figure 7. Complex Chromosome

```
and((not exists(select * from
stu_admission c where
(s.STU_ID = b.STU_ID) and
(c.ACAD_CAREER = b.ACAD_CAREER) and
(c.STDNT_CAR_NBR = b.STDNT_CAR_NBR) and
(c.ADM_APPL_NBR = b.ADM_APPL_NBR) and
(ED_LEVEL_MOTHER in
('1','2','3','8'))))
or (exists (select * from ERSA c
where (c.STU_ID = b.STU_ID) and
(c.year||term = b.period) and
(c.HS_GPA > 3.21 ))))
```

Attributes for each chromosome are randomly selected from an attribute table that holds one or more profiles for each one. These profiles contain the pseudo-data types (numerical range and string list) and other references, information, or templates that are required to generate a chromosome.

3.3 Scoring

In our solution, individual scoring provides us with the selection criteria. The scoring rewards individuals for each student in their result list that graduates and punishes them for each student that does not. A score in the upper 50th percentile by an individual allows the possibility of the individual contributing to the next generation. Figure 8 depicts the actual scoring equation as well as variable definitions.

Figure 8. Scoring Criteria

- Positives = correct predictions (predicted students that graduate)
- Negatives = incorrect predictions (predicted students that do not graduate)
- Totpos = all students from the cohort that graduated (inserted as part of configuration)
- Totneg = all non-graduating students from the cohort (inserted in configuration)
- Individual score = Positives /Totpos – Negatives/Totneg

Providing the proper incentive meant both rewarding correct predictions and punishing faulty ones. For this reason, we chose the individual score equation deliberately. First, in any random sample the negatives and positives tend to cancel each other out, resulting in a near zero score. Second, division by Totpos and Totneg, 936 and 1668 for 1999 FTF, normalizes both in the positive and usually in the negative direction, and this eases comparisons between training scores and validation scores. Finally, of course, this function properly motivates reproduction.

3.4 Crossover

After individual scoring occurs, all the individuals in the local population are sorted by their scores. Those individuals with scores in the upper 50th percentile are eligible to reproduce. This conserves successful individuals are through elitism [1, 10]. We copy the most successful individuals into the next generation as well as allowing them to take part in the normal crossover process. A configuration table determines the number of elite individuals copied. This approach

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

is something of a trade off. As more previous successes are preserved, fewer novel the innovations are allowed with each new generation.

During crossover, we start with the highest scoring individual. This individual selects an eligible mate at random and produces two children. We repeat this with each eligible breeder and continue until the new generation is complete. Most mating takes place at the chromosome granularity. Since the order of the chromosomes is unimportant, we use this to enhance genetic variability by mixing chromosome order during crossover [3].

During this process, the about 0.15% of children are selected for mutation, and another 0.2% are selected for deep crossover where individual genes are exchanged rather than whole chromosomes. These modifications are made in an attempt to avoid local maxima in the search space.

4 EXPERIMENTS

Due to time constraints, experimentation and operational testing were included in the same step. The insight and knowledge gained from experimentation and testing allowed us to invest back into the application design and implementation where feasible. To provide consistency among the trials, we used the same configuration values with each run. The configuration values are shown in Figure 9.

Figure 9. Configuration Values

All Trials:
Mutation rate = 0.0015
Deep crossover rate = 0.002
Termination score = 0.9
Termination generation = 1000
Trial 1–4: Elites = 2
Trial 5-7: Elites = 4

Prior to the first trial, we ran limited data runs to examine the application in depth with the debugger. In the original design, the outer gene's logical operator that connected the chromosome to the rest of the query was not limited to AND. This elevated an individual's chromosomes order to a high level of importance resulting in (1 and 0 or 1) ≠ (1 or 1 and 0). Therefore, we discovered we must conserve chromosome order during crossover, and we implemented this strategy. In addition, it quickly became apparent that joining sub-queries to the base query by OR operators was not a practical approach. Generated queries exploded and returned lists of student applicants that spanned decades. Query individual results were constrained to the proper cohort by imposing the AND operators to all outer queries just prior to the first trial.

In spite of the time limitations, we did gain useful data over the course of seven trials. Trials one through three tested basic application functionality. This allowed us to confirm the correctness of basic operations for both simple and complex chromosome scenarios. During the first trial, we noticed the drawbacks of using the fixed position crossover method. By the end of the trial, not only had the elite individuals stagnated; the mean scores of the entire population were reduced to just a few repeating values. The population's genetic variability was almost non-existent. We implemented a new crossover strategy that allowed later trials to avoid this pitfall.

Unfortunately, there were still more growing pains with the later trials. Configuration value errors caused an inflated Totneg (total non-graduates in the test cohort) to be used in scoring and selection until trial three. Although, trial three successfully proved both the use of complex genes and the validation procedure, a network hiccup terminated it prematurely. After the next trial, a researcher error destroyed the validation data, and during trial five, it became apparent that the gender attribute was being tested against 'Y' and 'N' rather than 'M' and 'F'. During trial six, a review of the administrative database revealed that the SAT math and SAT composite attribute were not loaded. We resolved all the previously mentioned issues for the seventh trial.

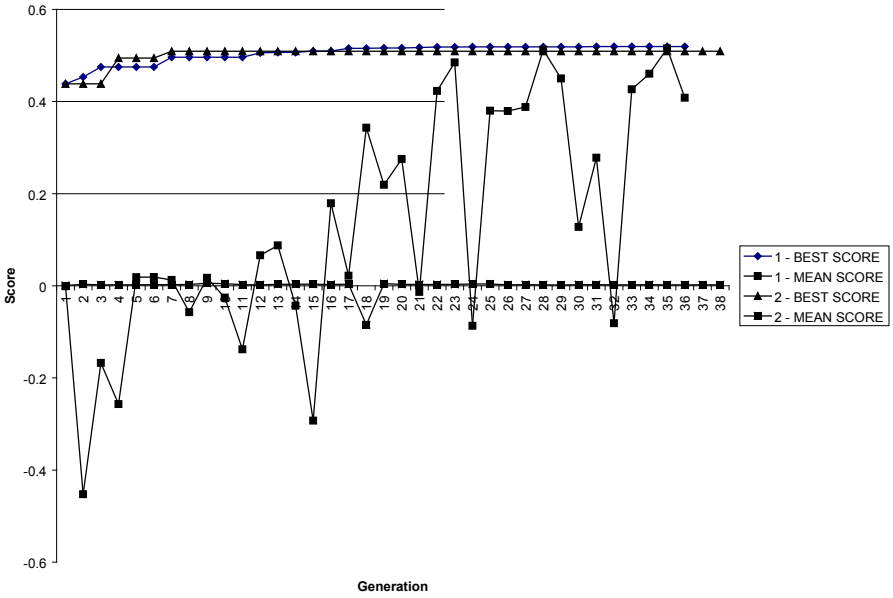
After trial three, we shifted the focus from application testing to evaluation of the resulting data. We conducted trials four thru seven in pairs in order to observe the effects of complex versus simple chromosome scenarios as the population increased from trial to trial until a population of 1000 was attained. This allowed us to obtain initial observations of the effects of individual complexity and population size.

Table 1. Summary of Trials

	Training (fall 1999) Cohort								Validation (fall 2000) Cohort		
Trial	Pop.	Best Score	No. chrom	Max genes	Pos	Neg.	In Gen	of	Pos	Neg.	Score
1	100	0.2937	6	1	455	321	21	60	N/A	N/A	N/A
2	100	0.4072	6	1	765	684	7	52	N/A	N/A	N/A
3	20	0.2157	4	4	685	861	46	93	767	891	0.1864
4.1	140	0.5334	6	4	823	577	25	212	N/A	N/A	N/A
4.2	140	0.2373	6	1	905	1217	38	268	N/A	N/A	N/A
5.1	500	0.5461	6	4	826	561	27	65	890	556	0.4926
5.2	500	0.4780	6	1	889	797	64	76	984	775	0.4520
6.1	1000	0.5424	6	4	841	594	29	89	921	574	0.5107
6.2	1000	0.4620	6	1	800	655	2	120	877	662	0.4195
7.1	1000	0.5194	6	4	838	627	31	36	873	563	0.4731
7.2	1000	0.5094	6	1	823	617	7	38	855	552	0.4630

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 10. Score activity from trial 7.1 and 7.2



4.1 Analysis

We asked ourselves, "Does this technique for data attribute selection predict significant results?" The answer is yes. Even with the small population and poor results from trial three, the observed values of 685 graduates and 861 non-graduates beat the expected values of 556 graduates and 990 non-graduates by a wide margin. We obtained a chi-squared value of 46.74 at one degree of freedom, which is nearly six times the 0.5% critical value of 7.89. Validation of this query individual with the fall 2000 data produced a chi-squared value 39.72 that is about five times the 0.5% critical value.

Table 1 summarizes results from the first 7 trials. It provides the best score in each trial from both the complex and simple populations. It also provides the population size during the trial along with the number of chromosomes. It also compares the results of the training data from the Fall 1999 cohort with the results of the validation data from the Fall 2000 cohort.

In the matter of accurately predicting the graduation rate for first time freshmen, we found increasing the complexity of the attribute relationships increased the scores obtained. This was true for trials 4.x, 5.x, 6.x and 7.x. We saw an average increase of 26% during the four trials. Increasing the generation size also helped improve scores to a point. We did see scores decrease in trial 6.x both for simple and complex populations. Using the training data, the best score for a simple population decreased 3.3%, and the best score for a complex population decreased 0.6%. We need to evaluate the cause of this. It may be a reflection of technique or a natural limit of the problem and attributes available.

We also noticed an increase in complexity seemed to cause a more dynamic process. As an example, Figure 10 shows the activity of both the mean and the best scores for trials 7.1 and 7.2. The characteristics presented here seem to be the norm rather than an exception.

Finally, in the area of genetic variability, we noticed little if any improvement after the 30th generation. We need to determine if this is the nature of the data or a flaw. It is also worth considering widening the scope to include either another problem or a different variation of the graduation rate problem. This would aid in discriminating between domain specific issues and those relating to the proposed solution.

In addition to validating the proposed solution for data attribute selection, we made several other basic observations. During the week, which we ran these trials, the application created nearly 400,000 SQL queries either out of whole cloth or because of crossover. This algorithm ran on a database in normal use, and it did not seem to have a noticeable effect on the database. We neither noticed any performance degradation, nor were we informed about one. This lends us to believe that is a practical solution to the problem of data attribute selection. Although the database handled the load, we must point out that large populations with complex genes can take more than an hour to process a generation. In this situation, a network or power problem could cost days of effort. Later versions of this tool should address this lack of robustness.

Another of our concerns was flexibility, but the template system building the queries was more than adequate. This tool allowed tests, functions, decodes and even sub-queries to be constructed as part of the attribute or the value it was tested against. For example, rather than using raw majors, we were able to use a sub-query and test against the department that owned the major.

5 CONCLUSION AND FUTURE WORK

The ability to select the correct data attributes for inclusion in the data warehouse has a direct impact on the success of later data mining. We need an efficient accurate method for selecting the data attributes. We present a solution to this problem by using genetic programming. We accomplish

this by using SQL queries that mutate and improve through multiple generations. Based on the results of several trials, our application provided a flexible solution for data attribute selection. The domain expert or data architect can configure this application while using an existing relational database to find appropriate attributes and data relationships for inclusion in the data warehouse. We intend for this tool to make possible an understanding of complex relationships that would be difficult to obtain in some other manner.

While GDASS shows promise, it is very much a work in progress. We will continue to investigate the areas where the application seemed to plateau specifically determining the exact cause of the lack of improvement after the 30th generation. We would also like to test this solution on multiple size data sets covering several varied problems. This will give us the opportunity to separate solution specific issues from domain specific issues.

Possible future research can be directed towards improving the robustness of GDASS to allow the application to recover from problems that might arise from user, application, and network issues. We recommend adding the ability to save and recover a population. This will make it possible to extract the best performers from a population to introduce into later trials. Another avenue would be to allow the modification population rules or constraints on the fly, or even allow the researcher to create and customize hypotheses in the form of individuals.

