ICT&A 2011

DeMset 2011

November 29th - December 2nd, 2011 – Orlando, Florida, USA

International Conference on
Information and Communication Technologies and Applications

International Conference on
Design and Modeling in Science, Education, and Technology

PROCEEDINGS

Post-Conference Edition

Edited by:

Nagib Callaos
Michael Savoie
Mohammad Siddique
C. Dale Zinn

Organized by
International Institute of Informatics and Systemics
Member of the International Federation for Systems Research (IFSR)

COPYRIGHT

The papers of this book comprise the proceedings of the conference mentioned on the title and the cover page. They reflect the authors' opinions and, with the purpose of timely disseminations, are published as presented and without change. Their inclusion in these proceedings does no necessarily constitute endorsement by the editors.

ISBN- 978-1-936338-51-1

PROGRAM COMMITTEE

Chairs: C. Dale Zinn (USA)
Hsing-Wei Chu (USA)

Abdel-Qader, Ikhlas	Western Michigan University	USA
Abusitta, Adel	Ajman University of Science and Technology	UAE
Acharya, Sushil	Robert Morris University	USA
Adamopoulou, Evgenia	National Technical University of Athens	Greece
Affenzeller, Michael	Upper Austrian University of Applied Sciences	Austria
Aguirre-Muñoz, Zenaida	Texas Tech University	USA
Aksoy, M. S.	King Saud University	KSA
Al Obaidy, Mohaned	Gulf College	Oman
Alhamouz, Sadeq	Umm Al-Qura University	KSA
Als, Adrian	University of the West Indies	Barbados
Alshara, Osama	Higher Colleges of Technology	UAE
Alvarado Moore, Karla	University of Central Florida	USA
Alzamil, Zakarya	Riyadh College of Technology	KSA
Amaral, Luis	University of Minho	Portugal
Ammann, Walter J.	SLF	Switzerland
Anantharaj, Valentine	Mississippi State University	USA
Andina, Diego	Technical University of Madrid	Spain
Anton, José M.	Technical University of Madrid	Spain
Anunciação, Pedro	Polytechnic Institute of Setubal	Portugal
Aoki, Toru	Shizuoka University	Japan
Aruga, Masahiro	Tokai University	Japan
Assaf, Mansour	University of Trinidad and Tobago	TRI
Astakhov, Vadim	University of California San Diego	USA
Aukstakalnis, Nerijus	Kaunas University of Technology	Lithuania
Auvinen, Anssi	University of Tampere	Finland
Bangert, Patrick	Algorithmica Technologies GmbH	USA
Barkana, Atalay	Anadolu University	Turkey
Barkstrom, Bruce	retired	USA
Barros-Justo, José Luís	University of Vigo	Spain
Baruah, Debendra C.	Tezpur University	India
Basso, Giuliano	Institute of Electrical and Electronics Engineers	Belgium

Belcher, E. Christina	Trinity Western University	Canada
Benbouziane, Mohamed	University of Tlemcen	Algeria
Benedicenti, Luigi	University of Regina	Canada
Bennett, Leslie	University of Louisville	USA
Bermeo, José	University of the Andes	Colombia
Bernardino, Jorge	Coimbra Institute of Engineering	Portugal
Bezuglov, Anton	University of South Carolina	USA
Bhat, Talapady N.	National Institute of Standards and Technology	USA
Bhattacharyya, Siddhartha	University of Kentucky	USA
Bidarra, José	University of Aberta	Portugal
Blair, Madelyn	Pelerei, Inc.	USA
Blanchard, Richard	Loughborough University	UK
Boguslavsky, Andrey A.	Russian Academy of Sciences	Russian Federation
Bolboaca, Sorana Daniela	University of Medicine and Pharmacy	Romania
Bönke, Dietmar	Reutlingen University	Germany
Borchers, Carsten H. J.	University of Hanover	Germany
Botto, Todd	Quinnipiac University	USA
Boukachour, J.	Le Havre University	France
Bradl, Peter	University of Würzburg	Germany
Branski, L. K.	University of Texas Medical Branch at Galveston	USA
Broussard, Randy P.	United States Naval Academy	USA
Bubnov, Alexej	Academy of Sciences of the Czech Republic	CZE
Bueno, Newton Paulo	Federal University of Viçosa	Brazil
Burke, David	Robert Morris University	USA
Burnett, Andrea	University of the West Indies	Barbados
Burton, Gideon O.	Brigham Young University	USA
Butrous, Nasir	Australian Catholic University	Australia
Byun, Juman	George Washington University	USA
Caldera, Lizeth	Florida International University	USA
Calenzo, Patrick	IM2NP	France
Cano, Julio	Charles III University of Madrid	Spain
Cao, Hong	Southern Medical University	China
Cárdenas, Henry E.	Louisiana Tech University	USA
Cardoso, Eduardo	Monterrey Institute of Technology and Higher Education	Mexico
Carrasquero, José Vicente	Simon Bolivar University	Venezuela
Carvalho, Marco	Institute for Human and Machine Cognition	USA
Cázares-R., Víctor M.	Monterrey Institute of Technology and Higher Education	Mexico
Cerny, Václav	University of West Bohemia in Pilsen	CZE
Cha, Seung Tae	Korea Electronic Power Research Institute	South Korea
Cha, Sung-Hyuk	Pace University	USA
Chandra, Vigyan	Eastern Kentucky University	USA

Chen, Huei-Huang	Tatung University	Taiwan
Chen, Lisa Y.	I-Shou University	Taiwan
Chen, Shih-Chih	National Taiwan University	Taiwan
Chen, Yil	Huafan University	Taiwan
Chen, Yuhua	University of Houston	USA
Cheng, Kuo-Sheng	National Cheng-Kung University	Taiwan
Cherinka, R.	The MITRE Corporation	USA
Cherry, Barbara	Indiana University	USA
Chien, Steven	New Jersey Institute of Technology	USA
Chiou, Richard	Drexel University	USA
Chiou, Yin-Wah	Nanhua University	Taiwan
Cho, Kyoung-Rok	Chungbuk National University	South Korea
Cho, Tae Won	Chungbuk National University	South Korea
Cho, Vincent	The Hong Kong Polytechnic University	Hong Kong
Choi, Seung-Seok	Pace University	USA
Choi, Yu-Lee	Chungbuk National University	South Korea
Choo, Jinboo	Korea Electric Power Research Institute	South Korea
Chou, Andrew	Kainan University	Taiwan
Chou, Hsueh-Cheng	National Taiwan Normal University	Taiwan
Chowdhury, Masud H.	University of Illinois at Chicago	USA
Cipolla Ficarra, Francisco	Multimedia Communications Corp.	Italy
Cirella, Jonathan	Research Triangle Institute	USA
Clarke, Tim	University of Wales Institute Cardiff	UK
Coffman, Michael G.	Souther Illinois University Carbondale	USA
Cohen, Bernard	City University London	UK
Contreras, Sebastián	University of the Andes	Colombia
Cote, Paul	Benet Laboratories	USA
Cowan, Jimmy	Florida Institute of Technology	USA
Cripe, Billy	Oracle Corporation	USA
Curran, Kevin	University of Ulster	UK
Dawoud, Dawoud	University of KwaZulu Natal	South Africa
De Volder, Dennis	Western Illinois University	USA
Demestichas, Konstantinos	National Technical University of Athens	Greece
Desa, Shakinaz	Sultan Idris University of Education	Malaysia
Dhall, Sudarshan	University of Oklahoma	USA
Diallo, Saikou Y.	Old Dominion University	USA
Dierneder, Stefan	Linz Center of Mechatronics GmbH	Austria
Djukom, C. D.	University of Texas Medical Branch at Galveston	USA
Dolan, Dan	South Dakota School of Mines & Technology	USA
Dosi, Vasiliki	University of Ioannina	Greece
Duman, Hakan	University of Essex	UK

Dunning, Jeremy	Indiana University	USA
Dziong, Zbigniew	University of Québec	Canada
Edwards, Stephen H.	Virginia Tech	USA
El-Halafawy, Farag Z.	Menoufiya University	Egypt
Elmahboub, W. M.	Hampton University	USA
Emdad, F.	University of Texas Medical Branch at Galveston	USA
Erickson, Sarah	Florida International University	USA
Erkollar, Alptekin	University of Applied Sciences	Austria
Eshraghian, Kamran	Chungbuk National University	South Korea
Estévez, Leonardo	Texas Instruments	USA
Fang, Rong Jyue	Southern Taiwan University of Technology	Taiwan
Fisher, Wendy	The Open University	UK
Florescu, Gabriela C.	ICI	Romania
Fogelholm, Mikael	Academy of Finland Communications	Finland
Fougeres, Alain-Jerome	University of Technology of Belfort-Montbéliard	France
Fox, Kelly	Texas Tech University	USA
Freund, Rudolf	Vienna University of Technology	Austria
Fu, Shih-Lung	National Taiwan University	Taiwan
Fu, Xiuju	Institute of High Performance Computing	Singapore
Fu, Yonggang	Shanghai Jiao Tong University	China
Fuhrer, Patrik	University of Fribourg	Switzerland
Fujikawa, Takemi	University of Western Sydney	Australia
Fujita, Naoyuki	Japan Aerospace Exploration Agency	Japan
Fumizawa, Motoo	Shonan Institute of Technology	Japan
Fúster-Sabater, Amparo	Spanish Council for Scientific Research	Spain
Gagnon, Francois	School of High Technology	Canada
Ganapathi, Nanthini	Research Triangle Institute	USA
Ganchev, Ivan	University of Limerick	Ireland
García-Atanacio, César	National Nuclear Research Institute	Mexico
Gardezi, A. K.	Colegio de Postgraduados	Mexico
Garibaldi, Jonathan M.	University of Nottingham	UK
Georgescu, Vasile	University of Craiova	Romania
Gesekus, Tim	Aeronautical Information Service	Germany
Glotzbach, Ronald J.	Purdue University	USA
Godavarty, Anuradha	Florida International University	USA
González, Jean	Florida International University	USA
Gordon, Richard	University of KwaZulu Natal	South Africa
Goriachkin, Oleg	Volga State Academy of Telecommunication and Informatics	Russian Federation
Gosselin, Clément	Laval University	Canada
Goulding, Tom	Daniel Webster College	USA

Gregory, Mark A.	RMIT University	Australia
Guadarrama, Javier de J.	Technological Institute of Toluca	Mexico
Guevara L., Miguel A.	University of Porto	Portugal
Guiasu, Radu Cornel	York University	Canada
Guiasu, Silviu	York University	Canada
Gulez, Kayhan	Yildiz Technical University	Turkey
Guo, Dagang	National University of Singapore	Singapore
Gustavsson, Rune	Blekinge Institute of Technology	Sweden
Haba, C. G.	Technical University of Iasi	Romania
Hallot, Frédéric	Royal Military Academy	Belgium
Hamanaka, Masatoshi	University of Tsukuba	Japan
Hammond, Bruce R.	Saint Leo University	USA
Hangai, Seiichiro	Tokyo University of Science	Japan
Hao, Tianyong	City University of Hong Kong	China
Hardy, Frank	University of South Carolina Upstate	USA
Hardy, Leon C.	Embry Riddle Aeronautical University	USA
Hashimoto, Shigehiro	Osaka Institute of Technology	Japan
Heiserich, Gerd	Institute of Transport and Automation Technology	Germany
Hemmelman, Brian	The South Dakota School of Mines and Technology	USA
Hendel, Russell Jay	Towson University	USA
Henninger, Michael	University of Education	Germany
Herget, Josef	University of Applied Sciences of Eastern Switzerland	Switzerland
Herr, Stephan	Aeronautical Information Service	Germany
Hetzer, Dirk	T-Systems International GmbH	Germany
Higashiyama, Yoichi	Ehime University	Japan
Hochin, Teruhisa	Osaka Prefecture University	Japan
Hodge, Diane M.	Radford University	USA
Hofmeister, Helge	BASF IT Services	Belgium
Hong, Yun-Ki	Chungbuk National University	South Korea
Horne, Jeremy	Independent	Mexico
Hovakimyan, Anna	Yerevan State University	Armenia
Hsu, Pai-Hui	National Taiwan University	Taiwan
Hsu, Shu-Mei	Tatung University	Taiwan
Huang, Pin Chia	I-Shou University Kaohsiung	Taiwan
Huang, Ruhua	Wuhan University	China
Huang, Sheng He	University of Southern California	USA
Hvass, Michael	IBM Denmark	Denmark
Iembo, Rosanna	University of Calabria	Italy
Imai, Michita	Keio Uiversity	Japan
Imamura, Nobuaki	Kobe City College of Technology	Japan
Ishikawa, Hiroshi	NUIS	Japan

Ito, Akinori	Tohoku University	Japan
Jäntschi, Lorentz	Academic Direct Organization	Romania
Jiménez R., Lourdes	University of Alcala	Spain
Johnson, Mark	Benet Laboratories	USA
Jones, Paul	University of Cincinnati	USA
Jonson, Mark	University of New Mexico	USA
Jung, Kyung Im	Samsung Electronics, R. O. Korea	South Korea
Jung, Kyung Kwon	Dongguk University	South Korea
Jung, Seul	Chungnam National University	South Korea
Kadoch, Michel	University of Québec	Canada
Kamejima, Kohji	Osaka Institute of Technology	Japan
Kao, Chih-Yang	Ming-Chuan University	Taiwan
Karamat, Parwaiz	Open Polytechnic of New Zealand	NZL
Kasapoglu, Ercin	Hacettepe University	Turkey
Kaszubiak, J.	University Magdeburg	Germany
Katsikas, Sokratis	University of Piraeus	Greece
Kawaguchi, Masashi	Suzuka National College of Technology	Japan
Kehtarnavaz, Nasser	University of Texas at Dallas	USA
Khaled, Pervez	University of Illinois at Chicago	USA
Khan, Faisal	Khalifa University of Science	UAE
Khatri, Anil	Bowie State University	USA
Kim, Jeongdae	Information and Communications University	South Korea
Kim, Kyungwoo	University of Florida	USA
Kim, Seok-Man	Chungbuk National University	South Korea
Kim, Yeo Jin	Samsung Electronics, R. O. Korea	South Korea
Kim, Yeon-Ho	Chungbuk National University	South Korea
Kim, Yong Hak	Korea Electric Power Research Institute	South Korea
Kincaid, Rex K.	The College of William and Mary	USA
Kira, Dennis S.	Concordia University	Canada
Kizirian, Robin	Drexel University	USA
Klapp, Jaime	National Nuclear Research Institute	Mexico
Kobal, Damjan	University of Ljubljana	Slovenia
Komatsu, Takanori	Shinshu University	Japan
Kouki, A.	School of High Technology	Canada
Kozma-Bognár, Veronika	University of Pannonia	Hungary
Krakowska, Monika	Jagiellonian University	Poland
Kreisler, Alain	University of Paris-Sud	France
Kromrey, Jeffrey D.	University of South Florida	USA
Kronreif, Gernot	Austrian Research Centers	Austria
Krothapalli, Sreenivasa Rao	Indian Institute of Technology Kharagpur	India
Kuftin, Felix A.	Kolomna Teacher Training Institute	Russian.