REFERENCES

1. Siddhartha Bhattachatyya, 2000, "Evolutionary Algorithms in Data Mining: Multi-Objective Performance Modeling for Direct Marketing", Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, Massachusetts, United States, Pages: 465 - 473.
2. Brian Connolly, April 2004 , "SQL, Data Mining & Genetic Programming – The practical side of evolutionary algorithms", Dr. Dobb's Journal,
3. Carlos Fernandes, Rui Tavares, Cristian Munteanu, Agostinho Rosa, 2001, "Using Assortive Mating in Genetic Algorithms for Vector Quantization Problems", Proceedings of the 2001 ACM symposium on Applied computing, Las Vegas, Nevada, United States, Pages: 361 - 365.
4. Joel Hickman, 2005 ,"A Parallel Implementation of a Genetic Algorithm", California State University, Northridge, California.
5. David Jackson, Andrew Fovargue, March 1997, "The Use of Animation to Explain Genetic Algorithms", ACM SIGCSE Bulletin, vol. 29, issue 1, San Jose, California, United States, Pages: 243 - 247.
6. George H. John, Ron Kohavi, Karl Pfleger, 1994, "Irrelevant features and the subset selection problem", Proceedings of ICML-94, 11th International Conference on Machine Learning, New Brunswick, NJ, Pages 121 - 129.
7. R. Colin Johnson, June 1996 ,"Genetic program auto designs analog circuits", Electronic Engineering Times, issue 904.
8. S. B. Kotsiantis, D. Kanellopoulos and P. E. Pintelas, 2006, "Data Preprocessing for Supervised Leaning", International Journal of Computer Science, vol. 1, no. 2 .
9. Michael Larson, April 2004, "Genetic Algorithms & Optimal Solutions – Solving for the best congressional redistricting in Texas", Dr. Dobb's Journal.
10. Adam Marczyk, 2004, "Genetic Algorithms and Evolutionary Computation", http://www.talkorigins.org/faqs/genalg/genalg.html, retrieved on April 23, 2006.
11. Marek Obitko and Slavík, Pavel, 1999 , "Visualization of Genetic Algorithms in a Learning Environment", Spring Conference on Computer Graphics, Comenius University, Bratislava, Pages 101–106.
12. Antonio Gomez-Skarmeta, Fernando Jiménez, Jesús Ibáñez, 1999, "Data Preprocessing in Knowledge Discovery with FuzzyEvolutionary Algorithms", To appear in the VIII International Fuzzy Systems Association Word Congress (IFSA'99), Taiwan, Agosto.
13. Z. Skolicki, 2005, "An analysis of island models in evolutionary computation", Proceedings of the 2005 Workshops on Genetic and Evolutionary Computation, Washington D.C., Pages 25 - 26.
14. Sam Talaie, Ryan Leigh, Sushil J. Lois, 2005, "Predicting Mining Activity with Parallel Genetic Algorithms", Proceedings of the Genetic and Evolutionary Computing Conference, Washington D.C., Pages: 2149 - 2155.
15. Alexandra Takac, 2004, "Cellular Genetic Programming Algorithm Applied to Classification Task", Neural Network World, Vol 14; Numb 5, The IDG Company, Czechoslovakia, Netherlands, Pages 435-452 http://dent.ii.fmph.uniba.sk/~sefranek/group/sasaJourn.pdf, retrieved on April 23, 2006.
16. Scott M. Thede, October 2004 , "An Introduction to Genetic Algorithms", Journal of Computing Sciences in Colleges, vol. 20, issue. 1, Pages: 115 - 123 .
17. Jihoon Yang and Vasant. G. Honavar, March 1998, "Feature Subset Selection Using a Genetic Algorithm", Intelligent Systems and Their Applications, vol. 13, issue 2, Pages: 44 - 49.
18. "Office of Institutional Research, California State University, Northridge", http://www.csun.edu/~instrsch/index.html, retrieved on April 23, 2006.

Proceedings of the **International Conference on Information Society (i-Society 2007)**, October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 **www.i-society.org/2007** © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Knowledge management: problems and prospects

Mohammad Poorsartep
Junainah Mohd Mahdee

Multimedia University (MMU)

Abstract Knowledge Management as a discipline to enable organizations leveraging their intellectual capital and achieving their business objectives has drawn great deal of attention during last decades and organizations have been investing on KM to stay ahead of their competitors. However, many organizations suffer due to lack of proper management of their KM system. Even though it may seems that corporate and human capitals are properly planned and managed to achieve a successful KM project, yet in most of the cases social aspects of the system are overlooked which most likely lead to failure of the system. Indeed, any knowledge management system is an "open system" which needs collaboration and interaction of its user in order to be effective. However, not always such a knowledge sharing culture exists in an organization. Therefore, to overcome this problem, there is a need to have a subtle social plan and deep insight to the barriers which may hinder success of a KM effort. In this paper the authors discuss the problems caused due to lack of social planning and propose a model as a framework of analysis that is associated with possible solutions.

Keywords Knowledge Management, Social Planning, Cultural Problems, Trust, System Acceptance.

1 INTRODUCTION

Peter Drucker [1] clearly states that the only or at least the most important source of wealth in contemporary post-capitalist society is knowledge and information rather than capital or labor .His statement is in light of the trends in technology, globalizations and emerging knowledge economies which are creating new landscape of competition. Economists have been also discussing the importance of knowledge and technology for achieving economic growth over decades [2]-[3]-[4]. As results, concepts such as knowledge economy or resource-based theory of the firm are emerged which putting in nutshell, emphasize on investment in knowledge.

Looking from micro-level (firm) perspective, knowledge and other intellectual capital components serve two vital functions within organization. They form the fundamental resources for effective functioning and provide valuable assets for sale or exchange. However, utilizing potential benefits of knowledge requires access to relevant and interconnected information and knowledge in easy and timely manner. The aim of knowledge management is to provide users and enterprises with appropriate knowledge in the most efficient and effective format to assist the performance of their roles. There are different perspectives towards the definition of knowledge management. Regarding the definition of KM, a total of 73 percent of 260 UK and European corporations voted for the business definition of KM as the "collection of process that governs the creation, dissemination and utilization of knowledge to fulfill organizational objectives" [5]. However, as Kakabadse et al. [6] mentioned, central to all different definitions, KM provides a framework that builds on past experiences and creates new mechanism for exchanging and creating knowledge.

Although many organizations recognize the imperative of knowledge management and are engaged in projects and activities to enhance their capabilities for knowledge processing; only few have been able to realize and embrace the benefits. By date, sheer number of publications have addressed issues and factors which need to be taken into consideration, regarding successful implementation of knowledge management [7]-[8]-[9]-[10]-[11]-[12].

Nonetheless, aforementioned studies provide constructive and valuable insight for managers and enterprises related to the implementation of knowledge management system, but still there are great number of enterprises fail to effectively exploit the system since cultural problems emerge during this phase especially while they are not foreseen during design and implementation phase. Hence, it will be of use to develop a holistic framework to describe the fundamental cultural problems arising, due to lack of social planning in earlier stages, during exploitation stage and their critical issues for depicting possible obstacles.

2 SOME BUILDING BLOCKS

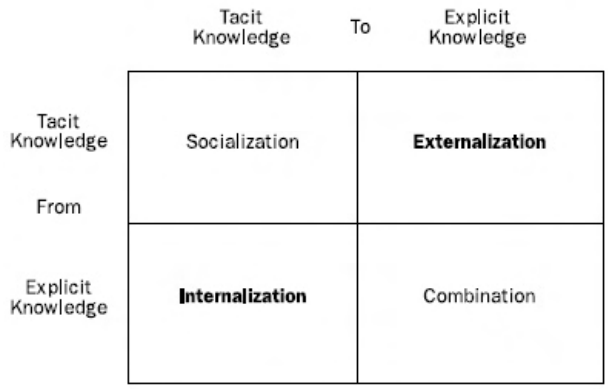
One basic understanding of knowledge management is the use of knowledge as organization's resource towards fulfilling the objectives of organization. However, to tap this resource KM should be initiated. KM as any other project should be built using capitals namely, corporate, human, and social capitals and undoubtedly every resource needs to be care-

fully planned. Social aspects of KM have been overlooked or partially discussed by practitioners or academia though this is a vital and determinant factor of any KM success. Social capital refers to communications, relationships, and interactions among the people across the organization. Primarily, social capital is a matter of culture in any organization. The extent to which communications, relationships, and interactions occurs is influenced by the culture resides in each enterprise. Therefore, for the proper social planning, it is imperative to have a clear picture of problems associated with culture. As McDermott [13] noted, problems associated with merging individual experiences and skills and codified instructions can be often traced into cultural and social aspects of organization.

2.1 Imperatives of Knowledge Creation

Pivotal idea of KM is the creation and dissemination of knowledge throughout the organization which consequently lead to the success of organization [14]. However, the challenge is how to create the knowledge while it is resided in human mind. Their idea is built upon the theory established by Polanyi [15] that is used to explain how personal knowledge can be created and then be converted into explicit knowledge and become a useful organizational resource. Nonaka and Takeuchi [14] constructed a dynamic model for knowledge creation. In their discussion, they elucidated a critical assumption that social interaction between tacit and explicit knowledge create and expand the knowledge. This interaction is called "knowledge conversion" by them and consists of four different modes of knowledge creation (see Figure 1).

Figure 1. Source: Nonaka and Takeuchi (1995, p .62)



Nonaka et al. [9] based on the theory of Nonaka and Takeuchi expanded a model to discuss dynamic process of organizational knowledge creation, maintaining and exploitation. His model comprises of three elements which are; 1) process of knowledge creation and conversion, 2) context conducive and supportive to knowledge creation and conversion, and 3) leadership that directs the process and designs the context. This is a social, cultural, and historical context in which people are living and experiencing. This context is dependent on time which imposes limitations for social interactions among people within that context.

Although the road has been paved for knowledge creation by proposing models and suggestions, yet there are frustrating records in achieving this goal. Ray and Clegg [16] discussed that this problem could be due to organizational culture and misunderstanding of Polanyi's idea. Regarding the first problem, they referred to the theory of low and high context cultures and derived that in high-context countries such as Japan, it is easier to create the knowledge since the culture already exist, unlike the low-context countries such as US which culture needs to be developed. Regarding the second problem, they outlined that tacit knowledge is not easy to be transferred as it is feasible in case of explicit knowledge. Polanyi recognizes the challenge of disengaging tacit knowledge from its origin, and mentioned:

"But suppose that tacit thought forms an indispensable part of all knowledge, then the ideal of eliminating all personal elements of knowledge would, in effect, aim at the destruction of all knowledge" [15].

2.2 Importance of Trust

It has been widely accepted among academia and practitioners that culture is the most demanding challenge which organizations encounter in creating a knowledge-based enterprise. As one important component of culture, trust has been regarded as a necessity for cooperation among groups. The development of trust, and the existence of mutually reciprocal relationships, impacts strongly on the way in which knowledge is used, shared and developed within the organization [17]. If trust as a fundamental aspect of culture does not exist among the employees, they will become sceptical about the intention and behavior of others and consequently they are more likely to refuse sharing their knowledge [18] .Even though trust has been highly emphasized for the success of KM due to its critical role in building the foundation of knowledge creation and sharing, but it is too intricate to achieve. One of the top managers of Buckman Laboratories highlighted that trust is the most difficult aspect of knowledge sharing to achieve. If you can't do it, you can't succeed [19].

2.3 Resistant Culture to Change

Effective knowledge sharing and learning require cultural change within the organization, new management practices, senior management commitment and technological support. Organizations can realize the full value of their knowledge assets only when they can be effectively transferred between individuals. A major problem is how to convince, coerce, direct or otherwise get people within organizations to share their information [20]. it is emphasized by Walczak [21] that creating and managing the corporate culture is the deriving factor which facilitate and encourage creation, sharing and utilization of knowledge even though organizational culture shifts is challenging and demanding which must be tackled [22]. Gupta et al, [20] mentioned that Knowledge Management requires a major shift in organizational culture and a commitment at all levels of a firm to make it work. However, often, organizational culture itself prevents people from sharing and disseminating their know-how in an effort to hold onto their individual powerbase and viability. The culture and other aspects of the organizational environment

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

are conducive to more effective knowledge creation, transfer and use. This involves tackling organizational norms and values as they related to knowledge.

3 CONCEPTUAL FRAMEWORK

Sheer number of publications has been published regarding knowledge management and its practices. However, there are handful guidelines and framework to address the issues pertaining cultural and social challenges in the way of getting the users to utilize the system. The concept of KM can be viewed as an interaction between information technology and social system which leads to the creation and dissemination of knowledge. As mentioned by Holsthouse [23], organizational KM system's social and technological attributes drive the success of knowledge exchange within the organization.

The critical role of information technology for enterprise KM initiatives as an enabler is discussed over the last decade [24]-[25]. Definitely the success of any information technology application is contingent upon acceptance and usage by its users. In this sense, Fred Davis conceived a model called Technology Acceptance Model (TAM) to study this phenomenon in year 1986. In this section, we have adopted TAM model, modified and further developed in order to cover the social aspect of KM as well, since the TAM is a general model to study the behavior of user in the context of information technology acceptance [26]. The underlying theory of TAM is the Theory of Reasoned Action (TRA) which explains individuals will adopt a specific behavior if they perceive it will lead to a fruitful outcome [27]. Hence, the factor of "Trust" Is added to tailor and further empower the TAM model to be used for better understanding of KM pertaining KM is not merely an IT application, and social interactions among the users play a crucial role as well (see Figure 2.).