Kung, C. M.	Shih Chien University	Taiwan
Kuragano, Tetsuzo	The American Society of Mechanical Engineers	Japan
Kutter, Anna K.	University of Education	Germany
Kwon, Yongjin (James)	Ajou University	South Korea
Lahlouhi, Ammar	University of Biskra	Algeria
Lai, James C. K.	Idaho State University College of Pharmacy	USA
Laksmivarahan, S.	University of Oklahoma	USA
Larab, Ali	Champollion University	France
Latawiec, Krzysztof J.	Opole University of Technology	Poland
Latchman, Haniph A.	University of Florida	USA
Lau, Newman	Hong Kong Polytechnic University	Hong Kong
Lee, Hwajung	Radford University	USA
Lee, Jae-Suk	Gwangju Institute of Science and Technology	South Korea
Lee, Kangsun	MyongJi University	South Korea
Lee, Nam Ho	Korea Electric Power Research Institute	South Korea
Lee, Sang-Jin	Chungbuk National University	South Korea
Lee, Suk Bong	Samsung Electronics, R. O. Korea	South Korea
Lee, Yih-Jiun	Chien Kuo Technology University	Taiwan
Lee, Yusin	National Cheng Kung University	Taiwan
Letellier, T.	Victor Segalen Bordeaux 2 University	France
Li, Lihong	City University of New York	USA
Lin, Feng-Tyan	National Taiwan University	Taiwan
Lin, Huan Yu	National Chiao Tung University	Taiwan
Lin, Shu-Chiung	Tatung University	Taiwan
Linares, Oscar	Francisco Jose of Caldas District University	Colombia
Lind, Nancy	Illinois State University	USA
Lipikorn, Rajalida	Chulalongkorn University	Thailand
Lipinski, Piotr	Technical University of Lodz	Poland
Litvin, Vladimir	California Institute of Technology	USA
Liu, Jun	University of Ulster	UK
Liu, Kuo-Shean	Tatung University	Taiwan
Liu, Shih-Chi	Tatung University	Taiwan
Livne, Nava L.	University of Utah	USA
Livne, Oren E.	University of Utah	USA
Long, Changjiang	Huazhong University	China
López Román, Leobardo	University of Sonora	Mexico
Lou, Shi Jer	National Pingtung University of Science and Technology	Taiwan
Loutfi, Mohamed	University of Sunderland	UK
Love C., Gloria	Dillard University	USA
Lowe, John	University of Bath	UK
Lowry, Pam	Lawrence Technological University	USA

Luh, Guan Chun	Tatung University	Taiwan
Lui, Wen Lik Dennis	Monash University	Australia
Lyell, Margaret	Intelligent Automation, Inc.	USA
Ma, Yongqing	Victoria University of Wellington	NZL
Machotka, Jan	University of South Australia	Australia
Mahendran, Francis	Motorola Software Group Singapore	Singapore
Mahgoub, Ahmed G.	Alexandria University	Egypt
Mak, Peng Un	University of Macau	Macau
Manley, Denis	Dublin Institute of Technology	Ireland
Mansikkamäki, Pauliina	Tampere University of Technology	Finland
Marino, Mark	Erie County Community College	USA
Marlowe, Thomas	Seton Hall University	USA
Martínez, Pablo	Laboratory for Advanced Brain Signal Processing	Japan
Martínez, Sergio	Florida International University	USA
Martínez Madrid, Natividad	Carlos III University of Madrid	Spain
Marwala, Tshilidzi	University of Johannesburg	South Africa
Mascari, Jean-François	National Research Council	Italy
Masoumi, Nasser	University of Tehran	Iran
Mathews, Brian	University of Bedfordshire	UK
Matsumoto, Kazunori	KDDI R&D Laboratories Inc.	Japan
Matsuno, Akira	Teikyo University	Japan
Mayer, Daniel	University of West Bohemia	CZE
Mbobi, Aime Mokhoo	Centennial College	Canada
Medina, Omar	University of Guanajuato	Mexico
Mehrabian, Ali	University of Central Florida	USA
Mellouli, Sehl	Laval University	Canada
Meyer, Heiko	Munich University of Applied Sciences	Germany
Michaelis, B.	University Magdeburg	Germany
Migliarese, Piero	University of Calabria	Italy
Miller, R.	The MITRE Corporation	USA
Minnaert, Eric	South Dakota School of Mines & Technology	USA
Mirza Sebzali, Yussef	Kuwait Institute for Scientific Research	Kuwait
Mistry, Jaisheel	University of Witwatersrand	South Africa
Mitchell, Charles	Grambling State University	USA
Moin, Lubna	Pakistan Navy Engineering College	Pakistan
Moreau, Alban	University of Brest	France
Moreno, Carlos	Central University of Venezuela	Venezuela
Moreno S.-C., Ana María	Polytechnic University of Madrid	Spain
Morii, Hisashi	Shizuoka University	Japan
Mostafaeipour, Ali	Yazd University	Iran
Muknahallipatna, Suresh	University of Wyoming	USA

Muraleedharan, Rajani	Syracuse University	USA
Naddeo, Alessandro	University of Salerno	Italy
Nadworny, Margaret	Global Software Group, Motorola	USA
Nagai, Yasuo	Tokyo University of Information Sciences	Japan
Nagalakshmi, Vadlamani	Gandhi Institute of Technology and Management	India
Nagaoka, Tomoyuki	Tokyo Metropolitan University	Japan
Nahmens, Isabelina	University of South Florida	USA
Nair, V. S. Sukumaran	Southern Methodist University	USA
Nakashima, Takuya	Shizuoka University	Japan
Nam, Su Chul	Korea Electronic Power Research Institute	South Korea
Nave, Felecia M.	Prairie View A&M University	USA
Nawawi, S. W.	University Technology of Malaysia	Malaysia
Nazmy, Taymoor M.	Ain Shams University	Egypt
Nedic, Zorica	University of South Australia	Australia
Nelwamondo, Fulufhelo V.	University of the Witwatersrand	South Africa
Neo, Yoichiro	Shizuoka University	Japan
Nguyen, Mai	RTI International	USA
Nguyen, Patricia	RTI International	USA
Ni, Xingliang	University of Science and Technology of China	China
Noh, Bong Nam	Chonnam National University	South Korea
Nonaka, Hidetoshi	Hokkaido University	Japan
Nousala, Susu	Royal Melbourne Institute of Technology University	Australia
Novikov, Oleg	Tomko, Inc.	Russian
Nyan, M. N.	National University of Singapore	Singapore
O'Brien, Ann	Loughborough University	UK
Obrebski, Jan B.	Warsaw University of Technology	Poland
Ocelka, Tomas	Institute of Public Health	CZE
Oh, Yun Sang	Samsung Electronics, R. O. Korea	South Korea
Ojala, Pasi	University of Oulu	Finland
Olla, Phillip	Madonna University	USA
Olson, Patrick C.	National University and Aware Consulting Group	USA
Ong, Soh-Khim	National University of Singapore	Singapore
Ong, Vincent Koon	University of Bedfordshire	UK
Osadciw, Lisa Ann	Syracuse University	USA
Ostergaard, Soren Duus	IBM Europe	Denmark
Overmeyer, Ludger	Institute of Transport and Automation Technology	Germany
Oya, Hidetoshi	Shonan Institute of Technology	Japan
Ozdemir, Ahmet S.	Marmara University	Turkey
Palesi, Maurizio	University of Catania	Italy
Paré, Dwayne E.	University of Toronto Scarborough	Canada
Park, Kyung-Chang	Chungbuk National University	South Korea

Park, Young C.	Baekseok University	South Korea
Parks, Gary G.	National University	USA
Pécatte, Jean-Marie	University Paul Sabatier-Toulouse 3	France
Pedrycz, Witold	University of Alberta	Canada
Peng, Jian	University of Saskatchewan	Canada
Petkov, Emil	Northumbria University	UK
Pfeifer, Michael	Technical University of Dortmund	Germany
Phillips, C. Dianne	NorthWest Arkansas Community College	USA
Phillips, James	Louisiana Tech University	USA
Pierce, Ryan M.	Arkansas State University	USA
Pieters, Cornelis P.	University for Humanistics	Netherlands
Pitzer, Erik	University of Applied Sciences Hagenberg	Austria
Platt, Glenn	Commonwealth Scientific and Industrial Research Organisation	Australia
Podaru, Vasile	Military Technical Academy	Romania
Polanski, Andrzej	Silesian University of Technology	Poland
Poon, Gilbert	University of Waterloo	Canada
Popentiu, Florin	University of Oradea	Romania
Postolache, Octavian	Institute of Telecommunications	Portugal
Poulin, Régis	Laval University	Canada
Pulcher, Karen L.	University of Central Missouri	USA
Pun, Daniel	Central Queensland University	Australia
Putnam, Janice	University of Central Missouri	USA
Quan, Xiaojun	University of Science and Technology of China	China
Quintyne, Vasco	University of the West Indies	Barbados
Rachev, Boris	Technical University of Varna	USA
Rahman, Anuar Abdul	Pusat Tenaga Malaysia	Malaysia
Rahman, Mohammad	University of Texas at Dallas	USA
Ramachandran, S.	Indian Institute of Technology Madras	India
Ramírez C., Guillermo H.	Soka University	Japan
Ratkovic Kovacevic, Nada	University of Belgrade	Serbia
Ren, Jianfeng	Qualcomm Incorporated	USA
Revetria, Roberto	University of Genoa	Italy
Reyes-Méndez, Jorge Joel	Metropolitan Autonomous University	Mexico
Riaz Moghal, Mohammad	University College of Engineering and Technology	Pakistan
Rigaud, Bernard	Champollion University	France
Rodríguez, María Dolores	University of Alcala	Spain
Ropella, Glen E.	Tempus Dictum Inc.	USA
Rossignol, R.	Victor Segalen Bordeaux 2 University	France
Rossmann, Jürgen	RWTH Aachen University	Germany
Ruan, Tongjun	New Mexico Tech	USA

Rutkauskas, Aleksandras V.	Vilnius Gediminas Technical University	Lithuania
Saadane, Rachid	Institut Eurecom	France
Sahara, Tomohiro	Osaka Institute of Technology	Japan
Sala, Nicoletta	University of Italian Switzerland	Switzerland
Salazar, Dora	Texas Tech University	USA
Saleh, Magda M.	Alexandria University	Egypt
Sanna, Andrea	Polytechnic of Torino	Italy
Šaruckij, Mark	Cracow University of Economics	Poland
Sateesh Reddy, J.	Indian Institute of Technology Madras	India
Sato, Tomoaki	Hirosaki University	Japan
Sax, Eric	Mercedes-Benz Technology Gmbh	Germany
Schaeffer, Donna M.	Marymount University	USA
Schlette, Christian	RWTH Aachen University	Germany
Schluse, M.	RWTH Aachen University	Germany
Schrader, P. G.	University of Nevada	USA
Schulz, Lennart	Institute of Transport and Automation Technology	Germany
Schumacher, Jens	University of Applied Sciences Vorarlberg	Austria
Seepold, Ralf	Carlos III University of Madrid	Spain
Segall, Richard S.	Arkansas State University	USA
Seitzer, Jennifer	University of Dayton	USA
Sert, Yasemin	University of South Florida	USA
Shaw, Jill	The Open University	UK
Shayeghi, Hossien	Iran University of Science	Iran
Shim, Eung Bo	Korea Electric Power Research Institute	South Korea
Shin, Jeong Hoon	Korea Electric Power Research Institute	South Korea
Shin, Jungpil	University of Aizu	Japan
Shiraishi, Yoshiaki	Nagoya Institute of Technology	Japan
Shklyar, Benzion	Holon Academic Institute	Israel
Shum, Kwok	Stanford University	USA
Sim, Sang Gyoo	Samsung Electronics, R. O. Korea	South Korea
Sinkunas, Stasys	Kaunas University of Technology	Lithuania
Sleit, Azzam Talal	University of Jordan	Jordan
Soeiro, Alfredo	University Porto	Portugal
Song, Hong Jun	University of Sydney	Australia
Soundararajan, Ezekiel	Indiana University of Pennsylvania	USA
Srisawangrat, Songsiri	University of Regina	Canada
Starosolski, Zbigniew	Silesian University of Technology	Poland
Stasytyte, Viktorija	Vilnius Gediminas Technical University	Lithuania
Stößlein, Martin	University of Erlangen-Nuremberg	Germany
Stubberud, Stephen	The Boeing Company	USA
Su, J. L.	Shanghai University	China

Su, Wei	National University of Defense Technology	USA
Su, Wen-Ray	NCDR	Taiwan
Sun, Baolin	Wuhan University	China
Swart, William	East Carolina University	USA
Szygenda, Stephen A.	Southern Methodist University	USA
Takemi, Fujikawa	University of Western Sydney	Australia
Tam, Wing K.	Swinburne University of Technology	Australia
Tay, Francis E. H.	National University of Singapore	Singapore
Taylor, Stephen	Sussex University	UK
Tchier, Fairouz	King Saud University	USA
Teng, Chia-Chi	Brigham Young University	USA
Teshigawara, Yoshimi	Soka University	Japan
Thakur, Amrit' Anshu	Mississippi State University	USA
Theologou, Michael	National Technical University of Athens	Greece
Thornton, Mitchell A.	Southern Methodist University	USA
Till, Robert	John Jay College	USA
Trahan, Shane	RTI International	USA
Traum, Maria	Johannes Kepler University	Austria
Trimble, Robert	Indiana University of Pennsylvania	USA
Trzaska, Mariusz	Polish Japanese Institute of Information Technology	Poland
Tsai, Li-Hung	Tatung University	Taiwan
Tsaur, Woei-Jiunn	Da-Yeh University	Taiwan
Tseng, Ko Ying	Tatung University	Taiwan
Tseng, Tzu-Liang (Bill)	University of Texas at El Paso	USA
Tsiligaridis, John	Heritage University	USA
Tsubaki, Michiko	University of Electro-Communications	Japan
Turnitsa, Charles D.	Old Dominion University	USA
Väänänen, Kalervo	University of Turku	Finland
Valakevicius, Eimutis	Kaunas University of Technology	Lithuania
Van Delden, Sebastian	University of South Carolina Upstate	USA
Vanka, Sita	Kakatiya University	India
Vargoz, Erik P.	The College of William and Mary	USA
Vasinek, Vladimir	Technical University of Ostrava	CZE
Venkataraman, Satyamurti	All India Association for Micro Enterprise Development	India
Verlinde, Patrick	Royal Military Academy	Belgium
Verma, Pramode	University of Oklahoma	USA
Viloria, Orlando H.	Simon Bolivar University	Venezuela
Vitral, Renan	Federal University of Juiz de Fora	USA
Vodovozov, Valery	St. Petersburg State Electrotechnical University	Russian
Volkov-Husovic, T.	University of Belgrade	Yugoslavia
Voss, Andreas	Dortmund University of Technology	Germany

Wagner, Stefan	University of Applied Sciences Upper Austria	Austria
Wang, Lei	University of Houston	USA
Wang, Ling	University of Oklahoma	USA
Warwick, Jon	London South Bank University	UK
Wei, Xinzhou	City University of New York	USA
Wells, Harvey	King's College London	UK
Whitbrook, Amanda M.	University of Nottingham	UK
Whiteley, Rick	Calabash Educational Software	Canada
Woodhead, Steve	University of Greenwich	UK
Woodthorpe, John	The Open University	UK
Wu, Chun Yin	Tatung University	Taiwan
Wu, Shang-Yu	National Science & Technology Center for Disaster Reduction	Taiwan
Xu, L.	National University of Singapore	Singapore
Yamada, Seiji	National Institute of Informatic	Japan
Yamaguchi, Akira	Meisei University	Japan
Yan, Kuo Qin	Chaoyang University of Technology	Taiwan
Yan, Mu Tian	Huafan University	Taiwan
Yang, Huiqing	Virginia State University	USA
Yang, Hung Jen	National Kaohsiung Normal University	Taiwan
Yang, Ming	Jacksonville State University	USA
Yang, Sung Un	Syracuse University	USA
Yang, Yueh-Ting	Drexel University	USA
Yap, K. L.	National University of Singapore	Singapore
Yasser, Muhammad	Chiba University	Japan
Yatsymirskyy, Mykhaylo	Technical University of Lodz	Poland
Yazawa, Toru	Tokyo Metropolitan University	Japan
Yilmaz, Levent	Auburn University	USA
Yoon, Changwoo	Electronics and Telecommunications Research Institute	South Korea
Yoshida, Eri	Toyohashi University of Technology	Japan
Yoshida, Takahiro	Tokyo University of Science	Japan
You, Younggap	Chungbuk National University	South Korea
Yu, Chong Ho	University of California at Berkeley	USA
Yu, Xin	University of Bath	UK
Yuen, Fei Lung	University of Hong Kong	Hong Kong
Zadeh, Jeff	Virginia State University	USA
Zarama, Roberto	University of the Andes	Colombia
Zaretsky, Esther	Givat Washington College of Education	Israel
Zelinka, Tomas	Czech Technical University in Prague	CZE
Zeller, Andrew J.	Norisbank	Germany
Zhang, Jinfang	University of Waterloo	Canada

Zhang, Linda L.	University of Groningen	Netherlands
Zhang, Qingyu	Arkansas State University	USA
Zhang, Xiaozheng Jane	California Polytechnic State University	USA
Zhonghua, Fang	Shanghai Institute of Technical Physics	USA
Zhou, Dan	University of California San Diego	USA
Zhu, Hui	Soochow University	China
Zobaa, Ahmed	Cairo University	Egypt

.