The value of trust in organization encourages the employees and workers for collaboration and knowledge sharing. Thus, development of trust can be taken into account as a critical factor for organizations to lead the KMS into success and bright horizon. Looking at trust, Cook and Wall [28] distinguish two components of trust namely Faith and Confidence in peers (co-workers) and management (supervisors). They defined faith as trustworthy intentions of others while confidence refers to the ability of others. Therefore, pertaining these two components of trust, three factors are derived which can hamper its formation and expansion namely, fear of losing job (Job Security), fear of being criticized or misleading the community (Criticism/Misleading), and absence of knowledge champion to build or smoothen the interaction among the KMS users (K-champion).

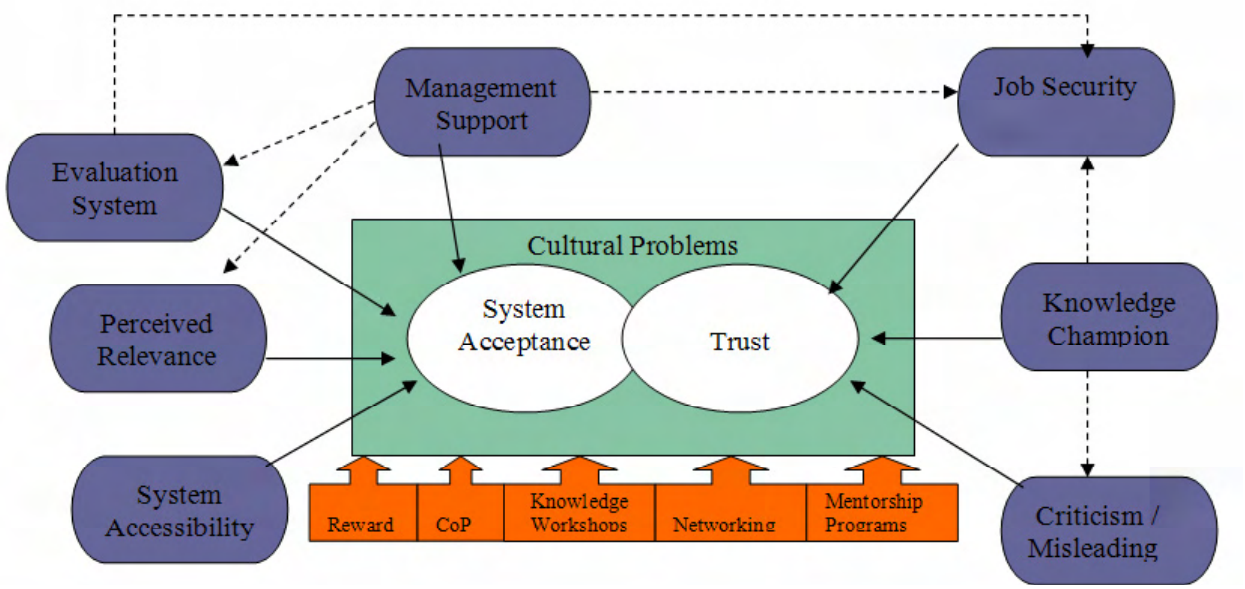
3.1 Job Security

As it is stated by Pan and Scarbrough [29], the organization's culture must provide a "climate of continuity and trust". They pointed out that unless a company trusts its employees and the employees believe that it is safe to share knowledge, effective KM will not happen. The intellectual assets of a company are its most valuable resource and these are largely held in the minds of individual employees. This knowledge can only be effectively used or managed in a corporate culture which promotes mutual trust and facilitative behavior.

"Trust is essential . . . you must trust your employees. Employees must trust that sharing enhances employment status and does not undermine the business's need for them "[30].

A willingness to share knowledge with others may be driven by a desire to contribute to organizational performance or to receive status and rewards from being seen to use personal knowledge, whereas a reluctance to share knowledge may be due to concerns that one is giving away what makes one powerful, or from a desire to prevent certain individuals/groups gaining access to one's knowledge [31]. As it can be seen, knowledge is perceived as a source of power for individuals to hedge themselves of losing the ground to others. On the other hand, human is basically self-interest and will trust when believe trusting can enhance or contribute to his/her

Figure 2. Acceptance-Trust Model



interest. However, while the fear of losing job due to sharing the knowledge with peers and management is wiped out and the sense of secure job is established, job security will create a sense of belonging for individuals which contribute to the promotion of trust. Therefore, there is a need for institution to block the practice of destructive untrustworthiness to ensure the job security among the users and use the methods to promote the culture of trust. During the 10th International Advisory Panel (IAP) which was held in Putrajaya, Malaysia, Stephen McGuckin, Managing Director, IT services of DHL ISSC Europe, emphasized on the importance of making the desire among the employees to work for the company, not to merely work for salary. Pertaining to his statement and also our observation, DHL knowledge bank (KB) is fully utilized by employees since they are passionate to serve their company and it has tackled the issue of job security.

3.2 Knowledge Champion

Here by k-champion we refer to CKO, change agent, or knowledge manager who in general is responsible to facilitate the acquisition, storage, and dissemination of new knowledge from external or internal sources and also keeping the users and employees focused on organizational goals. Davenport and Prusak [32] advocated that knowledge champions can lead the changes in organizational cultures and individual behaviors relative to knowledge as well as giving the professional knowledge managers the sense of community. Furthermore, they noted most of the k-champions are familiar with the culture and the business of the corporation they are working for, through their personal experiences and all of them are established figures in their organization. In this sense, K-champions are often perceived as social models for those who are involved with KMS. Hence, as any other social system, users of KMS try to imitate the behavior of k-champions. Given this argument, k-champions possess an important role to influence the behavior of users and direct the people towards using the system by building trust among them. This is an advantage which can be used by k-champions to facilitate and smooth the interactions among the user through establishing trust.

Additionally, k-champions should ensure the users that by contributing to KMS not only they will not endanger their profession, but also it can be a source for getting promoted or receiving rewards. Nonetheless, they might be still reluctant to contribute whilst there is the fear of criticism or misleading the community which k-champion should overcome these two obstacles too. Referring to the DHL Malaysia, evaluation on employees' contribution to KB is often a basis for considering contract-base staffs for being promoted to permanent employee. It happened when one of the managers who also serves as k-champion promoted a contract-base, introvert employee to a better organizational position because of his contribution to the organizational knowledge. Consequently, other employees became keener to contribute their know-how to the KB. This example simply illustrates how a knowledge champion can build the trust and give the sense of job security to the employees.

3.3 Fear of Criticism and Misleading

The fear of criticism often reaches to the point that employees think their ideas will be ridiculed which can eventually affect the trust towards KMS. In this regard, we argue that the fear of criticism should be mitigated through deployment of ability to listen to ideas in organizations, even though most companies are far better at snubbing or suppressing workers' ideas than promoting them. As it is stated by Dean Call [30] employees should perceive that experimentation and well-intentioned failure are acceptable. There should be no such thing as failure; every perceived failure should be turned into a success, by allowing the organization to learn from it.

On the other hand, employees will lose faith about KMS if they become afraid the wrong knowledge is transferred and will harm others or the community. However, the fear of misleading is not merely from employees' side. It happen while management don't trust their employees and is afraid they will put content in the wrong place. This attitude also hampers the system by adding cumbersome layers of approval for contributing or accessing information.

To overcome these barriers, management should provide an environment conducive for sharing the knowledge and k-champions should provide directions for users that what content is needed or appropriate to be sent into knowledge base. Consequently, the culture of trust will be developed while criticism is perceived as a constructive approach and also it has been assured the knowledge and information inside the system as relevant and accurate.

3.4 Management Support

Steven Walczak [21] point out that those organizations which embark to introduce a knowledge management initiative before having a supportive managerial structure will soon realize that their investment in KM does not produce any perceived benefits. Therefore, Management support is vital according to many models on information system development. The knowledge management project has management support. This implies that resources are available to conduct a thorough implementation.

To reach a knowledge friendly culture three managerial area should take into consideration. First one is preparing the tangible capitals, second is preparing the pragmatic mindset towards KM, and finally managing the cultural change throughout the organization. The preparations are the initial steps that every organization should take to develop their knowledge culture. Performing these initiatives often result in changing the culture of organization, changing the way employees work and interact, which needs to be carefully managed. Adoption of a new organizational structure or knowledge culture is most likely to face resistance within the organization. Resistance to change may be minimized by reducing the perception of change for the stakeholders. However, as an important fact, no doubt workers reject to use the system if the managers also do not use the system. In other words, management not only should support the system by providing adequate resource and directing the prepa-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

rations, but also they should utilize the system. Multimedia university (MMU) as the first private university in Malaysia initiated its knowledge bank to realize the target of materializing a campus-wide knowledge sharing culture as well as addressing the need for an effective knowledge management system. During its infancy period, the acceptance and usage rate among academic and administrative staffs were at its lowest. However, the idea of getting the top management involved with the system was put into the action and president of MMU initiated storytelling sessions based on real life experiences of sharing knowledge. As a result, MMU knowledge bank started to witness increasing number of administrative and academic users.

3.5 Perceived Relevance

Employees, who are to use the system, should perceive the knowledge management as relevant. Since it is possible for workers to work without using the system, it has to be obvious that usage of the system implies adding value to the work result. Their initial mindset of using KM is just an extra work which takes time. However, they will start using the system once they realize the benefits and actual results of the system in their functional areas. Hence "buy-in" will occur and system will be embedded in their daily work practice [33]. An additional aspect of relevance related to perceived relevance is how the system should be integrated in running work, that is to make the system an integrated part of the workers' work practice. The perception of users of the system is to reduce operational disturbance. The workers perceive the system as positive since it relieves him or her from unnecessary problems. Still, it has to be proven that this really is the case and that it is not just an idea from management to increase control and tempo in the workers' working situation. In the case of SCICOM (MSC) Berhad, which offers business process outsourcing for companies like Nokia, technical executives during their training period, realize the importance of knowledge bank in its relation to their routine daily job. Hence, when after training period they stationed and started their actual job, using knowledge bank becomes an indispensable part of their job.

3.6 Accessibility

If the knowledge management system is to be accepted, accessibility has to be satisfactory. Accessibility is a question of who is to be the user, what action the system is to support, where users get access to the system, when the system is ready to use and how the system's interface fulfills the goal of the system. The latter is also related to how the system's interface takes users' preferences into account.

Who is to be the user? It is vital to know who the user of KMS is. It is the worker or management? The relevance of knowing about who the user, is a question about if the workers themselves should enter, search and retrieve the knowledge represented in the system or if someone else is to do it. The strategy could be to make it possible for the workers themselves to enter, search and retrieve knowledge in order to make it as accurate and relevant as possible. The aim has been to make the knowledge represented in the system as close to the workers' work practice as possible by letting the workers themselves enter their knowledge in the system. However, this has to be done with consideration to the other aspects about accessibility.

Where users get access to the system is a question of the system's physical location. The physical location is an important decision that affects how the system can be used. The chosen physical location is highly dependent on who is to be the user. If the worker is to be the system user, the system has to be physically placed close to the working place. It is also relevant to account for the system's physical location to make the system an integrated part of the workers' work practice.

An important issue of systems accessibility is to decide how to design the interface of the system in order to fulfill the goals of the system. Design issues are dependent on who is to be the user of the system and when the system is to be used. Most of the time, the workers are to use the system continually, which implies simple design that meets with workers' preferences. Further, the design should clarify the functionality and relevant concepts found in the system. Functionality is related to entering, processing, and searching and retrieving knowledge. Knowledge captured by the system should be readily available to the users who require the information, and system should provide easy searches to help cull the information.

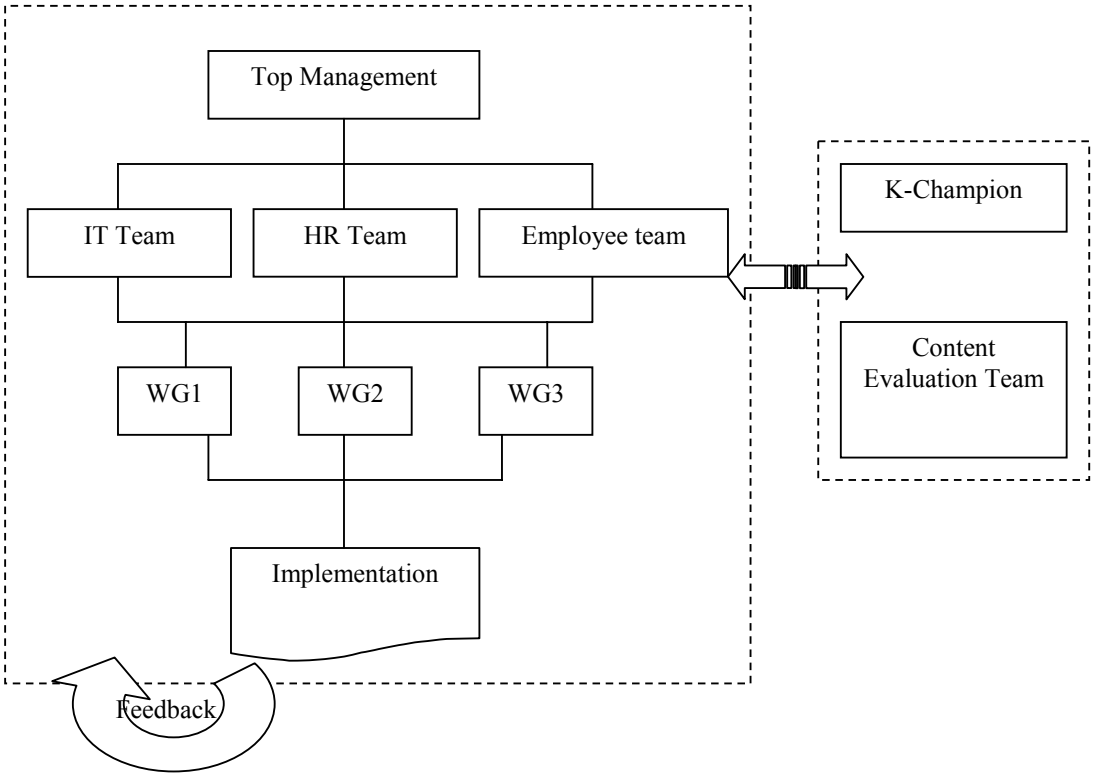
KM software should be designed around the way people work [34]. The tools used for capturing, analyzing and distributing knowledge do not have to be very high tech at all. While technology surely facilitates all of these actions, knowledge management should not be undertaken for the sake of technology. Rather the technology should address the needs of the knowledge management system's goals. In SCICOM (MSC) Berhad system is designed in such a way which it protects the critical information inside the KB from being edited or deleted. Users can add, edit or modify the content of the KB only through an authorised person who first evaluates the proposed content and if appropriate and correct, proper action will be taken by him/her. On the other hand, to facilitate the free flow of information and knowledge among the users, Lotus Note is used as a platform to meet the immediate needs of frontline users (technical executives) and provide more flexibility to the system. This subtle design which is a combination of two rigid and flexible systems is designed around the daily routine process of work inside the SCICOM as it was emphasized earlier.

3.7 Evaluation System

Wong [35] states that unprovoked employees can not be tapped at organization's intellectual resources unless; they become motivated through different mechanisms. On the other hand, as with many physical assets, the value of knowledge erodes over time. Since knowledge can get stale fast, the content should be constantly updated, amended and deleted. The relevance of knowledge at any given time changes, as do the skills of employees. Therefore, there is need for an appraisal system to evaluate the relevance and the value of the employees' contribution to KMS. Consequently, a mo-

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 3. Implementation Guideline



tivation schema should be developed based on the result of that evaluation system to further motivate employees for accepting, using, and posting updated contents to the system.

The evaluation system eventually affects the perception of users towards the issue of their job security. As long as they are getting rewards, either tangible (i.e, monetary rewards) or intangible (i.e, social recognition), they will accept and trust the system which is not going to take over their position in the organization.

A mechanism used by Multimedia University to encourage user towards contributing and utilizing the KB was to make it as mandatory for staff's annual performance assessment. They categorised the contributions into two groups namely Academic (publications, research, consultancy, etc.) and Non-Academic (assignments, training reports, work-related duties, etc.). Finally, all the submitted contents will be assessed and graded by a committee appointed by the president which the result of their evaluation will become a basis for employees' annual performance assessment. This mechanism has encouraged staffs for submitting quality write-ups as well as increasing the usage rate among them due to the rewards and increments which they will be entitled for later on.

4 CONCLUSION

The importance of managing the intellectual capital is clear for every new and old organization due to the fact that the economy is switching to knowledge economy; therefore, organizations need to move in this direction to stay ahead of their competitors and survive. The role of KM to fulfill this goal has been recognized by SMEs and large businesses. However, many knowledge management systems have led to failure since they did not take the culture of users and organizations into consideration before implementation of the system. When a KMS is created without concerning the cultural issues, users reject the system and do not trust it; hence, the system will be ineffective and a waste of time. Here, we tried to identify and address the issues concerning the social aspects of KM which should be carefully planned and managed to lead the KM into the success. To overcome these two major cultural obstacles and make the system practical, creating communities of practice (CoP) , establishing a reward system, developing a network via social channels, inaugurating mentorship programs to coach and support employees by key managers, and introducing knowledge workshops to promote organizational learning are suggested as social tools to boost the cultural adoption of KMS.

In our proposed model, we tried to identified causes which result in futile effort to implement and utilize KMS. Even though these factors may have been identified before, but the relation between them on how they could influence each other were not clear. Using the proposed model as a framework of analysis can help organization to identify their weak spots in the system and overcome them. In order to implement the model into the context of an organization, an implementation guideline is suggested as illustrated in Fig. 3. as it can be seen from the guideline, top management basically should provide direction and set goals for the system in order to make sure KMS will serve the organization towards its goals. At the next level it can be teams from human resource (HR) and IT department who should be responsible for issues such as reward schema, system accessibility and design, trainings, and etc,. There must be also a team consisting of employee's representatives to work closely with IT and HR team since they are users of system and this close collaboration can shade light on preferences and needs

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

of users. Bottom level of this chart will be interdisciplinary working groups which is responsible for implementation of ideas developed through the collaboration of IT, HR, and employee teams. This whole process should be supervised and monitor by knowledge champion who serves the role of consultant and also the content evaluation team. To make sure right decisions are made and appropriate steps are taken, a feedback mechanism should be designed to measure the impact of changes.

At the end, this point should be highlighted that knowledge management utilizes Double-Loop learning. Double-Loop learning occurs when error is detected and corrected in ways that involve the modification of an organization's underlying norms, policies and objectives [36]. In this regard, to successfully overcome the problems of rejecting the system and lack of trust, aforementioned practical solutions are not sufficient and causes of problems (job security, absence of k-champion, unsupportive management, etc.) should be improved. It is expected that by implementing this model, cultural changes towards promoting an environment conducive for sharing knowledge will be initiated and organizations will be able to tap the knowledge resides in employees' mind.

REFERENCES

1. Drucker, P. (1993), Post Capitalist Society, Harper Row, New York, NY.
2. Hayek, F.A. (1945), The uses of knowledge in society, American Economic Review. Vol. 35, pp. 1-18.
3. Arrow, K. (1962), economic welfare and the allocation of resources for invention, in National Bureau of economic Research (Ed.), The Rate and Direction of Inventive Activity, Princeton University Press, Princeton, NJ, pp. 609-25.
4. Marshall, A. (1965), Principles of Economics, Macmillan, London.
5. Murray, P. and Myers, A. (1997), The facts about knowledge, Information Strategy. Vol.2 No. 7, September, pp. 29-33.
6. Kakabadse, N.K., Kakabadse, A., and Kouzmin, A. (2003) Reviewing the knowledge management literature: towards a taxonomy, Journal of Knowledge Management. Vol. 7 No. 4, pp. 75-91
7. Nonaka, I. (1991), The knowledge-creating company, Harvard Business Review, Vol. 69 No. 6, pp. 96-104.
8. Barney, J. (1995), Looking inside for competitive advantage, academy of Management Executive, Vol. 9 No. 4, pp. 49-61.
9. Nonaka, I., Toyama, R. and Konno, N. (2000), SECI, ba and leadership: a unified model of dynamic knowledge creation, Long Range Planning, Vol. 33 No. 1, pp. 55-34.
10. Ndlela , L.T. and Toit, A.S.A. (2001), Establishing a knowledge management programme for competitive advantage in an enterprise, International Journal of Information Management, Vol. 21 No. 2, pp. 151-165.
11. Tiwana, A. (2001), The Knowledge Management Toolkit: Practical Techniques for Building Knowledge Management Systems, Prentice-Hall, Englewood Cliffs, NJ.
12. Rowley, J. (1999), What is knowledge management?, Library Management Vol. 20 No. 8 pp. 416-419.
13. McDermott, R. (1999), Why Information Technology Inspired but Cannot Deliver Knowledge Management, California Management Review, Vol. 41, No. 4., pp. 103-117.
14. Nonaka, I. and Takeuchi, H. (1995), The Knowledge-creating Company, Oxford University press.
15. Polanyi, M. (1966), The Tacit Dimension, Gloucester, Mass.: Peter Smith.
16. Ray, T. and Clegg, S. (2005), Tacit Knowing, Communication and Power: Lessons from Japan? Managing Knowledge: An Essential Reader, Second Edition, Edited by S. Little and T. Ray, The Open University, Sage Publications Ltd, pp. 319-347.
17. Kelly, C. (2006), Managing the relationship between knowledge and power in organizations, Aslib Proceedings: New Information Perspectives Vol. 59 No. 2, 2007 pp. 125-138
18. Lee, H, and Choi, B. (2003), knowledge management enablers, processes and organizational performance: An integrative view and empirical examination, Journal of Information Management, Vol. 20, No. 1, pp. 179-228.
19. Pan, S. and Scarbrough, H. (1999), Knowledge management in practice: an exploratory case study, Technology Analysis and Strategic Management, Vol. 11 No. 3, pp. 359-74.
20. Gupta, B., Iyer L. S., and Aronson, J. E. (2000), Knowledge management: practices and challenges, Industrial Management & Data Systems Vol. 100 No. 1, pp. 17-21
21. Walczak, Steven (2005), Organizational knowledge management structure, The Learning Organization Vol. 12 No. 4, pp. 330-339
22. Roth, G. (2004), Lessons from the desert: integrating managerial expertise and learning for organizational transformation, The Learning Organization, Vol. 11 No. 3, pp. 194-208.
23. Holthouse, D. (1998), Knowledge management research issues, California Management Review, Vol. 40, No 30, pp. 227-80.
24. Davenport, T., Grover, V. (2001), General Perspectives on Knowledge Management: Fostering a Research Agenda, Joumal of Management Information Systems, Vol. 18, No. 1.
25. Alavi, M., Leidner, D. (1999), Knowledge Management Systems, Issues, Challenges, and Benefits,Communications of the Associations of Information Management Systems, February
26. Davis, F. Bagozzi, R., and Warshaw, P. (1989), User Acceptance of Computer Technology: A Comparison of Two Theoretical Models, Management Science, August.
27. Compeau, D. Higgins, C. (1995), Computer Self-Efficacy:Development ofa Measure and Initial Test, MIS Quarterly, June
28. Cook, J.D. and Wall, T.D. (1980), New work attitude measuers of trust, organizational commitment and personal need non-fulfillment, Journal of Occupational Psychology, Vol. 53, pp. 39-52.
29. Pan, S.L. and Scarbrough, H. (1998), A socio-technical view of knowledge-sharing at Buckman Laboratorie', Journal of Knowledge Management, Vol. 2 No. 1, p. 59, 62.
30. Call, D. (2005), Knowledge management, not rocket science, Journal of Knowledge Management, Volume 9 Number 2 2005 pp. 19-30
31. Hislop, D. (2005), Knowledge Management in Organizations: A Critical Introduction, Oxford University Press, Oxford.
32. Davenport, T. and Prusak, L. (1998), Working Knowledge, Harvard Business School, Boston, MA.
33. Hariharan, A. (2005), Critical success factors for knowledge management: fifteen common challnegs and how to overcome them, KM Review, Vol.8, No. 2, pp, 16-19
34. Hackett, B. (2000), Beyond knowledge management: new ways to work and learn, pp. 42, 21, 63, available at: http://www.ispi-van.org/htm/articles/conference%20board.pdf (access date, 2007 Feb.)
35. Wong, K. Y., (2006), Critical success factors for implementing knowledge management in small and medium enterprises, Industrial Management & Data Systems Volume 105 3 2005 pp. 261-279
36. Smith, M.K. (2002), Chris Argyris: theories of action, double-loop learning and organizational learning, 14 July, available at: www.infed.org/thinkers/argyris.htm (access date, 2007 Feb.)

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007

Blind detection of statistical watermark using extreme learning machine

Anurag Mishra
Rampal Singh
Deendayal Upadhyay College, University of Delhi, Shivaji Marg, New Delhi. India
anurag_cse2003@yahoo.com, rpsrana@ddu.du.ac.in

S Balasundaram
School of Computer Systems and Sciences, Jawaharlal Nehru University, New Delhi, India
balajnu@hotmail.com

Abstract This research study, for the first time, perceives the blind zero – bit, soft decision detection of an statistical watermark within still digital images as a classification problem based on multiclass data trained and testing by using a recently proposed machine learning algorithm known as Extreme Learning Machine (ELM). Till now, researchers are using Support Vector Machines (SVMs) as binary and multiclass classifiers to detect the presence of a watermark in a digital image. However, for a large multiclass data set available to train the machine such as in case of digital images, the SVMs tend to complete the job in a larger time frame. In the present case, the training and testing procedures are executed for two different activation functions: "Sigmoid" and "Sin". In both cases, training time variation w. r. t. number of hidden neurons is identical and is of the order of a few seconds that makes an ELM based detector a faster alternative to its SVM counterpart. The testing accuracy values are drastically different when a watermarked image is tested with the ELM on one hand and the un-watermarked image on the other. The large differential in these numerical values is concluded as the classification executed by the ELM based on the input multiclass data set. It is also concluded that the testing accuracy parameter itself is sufficient to establish the basis of distinction between watermarked and un-watermarked images which is considered as a main objective of a blind zero – bit detector.