ADDITIONAL REVIEWERS

Abd Wahab, Mohd Helmy	University Tun Hussein Onn Malaysia	Malaysia
Aboalsamh, Hatim	King Saud University	Saudi Arabia
Al Sabbagh, Abdallah	University of Technology Sydney	Australia
Al-Fahoum, Amjed	Yarmouk University	Jordan
Al-Yasiri, Adil	University of Salford	UK
Amborski, Krzysztof	Warsaw University of Technology	Poland
Anschel, David	Stanford University	USA
Anthony, Anish	University of Alabama at Birmingham	USA
Arnold, Eckhard	University of Stuttgart	Germany
Arya, Ali	Carleton University	Canada
Asproni, Giovanni	Association for Computing Machinery	UK
Asterio de Castro G., Paulo	State University of Campinas	Brazil
Baffa, Carlo	Arcetri Astrophysical Observatory	Italy
Bamidis, Panos	Aristotle University of Thessaloniki	Greece
Barnabas, Bede	DigiPen Institute of Technology	USA
Baro, Jesús A.	University of Valladolid	Spain
Bartolini, Sandro	University of Siena	Italy
Baudin, Veronique	Laboratory for Analysis and Architecture of Systems	France
Behmard, Hamid	Western Oregon University	USA
Behnam, Hamid	Iran University of Science and Technology	Iran
Beigy, Hamid	Sharif University of Technology	Iran
Belala, Faiza	University Mentouri of Constantine	Algeria
Bernabé, Gregorio	University of Murcia	Spain
Bernardi, Ansgar	German Research Center for Artificial Intelligence	Germany
Bevinakoppa, Savitri	Melbourne Institute of Technology	Australia
Bubnov, Alexey	Academy of Sciences and the Czech Republic	CZE
Buraga, Sabin	Alexandru Ioan Cuza University	Romania
Calistru, Catalin Nicolae	Gheorghe Asachi Technical University	Romania
Callaos, Nagib	Simon Bolivar University	Venezuela
Campos, Ana	Institute for the Development of New Technologies	Portugal
Canalda, Philippe	University of Franche-Comté	France
Carnes, Patrick	Kirtland Air Force Base	USA
Carrasquero, José Vicente	Simon Bolivar University	Venezuela

Chouvarda, Ioanna	Aristotle University of Thessaloniki	Greece
Cipolla-Ficarra, Francisco	Latino Association of Human-Computer Interaction	Italy
Corrêa, Lilian	Mackenzie Presbyterian University	Brazil
Costa, Guilherme	University of Caxias	Brazil
Daneshmand M., Amin	Malayer Azad University	Iran
De la Puente, Fernando	University of Las Palmas de Gran Canaria	Spain
De Silva, Clarence	University of British Columbia	Canada
Delli Carpini, Michael	University of Pennsylvania	USA
Dias, Malcolm Benjamin	Unilever Corporate Research	UK
Dudas, Marek	Safarik University	Slovakia
D'Ulizia, Arianna	National Research Council	Italy
Durai Raj, K. Antony A.	Infosys Technologies Limited	India
Dutta, Arjun	Touro College of Pharmacy	USA
El Bakkali, Hanan	Mohammed V University	Morocco
Erbacher, Robert	Utah State University	USA
Erol, Cemil B.	Scientific and Technological Research Council	Turkey
Evesque, Pierre	Ecole Centrale Paris	France
Finch, Aikyna	Strayer University	USA
Florescu, Gabriela	ICI	Romania
Floyd, Raymond	Innovative Insights, Inc.	USA
Foglia, Pierfrancesco	University of Pisa	Italy
Frejlichowski, Dariusz	West Pomeranian University of Technology	Poland
Fúster-Sabater, Amparo	Spanish Council for Scientific Research	Spain
García Marco, Francisco J.	University of Zaragoza	Spain
Giakoumaki, Aggeliki	National Technical University of Athens	Greece
Giampapa, Joseph	Carnegie Mellon University	USA
Gravvanis, George A.	Democritus University of Thrace	Greece
Grouverman, Valentina	Research Triangle Institute	USA
Guillen, Alberto	University of Granada	Spain
Ho, Liangwei	Collex Communication Corporation	Taiwan
Hou, Chun-Ju	Southern Taiwan University of Technology	Taiwan
Hsu, Ching-Sheng	Ming Chuan University	Taiwan
Hsu, Chun-Fei	Chung Hua University	Taiwan
Huang, Chun Hong	Lunghwa University of Science and Technology	Taiwan
Hussain, Aini	National University of Malaysia	Malaysia
Ianigro, Massimo	Italian National Research Council	Italy
Iovan, Stefan	Informatica Feroviaria S.A.	Romania
Irtegov, Valentin	Institute of System Dynamics and Control Theory	Russian
Jwo, Dah-Jing	National Taiwan Ocean University	Taiwan
Karakos, Alexandros	Democritus University of Thrace	Greece
Kastania, Anastasia	Athens University of Economics and Business	Greece

Kausch, Bernhard	RWTH Aachen University	Germany
Kawa, Arkadiusz	Poznan University of Economics	Poland
Khalil, Mohamad	Lebanese University	Lebanon
Kim, Do-Hoon	Kyung Hee University	South Korea
Kimiaei, Sharok	Infogosoft AB	Sweden
Klyuev, Vitaly	University of Aizu	Japan
Kroumov, Valeri	Okayama University of Science	Japan
Kurubacak, Gulsun	Anadolu University	Turkey
Law, Rob	Hong Kong Polytechnic University	Hong Kong
Lemos, Rodrigo	Federal University of Goias	Brazil
Li, Jingyi	University of Maryland	USA
Liang, Chih-Chin	National Central University	Taiwan
Maccione, Alessandro	University of Genova	Italy
Macerata, Alberto	University of Pisa	Italy
Mansour, Ali	ENSIETA	France
Martins, Constantino	Polytechnic Institute of Porto	Portugal
Martins, Valeria	Mackenzie Presbyterian University	Brazil
Matsuno, Akira	Teikyo University Chiba Medical Center	Japan
Meyer, Heiko	Gefasoft AG	Germany
Michalik, Bartosz	Poznan University of Technology	Poland
Ming, Bao	Chinese Academy of Sciences	China
Mitra, Debojyoti	Sir Padampat Singhania University	India
Nahm, In Hyun	Sunmoon University	South Korea
Ogorodnikov, Dmitri	Mount Sinai School of Medicine	USA
Ophir, Dan	University Center Ariel in Samaria	Israel
Ortiz, Andrés	University of Malaga	Spain
O'Shaughnessy, Douglas	Institut National de la Recherche Scientifique	Canada
Osman, Abdalla	University of Calgary	Canada
Ostrowski, David	Ford Motor Company	USA
Pisarchik, Alexander	Center for Research in Optics	Mexico
Poobrasert, Onintra	National Electronics and Computer Technology Center	Thailand
Provaznik, Ivo	Brno University of Technology	CZE
Rabaea, Adrian	North University of Baia Mare	Romania
Rachid, Elias	Saint Joseph University	Lebanon
Rameshwar, Pranela	University of Medicine & Dentistry of New Jersey	USA
Reyes-Méndez, Jorge Joel	Metropolitan Autonomous University	Mexico
Riznyk, Volodymyr	Lviv Polytechnic National University	Ukraine
Robinson, Jean	Reseach Triangle Institute International	USA
Romero, Margarida	Universitat Autònoma de Barcelona	Spain
Rosa, Agostinho	Technical University of Lisbon	Portugal
Roschildt Pinto, Alex S.	Universidade Estadual Paulista	Brazil

Rot, Artur	Wroclaw University of Economics	Poland
Saitta, Francesco	University of Palermo	Italy
Samkin, Grant	University of Waikato	New Zealand
Santiago-Jiménez, María E.	Research Technological Institute of Puebla	Mexico
Sathyamoorthy, Dinesh	Science and Technology Research Institute for Defense	Malaysia
Simsek, Ali	Anadolu University	Turkey
St. Peters, Megan	University of Michigan	USA
Talukdar, Fazal Ahmed	National Institute of Technology Silchar	India
Tanaka, Hirokazu	Toshiba Corporation	Japan
Tsipouras, Markos	University of Ioannina	Greece
Turnage, Doris	U.S. Army Research Laboratory	USA
Vandeyar, Thiru	University of Pretoria	South Africa
Verber, Domen	University of Maribor	SOL
Vicente, Luís M.	Polytechnic University of Puerto Rico	Puerto Rico
Xin, Huolin	Cornell University	USA
Yahiaoui, Azzedine	Eindhoven University of Technology	Netherlands
Zemliak, Alexander	Autonomous University of Puebla	Mexico
Zhang, W. J. (Chris)	University of Saskatchewan	Canada
Zhong, Cheng	Guangxi University	China
Zissis, Dimitrios	University of the Aegean	Greece

ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

Acero, Diego	Pedagogical University	Colombia
Affonso, Thais	Mackenzie Presbyterian University	Brazil
Al Sabbagh, Abdallah	University of Technology Sydney	Australia
Alayón, Silvia	University of La Laguna	Spain
Anthony, Anish	University of Alabama at Birmingham	USA
Arguelles, Amadeo	National Polytechnic Institute	Mexico
Armitage, William	University of South Florida Polytechnic	USA
Arreche, Luís	Polytechnic University of Madrid	Spain
Ayewoh, Mike	West Chester University	USA
Azimifar, Zohreh	Shiraz University	Iran
Balitanas, Maricel	Hannam University	South Korea
Bardone, Emanuele	University of Pavia	Italy
Bhargava, Raj	Indian Institute of Technology Roorkee	India
Bii, Harrison	Moi University	Kenya
Bussola, Valéria	Mackenzie Presbyterian University	Brazil
Bytheway, Andy	Cape Peninsula University of Technology	South Africa
Camargo, Luz	Francisco Jose of Caldas District University	Colombia
Chee, Yi-Min	International Business Machines	USA
Choi, Yongho	Jungwon University	South Korea
Cofiel, Luciana	University of São Paulo	Brazil
Conte, Roberto	Center for Scientific Research and Higher Education	Mexico
Correa, Lilian	Mackenzie Presbyterian University	Brazil
Cruz, Nareli	National Polytechnic Institute	Mexico
Dalmadge, Cretson	Winston Salem State University	USA
Daniel, Evelyn	University of North Carolina	USA
Datta, Ajoy	University of Nevada, Las Vegas	USA
Defelice, Robyn	Indiana University of Pennsylvania	USA
Dominic, Dhanapal	Petronas University of Technology	Malaysia
Fossum, Ingerid	Buskerud University College	Norway
Frasier, Stephen	University of Massachusetts Amhers	USA
Garrard, Christopher	University of Oxford	UK

Garzón, Yamid	Francisco Jose of Caldas District University	Colombia
Geiger, Rebecca	International Performance Research Institute	Germany
Gluchowski, Peter	Chemnitz University of Technology	Germany
Gómez, Diego	Wood Group ESP	Colombia
Gómez, José	University of Murcia	Spain
González, Juan	CENIDET	Mexico
Hackley, Grant	Burleson LLP.	USA
Hansen, Hallstein	Buskerud University College	Norway
Harabagi Hanna, Vera L.	Mackenzie Presbyterian University	Brazil
Helil, Nurmamat	Xinjiang University	South Korea
Hiew, Pang	British Telecommunications	Malaysia
Hoh, Simon	British Telecommunications	Malaysia
Hsieh, Yen-Hao	Chia Nan University of Pharmacy and Science	Taiwan
Ikem, Fidelis	Winston Salem State University	USA
Imai, Masafumi	Toyohashi Sozo University	Japan
Infante, Willson	Francisco Jose of Caldas District University	Colombia
Ingham, John	Université de Sherbrooke	Canada
Ito, Koichi	Chiba University	Japan
Iwatsuki, Katsumi	Nippon Telegraph and Telephone Corporation	Japan
Jeong, Dong Hyun	University of the District of Columbia	USA
Johnson, Michelle	Sacred Heart University	USA
Jung, Yung-Joon	Electronics and Telecommunications Research Institute	South Korea
Kahlert, Dirk	Leipzig University of Applied Sciences	Germany
Kang, Jin-Suk	Jangwee Research Institute for National Defense	South Korea
Kang, Kyu Chang	Electronics and Telecommunications Research Institute	South Korea
Katsuyama, Tomoo	Numazu National College of Technology	Japan
Kieny, Christophe	Laboratoire de Recherche en Génie Électrique	France
Kilikauskas, Michelle	Naval Air Warfare Center	USA
Kim, Byung-Seo	Hongik University	South Korea
Kim, Sunhong	E-Commerce and Internet Application Laboratory	South Korea
Kreinovich, Vladik	University of Texas at El Paso	USA
Kruger, David	Tshwane University of Technology	South Africa
Kumar, Santhosh	Bits Pilani-Dubai	UAE
Kung, C. M.	Shih Chien University	Taiwan
Lawrence, David	University of North Dakota	USA
Lee, Sunguk	Research Institute of Industrial Science and Technology	South Korea
Lin, Pei-Jung	Hungkuang University	Taiwan
Lind, Nancy	Illinois State University	USA
Lindsay, Val	Victoria University of Wellington	New Zealand
Love C., Gloria	Dillard University	USA
Lutz, Robert	Applied Physics Laboratory at Johns Hopkins University	USA

Madian, Ahmed	German University in Cairo	Egypt
Mahanti, Ambuj	Indian Institute of Management Calcutta	India
Marcondes, María	Mackenzie Presbyterian University	Brazil
Mareboyana, Manohar	Bowie State University	USA
Marx, Jorge	University Oldenburg	Germany
Mccarroll, Christopher	Raytheon Corporation	USA
Melo, Carlos Thomas	Minas Gerais Court of Auditors	Brazil
Mikhaylov, Nikolay	MeraLabs	Russian
Mitchell, Charles	Grambling State University	USA
Mogotlhwane, Tiroyamodimo	University of Botswana	Botswana
Msuya, Athuman	National Bureau of Statistics	Tanzania
Mugiraneza, Theodomir	University of Rwanda	Rwanda
Navarro, Leandro	Technical University of Catalonia	Spain
Oh, Sejin	Hanyang University	South Korea
Oh, Seongjun	Korea University	South Korea
Oppenheim, Daniel	International Business Machines	USA
Pacheco, Enrique	Center for Scientific Research and Higher Education at Ensenada	Mexico
Palmgren, Juni	Karolinska Institutet	Sweden
Petros, Thomas	University of North Dakota	USA
Pieters, Cornelis P.	University for Humanistics	Netherlands
Purba, Ambrosius	Padjadjaran University	Indonesia
Ray, Kaushik	University of Windsor	Canada
Ray, Tane	University of the West Indies Cavehill Campus	Barbados
Reali, Gianluca	University of Perugia	Italy
Rebholz-S., Dietrich	European Bioinformatics Institute	UK
Ribeiro, Vilar	Mackenzie Presbyterian University	Brazil
Robles, Rosslin John H.	Science and Engineering Research Support Society	Philippines
Sánchez, Gerardo	National Autonomous University of Mexico	Mexico
Sathyamoorthy, Dinesh	Science and Technology Research Institute for Defense	Malaysia
Shin, Kihong	Andong National University	South Korea
Siqveland, Arvid	Buskerud University College	Norway
Sleit, Azzam Talal	University of Jordan	Jordan
Smal, Desiree	University of Johannesburg	South Africa
Sobhanmanesh, Faribоorz	Shiraz University	Iran
Sulaiman, Suziah	Petronas University of Technology	Malaysia
Tam, Wing K.	Swinburne University of Technology	Australia
Tammisto, Yulia	Aalto University School of Economics	Finland
Teng, Wei-Guang	National Cheng Kung University	China
Thomas, Little	Smart Lighting Engineering Research Center	USA