Keywords Blind zero – bit Detector, Soft Decision Detection, Training time, Testing Accuracy, Extreme Learning Machine, Generalization Ability.

1 INTRODUCTION

Watermark detection has achieved a significant position in the research going on in the area of digital watermarking of images and videos. Till recent past, the detection of a watermark is accomplished using statistical methods based on correlation similarity parameters such as $SIM(X,X^*)$[1]. Nikolaidis and Pitas[2] presented a good overview of the benchmarking of watermarking algorithms. They have put the watermark detector algorithms in two categories – Zero bit systems and Multiple bit systems. The zero bit systems are ones which simply identify the presence of the watermarks in a given image. On the other hand, the multiple bit systems are capable of decoding the mark or message within the image with or without the help of the source watermarked image. The authors have said that with respect to the output of the detector, systems are categorized as Hard decision detectors and Soft decision detectors. The former generates a binary (true/false) type output on the basis of a threshold function whereas the latter is responsible to produce a test statistic that may be used to verify the detector reliability. In this case, thresholding is done in a subsequent step.

Optimization methods such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) used as binary classifiers have drawn the attention of watermark researchers only recently [3,4]. Till now, there is hardly any work on watermark detection using these optimization tools as multiclass classifiers. SVMs generally do not suffer from inherent complexities such as longer training time, presence of local minima and imprecise learning rate etc like their NN counterparts. But, when used as a multiclass classifier tool, they also show large time complexities. To overcome these problems, a new learning algorithm known as Extreme Learning Machine (ELM) was proposed by Huang et al [5,6,7]. This algorithm may be used as a tool for regression analysis as well as a binary and multiclass classifier. It is used as a classifier in the present experiment and is applied on a multiclass data set obtained from five watermarked images and an un-watermarked image. For watermarking, the robust algorithm given by Cox et. al. [1] is used in the present study. In this

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

method, the authors have embedded a set of 1000 random numbers distributed normally with mean 0 and variance 1 in low frequency coefficients of the source image in transform domain. The watermarks were subsequently extracted also and the two sets of marks (original and recovered) were compared on the basis of an statistical parameter known as similarity correlation parameter $SIM(X,X^*)$. The $SIM(X,X^*)$ parameter was also used to detect the presence of watermark with in the image by equating it with a threshold value. If the set of extracted watermarks gives a numerical value for $SIM(X,X^*)$ greater than or equal to threshold, it is presumed that watermark is present with in the image. However, in the present case, it is shown that ELM algorithm can be used as a classifier tool to identify an un-watermarked image from a given set of un-watermarked and watermarked images. It produces distinct numerical results for testing accuracy when applied as a classifier on a multiclass data set using a watermarked image as a test image on one hand and an un-watermarked image as a test image on the other. It is concluded that ELM used as a classifier on a multiclass data set can be successfully applied for blind zero – bit soft decision watermark detection.

2 EXPERIMENTAL DETAILS

The present work is focused towards developing a blind zero – bit watermark detection algorithm in the category of soft decision watermark detectors using the newly developed ELM algorithm. ELM is used here as a classifier tool that consumes as its input, a data set comprising of five watermarked images and a single un-watermarked image. The source image is 128×128 pixel image "Lena" represented in 8 – bit bitmap format. These watermarked images are obtained by embedding 500 random numbers generated by five different seed values (keys 1-5). The embedding of the random numbers is done at the most relevant low frequency coefficients of the image in transform domain as suggested by [1]. The data values obtained from these images are normalized with mean 0 and variance 1 and labeled as class 1 to 6 respectively. Class 1 represents the un-watermarked image whereas classes 2 to 6 represent five watermarked images for keys 1-5. The data set thus obtained is used to train the extreme learning machine by using "Sigmoid" and "Sin" activation functions with respect to the number of hidden neurons $\left(\tilde{N}\right)$. Secondly, at one instance, the un-watermarked or any one watermarked image is tested with the learning machine, by varying $\tilde{N}$. Thus, two different parameters namely training time (seconds) and testing accuracy (%) are computed using the Matlab program of extreme learning machine [8, 9] with respect to $\tilde{N}$. These computed results are also plotted and analyzed in the light of watermark detection reported in the literature.

2.1 Watermark Embedding and Extraction

The embedding is performed in such a way that the perceptible quality of the image is not lost. The embedding scheme is mathematically represented as:

$$v_i' = v_i(1.0 + \alpha x_i) \tag{1}$$

where x_i is the watermark sequence to be embedded in image coefficients v_i' and as a result we obtain the coefficients of a signed image . The parameter α is known as embedding strength and is assumed as 0.1 for all our practical calculations. To ensure the robustness of the watermark embedding procedure, coefficients from the low frequency band in the transform domain are selected. The selected coefficients are subsequently modulated by the coefficients of a normally distributed random number sequence of length $n=500$ using (1). After embedding the watermarks, Inverse Discrete Cosine Transform (IDCT) of the image is computed to obtain the image back in the spatial domain. This image is a signed watermarked image. Studies related to extraction of the watermarks using the $SIM(X,X^*)$ correlation parameter and robustness studies done by computing MSE and PSNR have also been performed on these watermarked images and have been published elsewhere[10].

3 REVIEW OF EXTREME LEARNING MACHINE (ELM) MODEL

The Extreme Learning Machine (ELM) [5, 6, 7] is a Single hidden Layer Feed forward Neural Network (SLFN) architecture. Unlike traditional approaches such as Back Propagation (BP) algorithms which may face difficulties in manual tuning control parameters and local minima, the results obtained after ELM computation are extremely fast, have good accuracy and finally has a solution as that of a system of linear equations. For a given network architecture, ELM does not have any control parameters like stopping criteria, learning rate, learning epochs etc., and therefore, the implementation of this network is very simple and easy. The main concept behind this algorithm is that the input weights (linking the input layer to the hidden layer) and the hidden layer biases are randomly chosen based on some continuous probability distribution function such as uniform probability distribution in our simulation model and the output weights (linking the hidden layer to the output layer) are then analytically calculated using a simple generalized inverse method known as Moore – Penrose generalized pseudo inverse [9].

3.1 Mathematics of ELM Model

Given a series of training samples $(x_i,y_i)_{i=1,2,\ldots,N}$ and $\tilde{N}$ the number of hidden neurons where $x_i = (x_1,\ldots,x_{in}) \in \Re^n$ and $y_i = (y_1,\ldots,y_{im}) \in \Re^m$, the actual outputs of the single-hidden-layer feed forward neural network (SLFN) with activation function $g(x)$ for these N training data is mathematically modeled as

$$\sum_{k=1}^{\tilde{N}} \beta_k g\left(\langle w_k, x_i\rangle + b_k\right) = o_i, \forall i = 1,\ldots,N \tag{2}$$

where $w_k=(w_{k1},\ldots,w_{kn})$ is a weight vector connecting the k^{th} hidden neuron, $\beta_k = (\beta_{k1},\ldots,\beta_{km})$ is the weight vector connecting the k^{th} hidden neuron and output neurons and b_k is the threshold bias of the k^{th} hidden neuron. The weight vectors w_k are randomly chosen. The term $\langle w_k, x_i\rangle$ denotes the

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

inner product of the vectors w_k and x_i and g is the activation function.

The above N equations can be written as

$$H\beta = O \tag{3}$$

and in practical applications $\tilde{N}$ is usually much less than the number N of training samples and $H\beta \neq Y$,where

$$H = \begin{bmatrix} g(\langle w_1, x_1 \rangle + b_1) & \cdots & g(\langle w_{\tilde{N}}, x_1 \rangle + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\langle w_1, x_N \rangle + b_1) & \cdots & g(\langle w_{\tilde{N}}, x_N \rangle + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{\tilde{N}} \end{bmatrix}_{\tilde{N} \times m}, \quad O = \begin{bmatrix} o_1 \\ \vdots \\ o_{\tilde{N}} \end{bmatrix}_{\tilde{N} \times m} \text{ and } Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{\tilde{N}} \end{bmatrix}_{\tilde{N} \times m} \tag{4}$$

The matrix H is called the hidden layer output matrix. As analyzed by Huang et al [5,6,7] for fixed input weights $w_k = (w_{k1}, \ldots, w_{kn})$ and hidden layer biases b_k, we get the *least-squares solution* $\hat{\beta}$ of the linear system of equation $H\beta = Y$ with *minimum norm* of output weights β , which usually tend to have good generalization performance.

The resulting $\hat{\beta}$ is given by $\hat{\beta} = H^{+}Y$

where matrix H^{+} is the Moore-Penrose generalized inverse of matrix H[9]. The above algorithm may be summarized as follows:

3.2 The ELM Algorithm

Given a training set $S = \{(x, y) \in \Re^{m+n}, y_i \in \Re^m\}_{i=1}^{N}$, activation function $g(x)$ and the number of hidden neurons $\tilde{N}$;

Step1: For $k=1,\ldots, \tilde{N}$ randomly assign the input weight vector $w_k \in \Re^n$ and bias $b_k \in \Re$

Step2: Determine the hidden layer output matrix H.

Step3: Calculate H^{+}.

Step4: Calculate the output weights matrix $\hat{\beta}$ by $\hat{\beta} = H^{+}T$.

Many activation functions can be used for ELM computation. In the present case, Sigmoid and Sin activation functions have been used to establish soft decision detection of digitally watermarked images.

3.3 Computing the Moore-Penrose Generalized Inverse of a matrix

Definition 1.1: *A matrix G of order* $\tilde{N} \times N$ *is the Moore-Penrose generalized inverse of real matrix A of order* $N \times \tilde{N}$ *if AGA=A, GAG=G and AG, GA are symmetric matrices.*

Several methods, for example orthogonal projection, orthogonalization method, iterative methods and singular value decomposition (SVD) methods exist to calculate the Moore-Penrose generalized inverse of a real matrix. In ELM algorithm, the SVD method is used to calculate the Moore-Penrose generalized inverse of H. Unlike other learning methods, ELM is very well suited for both differential and non – differential activation functions. As stated above, in the present work, computations are done using "Sigmoid" and" Sin" activation functions.

4 RESULTS AND DISCUSSIONS

Embedding of 500 watermarks generated using normally distributed random number sequence with different seed values does not result in any perceptible difference in this work. The original / un-watermarked and signed images (for *n=500* with keys 1-5) are respectively shown in Figures 1(a)-(f).

Figure 1. (a) Un-watermarked Image Lena (128×128), (b)-(f) Signed Images with Normally Deviated Watermarks n=500 having keys 1-5 respectively

Table 1. Training Time (seconds) w. r. t. number of hidden neurons $(\tilde{N})$ for activation function "Sigmoid"

No of hidden neurons $(\tilde{N})$	Training Time (seconds)
50	0.0469
100	0.1406
150	0.2500
200	0.4375
250	0.7813
300	1.1719
350	1.8438
400	2.7188
450	3.5938
500	5.5313
550	6.9219
600	8.7969

The computed results for training time w. r. t. $\tilde{N}$ taking into account activation functions "Sigmoid" and "Sin" are respectively tabulated in table 1 and 3. Similarly, computed results for testing accuracy (%) w. r. t. $\tilde{N}$ taking into account the

Table 2. Testing Accuracy (%) w. r. t. number of hidden neurons $(\tilde{N})$ for activation function "Sigmoid"

No of Neurons $(\tilde{N})$	Testing Acc (%) for n=0 (Original)	Testing Acc (%) for n=500 with key = 1	Testing Acc (%) for n=500 with key = 2	Testing Acc (%) for n=500 with key = 3	Testing Acc (%) for n=500 with key = 4	Testing Acc (%) for n= 500 with key = 5
50	4.69	21.09	24.22	28.13	22.66	24.22
100	7.03	22.66	28.13	27.34	30.47	30.47
150	10.94	35.94	32.81	35.94	33.59	39.84
200	17.19	55.47	50.78	56.25	57.03	60.16
250	33.59	73.44	75.78	64.06	71.88	71.09
300	53.13	88.28	79.69	82.81	90.63	82.81
350	59.38	92.19	92.19	92.19	92.97	88.28
400	85.16	96.88	96.88	95.31	96.88	96.06
450	89.84	100	100	100	100	100
500	90.63	100	100	100	100	100
550	95.31	100	100	100	100	100
600	100	100	100	100	100	100

two activation functions for all six images are tabulated in tables 2 and 4 respectively.

We have plotted these results in Figures 2 through 5 respectively. From these figures we observe that as the number of hidden neurons increases so is the training time. However, the training time even for the case of N = 600 is just a few seconds clearly shows that ELM is an efficient algorithm to train a large multiclass data set submitted to the learning machine. This is specifically important for watermark detection in the real time application domain for videos. The efficiency exhibited by the ELM training algorithm places it in a class altogether different from that of ANNs or SVMs. Moreover, the variation of the training time with respect to $\tilde{N}$ is very much similar for both activation functions used in the present work (Figures 2 and 4).

Table 3. Training Time (seconds) w. r. t. number of neurons $(\tilde{N})$ for activation function "Sin"

No of hidden neurons $(\tilde{N})$	Training Time (seconds)
50	0.0469
100	0.1406
150	0.2656
200	0.4688
250	0.7500
300	1.2656
350	1.7813
400	2.8125
450	3.7500
500	5.875
550	7.2656
600	9.1875
700	12.875
750	14.4375

Figures 3 and 5 show that the testing accuracy curves for the un-watermarked image (*n=0* or class label 1) are clearly differentiable from all other curves obtained for the watermarked images with different keys. These results are also very much similar for both activation functions employed in the present work. From these figures we clearly observe that the testing accuracy curves for the watermarked images are distinctly obtained within one region of the figure with reference to the plot of testing accuracy values for the un-watermarked image. This is because, numerical values of the testing accuracy parameter for all watermarked images (class label 2 to 6) do not match with those obtained for the un-watermarked image. Therefore, a differential in the numerical values of testing accuracy is obtained between the un-watermarked image on one hand and all five signed images on the other in both the cases. This differential is interpreted as the actual classification being done by the Extreme Learning Machine on the basis of the multiclass data set provided to it for training and testing purposes. Thus, it can be concluded that the Extreme Learning Machine (ELM) algorithm can be used as a multiclass classifier tool for blind zero – bit detection within the category of soft decision detectors of the watermarks on the basis of the testing accuracy computation. This further indicates that testing accuracy as computed by the ELM algorithm can be used to establish the qualitative and quantitative distinction between a watermarked image used as a test image on one hand and the un-watermarked image used as a test image on the other.

It is interesting to note here that although Figures 3 and 5 show very much similar behavior along with an outcome of the differential as explained above, yet there is a minor difference between the obtained numerical values for training time and testing accuracy in the two cases. A close look at the tabulated values (Tables 1-4) indicate that in case of "Sigmoid" activation function,100% testing accuracy is achieved for $\tilde{N}$=450 only which could be achieved in case of "Sin" activation function only at $\tilde{N}$=600. Similarly, the training time obtained for "Sigmoid" and "Sin" activation functions do not match either. The training time for the "Sigmoid"

Figure 2. Plot of Training Time (seconds) w .r. t. number of hidden neurons $\left(N^{\hat{H}}\right)$ for activation function "Sigmoid"

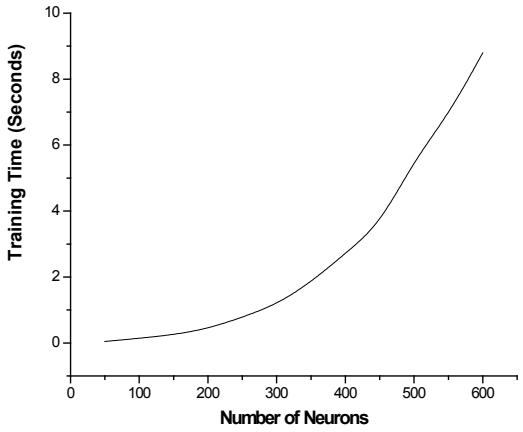


Figure 4. Plot of Training Time (seconds) w. r. t. number of hidden neurons $\left(N^{\hat{H}}\right)$ for activation function "Sin"

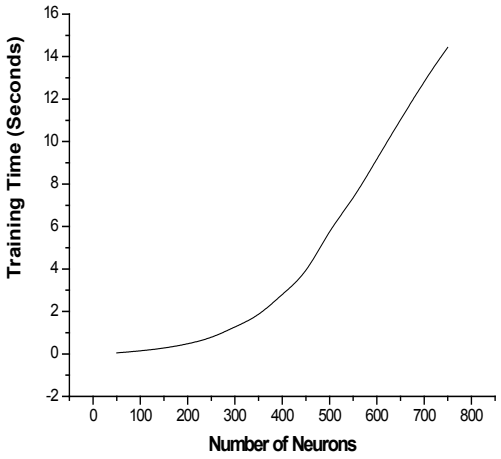


Figure 3. Plot of Testing Accuracy (%) w. r. t. number of hidden neurons $\left(N^{\hat{H}}\right)$ for activation Function "Sigmoid"

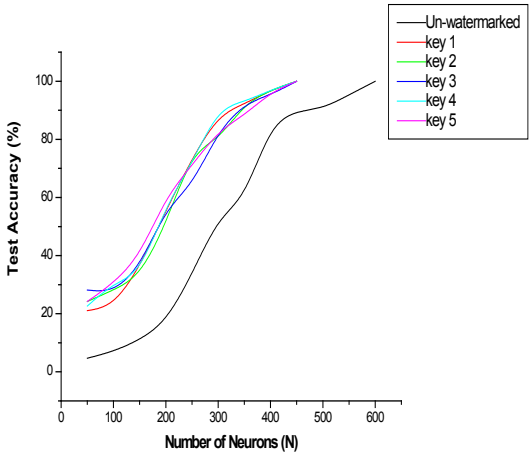


Figure 5. Plot of Testing Accuracy (%) w. r. t number of hidden neurons $\left(N^{\hat{H}}\right)$ for activationfunction "Sin"

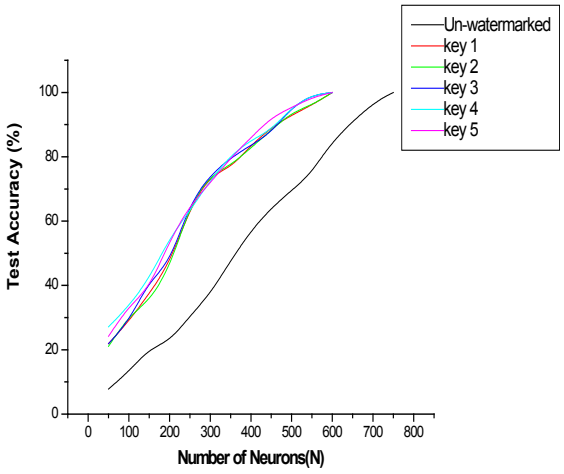


Table 4. Testing Accuracy (%) w. r. t. number of hidden neurons $\left(N^{\hat{H}}\right)$ for activation function "Sin"

No of Neurons $\left(N^{\hat{H}}\right)$	Testing Acc (%) for n=0 (Original)	Testing Acc (%) for n=500 with key = 1	Testing Acc (%) for n=500 with key = 2	Testing Acc (%) for n=500 with key = 3	Testing Acc (%) for n=500 with key = 4	Testing Acc (%) for n= 500 with key = 5
50	7.81	21.88	21.09	21.88	27.15	24.27
100	13.28	28.91	30.47	28.91	33.59	33.59
150	20.31	37.50	35.16	41.41	42.19	39.06
200	22.66	46.88	45.31	46.88	54.69	53.91
250	30.47	64.06	64.84	65.63	63.28	64.84
300	37.50	74.22	74.22	74.22	72.66	71.88
350	47.66	76.56	77.34	79.69	80.47	79.69
400	57.03	83.59	82.81	83.59	85.16	85.94
450	64.06	89.06	88.28	87.5	88.28	92.19
500	69.53	92.97	93.75	95.31	95.31	95.31
550	75.00	96.09	96.09	99.22	99.22	98.44
600	85.16	100	100	100	100	100
700	96.88	100	100	100	100	100
750	100	100	100	100	100	100

activation function is less than what is obtained in case of "Sin" activation function. Thus, a comparison between the activation functions used in the present work is done on the basis of the training time computation. It can, therefore, be concluded that among the two functions used here, "Sigmoid" is proved to be the better and efficient option as far as time complexity of the training algorithm is concerned.

5 CONCLUSIONS

This research paper presents a method to successfully identify an un-watermarked image from a given set of un-watermarked image and a few watermarked images using the ELM algorithm used as a multiclass classifier tool. Normalized data values belonging to a mixture of an un-watermarked and five watermarked images are first used as a multiclass data set with class label 1(for un-watermarked image) and class labels 2-6 (for watermarked images signed with keys 1-5) submitted to train the learning machine. First, the time (in seconds) consumed to train the learning machine is computed by varying the number of hidden neurons $\left(\tilde{N}\right)$ and by employing two different activation functions "Sigmoid" and "Sin". The numerical values obtained for training time are plotted with respect to $\tilde{N}$ and this result is used to establish a comparison between the efficacies of the two activation functions used in the present work. Secondly, an individual image (un-watermarked image or any watermarked image) is also tested with the trained learning machine for computation of testing accuracy (%) with respect to number of hidden neurons $\left(\tilde{N}\right)$.

The results obtained in this experiment are interesting. First, the numerical values obtained for the training time are just a few seconds. This makes ELM algorithm very much suitable to be used as a blind zero – bit watermark detector within the category of soft decision detectors especially for real time watermarking applications for video sequences.

Second, the machine is able to identify the un-watermarked image from the given mixed set of watermarked and un-watermarked images. Plots of testing accuracy with respect to $\tilde{N}$ indicate that there is a qualitative and quantitative difference between the tested un-watermarked image on one hand and the tested watermarked image on the other. This is established by the large differential in the numerical values of testing accuracy obtained in the two cases. This result is obtained for both the activation functions in the present experiment. Thus, it is concluded that the ELM algorithm as a multiclass classifier tool can be used for blind zero – bit watermark detector within the category of soft decision detectors to detect normally distributed robust watermarks embedded in digital images in the transform domain.

Third, a comparison in the numerical values for training time computed in case of two activation functions indicate that for "Sigmoid" function, 100% testing accuracy is obtained at $\tilde{N}$=450 which could only be obtained at $\tilde{N}$=600 for "Sin" function. Moreover, training time (in seconds) for "Sigmoid" is less than that obtained for "Sin". This proves that out of the two activation functions used in the present work, "Sigmoid" works efficiently for watermark detection.

6 FUTURE SCOPE

This work should be theoretically and experimentally examined thoroughly by taking into account more number of activation functions. Since, the multiclass data set submitted to train the learning machine is a big one, effort should be made to eliminate redundancies within the normalized data. This will help improving the performance of the training algorithm even better. Another important extension of this work is to design and implement the hard decision detector having multiple bit detection capability. That is, to decode the type and location of a watermark in any given image – available either in spatial or transform domain. The present work will be extended to satisfy this important criterion of the detection procedure. The reliability of the ELM algorithm used to detect a generic watermark will also be checked as a subsequent step.

REFERENCES

1. Ingemar J. Cox, Joe Kilian, F. Thomson Leighton and Talal Shamoon (1997), "Secure spread spectrum watermarking for multimedia", IEEE Transactions on Image Processing, vol 6(12), pp 1673-1687.
2. Nikolaos Nikolaidis and Ioannis Pitas (2004), "Benchmarking of watermarking algorithms", Intelligent Watermarking Techniques, World Scientific, New Jersey, vol(7), pp. 315-347.
3. Then Patrick H. H, Wang Y. C (2006), "Support vector machine as digital image watermark detector", Proceedings of the SPIE, vol (6064), pp. 478-489.
4. Madan M Gupta, Liang Jin and Noriyasu Homma (2002), "Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory", IEEE Press & Wiley Interscience, New Jersey.
5. M-B. Lin, G-B Huang, P. Saratchandran and N. Sudararajan (2005), "Fully complex extreme learning machine", Neurocomputing, vol (68), pp 306 - 314.
6. G -B Huang, Q -Y Zhu and C K Siew (2006), "Extreme Learning Machine: Theory and Applications", Neurocomputing, vol (70), pp 489-501.
7. G -B Huang, Q-Y Zhu and C K Siew (2006), "Real-Time Learning Capability of Neural Networks", IEEE Transactions on Neural Networks, vol 17(4), pp 863-878.
8. G-B Huang (2004), The Matlab code for ELM is available on: http://www.ntu.edu.sg/home/egbhuang (June 2007).
9. D. Serre (2002), "Matrices: Theory and Applications", Springer Verlag, New York Inc.
10. Anurag Mishra and S Balasundaram (2007), "Robustness studies on statistically distributed watermarked images", Abstract Proceedings of National Conference on Mathematical Modeling, Optimization and their Applications (Optima-2007), India, pp 45, ISBN 978-81-904526-1-8. Proceedings

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

A wiki-system with integrity support for structured data

Jaehui Park, Sang-goo Lee, Jonghoon Chun

School of Computer Science & Engineering, Seoul National University, Seoul, Republic of Korea,
Department of Computer Engineering, Myongji University, Seoul, Republic of Korea
{jaehui, sglee}@europa.snu.ac.kr, jchun@mju.ac.kr

Abstract A Wiki application allows people to create and modify any number of documents in a collaborative fashion. However, the information contained in the documents cannot be readily utilized by a computer program. In order for a program to do so, it would have to parse and analyze the contents of the documents just as it would any text or HTML document. We propose a way of using the Wiki for creating and updating structured data such as those in a relational database. We are able to validate the data by constraining the data types, which allows immediate utilization by a computer program. We implement the model for a music data management system called "WikiMusic".