Tolentino, Marcus	Ministry of Health	Brazil
Trifas, Monica	Jacksonville State University	USA
Tung, Wei-Feng	Fu-Jen Catholic University	Taiwan
Walcott, Paul	University of the West Indies Cavehill Campus	Barbados
Wan Ahmad, Wan Fatimah	Petronas University of Technology	Malaysia
Watada, Junzo	Waseda University	Japan
Wright, Tennyson	College of Behavioral & Community Sciences	USA
Xin, Huolin	Cornell University	USA
Yanagihara, Mitsuyoshi	Nagoya University	Japan
Yang, Bo	Bowie State University	USA
Yannakakis, Georgios	IT University of Copenhagen	Denmark
Yatsenko, Yuri	Houston Baptist University	USA
Yoshimura, Jin	Shizuoka University	Japan
Zissis, Dimitrios	University of the Aegean	Greece

PROGRAM COMMITTEE CHAIRS
C. Dale Zinn
Hsing-Wei Chu

GENERAL CHAIR
Nagib Callaos

ORGANIZING COMMITTEE CHAIRS
Andrés Tremante
Belkis Sánchez

HARDCOPY PROCEEDINGS PRODUCTION CHAIR
María Sánchez

CD PROCEEDINGS PRODUCTION CHAIR
Juan Manuel Pineda

SYSTEMS DEVELOPMENT, MAINTENANCE AND DEPLOYMENT
Dalia Sánchez
Keyla Guédez
Nidimar Díaz
Jesús Malavé

OPERATIONAL ASSISTANTS
Marcela Briceño
Cindi Padilla

HELP DESK
Louis Barnes
Sean Barnes

CONFERENCES PROGRAM MANAGER
Leonisol Callaos

TECHNICAL CONSULTANT ON COMPUTING SYSTEMS
Juan Manuel Pineda

META-REVIEWERS SUPPORT
Dalia Sánchez

PROCEEDINGS PRODUCTION SUPPORT
Dalia Sánchez
Keyla Guédez
Marcela Briceño

PROMOTIONAL SUPPORT
Keyla Guédez
Nidimar Díaz
Freddy Callaos

PROGRAM COMMITTEE

Chairs: C. Dale Zinn (USA)
Mohammad Siddique (USA)

Abdulwahed, Mahmoud	Loughborough University	UK
Acharya, Sushil	Robert Morris University	USA
Ahn, Kwang Il	Korea Atomic Energy Research Institute	South Korea
Alameh, Kamal	Edith Cowan University	Australia
Alhayyan, Khalid N.	University of South Florida	USA
Ali, Shahid	National University of Science and Technology	Pakistan
Al-Omari, Hussein	Applied Science University	Jordan
Alsayegh, Osamah	Kuwait Institute for Scientific Research	Kuwait
Ambrose, Jude Angelo	Northumbria University	UK
Annakkage, U. D.	University of Manitoba	Canada
Arimitsu, Yoshihiro	Mesei University	Japan
Aristarkhov, Vasily	Intel Corporation	Russian Federation
Arndt, Angela E.	University of Cincinnati	USA
Aslam, Dean M.	Michigan State University	USA
Bang, Duck Je	Chungbuk National University	South Korea
Bell, Colin A.	Brunel University	UK
Bertel, Lykke	Aalborg University	Denmark
Blanchard, Richard	Loughborough University	UK
Bo, Hao	Shenyang Ligong University	China
Butler, Cary	US Army Corps of Engineers	USA
Call, Anson B.	College of Design	USA
Campbell, David J.	Siemens Medical Systems	USA
Carnahan, Heather	University of Toronto	Canada
Castillo A., Alejandro	Autonomous University of Yucatan	Mexico
Chalupa, Milan	Brno Univerzity of Technology	Czech Republic
Chang, Julian	Ching Yun University	Taiwan
Chang, Jyun-Wei	National Chiao Tung University	Taiwan
Chang, Kuo-Hwa	Chung Yuan Christian University	Taiwan
Chen, Chin-Ti	Institute of Chemistry Academia Sinica	Taiwan
Chen, Huifen	Chung Yuan Christian University	Taiwan
Chen, Mei-Yung	National Taiwan Normal University	Taiwan
Chen, Zong	Fairleigh Dickinson University	USA
Cheng, Yi-Chang	National Chiao Tung University	Taiwan
Chiu, Ting-Lan	University of Minnesota	USA
Cho, Kyoung-Rok	Chungbuk National University	South Korea
Cho, Young Duk	Blink Studios	South Korea
Choi, Yu-Lee	Chungbuk National University	South Korea
Chou, Yu-Hur	Tung Nan University	Taiwan
Chowdhury, Masud H.	University of Illinois at Chicago	USA

Chudý, Peter	Brno University of Technology	Czech Republic
Chung, Won Jee	Changwon National University	South Korea
Ciastellardi, Matteo	Polytechnic of Milano	Italy
Costa, F. F.	University of Bahia	Brazil
Cowan, Mark	US Army Corps of Engineers	USA
Cruciani, Andrea	University of Toronto	Canada
Das, Asesh	Pennsylvania College of Technology	USA
Davis, Karen C.	University of Cincinnati	USA
Dawoud, D. S.	University of KwaZulu Natal	South Africa
Dawoud, Peter D.	University of KwaZulu Nata	South Africa
D'Cruz, Carmo	Florida Institute of Technology	USA
De Castro L., Antonio C.	Federal University of Bahia	Brazil
De Kerckhove, Derrick	University of Toronto	Canada
Dierneder, Stefan	Linz Center of Mechatronics	Austria
Domínguez, Miguel Á.	University of Vigo	Spain
Dubrowski, Adam	University of Toronto	Canada
Eilouti, Buthayna H.	Jordan University of Science and Technology	Jordan
Fontana, M.	Federal University of Bahia	Brazil
Fujita, Naoyuki	Japan Aerospace Exploration Agency	Japan
Glovnea, Romeo P.	Brunel University	UK
Greca, Ardian	Georgia Southern University	USA
Guimarães, Diego	Federal University of Bahia	Brazil
Gunn, Rod	University of Glamorgan	UK
Guo, Weidong	Purdue University	China
Harvey, Adrian	University of Toronto	Canada
Hashimoto, Shigehiro	Osaka Institute of Technology	Japan
Hasnaoui, Salem	National Engineering School of Tunis	Tunisia
Hendel, Russell Jay	Towson University	USA
Herrnstadt, Steven	Iowa State University	USA
Hirvonen, Juhani	Technical Research Centre of Finland	Finland
Hirz, Mario	Graz University of Technology	Austria
Honzik, Petr	Brno University of Technology	Czech Republic
Hook, Derek J.	University of Minnesota	USA
Horvath, Gabor	University of Glamorgan	UK
Howell, P. R.	Pennsylvania State University	USA
Hsu, Yung-Chi	National Chiao Tung University	Taiwan
Huang, C. J.	Ching Yun University	Taiwan
Huang, Hsuan-Kuan	Industrial Technology Research Institute	Taiwan
Ishikawa, Koichiro	Japan Advanced Institute of Science and Technology	Japan
Ismail, Napsiah	Putra Malaysia University	Malaysia
Iyyunni, Chakradhar	University of Houston	USA
Jamuar, Sudhanshu S.	Putra Malaysia University	Malaysia
Jun, Moon-Seog	Soongsil University	South Korea
Jung, Dong Won	Lee International	South Korea
Kao, Diana	University of Windsor	Canada
Katila, Sanda	Kent State University	USA
Kermanshahi, Shahab	Sharif University of Technology	Iran
Khaled, Pervez	University of Illinois at Chicago	USA
Kim, Chul	University of New South Wales	Australia
Kim, Dae-Jung	Soongsil University	South Korea
Kim, Ho Chong	Shinsung	South Korea

Kim, Hyung Nam	Virginia Tech	USA
Kim, Jae Min	Changwon National Univesity	South Korea
Kim, Jeong-Jai	Soongsil University	South Korea
Kim, Mijin	Duksung Womens University	South Korea
Kim, Seok-Man	Chungbuk National University	South Korea
Koo, Kil Mo	Korea Atomic Energy Research Institute	South Korea
Krylov, Vladimir	Intel Corporation	Russian Federation
Kucera, Pavel	Brno University of Technology	Czech Republic
Kulkarni, Arun	University of Texas	USA
Kuragano, Tetsuzo	Meisei University	Japan
Kurita, Takio	Hiroshima University	Japan
Lau, Newman	Hong Kong Polytechnic University	Hong Kong
Lee, Choon Man	Changwon National University	South Korea
Lee, Eun-Hee	Chungbuk National University	South Korea
Lee, Eunoak	Duksung Womens University	South Korea
Lee, Je-Hoon	Chungbuk National University	South Korea
Lee, Jun-Ho	Samsung	South Korea
Lee, Seung-Min	Soongsil University	South Korea
Lee, Won Chang	Changwon National University	South Korea
Lee, Young-Sang	Samsung	South Korea
Lesaja, Goran	Georgia Southern University	USA
Leung, Lin	City University of New York	USA
Li, Xiaoping	National University of Singapore	Singapore
Lin, Hong	University of Houston Downtown	USA
Lin, Huan Yu	National Chiao Tung University	Taiwan
Lin, J.	Ching Yun University	Taiwan
Lin, Sheng-Fuu	Nation Chiao Tung University	Taiwan
Liu, Yuyu	University Tsinghua	China
Locklin, R. H.	Penn State University	USA
López-M., Cuauhtemoc	University of Guadalajara	Mexico
Machado, I. D. R.	Federal University of Bahia	Brazil
Maheshwari, Bharat	University of Windsor	Canada
Mares, Cristinel	Brunel University	UK
Mariño, Perfecto	University of Vigo	Spain
Masunov, Artëm E.	University of Central Florida	USA
Mccann, Roy A.	University of Arkansas	USA
Merino, Miguel	Community Health Centre at Merino	Spain
Mikhaylov, Ivan A.	University of Central Florida	USA
Miranda de A., Cristina	University of the Basque Country	Spain
Mochizuki, Shuichi	Osaka Institute of Technology	Japan
Mohan, K. Krishna	Indian Institute of Technology Bombay	India
Moin, Lubna	National University of Science and Technology	Pakistan
Mokhov, Serguei A.	Concordia University	Canada
Morita, Yusuke	Osaka Institute of Technology	Japan
Moss, Gloria	University of Glamorgan	UK
Muñoz, Humberto	Southern University and A&M College	USA
Nagy, Zoltan K.	Loughborough University	UK
Nam, Chang Soo	University of Arkansas	USA
Ndagije, Charles	National University of Rwanda	Rwanda
Ng, M. C. F.	Hong Kong Polytechnic University	Hong Kong
Ng, Min Shen	Putra Malaysia University	Malaysia

Nooshabadi, Saeid	Gwangju Institute of Science and Technology	South Korea
Oanta, Emil	Constanta Maritime University	Romania
Obrebski, Jan B.	Warsaw University of Technology	Poland
Oh, Kil Nam	Gwangju University	South Korea
Oh, Paul	Drexel University	USA
Okada, Masahide	Osaka Institute of Technology	Japan
Ono, Kohei	Osaka Institute of Technology	Japan
Ostergaard, Soren Duus	IBM Software Group	Denmark
Otero, Santiago	University of Vigo	Spain
Paiano, Roberto	University of Salento	Italy
Park, Uchang	Duksung Womens University	South Korea
Petkov, Emil	Northumbria University	UK
Pisupati, Sarma V.	Penn State University	USA
Platanitis, George	University of Ontario Institute of Technology	Canada
Podaru, Vasile	Military Technical Academy	Romania
Pop-Iliev, Remon	University of Ontario Institute of Technology	Canada
Pulimeno, Enrico	University of Lecce	Italy
Ramírez C., Guillermo H.	Soka University	Japan
Raud, Zoja	Tallinn University of Technology	Estonia
Rebielak, Janusz	Wroclaw University of Technology	Poland
Riedling, Eveline	Technical University of Vienna	Austria
Robins, David	Kent State University	USA
Rojas-Moreno, Arturo	National University of Engineering	Peru
Romero, Pedro	Rafael Belloso Chacin University	Venezuela
Rossbacher, Patrick	Graz University of Technology	Austria
Rutherfoord, James K.	Chattahoochee Technical College	USA
Rutherfoord, Rebecca H.	Southern Polytechnic State University	USA
Rzucidlo, Pawel	Rzeszow University of Technology	Poland
Saad, Inès	University of Picardie Jules Verne	France
Sangiorgi, U. B.	Federal University of Bahia	Brazil
Sariyildiz, Sevil	Delft University of Technology	Netherlands
Sateesh Reddy, J.	Birla Institute of Technology and Science	India
Sato, Yoshishige	Tsuruoka National College of Technology	Japan
Scharfe, Henrik	Aalborg University	Denmark
Seetharaman, R.	National Academy of Excellence	India
Shafahi, Yousef	Sharif University of Technology	Iran
Shaikh, Muzaffar	Florida Institute of Technology	USA
Sharieh, Ahmad	University of Jordan	Jordan
Shyu, Hsin-Yih Cindy	Tamkang University	Taiwan
Sidek, Roslina	University Putra Malaysia	Malaysia
Sissom, James	Southern Illinois University	USA
Smith-Jackson, Tonya	Virginia Tech	USA
Snidvongs, Suravut	Thai Natural Products	Thailand
Snow, Richard K.	Embry-Riddle Aeronautical University	USA
Song, Yong Mann	Korea Atomic Energy Research Institute	South Korea
Srinivasan, S.	Indian Institute of Technology	India
Srivastava, Smriti	Netaji Subhas Institute of Technology	India
Steiner, Bernhard	University of Vienna	Austria
Su, J. L.	Quanzhou Normal University	China
Sulaiman, S.	Putra Malaysia University	Malaysia
Tang, S. H.	Putra Malaysia University	Malaysia

Teshigawara, Yoshimi	Soka University	Japan
Tomovic, Mileta M.	Purdue University	USA
Torres Román, D.	Center for Research and Advanced Studies of National Polytechnic Institute	Mexico
Tseng, Shian-Shyong	National Chiao-Tung University	China
Uddin, Vali	National University of Sciences & Technology	Pakistan
Vander Biest, Alexis	Free University of Brussels	Belgium
Vasili, M. R.	Putra Malaysia University	Malaysia
Ventä, Olli	Technical Research Centre of Finland	Finland
Verma, A. K.	Indian Institute of Technology Bombay	India
Vodovozov, Valery	St. Petersburg State Electrotechnical University	Russian Federation
Vrána, Stanislav	Czech Technical University in Prague	Czech Republic
Wagiran, Rahman	Putra Malaysia University	Malaysia
Wang, Chuan-Ju	National Taiwan University	Taiwan
Wang, S. W.	Ching Yun University	Taiwan
Wile, Gregory	Case Western Reserve University	USA
Wilson, Ralph	Florida State University	USA
Wong, Ben	Hong Kong Polytechnic University	Hong Kong
Xia, W. J.	Hong Kong Polytechnic University	Hong Kong
Yamaguchi, Akira	Mesei University	Japan
Yang, Joon Eon	Korea Atomic Energy Research Institute	South Korea
Yeh, Syh-Shiuh	National Taipei University of Technology	Taiwan
Yoon, Sang Hwan	Changwon University	South Korea
Yoon, Young Min	Shinsung Holdings	South Korea
Zaretsky, Esther	Givat Washington Academic College of Education	Israel
Zhang, Yuru	Beihang University	China
Zheng, Huiyong	University of Michigan	USA

ADDITIONAL REVIEWERS

Abbott, Mick	University of Otago	New Zealand
Abe, Norihiro	Kyushu Institute of Technology	Japan
Albayrak, Özlem	Bilkent University	Turkey
Alturas, Braulio	Lisbon University Institute	Portugal
Aruga, Masahiro	Teikyo Heisei University	Japan
Bajpai, Gaurav	Kigali Institute of Science and Technology	Rwanda
Bangert, Patrick	Algorithmica Technologies Corporation	USA
Benbouziane, Mohamed	University of Tlemcen	Algeria
Bennani, Samir	Mohammed Vth University Agdal	Morocco
Bhat, Narayan G.	University of Texas-Pan American	USA
Boutejdar, Ahmed	Otto von Guericke University of Magdeburg	Germany
Chang, Teng-Wen	National Yunlin University of Science & Technology	Taiwan
Chen, Hsiao-Ping	Grand Valley State University	USA
Chou, Shyan-Bin	National Taiwan Normal University	Taiwan
Deliyska, Boryana	University of Forestry	Bulgaria
Ferreira, Maria	University of Portucalense	Portugal
Gardezi, A. K.	Colegio de Postgraduados	Mexico
Giménez, Eduardo L.	University of Vigo	Spain
Goyal, Megh R.	University of Puerto Rico	Puerto Rico
Hernández Arias, Aymara	Lisandro Alvarado University	Venezuela
Huang, En-Hsin	Alcatel-Lucent Technologies	USA
Imaña, José Luís	Complutense University of Madrid	Spain
Iribarne, Luís	University of Almeria	Spain
Ivanov, Sergiu	University of Craiova	Romania
Jia, Lei	New York University	USA
Kamyshnikov, Vladimir	Tomsk State Architectural University	Russian Federation
Kobayashi, Tadashi	Aichi Institute of Technology	Japan
Lin, Chieh-Yu	Chang Jung Christian University	Taiwan
Lin, Kuei-Chih	Ming Chuan University	Taiwan
López Román, Leobardo	University of Sonora	Mexico
Magnani, Lorenzo	University of Pavia	Italy
Manikas, Theodore W.	Southern Methodist University	USA
Marur, Prabhakar	General Motors R&D Center	India
Merten, Pascaline	Haute Ecole de Bruxelles	Belgium
Minea, Alina Adriana	Gheorghe Asachi Technical University of Iaşi	Romania
Neaga, Elena Irina	Loughborough University	UK
Oh, MyeongHoon	Electronics and Telecommunications Research Institute	South Korea
Opletal, Sascha	University of Stuttgart	Germany
Popov, Lubomir	Bowling Green State University	USA
Qu, Junfeng	Clayton State University	USA

Ramos, Ana Luísa	University of Aveiro	Portugal
Reis, Arsénio	University of Trás-os-Montes and Alto Douro	Portugal
Reyes-Méndez, Jorge Joel	Metropolitan Autonomous University	Mexico
Rodera B., Ana María	Open University of Catalonia	Spain
Rodrigues, Leonardo	State University of Campinas	Brazil
Rodríguez L., Gloria I.	National University of Colombia	Colombia
Röning, Juha	University of Oulu	Finland
Rubín, Gustavo	Autonomous University of Puebla	Mexico
Sembera, Jan	Technical University of Liberec	Czech Republic
Sizikov, Valery	Saint-Petersburg State University	Russian Federation
Stanchev, Peter	Kettering University	USA
Strickfaden, Megan	University of Alberta	Canada
Sulema, Yevgeniya	National Technical University of Ukraine	Ukraine
Valdés-H., María del C.	University of Edinburgh	UK
Vizureanu, Petrica	Gheorghe Asachi Technical University of Iaşi	Romania
Whiteley, Rick	Calabash Educational Software	Canada
Xu, Guohua	Huazhong University of Science and Technology	China
Yang, George	Missouri Western State University	USA
Yun, Myung Hwan	Seoul National University	South Korea

ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

Abe, Akinori	NTT Communication Science Laboratories	Japan
Acosta Sánchez, Leopoldo	University of La Laguna	Spain
Alayón Miranda, Silvia	University of La Laguna	Spain
Alhayyan, Khalid N.	University of South Florida	USA
Amblard, Frederic	Toulouse 1 University Capitole	France
Ameta, Gaurav	Washington State University	USA
Bugdol, Marek	Jagiellonian University	Poland
Canbolat, Huseyin	Mersin University	Turkey
Carrington, Michael	Northern Virginia Community College	USA
Casallas, Luz	Francisco Jose of Caldas District University	Colombia
Chen, Wei	Eindhoven University of Technology	Netherlands
Christiansen, Ellen	Aalborg University	Denmark
Courdier, Rémy	University of La Réunion Saint-Denis	France
Das, Diganta	University of Maryland	USA
Dias da Cunha, Myrtes	Federal University of Uberlândia	Brazil
Erawan, Mahendrawathi	Sepuluh Nopember Institute of Technology	Indonesia
Filkov, Alexander	Tomsk State University	Russian Federation
Fontana, Giorgio	University of Trento	Italy
Forouzbakhsh, Farshid	Tehran University	Iran
Gardiol, Fred	Swiss Federal Institutes of Technology	Switzerland
Goyal, Megh R.	University of Puerto Rico	Puerto Rico
Guerrero C., José Fermi	Autonomous University of Puebla	Mexico
Hardt, Wolfram	Chemnitz University of Technology	Germany
Hassoun, Alain	Ministry of Higher Education and Scientific Research	France
Huang, Yudong	Harbin Institute of Technology	China
Ibrahim, Othman	Universiti Teknologi Malaysia	Malaysia
Imaña, José Luís	Complutense University of Madrid	Spain
Iyyunni, Chakradhar	University of Houston	USA
Kasiri, Norollah	Iran University of Science and Technology	Iran
Katashaya, Steven	University of North-West	South Africa
Kawabe, Tohru	Tsukuba University	Japan
Kent, Sedef	Istanbul Technical University	Turkey
Korzun, Dmitry	Petrozavodsk State University	Russian Federation
Kupelwieser, Hans	Graz University of Technology	Austria
Lawrence, David	University of North Dakota	USA
Luhanga, Pearson	University of Botswana	Botswana
Mishra, Rakesh	University of Huddersfield	UK
Møller Andersen, Nina	University of Copenhagen	Denmark
Nakamura, Yukinori	Tokyo University of Agriculture and Technology	Japan
Nguyen, Tien Giang	Vietnam National University	Vietnam

Nicolescu, Bogdan	University of Pitesti	Romania
O'Leary, Diane	University of Maryland	USA
Oosterhuis, Kas	Delft University of Technology	Netherlands
Ortega, Romeo	Laboratoire des Signaux et Systèmes	France
Palmgren, Juni	Karolinska Institutet	Sweden
Pedersen, Susan	Texas A & M University	USA
Petros, Thomas	University of North Dakota	USA
Pittet, Patrick	Université Claude Bernard Lyon	France
Plazas Nossa, Leonardo	Francisco Jose of Caldas District University	Colombia
Ramaswamy, Sriram	Creighton University	USA
Rebholz-S., Dietrich	European Bioinformatics Institute	UK
Regazzoni, Carlo	University of Genoa	Italy
Roy, Ram Naresh	Eastern Institute of Technology	New Zealand
Rubín, Gustavo	Autonomous University of Puebla	Mexico
Rubín Linares, Gustavo	Autonomous University of Puebla	Mexico
Salvatore, Distefano	University of Messina	Italy
Saracco, Roberto	Telecom	Italy
Scarpa, Marco	University of Messina	Italy
Sharma, Satish	San Diego State University	USA
Sheynin, Yuriy	Yerevan State University	Russian Federation
Sim, Alan	Mount Sinai School of Medicine	USA
Sira Ramírez, Hebertt J.	CINVESTAV	Mexico
Stephenson, Gary	Linquest Corporation	USA
Suomala, Jyrki	Laurea University of Applied Sciences	Finland
Swan, Karen	University of Illinois Springfield	USA
Tammisto, Yulia	Aalto University	Finland
Toller, Eric	Battelle Memorial Institute	USA
Tossavainen, Päivi	Laurea University of Applied Sciences	Finland
Vaninsky, Alexander	City University of New York	USA
Vargas, Edgar	Jorge Tadeo Lozano University	Colombia
Vicario, Enrico	University of Florence	Italy
Wei, Xinzhou	City University of New York	USA
Xiong, Ying	Northwestern University	USA
Yamakawa, Takeshi	Kyushu Institute of Technology	Japan
Yang, Chunhui	Harbin Institute of Technology	China
Young, Patrica	University of Maryland	USA

International Conference on Design and Modeling in Science, Education, and Technology: DeMset 2011

PROGRAM COMMITTEE CHAIRS
Mohammad Siddique
C. Dale Zinn

ORGANIZING COMMITTEE CHAIRS
Andrés Tremante
Belkis Sánchez

HARDCOPY PROCEEDINGS PRODUCTION CHAIR
María Sánchez

CD PROCEEDINGS PRODUCTION CHAIR
Juan Manuel Pineda

SYSTEMS DEVELOPMENT, MAINTENANCE AND DEPLOYMENT
Dalia Sánchez
Keyla Guédez
Nidimar Díaz
Jesús Malavé

OPERATIONAL ASSISTANTS
Marcela Briceño
Cindi Padilla

GENERAL CHAIRS
Nagib Callaos
Hsing-Wei Chu

HONORARY PRESIDENT
William Lesso

HELP DESK
Louis Barnes
Sean Barnes

CONFERENCES PROGRAM MANAGER
Leonisol Callaos

TECHNICAL CONSULTANT ON COMPUTING SYSTEMS
Juan Manuel Pineda

META-REVIEWERS SUPPORT
Dalia Sánchez

PROCEEDINGS PRODUCTION SUPPORT

Dalia Sánchez
Keyla Guédez
Marcela Briceño
Cindi Padilla

PROMOTIONAL SUPPORT

Keyla Guédez
Nidimar Díaz
Freddy Callaos

Number of Papers Included in these Proceedings per Country

(The country of the first author was the one taken into account for these statistics)

Country	# Papers	%
TOTAL	112	100%
United States	34	30,36%
Japan	9	8,04%
Finland	6	5,36%
Mexico	5	4,46%
Brazil	4	3,57%
Germany	4	3,57%
South Korea	4	3,57%
Taiwan	4	3,57%
New Zealand	3	2,68%
Spain	3	2,68%
Canada	2	1,79%
France	2	1,79%
Hungary	2	1,79%
India	2	1,79%
Iran	2	1,79%
Italy	2	1,79%
Malaysia	2	1,79%
Norway	2	1,79%
South Africa	2	1,79%
Sweden	2	1,79%
Switzerland	2	1,79%
United Kingdom	2	1,79%
Argentina	1	0,89%
Australia	1	0,89%
Austria	1	0,89%
Barbados	1	0,89%
Botswana	1	0,89%
China	1	0,89%
Lithuania	1	0,89%
Netherlands	1	0,89%
Poland	1	0,89%
Portugal	1	0,89%
Russian Federation	1	0,89%
Turkey	1	0,89%

Foreword

Information and Communication Technologies (ICT) are having an increasing impact in almost every scientific discipline and are facilitating the creation of integrative systems and processes, which are in turn supporting the creation of effective relationships among different academic activities and potentiating effective collaboration in research, design, and education. On the other hand the conceptual infrastructures of Systemics, Informatics, and Cybernetics (Communication and control) are increasingly being related to each other and are providing an effective intellectual platform for inter-disciplinary communication. Accordingly, the main purpose of the organizing committees of the collocated events organized by the International Institute of Informatics and Systemics (IIIS) on 11/29-12/2, 2011, in Orlando Florida is to bring together researchers, developers, practitioners, consultants and users of Information and Communication Technologies, for **intra- and inter-disciplinary communication**,

Consequently, three kinds of activities have been planned:

1. Regular traditional presentations in breakout sessions to support *intra-disciplinary* communication,

2. Plenary sessions where Keynote Speakers will address the *multi-disciplinary* audience, mostly with inter- or trans-disciplinary topics, and

3. Conversational sessions on inter- or trans-disciplinary topics in order to support *inter-disciplinary communications* and to foster the **analogical thinking** that might emerge in a multi-disciplinary forum based on trans-disciplinary concepts and/or multi-disciplinary tools, technologies, and methodologies. Ideas generated by analogical thinking might be a) applied to a diversity of areas and practical domains, and b) support a synergic combination of **analytical** and **synthetic** thinking.

The disciplinary *variety*, required for inter-disciplinary communications, analogical learning, and synergic analytical/synthetic thinking, is one of the motivation for organizing the following related events:

- International Conference on Information and Communication Technologies and Applications ICTA 2011
- Design and Modeling in Science, Education, and Technology: DeMset 2011
- International Symposium on Integrating Research, Education, and Problem Solving: IREPS 2011
- International Conference on Education, Informatics, and Cybernetics: icEIC 2011

The articles accepted for presentation that also have an author registered in the conference for the respective presentation, have been grouped in two volumes for their publications in the hard copy proceedings of the collocated events. Their grouping is based on the similarities of the respective topics. Consequently, papers of ICTA 2011 and DeMset 2011 have been grouped in one volume, and papers of IREPS 2011 and ICEIC 2011 have been grouped in another volume.

All papers to be presented at the collocated events were included in the electronic version of the proceedings as well.

Since different kinds of reviewing methodologies are applied in different disciplines we integrated the most used reviewing methods into a *systemic reviewing methodology* for the reviewing process of the papers submitted to the collocated events.

This methodology is based on three-tier reviews: open (or non-blind), double-blind, and participative reviews. Final acceptance depends on the three kinds of reviews. However, a paper should be recommended by non-blind reviewers AND blind reviewers in order to be accepted for presentation at any event and to be included in the respective proceedings. A recommendation to accept made by non-blind reviewers is a **necessary** condition, but it is not a **sufficient** one. A submission, to be accepted, should also have a majority of its double-blind reviewers recommending its acceptance. This double necessary conditions generate a **more reliable and rigorous** reviewing than those reviewing methods based on just one of the indicated methods, or just on the traditional double-blind reviewing.

Double-blind reviewing has been done by a random selection of 3-5 reviewers from about 20,000 IIIS reviewers who classified their research or expertise field in the same theme, area, or subarea where the author classified his/her submission. The random selection (made by a computer program) has been conceived in order to avoid any conscious, or un-conscious, bias that might be done by a human-being selection of the respective reviewers.

IIIS' non-blind reviewing is based on the essence of what Kaplan (2005, "How to Fix Peer Review", *The Scientist*, Volume 19, Issue 1, Page 10, Jun. 6) proposed in order to fix peer reviewing problems. Kaplan affirms that "Peer review subsumes two functions. First, peer reviewers attempt to improve manuscripts by offering constructive criticisms about concrete elements … The second function of peer review is to render a decision about the … significance of the findings so that the manuscript can be prioritized for publication. I propose reforming peer review so that the two functions are independent." With regards to the first function of peer reviewing, Kaplan proposes that "**Review of a manuscript would be solicited from colleagues by the authors.** The first task of these reviewers would be to identify revisions that could be made to improve the manuscript. Second, the reviewers would be responsible for writing an evaluation of the revised work. This assessment would be mostly concerned with the significance of the findings, and the reviewers would sign it" (emphasis added).

We try to achieve the first function via Kaplan's non-blind peer reviewing and the second function by the traditional means of double-blind review. This is why acceptance of submissions by the non-blind reviewers is a necessary condition but not a sufficient one. The submission should also have favorable recommendations by the majority of the double-blind reviewers in order to be accepted by IIIS for its presentation and inclusion in the respective conference proceedings.

A third reviewing tier is the participative peer reviewing, which complements the two tiers described above but is not a necessary condition for accepting a submission. An article submitted to a conference being organized by IIIS is immediately displayed for review to those authors

who submitted articles in the same theme, area, or sub-area. Accordingly, each submitting author has access to all submissions sent to the same area where he/she submitted his/her article and can comment and evaluate them. This is what we call at IIIS "Participative Peer-to-Peer Reviewing" or PPPR. This kind of reviewing provides additional input to the selection process and assists all participants in placing their presentations in context. It is not a necessary condition but it has a complementary function, especially in those cases where the non-blind reviewers have a strong disagreement or there is no majority of recommendations accepting, or not accepting, the article.

On behalf of the Organizing Committees, I extend our heartfelt thanks to the members of the four Program Committees (from 74 countries), and to the additional 847 reviewers, from 85 countries, each one of whom reviewed at least one of the submitted articles. 327 reviewers, from 65 countries, were suggested by the respective authors for the non-blind peer reviews. *Each registered author could get information about: 1) the average of the reviewers' evaluations according to 8 criteria, and the average of a global evaluation of his/her submission; and 2) the comments and constructive feedback made by the reviewers, who recommended the acceptance of his/her submission, so the author would be able to improve the final version of the paper.*

A total of 1792 reviews were made to the 303 submissions that were received, which means that an average of 5.91 reviews were made to each received submission, and an average of 2.12 reviews were made by each reviewer. The 112 papers included in these proceedings, from 33 countries, are 36.96% of the 303 submissions that were initially received. Details for each of the four events are summarized in the following table.