Keywords Web 2.0, Wiki, Media Wiki, forms

1 INTRODUCTION

It is difficult to define Web 2.0; it may be regarded as a trend, business model, or a set of technologies. We can also view it as a set of new services, and among them, Wikipedia [1] is perhaps the most well-known and significant application. It is the largest encyclopaedia harvesting the collective intelligence of millions of voluntary contributors. A document about a subject can be created by anyone and then be viewed and edited by anyone. Each document is the result of a collective effort of dozens, hundreds, or even thousands of people. The accumulated information can be used for various purposes and applications [2]. We call such a system a Wiki-system or a Wiki.

As with HTML documents in the World Wide Web, the documents in a Wiki are primarily for human reading; they have formatting tags and hyperlinks to other documents. The information within the documents cannot be used immediately by a computer program. In order for a program to utilize the information embedded in the documents, it would have to parse and analyze the contents and extract individual data and relationships in structured forms.

Furthermore, in an open environment where large number of (virtually anonymous) users modify the contents, it is impossible to ensure the validity of the resulting information. Incorrect data, out-of-bound values, and vandalisms are all sources of information corruption.

We present a modified Wiki-system where structured data elements can be defined. By allowing data type definitions for these elements, it is possible to validate the data values at entry time. The data collected for structured elements can be stored as database records in a relational database. The set of information can be easily utilized by other applications through a simple database access. Also, form-based user interfaces can be supported for more convenient data entries and updates.

In section 2, we introduce other works that identify problems in Wiki-systems and propose remedies for them. In section 3, we present structured elements and documents with structured elements. The database storage model for the structured elements is introduced in section 4. In section 5, the WikiMusic system is introduced. It is a music information collection and management system using our structured elements and database model. We present our conclusion in section 6.

2 PREVIOUS WORK

Wiki comes from "WikiWiki" which means "fast" in the Hawaiian language. A Wiki is a software that enables anyone to simply create and edit web documents [3]. Documents are written using a markup language that is a simple extension of HTML. Hyperlinks can be embedded in the documents. A document is stored as a text object in a database system. This allows a very simple management of documents and so the Wiki documents can be altered more dynamically than the documents in a conventional web site. Due to these simple concepts of interactions and managements, Wiki has been a popular utility in building collective intelligence services on the web.

But the documents in a Wiki are primarily for human reading. Computer programs cannot understand the contents of the documents in Wiki systems. There have been a few attempts to giving more semantics to Wiki documents.

2.1 Platypus

Semantic Web [4] language editor is used in Platypus [5] for enhanced data sharing and reusing of Wiki documents. Each Wiki document has an associated RDF [6] document which is a translated form of the original. The contents are viewed and edited simultaneously. RDF makes the contents more understandable for computer programs.

However, users have to move between the RDF document editor window and the Wiki editor window checking the consistency between the two manually. Also, users are required to know the RDF language which is difficult for most users of the Wiki-system.

Compatibility with existing Wikis is another weakness of the approach.

2.2 Semantic Wikipedia

Semantic Wikipedia [7] adds semantic constructs to Wikipedia in order to better utilize the information embedded in the knowledge base. Hyperlinks can be 'typed' so that meanings can be associated to the links. Certain parts of the document can be declared as attribute which is a way of separating a piece of data.

However, the information can be used only through an RDF export process. Also, there are no type checking of the attribute; they are freeform text data.

3 STRUCTURED ELEMENTS

3.1 What is Structured Data Element

Structured data elements are those whose values are formatted. Structured data includes atomic data (with predefined domain) and composite data which are uniform compositions of other structured data. Unstructured data are freeform data. For example, data in a relational database are mostly structured while free text documents are unstructured data.

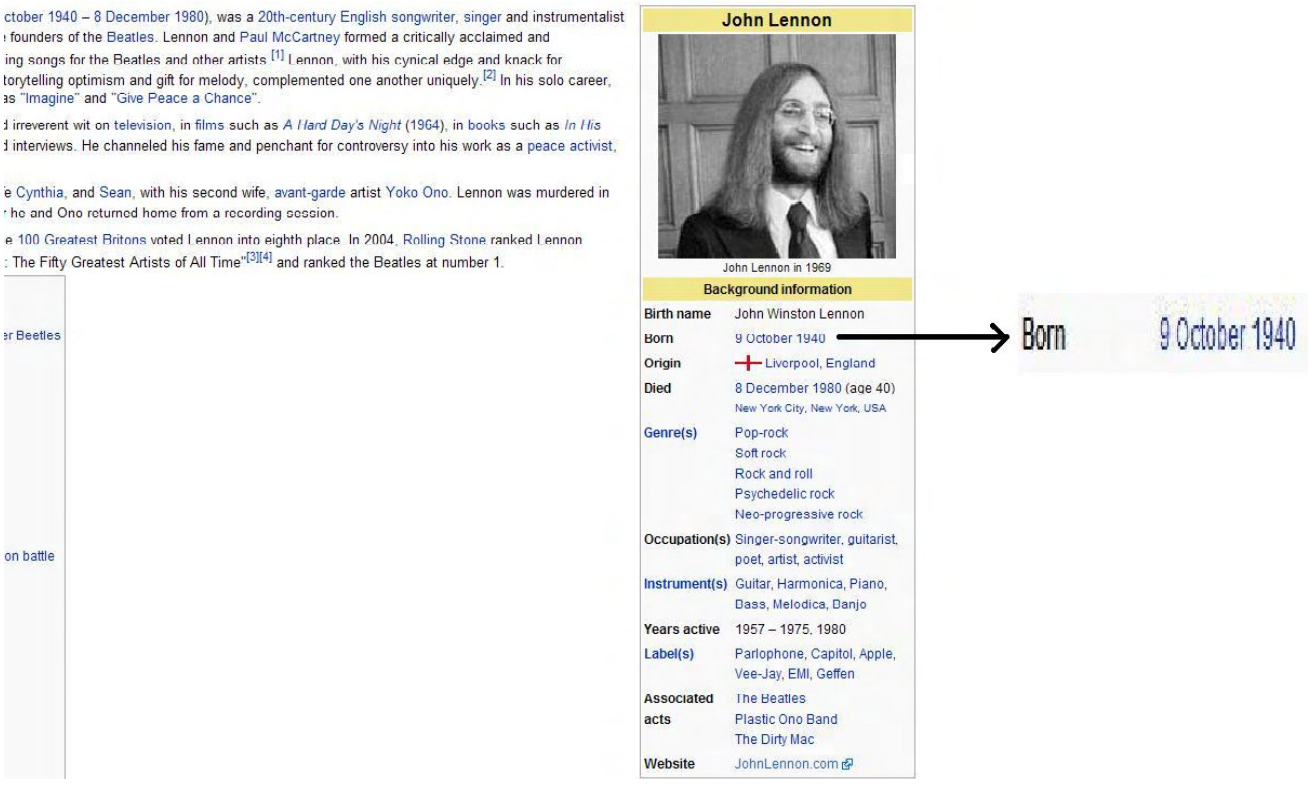
The distinction between structured composite data and unstructured data is whether a computer program can expect the way that the internal components are composed. For example, a date field is structured if all values are in "dd-mm-yyyy" format. It is unstructured if all forms of dates are allowed; for example, "Jan/21/2007" and "the 21st of January, 2007".

For our purpose, we regard HTML documents as unstructured data (rather than semi-structured as they are frequently called).

3.2 Use of Structured Data Elements

As the amount of its accumulated information increases at an alarming rate, Wikipedia introduces a form of synopsis called "infobox" for commonly used subject types such as musical artists, music, and country. For example, the document 'John Lennon' contains an infobox including 'Birth name', 'Born', 'Died', 'Genres', etc, as shown in figure 1. These data elements are defined for 'musical artists'.

Figure 1. The 'Born' data element for 'John Lennon' in Wikipedia



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 2. 'The Beatles' in Wikipedia

The Beatles

From Wikipedia, the free encyclopedia
(Redirected from Beatles)

The Beatles were an English musical group from Liverpool whose members were John Lennon, Paul McCartney, George Harrison, and Ringo Starr. They are one of the most commercially successful and critically acclaimed band in the history of popular music.[2]

The Beatles are the best-selling musical act of all time in the United States of America, according to the Recording Industry Association of America, which certified them as the highest selling band of all time based on American sales of singles and albums.[3] In the United Kingdom, The Beatles released more than 40 different singles, albums,

Although the basic format of its documents is free-form HTML, Wikipedia is introducing a certain degree of structure to better organize its information. Users are encouraged to fill in the infobox in the predefined form. As part of the WikiProject [8], the set of infobox data elements for specific document types are decided by a group of experts in the respective field.

The infobox mostly consists of information elements that are inherently structured. The birth date of a person, 'Born' element, is a date field. However, the responsibility to follow these formats is up to the user. You would edit the values as free text and then store them as a part of the HTML document.

3.3 Benefits of Structured Data Elements

By specifying the set of data elements for a document (type), the type of the information that is collected for the document can be standardized. By requiring a minimal set of structured information while accommodating freeform elements, we can greatly enhance the quality of information collected in a Wiki-system.

If an information element is of a structured data type, it is possible to validate the data value entered by a user. We can even specify the range of values that can occur for a specific data element. If the Wiki-system presents the user with form-based user interface that reflects the set of structured elements, the user can conveniently update the respective values just as she/he use any form-based database application.

4 STORING DOCUMENTS

4.1 Data Integrity

Collaborative system like Wikipedia runs on the premise that the more people participate in editing a document, the more accuracy it gets. While this is generally true for free-form natural language documents, it is error-prone for simple data values. System level data validation has proven to be essential for enterprise database applications, especially for structured data elements.

If the structured data element is defined with data type and domain information, the value of the data can be validated and the integrity of that value will be guaranteed when the documents are stored. Also, the structured data values of a document can be stored as a record in a table in relational database.

A domain expert can predefine a document type by specifying the data elements. By specifying the structured data elements, she/he is in fact designing the database schema for the document type. So, it is possible to store the document in the Wiki-system and the structured elements in a separate relational database system (redundantly). Other computer programs can make use of the database, and hence, the utility of the knowledge base increase greatly.

4.2 Data Model

The data model for a conventional Wiki-system can be modeled as follows.

$System = \{Article_1, Article_2, \ldots, Article_n\}$

$Article_i = \{(v, l) \mid v \text{ is text, } l \text{ is hyperlink to } Article_j \text{ or } l \text{ is null}\}, i \neq j$

System represents the entire document set of the Wiki-system which consists of a set of *Articles*. An *Article* is a set of pairs of values and links. A value v can be the label of a hyperlink or a chunk of freeform text (where the associated link l is null). The following text from Wikipedia(Figure 2) is decomposed into value and link pairs.

System = {The Beatles, England, Liverpool, John Lennon, Paul McCarthney, George Harrison, Ringo Star, History of music, Musical ensemble, United States, …} // set of articles

The Beatles = { ("The Beatles were an", null), ("English", Link(England)), ("musical group from", null), ("Liverpool", Link(Liverpool)), ("whose members were", null), ("John Lennon", Link(John Lennon)), ("Paul McCarthney", Link(Paul

Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

McCarthney)), ("George Harrison", Link(George Harrison)), ("and", null), ("Ringo Star", Link(Ringo Star)), ("They are one of the most commercially successful and critically acclaimed band in th", null), ("history of popular music", Link(History of music)), … }

An *Article* can have several links, so the document collection of a Wiki-system is a graph whose nodes represent documents (articles). As the content of a node is free text, it is difficult for a computer program to use it for other purposes.

We propose a data model where every *value* in a document is associated with a predefined structured element. In effect, we are maintaining semantic information related to structured elements.

First we define a document type by specifying the schema consisting of data elements. Each data element has a domain which can be simple data types such as integer and string or a Wiki-text type that is similar to the article type of a conventional Wiki document. A document type is given a namespace.

ns_i is a namespace with an associated schema $\{E_{i1}, E_{i2}, \ldots Ei_{ni}\}$, for $i > 0$.

E_{ij} = <$label_{ij}$, $domain_{ij}$>,
where $label_{ij}$ is the name of the data element and $domain_{ij}$ is its domain such as *integer*, *string*, *range(0, 100)*, or *wiki-text*, for $1 \le j \le n_i$.

A wiki-text is a part of article of the form {(v, l) | v is text, l is hyperlink to Articlek or l is null}

The system now consists of *typed* articles.

System = $\{ns_i$:$Article_k \mid i, k > 0\}$

ns_i:$Article_k$ = $\{T_{i1}^k, T_{i2}^k, \ldots T_{ini}^k\}$

T_{ij}^k = < v, l > where v≡$domain_{ij}$, l is hyperlink to $Article_m$ or l is null }, $1 \le j \le ni, j \ne m$

The data model uses the same graph structure as the existing model in Wiki-system but has richer semantics. For example, suppose there is a document about a person whose name is 'John Lennon'. It has structured element called 'has Album' which has a link to another document 'Imagine' of type 'Album'. In the proposed data model, 'Artist:John Lennon' is the document title meaning that its namespace is 'Artist'. The other document will be 'Album:Imagine'. So, the semantics of this document is "Artist whose name is 'John Lennon' has an Album titled 'Imagine'".

We present another example.

System = {Artist:Michael Jackon, Music:Billie Jean, Music: Thriller, Artist:Bon Jovi, …}

Artist Schema = { <Name, string>, <Picture, binary-object>, <Gender, {male, female}>, <Artist Type, {solo, group,…}>, <Birth date, date>, <Country, string>, …, <Introduction, wiki-text>, … }

Artist:Michel Jackson = {$Name^{MJ}$, $Picture^{MJ}$, $Gender^{MJ}$, Artist $Type^{MJ}$, Birth $date^{MJ}$, $Country^{MJ}$, …, $Introduction^{MJ}$, … }

$Name^{MJ}$ = < "Michael Jackson", null >

$Gender^{MJ}$ = < "male", null, >

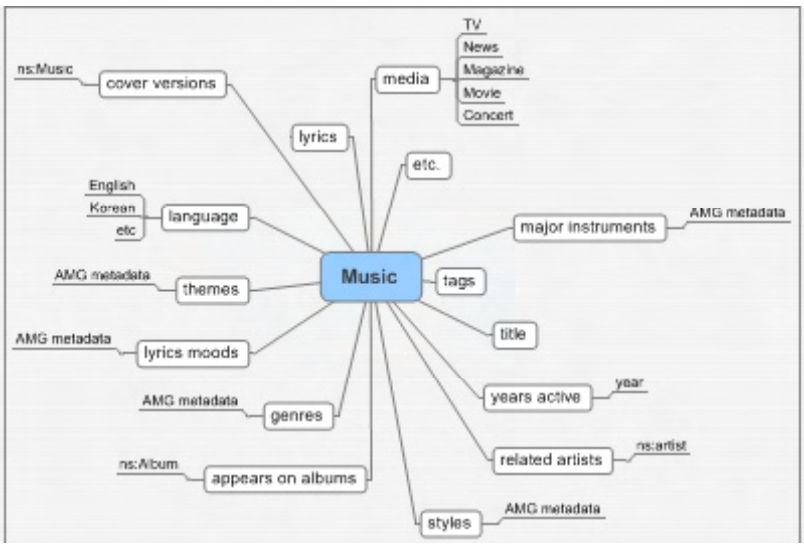
Artist $Type^{MJ}$ = < "solo singer", null >

Brith $date^{MJ}$ = < "August 29, 1958", null >

$Country^{MJ}$ = < "USA", Link(main:USA) >

Introduction = < { ("Jackson began his musical career at the age of seven as the lead singer of, null), ("The Jackson 5", Link(Artist:The Jackson 5)), ("He released his first solo recording", null), ("Got to Be There", Link(Album:Got to Be There)), …}, null >

Figure 3. Data elements for namespace Music



Proceedings of the International Conference on Information Society (i-Society 2007), October 7–11, 2007, Merrillville, Indiana, USA. ISBN 0-9546628-1-4 www.i-society.org/2007 © e-Centre for Infonomics

Figure 4. The form-based user interface (for namespace Artist)

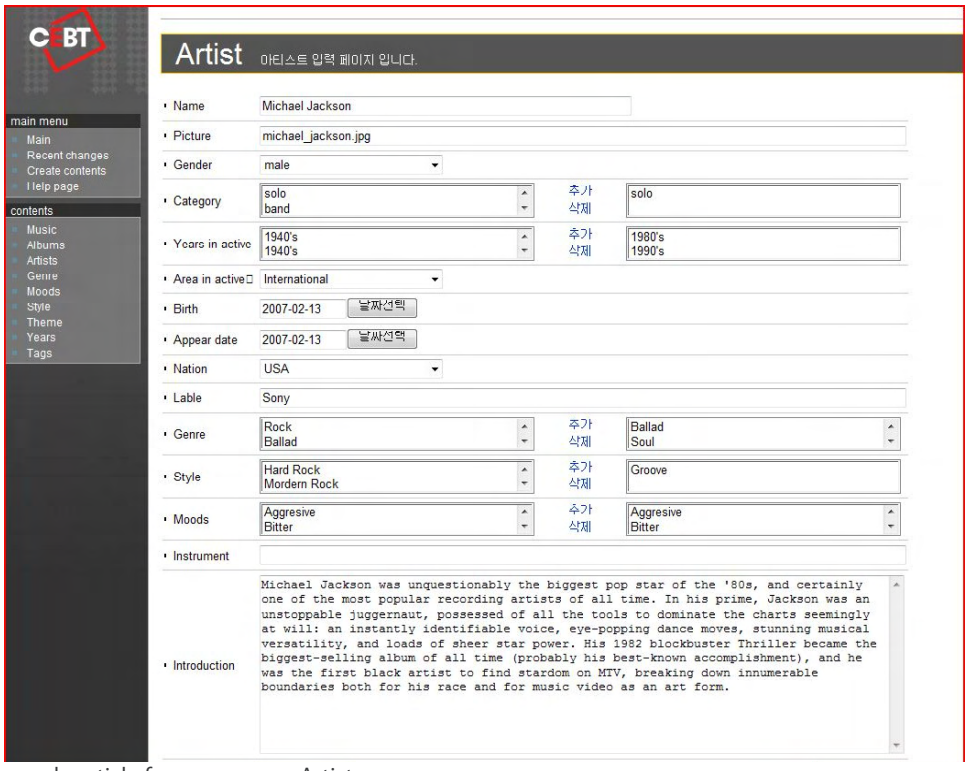
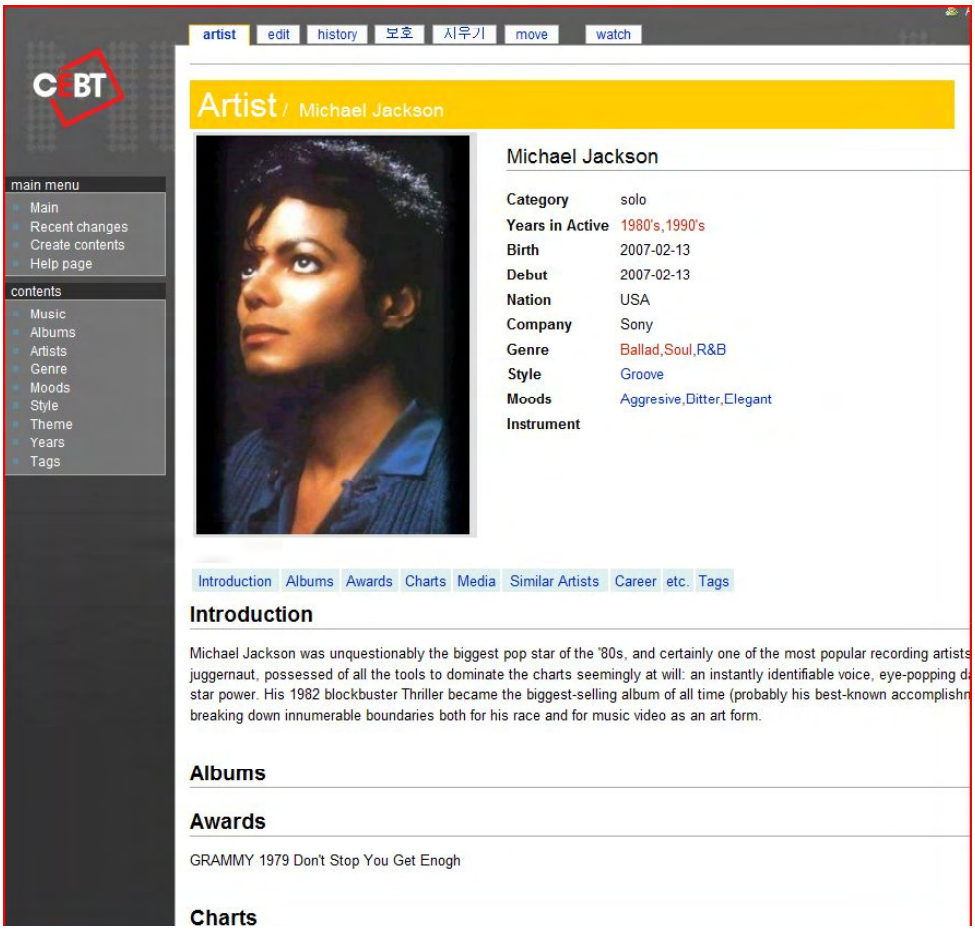


Figure 5. An example article for namespace Artist



5 IMPLEMENTATION

We have implemented a Wiki-system for music. The music data model is based on All Music Guide (AMG) [9].

5.1 Environment

Documents are managed by the Mediawiki Engine [10] which is the software used by Wikipedia. The Oracle DBMS is used for storage of tables representing the structured elements. We referenced the All Music Guide definitions for

the data model, i.e., the set of structured data elements for music.

The form-based extension to the editor is implemented in Java scripts and php languages.

5.2 Implementations

Three document types (namespaces) have been defined; *Music*, *Album*, and *Artist*. They are defined using our extended model presented in section 4. Figure 3 is the metadata diagram related to Music namespace.

There is a specific editor for each namespace. When creating or editing an article, the appropriate namespace is chosen. The edit page of the document is form-based determined by the namespace. The form-based editor interface for Music is shown in Figure 4. The navigation sidebar is provided for added user convenience. The article view for a Music namespace document is shown in figure 5.

6 CONCLUSIONS

We have presented a way of introducing structured data elements into a Wiki-system. By allowing data type definitions for these elements, it is possible to validate the data values at entry time. While the combined information is stored as one document in a Wiki-system, the data collected for structured elements can be exported as database records into a relational database, upon which additional applications can be built easily. We believe our system has the following characteristics.

- Usability
  Since the interface is form-based, users need not to learn the language constructs such as Wiki markups and RDF. Users only need to concentrate on the actual contents. Still, the user can always use the conventional markup language in any of the free formatted elements.
- Expressiveness
  We require the domain experts decide on the data elements for a document (type) beforehand. This may be seen as a limit. However, by defining these information elements, we are actually enriching the article; similar to the infobox of Wikipedia.
- Flexibility
  Because a document type has associated schema, documents in our system are less flexible compared to conventional Wiki-systems. But for validation of data and rich semantic information, a certain degree of flexibility loss is inevitable.
- Compatibility
  We use the MediaWiki Engine and implemented extensions. The typed data elements will appear as elements with special tags. Thus, it is possible to browse and edit (in a freetext form) documents from our system in Wikipedia and other MediaWiki's. The WikiMusic system which implements our approach is used as a metadata collection medium for a music recommendation service. As more validated information is collected from the Wiki, the recommendation engine will make more informed decisions.

Future works include semantic validation for atomic values.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program.

REFERENCES

1. Wikipedia; http://en.wikipedia.org/wiki/Main_Page (July 31 2007)
2. J. Wales. Wikipedia and the free culture revolution, OOPSLA/ WikiSym Invited Talk, 2005.
3. Bo Leuf, and Ward Cunningham(2001), The Wiki Way: Collaboration and Sharing on the Internet, Addison-Wesley, 1st edition.
4. Semantic Web, http://www.w3.org/2001/sw/ (July 31 2007)
5. S. E. Campanini, P. Castagna, and R. Tazzoli (2004). 'Platypus wiki: A semantic wiki wiki web', In Semantic Web Applications and Perspectives, Proceedings of 1st Italian Semantic Web Workshop
6. Resource Description Framework (RDF), http://www.w3.org/RDF/ (July 31 2007)
7. Max Völkel, Markus Krötzsch, Denny Vrandecic, Heiko Haller, Rudi Studer (2006), 'Semantic Wikipedia', In Proceedings of the 15th International Conference on World Wide Web, 585-594
8. WikiProject, http://en.wikipedia.org/wiki/Wikipedia:WikiProject (July 31 2007)
9. All Music Guide, http://www.allmusic.com (July 31 2007)
10. MediaWiki, http://www.mediawiki.org (July 31 2007)