Conference	# of submissions received	# of reviewers that made at least one review	# of reviews made	Average of reviews per reviewer	Average of reviews per submission	# of papers included in the proceedings	% of submissions included in the proceedings
icEIC 2011	58	210	400	1.90	6.90	22	37.93%
ICTA 2011	115	323	758	2.35	6.59	41	35.65%
DeMSET 2011	56	142	326	2.30	5.82	22	39.29%
IREPS 2011	74	172	308	1.79	4.16	27	36.49%
TOTAL	**303**	**847**	**1792**	**2.12**	**5.91**	**112**	**36.96%**

We are also grateful to the co-editors of these proceedings for the hard work, energy, and eagerness they displayed in preparing them. We express our intense gratitude to Professor William Lesso for his wise and opportune tutoring, for his eternal energy, integrity, and continuous support and advice as Honorary President of IIIS' conferences, as well as for being a very caring old friend and intellectual father to many of us. We also extend our gratitude to Professor Belkis Sanchez, who brilliantly managed the organizing process. Special thanks to doctors C. Dale Zinn and Jeremy Horne, and to professors Hsing-Wei Chu, Friedrich Welsch, Michael Savoie, Andrés Tremante, Jorge Baralt, Mohammad Siddique, and José Ferrer for chairing, or co-chairing the respective program committees.

We also extend our gratitude to doctors Robert Baker, Joseph Finkelstein, Jeremy Horne, Daniel Katz, Lisbeth Amhag, Merja Bauters and to professors Juha Kettunen, T. Grandon Gill, Bodil Ask, Harald Haugen, Mohamed El-Sayed, and Donald Poochigian, for accepting to address the audience of the General Joint Plenary Sessions with keynote addresses.

We also wish to thank all the authors for the quality of their papers.

We extend our gratitude as well to Juan Manuel Pineda, Leonisol Callaos, Dalia Sánchez, Keyla Guedez, Nidimar Díaz, Marcela Briceño, Cindi Padilla, Louis Barnes, Sean Barnes, Abrahan Marin, and Freddy Callaos for their knowledgeable effort in supporting the organizational process producing the hard copy and CD versions of the proceedings, developing and maintaining the software supporting the interactions of the authors with the reviewing process and the Organizing Committee, as well as for their support in the help desk and in the promotional process.

Professor Nagib C. Callaos,
General Chair

CONTENTS
(Post-Conference Edition)

Contents

Communication and Network Systems and Technologies

Computer Science and Engineering

Computer-Based Training: Web-Based Training, Internet-Based Teaching, etc.

Control Systems, Technologies and Applications

ICT applications in Education and Training

ICT applications in Science and Engineering

ICT Applications in Social and Political Systems

ICT Applications of in Health Care and Bio-Medical ICT

Image, Acoustic, Speech and Signal Processing

Information and Communication Technologies and Applications

Information Systems and Technologies

Information Systems Management

Tele-Communication Systems, Technologies and Applications: Wireless Networking, Mobile Computing. Wireless, Mobile Software Engineering and Applications, etc.

Computing, Communications and Control Technologies

Design and Modeling in Science, Education, and Technology

Design and Modeling Methods and Methodologies

Design of Complex Systems or Environments for Living, Playing, and Learning

Design Research. Research through, into, by, and/or for Design

Educational Research, Design, and Modeling

Qualitative Models and Modeling in Science and Engineering

Quantitative Models and Modeling in Science and Engineering

Performance Evaluation of DV-Hop and NDV-Hop Localization Methods in Wireless Sensor Networks

Manijeh Keshtgary
Dept. of Computer Eng. & IT
ShirazUniversity of technology
Shiraz,Iran,
Keshtgari@sutech.ac.ir

And

Mohammad Hadi Tabeshfaraz
Dept. of Computer Eng. & IT
Tehran University
Kish, Iran,
Tabeshfaraz@ut.ac.ir

And

Masoud Fasihy
Dept. of Computer Eng. & IT
Amirkabir University
Tehran, Iran,
Mfasihy@gmail.com

ABSTRACT

Knowledge of nodes' locations is an important requirement for many applications in Wireless Sensor Networks. In the hop-based range-free localization methods, anchors broadcast the localization messages including a hop count value to the entire network. Each node receives this message and calculates its own distance with anchor in hops and then approximates its own location. In this paper, we review range-free localization methods and evaluate the performance of two methods: "DV hop" and "NDV-Hop" by simulation. Recent papers are mostly concentrated on the number of anchors for NDV-Hop and no other parameters like network area and distribution model. Here, we consider the effect of area and distribution model on localization accuracy.

Keywords: Hop-based, Accuracy, Localization, Wireless Sensor Networks.

1. INTRODUCTION

WSN is usually constructed with large number of sensor nodes, which are heavily organized within the monitored area [1]. Algorithm in localization technology is an important issue in WSN. In WSN technology, knowledge of where the information has been obtained is very vital and important. When an event takes place, it can only be recognized and processed if the position of the place that generates the information is known.

At present, the widely used poisoning system is GPS [2] (Global Position System). The nodes with GPS receiver is localized by receiving the real time information from GPS satellite. The GPS has advantages of high accuracy, good real time, and high anti-jamming performance, etc. However, this method is not a suitable solution for localization because of the below reasons [11].

1. GPS can't be used as an in indoor solution since it can't be placed in a satellite line of sight.
2. GPS consumes lots of energy and reduces the network lifetime.
3. GPS is expensive and can't be used in large-scale sensor nodes.

In the most of localization methods, limited numbers of nodes are location-aware which are called "Anchors". Anchors can be fixed or mobile. Some methods like mobile beacon, mobile anchor

[3], and single mobile beacon [4] use mobile anchors. The advantage of these methods is that a mobile anchor can be used instead of many fixed anchors, and in each location, it can be considered as a fixed anchor. However, the mobility could cause some problems.

Other localization methods use "Fixed Anchors". In this method, localization can be commenced by anchors. They broadcast their own location to the network and three other anchors in order to calculate their own location. In this method, the accuracy of localization depends on the approximate value of own location. The localization algorithm is generally divided into two categories: range-based and range-free algorithm. Range-based methods use some measurement equipment for calculation of distance. Some of these methods are DV-Distance [5], Euclidean [5], N-hop Multitration [6], and Robust Sensor Localization [7]. Extra hardware and additional energy consumption are important bottleneck for these methods. Another category of localization methods named range-free, don't use measurement equipment, so they have lower cost. In hop-based range-free methods, nodes estimate the distance by counting the hops to the anchors.

Hop-based range-free methods have no extra hardware and flexible. Localization precision is more practical for implementation in any environment, so we concentrate on this category of localization method. We consider DV-hop [5] and NDV-HOP [8] as typical algorithms of this category and evaluate their localization error using simulation.

Recent researches [9, 10], improve the DV-Hop to increase the accuracy of localization method. Here we consider classic DV algorithm. Therefore, classic DV-HOP can be seen and used as a benchmark for evaluating the errors and accuracy range in NDV-HOP Localization method.

The rest of this paper is organized as follow. We discuss DV-Hop and NDV-HOP methods in sections 2 and 3. We evaluate the efficiency of these two methods focusing on their localization errors in section 4. We conclude the paper in section 5.

2. DV-HOP LOCALIZATION METHOD

Niculescu and Nath [5] introduce a classic DV-Hop method which is similar to the distance vector routing schemes. This method includes three stages: First, Anchors distribute their localization message to the entire network. This message may include fields such as Anchor ID, Location, and Hop-count. Hop-Limit is defined as the maximum number of hops that a localization message can be alive. Each node receives this message and stores it in its memory, increment the hop-count by one and forwards this message for neighbors if the hop-count is less than or equal to the hop-limit, otherwise the message will be killed by sensor node. If a node receives more than one message from an anchor, the shortest path will be considered; the new message will be ignored and the sensor node only forwards the message. According to this fact that the first message from an anchor include the shortest hop-length, each node considerers only the first message from each anchor and ignore the next messages. When a node receives localization messages from 3 anchors in hops, this step will be terminated. [11]

If any anchor receives a localization message from an anchor, store it to calculate the AHL (Average hop Length). For example in figure 1, we have 3 anchors (a1, a2 and a3).

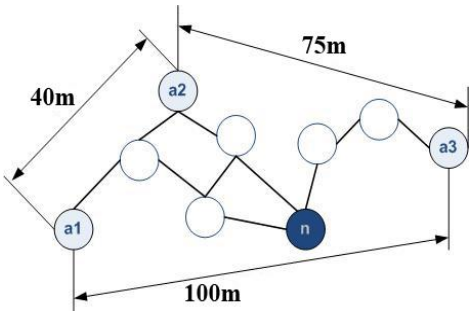


Figure 1: DV-Hop example

The distance from a1 to a2 is 40m, a2 to a3 is 75m and a1 to a3 is 100m. The hop-count from a1 to a2 is 2, a2 to a3 is 5 and a1 to a3 is 6. Each anchor calculates own AHL as follows:

AHL (a1) = (100+40) / (6+2) =17.5

AHL (a2) = (40 + 75) / (2 + 5) = 16.42

AHL (a3) = (75 + 100) / (5 + 6) = 16.42

In the second step, each anchor distributes the AHL to the entire network. Each anchor receives this packet and considers the first AHL receives from nearest anchor. For example in figure 1, for node n, the first AHL received from a2 is 16.42, so the distance from anchors for node n is calculated by multiplying the minimum hop number and average distance of each hop:

n-a1=3*16.42=49.26

n-a2=2*16.42=32.84

n-a3=3*16.42=49.26

Now, each node is able to estimate the location by Multilateration method. Let's see how the location could be estimated. Assume that (x_i,y_i) is the location of ith anchor and r_i is the distance of a node to this anchor. If (x_u,y_u) is the location of the node, then we have from Euclidean distance equation:

$$(\mathbf{x_i + x_u})^2 + (\mathbf{y_i + y_u})^2 = \mathbf{r_i^2} \quad \mathbf{for\ i = 1, 2, 3} \quad (\mathbf{1})$$

We subtract this equation with i=3 from equation with i=1 and i=2 and simplify equation based on x_u and y_u then:

$$(\mathbf{x_1 - x_u})^2 - (\mathbf{x_3 - x_u})^2 + (\mathbf{y_1 - y_u})^2 - (\mathbf{y_3 - y_u})^2 = \mathbf{r_1^2 - r_3^2} \quad (\mathbf{2})$$

$$\mathbf{2(x_3 - x_1)x_u + 2(y_3 - y_1)y_u = (r_1^2 - r_3^2) - (x_1^2 - x_3^2) - (y_1^2 - y_3^2)} \quad (\mathbf{3})$$

$$\mathbf{2(x_3 - x_2)x_u + 2(y_3 - y_2)y_u = (r_2^2 - r_3^2) - (x_2^2 - x_3^2) - (y_2^2 - y_3^2)} \quad (\mathbf{4})$$

By converting these equations into a matrix we have:

$$2\begin{bmatrix} \mathbf{x_3 - x_1} & \mathbf{y_3 - y_1} \\ \mathbf{x_2 - x_1} & \mathbf{y_3 - y_2} \end{bmatrix}\begin{bmatrix} \mathbf{x_u} \\ \mathbf{y_u} \end{bmatrix} = \begin{bmatrix} \mathbf{(r_1^2 - r_3^2) - (x_1^2 - x_3^2) - (y_1^2 - y_3^2)} \\ \mathbf{(r_2^2 - r_3^2) - (x_2^2 - x_3^2) - (y_2^2 - y_3^2)} \end{bmatrix} \quad (\mathbf{5})$$

Now x_u and y_u can be calculated easily [11]

3. NDV -HOP LOCALIZATION METHOD

X. Zhang, H. Xie and X. Zhao [12] proposed some improvements over DV-Hop from following three aspects:

Average hop distance is calculated by selecting all anchors. Classic DV-Hop uses only two anchors to calculate average Hop length (AHL), But, experiments shows that the more anchors are considered, the lower localization error is. For the applications which need high accuracy, all the anchors can be considered when calculating AHL. This method is adopted in order to obtain higher accuracy of localization.

AHL is corrected by anchors. In this method, the distance between nodes Beacon can be calculated by multiplying the distance between each hop and the recorded hops. For example; the approximate distance can be estimated and shown from the node Beacon i till node Beacon j by $d_{est}^{i,j}$:

$$\mathbf{d_{est}^{i,j} = HopSize_i * hop_{i,j}} \quad (\mathbf{6})$$

where $HopSize_i$ denote the average hop distance calculated by beacon node i , and $hop_{i,j}$ is the number of hops between beacon nodes i and j .At the same time ,the actual distance between beacon node i and j can be calculated according to the self-position information and that of beacon node j obtained in the first phase. The actual distance between i and j can be estimate by formula (7):

$$\mathbf{d_{rel}^{i,j}} = \sqrt{(\mathbf{x_i - x_j})^2 + (\mathbf{y_i - y_j})^2} \quad (\mathbf{7})$$

At the same time the approximate error of hop between i,j node can be calculated. Difference between estimated distance $d_{est}^{i,j}$ and the real distance $d_{rel}^{i,j}$ can be divided to the number of hops between them. $hopmod_i^{i,j}=(d_{est}^{i,j}-d_{rel}^{i,j})/hop_{i,j}$, wher eh $hopmod_i^{i,j}$ is the average hop error.

Finally, the location of boundary nodes is corrected by considering the network bounds.

This is a fact that the estimated location for any nodes should be satisfies some conditions, such as

xmin≤x≤xmax (9)
ymin≤y≤ymax (10)

Where xmin and xmax are the minimum and maximum value of along the x axis in the monitored area respectively. So for the x-axis, if x<xmin, replace x with xmin, and if x>xmax, replace x with xmax. For the y-axis, the same replacement is done.

So, the out-of-scope problem has been avoided and relative error decreases.

Therefore, we can say that NDV-HOP method has advantages over the classic DV-HOP because it uses all the distance between node Beacons in the network space, in order to find an accurate location compared to classic DV-Hop.

4. SIMULATION RESULTS

We have used OMNET++ 3.2 to simulate these methods. OMNET is an object oriented and event-driven software based on C++. Implemented model is formed by hierarchical modules communicating by messages.

Simulation parameters are:

- Network size (the number of sensor nodes)
- The number of anchors.
- Range of sensor nodes.
- Range of anchors.
- Distribution model of sensor nodes

The network model is set as follows:

- All the nodes in the network are homogenous, and there is no influential factor in the communicating procedure with one another.
- All the nodes have the same communication radius, success probability of the communication and flooding routing protocol.
- 100 nodes are randomly distributed in different area (50*50,100*100,150*150,200*200…)

The evaluation parameter is:

- **Localization error:** The difference between estimated coordinates and real coordinates divided by communication radius, which can be described as $RE=\|R_i-E_i\|/R$ .where R_i denotes real coordinates, E_i is estimated coordinates and R is the communication radius.

4.1. Localization Error Evaluation

At first, we will evaluate the performance of DV-hop and NDV-hop algorithm performance with respect to different area in a uniform model. Then we do the same evaluation for normal model. Figures 2, 3 and 4 show the localization errors in uniform model for anchor nodes of 5, 10 and 20 when there has been an increase in area. Xmax shows the area of network in figures. In the figure 2, the localization error for 5 anchors nodes in the area of (50*50, 100*100,150*150…) is shown. As it can be seen, the localization error in the NDV-HOP is much less than DV-HOP although. With an increase in the number of anchor nodes, the error is reduced in the NDV-HOP and difference between accuracy of these two algorithms is more. For instance; in the area of 500*500, the localization errors of NDV-HOP localization for 10 anchor nodes is approximately2.3637 where as in DV-HOP is around 11.8243.

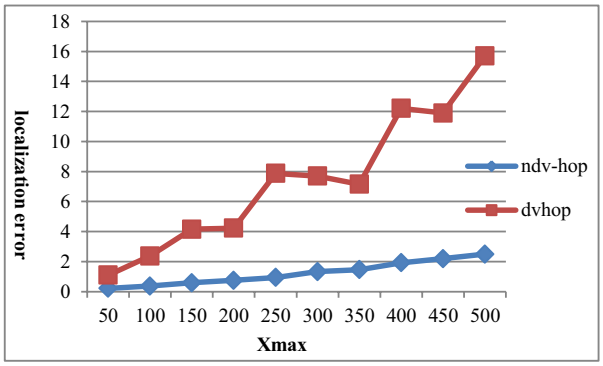


Figure 2: Localization error in uniform distribution (for 5 anchors)

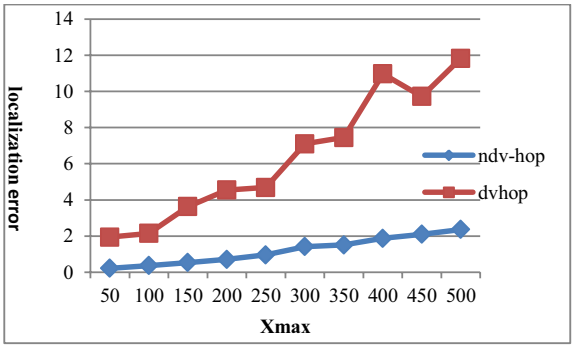


Figure 3: Localization error in uniform distribution (for 10 anchors)

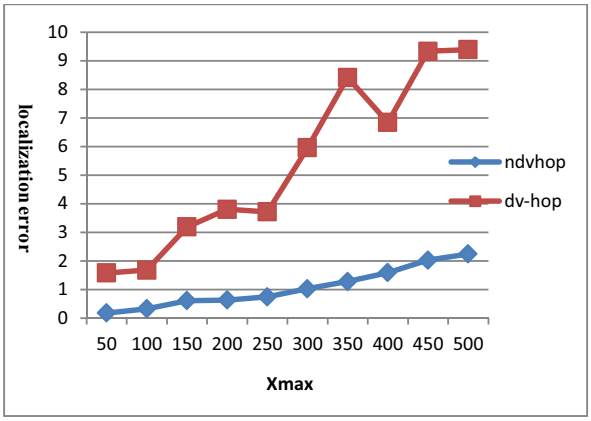


Figure 4: Localization error in uniform distribution (for 20 anchors)

Until now, we assume that nodes are distributed in a uniform model. Now we study the localization error in non-uniform networks. We consider normal distribution model because this model is more feasible than the others. Figure 5 compares the localization error for the case of 20 anchors and NDV-Hop algorithm with two different distribution models, Uniform and normal. As we can see, Localization error in normal distribution is more than uniform networks. It means that the performance of NDV-Hop method in uniform networks is better than other distribution models such as normal.

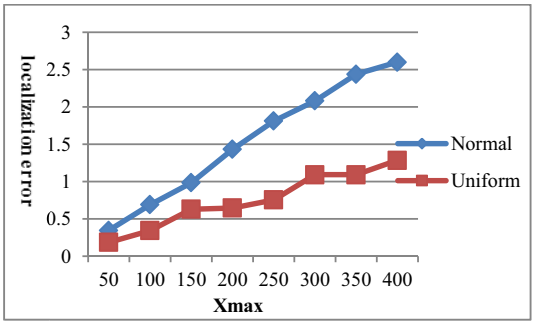


Figure 5: Localization error of NDV-Hop in normal and uniform distribution (for 20 anchors)

5. CONCLUSIONS

Localization is an important challenge in the wireless sensor networks. In this paper, we evaluate performance of two hop-based range-free methods that are more practical in WSN: classic DV-Hop and NDV-Hop. Localization error of NDV-hop is less than DV-Hop because it uses different error correction scheme. In both methods, localization error increases with increasing the area, but NDV-Hop has better accuracy in large-area networks. Also we found that accuracy in uniform distribution is better than normal distribution. As the result, NDV-Hop method has low localization error in uniform distribution. For future work, we will consider other parameters for performance evaluation of these two methods such as energy consumption and fault tolerance.

6. REFERENCES

[1] Z. Li, R. Li, Y. Wei and T. Pei, "Survey of localization Techniques in Wireless Sensor Networks", Information Technology Journal 9 (8), 2010, pp. 1754-1757.

[2]A. M. Youssef, M. Youssef, “A Taxonomy of Localization Schemes for Wireless Sensor Networks", Proc. Of International Conference on Wireless Networks (ICWN 07), Las Vegas, USA, 2007, pp. 444-450.

[3] K. F. Ssu, C. H. Ou and H. C. Jiau, "Localization with Mobile Anchor Points in Wireless Sensor Networks", IEEE Transactions on Vehicular Technology, vol. 54, No.3, 2005, pp. 1178-1197.

[4] A. Galstyan, B. Krishnmachari, K. Lerman and S. Pattem, "Distributed Online Localization in Sensor Networks Using a Moving Target", Proceedings of third international symposium on information processing in sensor networks (ISPN). Berkeley, 2004, pp. 61-70.

[5] D. Niculescu and B. Nath. "Ad Hoc Positioning System (APS)," Proc. of the IEEE GLOBECOM 2001, San Antonio, 2001, pp. 2926-2931.

[6] A. Savvides, H. Park and M. Srivastava, "The bits and flops of the N-hop multilateration Primitive For Node Localization Problems", First ACM International Workshop on Wireless Sensor Networks and Application (WSNA), Atlanta Georgia. USA, 2002, pp. 112-121.

[7] X. Ji and H. Zha, "Robust Sensor Localization Algorithm in Wireless Ad-hoc Sensor Networks", Proceeding of 12th International Conference on Computer Communications and Networks, Dallas, Texas, USA, 2003, pp. 527-532.

[8] R. Nagpal, "Organizing a Global Coordinate System from Local Information on an Amorphous Computer", A.I. Memo1666, MIT A.I. Laboratory, August 1999.

[9] L. Wu, Max Q.-H. Meng, Z. Dong and H. Liang, "An Empirical Study of DV-Hop Localization Algorithm in Random Sensor Networks", Second International Conference on Intelligent Computation Technology and Automation, ICICTA , China, 2009.

[10] J. Li, J. Zhang and L. Xiande, "A Weighted DV-Hop Localization Scheme for Wireless Sensor Networks", 2009 International Conference on Scalable Computing and Communications; Eighth International Conference on Embedded Computing, 2009.

[11] M. Keshtgar , M. Fasihy, Z. Ronaghi, “Performance Evaluation of Hop-Based Range-Free Localization Methods in Wireless Sensor Network", ISRN Communications and Networking, Volume 2011, 2011.

[12] X. Zhang, H. Xie and X. Zhao, "Improved DV-Hop localization Algorithm for Wireless Sensor Networks", Computer Applications, vol. 27, 2007, pp. 2672-2674.

Investigating O:MIB-Based Distributed Active Information Model (DAIM) for Autonomics

Pakawat Pupatwibul, Bahram Jozi, and Robin Braun, *Faculty of Engineering and IT, University of Technology Sydney, New South Wales 2007, Australia*

***Abstract*—Technological innovations in communication networking, computing applications, and information modeling have played a significant role in managing complex distributed electronic systems. Autonomic Computing (AC) is a concept to deal with the over growing complexity of distributed networks; this term gives systems the ability of self-management, which mean each component in AC can adapt itself to changing conditions of the dynamic environment. In this paper we investigate a new nature-inspired Distributed Active Information Model (DAIM) to allow the local decision making process, that will essentially contribute to complex distributed network environments. Details of the DAIM structure are also described, which will hopefully address the schemes of some previous network management protocols such as Simple Network Management Protocol (SNMP), Common Information Model (CIM), and mechanism like Policy-Based Network Management. Finally, we will introduce a benchmark networking system called OpenFlow for applying the DAIM model to enhance autononmic fuctions.**

***Index Terms*—Distributed Network Management, Information Model, Management Information Base, DAIM, Self-management, Artificial Intelligece.**

I. Introduction

COMPLEX electronic environment refers to a group of electronic devices connected together (wire or wireless networks) to share information and resources. Any network system needs a management protocol which performs different kinds of tasks such as operation (monitoring the performance of the network and detect any occurred problem as soon as possible), and maintenance (fix any occurred problem or in better word, always tries to keep network perform better). Most of the proposed network management protocols are based on the International Organization for Standardisation (ISO) definition for management model. Performance management, Configuration management, Accounting management, Fault management, and Security management are five characteristics of this definition[1]. Performance management refers to operation task by means of collecting and processing important management information. Configuration management refers to monitoring and controlling the effects of any device on network performance. This characteristic can be use to search any useful information when any problem occurs. Accounting management guarantees a fair usage of network resources by any user. Fault management refers to maintenance task which can detect the problem, isolate and fix it in the network. Security management controls the access level of all users to the resources of the network based on specific policies[2].

In this paper, we introduce the new information model, Distributed Active Information Model (DAIM) to allow the local decision making process, that will essentially contribute to complex distributed network environments. DAIM offers adaptation algorithms embedded with intelligent agents and information objects to be applied to such complex systems. By adopting the DAIM model and these adaptation algorithms, managing complex systems in any distributed network environment can become autonomous, adaptable, and scalable. This DAIM model can enhance objects to make their own local decisions through its active performance, and thus significantly reduce the workload of centralized decision-maker. In order to achieve the system's goal, a large amount of distributed objects in the DAIM model needs to be highly integrated.

Pakawat Pupatwibul, e-mail: pakawat.pupatwibul@uts.edu.au.
Bahram Jozi, e-mail: bahram.jozi@uts.edu.au.
Robin Braun, e-mail: robin.braun@uts.edu.au.

II. Needs for New Protocol

Advanced technologies have dramatically escalated over the past few decades, especially distributed networks, and play a significant role in providing management services for large and complex networks. Using human operator as manager is not economical and also error-prone. Moreover, as the complexity of distributed system grows over time, an effective computing environment is needed to ensure good quality of network services and performance. Currently, large-scale electronic systems like Wireless Sensor Networks (WSNs) or Bush fire alarm system are becoming more difficult to manage, configure, operate, maintain, and re-structure. It is important to propose a new OSS (Operations Support System) management structure to cope with such complex distributed networks (systems). In order to deal with this problem, IBM introduced the term Autonomic Computing in 2001, which can create its own strategies so it is able to constantly adapt itself with dynamic conditions of environment. Autonomic computing can also free networks manager from some management tasks especially low-level tasks, and in the meantime bringing better system behaviour. In this regard, each Autonomic Computing System (ACS) should have two main capabilities; adapting itself quickly to dynamic environment,

and it should have self-x management properties. The main purpose of the self-x management framework is to work at a high-level goal driven functions to deal with the increasing challenges from managing distributed network environments. It requires important capabilities of (1) gathering related information, (2) modifying the attributes of network nodes, and (3) managing its own functions in order to adapt itself to ever changing environment of the network, which is defined as network autonomy [3]. Scholars also believed that in order to establish a comprehensive Autonomic Communication, the systems should have a sustainable and maintainable information model, and have the ability to make local decisions when collecting information. The main properties proposed by IBM as the basics of AC are as follows [4]:

1) Self-optimization: system software and hardware should use resources maximally to provide optimized functioning and performance of communications, as well as to detect optimal behaviors in order to improve the systems' performance.
2) Self-healing: This means that the system should be able to automatically detects and recover from potential problem that might occur to the local software and hardware for example restarts or reboots a failed element.
3) Self-protection: This means that the system should be able to automatically detect and prevent from any malicious attacks within the network, or take the resources offline in case of severe threats. This property also includes the ability to maintain the systems' overall security.
4) Self-configuration: This means that the system should be able to automatically adapt to the ever-changing environment. Moreover, the system should automate configuration of components and systems where high-level objectives are defined.

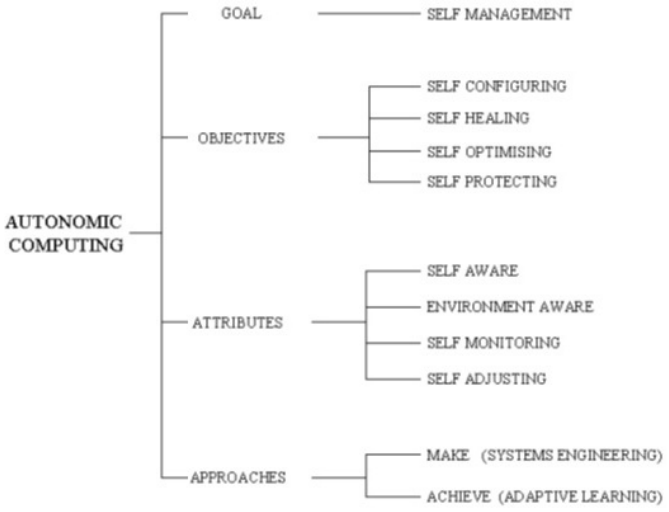


Figure 1. Autonomic Computing tree. [5]

Figure 1. Illustrates the general properties of autonomic systems. The objectives describe the broad system requirements, whereas the attributes identify the primary implementation mechanisms. In order to meet these objectives, the network system must be aware of its actual state (self-awareness) and the current external operating flows (environment-aware). If the conditions change, it will be detected via self-monitoring and adaptations are made consequently (self-adapting). In addition, this means that the network environment has some knowledge of its available resources, its components, its on-going status, and the status of communications with other systems[6].

In the autonomic research field, key researchers have defined many different attributes and properties for ACs. Each of the self-x literature on ACs is mainly categorised base on the working areas of the self-x functions itself. Since 2001, the list of self-x properties has grown dramatically. Currently, self-x framework includes functions such as self-definition, self-organisation, self-adjustment, self-monitoring, self-regulating and so on. For example, "Sabio" a program which classifies large number of documents automatically apply self-organisation and self-awareness [7]. IBM and other independent research centres have been recently proposed a model to measure the degree of these autonomic systems. Examples of distributed systems and applications of distributed computing include the following:

- Telecommunication networks.
- Telephone and cellular networks.
- Internet and other computer networks.
- Wireless sensor networks.
- Network applications.
- Peer-to-peer networks.
- Massively multiplayer online games and virtual reality communities. Assuming every player a node, autonomous computing can be mostly used in controlling interactions between these nodes.
- Real-time process control
- Industrial control systems. Assuming a industrial system as a autonomous system consisting of many sensors, control system, and actuators (e.g. robots), each actuator can be assumed as a node which can decide for itself in a way that the whole system reach a specific goal.
- Robots' control systems. Just like an industrial system, one can assume a robot as a system reaching a specific goal consisting of many nodes.
- Aircraft or train control systems. Assuming each train or airplane a node which can decide for itself, so trafficking control centre or network manager can be freed from so many tasks.

III. Distributed Active Information Model (DAIM)

In this section we will focus on active O:MIB, which is one of the main parts of the DAIM Model. An

object-oriented Management Information Base (O:MIB) is considered as a theoretical framework with the hope to replace the traditional Management Information Base (MIB), whereas hybrid O:XML is proposed as a practical way to implement O:MIB, and can be implemented with platform-independent JAVA agents (e.g. Jade and JadeX). However, the details of O:XML will not be mentioned in this paper.

DAIM model can be applied with distributed communication networks to enable autonomic functions. One of the most significant barriers when dealing with large-scale and complex distributed systems is insufficient centralized service management. Because the development of agent-based in the field of distributed artificial intelligent (DAI) has grown rapidly, autonomous decentralized systems (ADSs) and multi-agent technology are by far the best solution for complex network environments. The DAIM model consists of adaptation algorithms for adapting the intelligent agents and information objects to be applied to large-scale distributed electronic systems. The main purpose of this model is to re-engineer the structure of network information model, so that the new structure can effectively cope with the next generation communication networks. It also aims to redesign the traditional MIB structure in a way of object-oriented called object-oriented management information base (O:MIB) is required to fulfil management services such as configuration management, topology discovery, activating application process, and assigning resource process. Furthermore, O:MIB can play as a part of the distributed information model to enable autonomic software agents that act as the network elements (other routers, switches, hosts, etc). These autonomic agents (AAs) inherit the surrounding agent's behavior and also make local decisions based on the state of the network. The agents of distributed O:MIB technology will allow the richness of self-organized management. For example, dynamic software configurations, service activation, and service discovery. Furthermore, DAIM model is developed specifically with embedded smart algorithms for distributed elements to improve the efficiency of local execution abilities.

This O:MIB model is expected to be used in peer-to-peer networks, mobile technology, and wireless ad-hoc sensor networks (WASNs) as well as to address other complex issues. O:MIB adopts the object-oriented principles to manage the MIB objects. It has multiple distributed agents that remain in every network component and node, which functions with its own O:MIB as a way to activate applications when required. These network components can also analyse the important data, learning the systems environment, calculate situations, and perform adapting capability. Therefore, a full understanding of autonomic communication will be obtained. Object or element is the basic information unit of the O:MIB. Each important element comprises [8]:

- Attributes: It specifies the information values that represent the characteristics of the managed object identifiers (OIDs).
- Method behaviours: An action helping to achieve autonomic communications. This can includes the self-awareness function in real time and intensive and spatial data.
- Algorithms: These are the algorithms that will support a specific network task to be embedded into O:MIB domains. It also represents a set of predefines uses of the total available method calls. For example, humidity of the network environment, temperature monitoring, and predicting the level of raising alarms and risks in autonomic communication networks.
- Messages: In a response of the on-demand requests, local messaging daemon action can invoke messages in order to obtain general information like network topology or mapping discovery.

Figure 2, indicates the implementation process of O:MIB via O:XML. The software agents remain on each node having O:XML employed to populate the recorded data to the corresponding agents. In addition, JAVA agents are also involved because it is platform independent, and due to other agent development tools are mainly based on JAVA technology as well. Each agent is defined from instantiated agents according to their electronic environment. The O:MIB algorithms are invoked by the instantiated agents, whereas information values are reconfigured by java-based agents. Ultimately, the agent's lifecycle are accomplished while the program is operating.

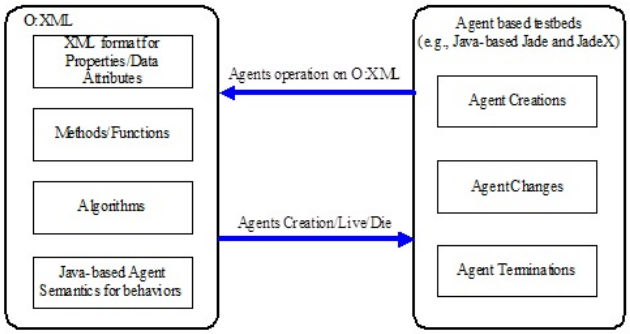


Figure 2. Self-maintained process. [8]

The overall stages of this approach can be described as the follows[8]:

1) Observing the current agents on distributed nodes and noticing the environments.
2) Generating new agents when a new environment is identified through the adaptation and learning strategies.
3) Functioning the local O:MIB by invoking the algorithms and methods instructed into the local O:MIB systems by default.
4) Adapting the node in regard to the awareness of the system-level objectives.
5) Rap up the agent's lifecycle until the next round of

process is ready.

This efficient O:MIB-based DAIM model approach is introduce to cope with managing autonomic communications in terms of self-configuring, self-adapting, self-optimizing, self-learning, self-awareness, and so on. This new information model scheme can also be applied into other self-x properties in ACNs. The attributes of each information object in O:MIB-based DAIM model can be implemented in one O:XML file. This brings the possibility of embedding DAIM agents into portable communication devices as well as applying into real networks in the future such as wireless networks, including WASNs, MANET, Peer-to-Peer networks, and Mesh networks.

IV. Network Management Protocols

Some of the important network management protocols are explained in this section. We proposed that the new DAIM model will outperform these conventional schemes.

A. SNMP

First protocol for network management was Simple Network Management Protocol (SNMP) proposed in 1988, which was easy to use and does not need complex management support, flexible and could be used in most of the devices. Moreover, it was the first protocol that was widely accepted and used for a long period of time. SNMP protocol consists of three parts: managed devices, agents, and network management system. Agents are software modules run on each managed device and can provide the management information of the device asked by manager through a communication interface called Management Information Base (MIB which is database holding important management information in each device). As this communication is in-band, if any problem occurs in the network, it is almost impossible to diagnose and recover the network without using external device. management information in devices are always different from MIB so a table called method routines is defined in SNMP in order to implements the access mechanism to management information on each device, another disadvantages of SNMP is the lack of standard definition of mentioned implementation. There are also other shortcomings as follows: Completely centralized and agents have rarely active role, Inferior scalability and inflexible, Cannot handle the massive increasing size and complexity of networks[2], [9].

B. NETCONF

The limitation of SNMP has lead to implementing alternative approaches to manage large-scale and complex network environment. One of the newly approved network management protocol proposed by IETF in December 2006 is the Network Configuration Protocol (NETCONF). It is a document-oriented approach based on XML technology that aims to address the weaknesses of SNMP, especially the application in configuration management. NETCONF protocol is regarded as the next generation of automated XML-Based network management system. This is because the communication between the NETCONF manager and the agents are formed in a XML document, and based on XML-encode Remote Procedure Call (XML-RPC) [10]. Moreover, NETCONF can also upload and retrieve configuration data of the network devices separately with high-level configuration operations. In order to assure the security of message transmissions, NETCONF adopts a transport independent protocol so-called Simple Object Access Protocol (SOAP) [11]. The NETCONF protocol brings many great advantages when compared to SNMP. For example, NETCONF provides more advanced functionalities, more effective transactions of complex configuration data, more secure and much easier to develop new applications than SNMP. Although NETCONF protocol is better than SNMP in some aspects for instance in configuration management, but there are also some important drawbacks associated with this approach as follows:

- One major issue related to NETCONF is the lack of support from industries and because of few publications regarding NETCONF implementation.
- New elements of NETCONF security aspects should be added, especially in Access Control. For example, using an XML-based access control standard, the eXtended Access Control Markup Language (XACML) as a good open source support[12].
- Needs new data model and new data modelling language as it plays a significant role in universality of NETCONF.

C. CIM

The standard of Common Information Model (CIM) has developed by DMTF, with the goal to produce an object-oriented scheme to model the hierarchical data of the managed IT environment. CIM is a conceptual view of the managed environment that attempts to unify and extend the existing traditional management standards such as SNMP using object-oriented constructs and design[13]. Moreover, CIM can provide a consistent definition and structure of data by presenting managed elements as a basic set of objects and relationship between objects. This standard includes the CIM infrastructure specification and the CIM schema. In regard to its infrastructure, the managed objects are described as class, and the relationship between them are represented by associations. In addition, CIM applies object-oriented concepts of inheritance to effectively define the common framework of managed objects and inherited sub-objects [14]. The values of object orientation techniques from CIM also provide support for object design with the following capabilities:

- Classification – High-level and fundamental concepts are defined when objects are grouped into types (class), identify common features and characteristics (properties), relationship (associations), and behavior (methods).
- Object inheritance – Sub-classing the high-level and fundamental objects. A sub-class inherits all the information (properties, associations, methods) defined for its higher level objects. Sub-classes are created to manage the same level of detail and complexity at the same level in the model.
- Ability to show dependencies, component and the connection or relationship between objects
- Standard inheritable methods – The capability to identify standard object behavior (methods) and encapsulate standard methods with an object's data Regarding the structure of data modelling, CIM scheme can provide greater representation of information than SNMP static MIBs. Conventional SNMP MIBs, on the other hand, have been used in IT industries for decades ago since the ISO layer model was introduced. Moreover, SNMP MIBs describe the information of managed objects from a different view in contrast with CIM models.

D. Policy-Based Network Management

Policy- based management framework was proposed by Internet Engineering Task Force (IETF) in 2001 [15]. Applying this method can make a network highly automated specially in terms of configuration; policies that are defined by administrator in this method will be used to configure all of existing or future devices of the network, and there would be no need for admin to configure any device itself, as a result the administrator's configuration task will be simplified and this task will be done automatically. There are four elements in this framework (Figure 3.) as follows:

- Policy repository which is using to save the policies made by administrator.
- Policy enforcement point which refers to a controlled device in a network.
- Policy decision point which is a communication element between enforcement point and repository, as these two elements can be placed in same or different devices; previous protocols such as SNMP can be used to make communication between these two elements.
- Policy management tool that enable the administrator to define his desired policies, in this element two levels of policies were defined, business level which is based on business needs and also is so close to the user language (does not include technical terms and specifications), and technology level policy which is the interpretation of business terms into a technical applicable policies for devices.

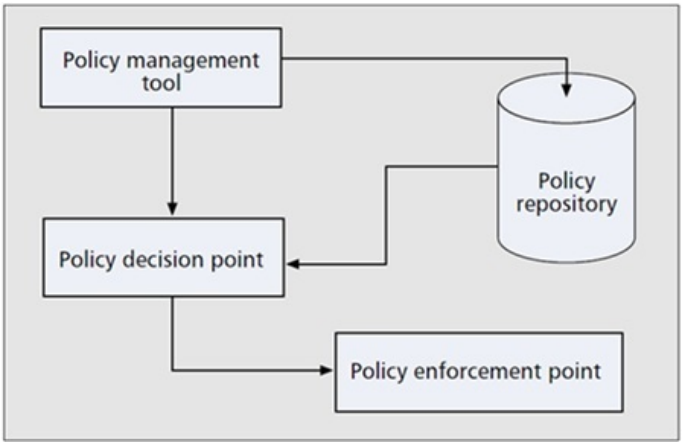


Figure 3. Four elements of policy based framework. [16]

Policy management tool consists of four basic elements as follows: user interface which is a tool that administrator can input its policies in a form of business level, resource discovery verifies the network topology, policy transformation logic which is most important section of this framework which guarantees reliability, suitability and practicability of the administrator's policies with a ability of the network, and finally policy distributor which distribute the policies to all devices[16].

Despite of so many benefits of this method such as being automatic and highly user friendly, the main disadvantage of this method is being centralized, as all policies are defined by central manager, so by increasing the complexity of networks, conflictions between policies will increase, and as a result, the framework performance will decrease or it will be almost unable to work.

V. OpenFlow System

OpenFlow is a system, developed by Stanford University in 2008, that allows researchers to operate experimental protocols in switches and routers in a uniform way, without having to expose the vendor's internal work of their products. OpenFlow is based on an Ethernet switch with the goal to encourage networking vendors to apply OpenFlow features to their product devices for deployment in university backbone networks and wiring closets. It also offers researchers to evaluate their ideas in real world traffic settings. The basic idea of OpenFlow is providing an open protocol to program the internal flow-table and standardized interface in different switches and routers to add and remove flow entries. In addition, researchers can control a portion and a flow of their local network by choosing the routes of the packets and also the processing they retrieve. By doing this, researchers can run experiments on new networking protocols, security testing, and addressing schemes without disrupting others who depend on the production traffic. However, the flow-tables are controlled by a remote centralized controller via the Secure Socket Layer (SSL) connection. This notion

has raised legitimate questions to ask on performance, scalability, and reliability of a centralized controller[17].

VI. Conclusion and Future Works

In this paper, we have described the needs of new network protocol in terms of self-x management properties and its capabilities to adapt itself with the ever changing environment. In this regard, an efficient object-oriented DAIM information model has been proposed with the hope to overcome the shortcomings of some traditional network protocols such as SNMP, NETCONF, and CIM. We also mentioned some of the previous network management protocols and their advantages and disadvantages. Finally, we introduced a benchmark networking system called OpenFlow, which enables researchers to test their protocols in real world networks without disturbing it. The future work will be employing and extending functionality of DAIM model into the OpenFlow system, focusing specially on self-healing and self-protection properties, as most of the recent researches focused on self-configuration property.

References

[1] "International organization for standardization, http://www.iso.org," tech. rep.

[2] H. Frenandez Rodriguez, "Active MIB, an object oriented solution for network management," Chalmers University of Technology, May 2007.

[3] F. Chiang, H. Fernandez, R. Braun, and J. Agbinya, "Integrating object-oriented O:XML semantics into autonomic decentralised functionalities," in *International Symposium on Communications and Information Technologies, 2007. ISCIT '07*, pp. 768–773, IEEE, Oct. 2007.

[4] F. Chiang, *Self-adaptability, Resilience and vulnerability on Autonomic Communications with Biology-inspired Strategies.* PhD thesis, University of Technology Sydney, Australia, 2008.

[5] R. Sterritt and D. Bustard, "Towards an autonomic computing environment," in *14th International Workshop on Database and Expert Systems Applications, 2003. Proceedings*, pp. 694– 698, IEEE, 2003.

[6] R. Sterritt and D. F. Bantz, "Personal autonomic computing reflex reactions and self-healing," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 36, pp. 304–314, May 2006.

[7] B. Otero, S. Sahuquillo, P. Barlet-Ros, S. Spadaro, and J. Sole-Pareta, "Self-* algorithms and autonomic communication system," *Internal publication of Technical university of Catalonia*.

[8] R. Braun and F. Chiang, "A distributed active information model enabling distributed autonomics in complex electronic enviornments," in *2008 Third International Conference on Broadband Communications, Information Technology & Biomedical Applications*, pp. 473–479, IEEE, Nov. 2008.

[9] J. Davin, J. D. Case, M. Fedor, and M. L. Schoffstall, "Simple network management protocol (SNMP)." http://tools.ietf.org/html/rfc1157, May 1990.

[10] J. Yu and I. Al Ajarmeh, "An empirical study of the NETCONF protocol," in *2010 Sixth International Conference on Networking and Services (ICNS)*, pp. 253–258, IEEE, Mar. 2010.

[11] Y. Chang, D. Xiao, H. Xu, and L. Chen, "Design and implementation of NETCONF-Based network management system," in *Second International Conference on Future Generation Communication and Networking, 2008. FGCN '08*, vol. 1, pp. 256–259, IEEE, Dec. 2008.

[12] B. Wu and Y. Chang, "Integrating SNMP agents and CLI with NETCONF-based network management systems," in *2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, vol. 1, pp. 81–84, IEEE, July 2010.

[13] "Cim concepts whitepaper, cim version 2.4+, available at http://www.dmtf.org/standards/cim. technical report, dmtf, 2003.,"

[14] F. Chiang and R. Braun, "Self-adaptability and vulnerability assessment of secure autonomic communication networks," in *Managing Next Generation Networks and Services* (S. Ata and C. S. Hong, eds.), vol. 4773, pp. 112–122, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[15] "The IETF policy framework working group: Charter available at http://www.ietf.org/html.charters/policy-charter.html," tech. rep.

[16] D. C. Verma, "Simplifying network administration using policy-based management," *IEEE Network*, vol. 16, pp. 20–26, Apr. 2002.

[17] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: enabling innovation in campus networks," *SIGCOMM Comput. Commun. Rev.*, vol. 38, p. 69âĂŞ74, Mar. 2008. ACM ID: 1355746.

A Study on the application of Data Mining Methods in the analysis of Transcripts

Luis Raunheitte*, Rubens de Camargo*, TAkato Kurihara*, Alan Heitokotter*, Juvenal J. Duarte*
*School of Computer and Informatics (FCI) – Universidade Presibiteriana Mackenzie – São Paulo – SP - Brasil

Abstract. Schools always had an essential role in the formation of students' intellect; however, the constant incorporation of knowledge to improve techniques and technologies used in the production of goods and services has caused a major demand for highly qualified professionals and, in order to meet that need, the teaching process must understand and adapt to the profile of the students. The transcript is the most used document to measure the performance of a student. Its digital storage combined with data mining methodologies can contribute not only to the analysis of performances, but also to the identification of significant information about student's profiles and deficiencies in the structure of a course. This study shows an example of the application of data mining techniques in transcripts, based on the real Computer Sciences course of Universidade Presbiteriana Mackenzie, and the use of the open source tool WEKA.

Introduction

One of the major challenges encountered by schools in several countries, mainly within the academic sphere, is to handle the difficulties of students during the learning process, which in many cases might result in the lack of motivation and even university drop-out. In this sense, better understanding the students and their characteristics is crucial for the application of pedagogical techniques with a specific focus, aiming at reaching optimum productivity during the learning process.

Since that up to recent times the storage of transcripts used to be made on paper and their analysis had to be manually processed, a deeper evaluation on these documents would be rejected and deemed unfeasible. However, the implementation of computer processing technology in several branches has been contributing to the increasing usage of digital data storage in view of the savings on material resources and space, as well as the ease of handling allowed by this format, which also foments environmental sustainability.

Data mining offers a method to benefit from the computer processing efficiency in the analysis of data with a view to search for tacit information [6]. Through data mining, it is possible to avoid massive data exploitation and to use the knowledge acquired by the interpretation of ascertained patterns.

The transcript represents a source of information that allows depicting not only the individual performance characteristics of students, but also their profiles, in addition to providing several details about the course at issue. Among the several potentials of data mining, it is possible to recognize distinct groups of students, distinguishing their difficulties and virtues; to evaluate academic disciplines that concentrate a greater need for efforts aiming at interlinking them; and to set forth right at the beginning the academic disciplines in which each student, in average, will face greatest challenges, as well as to establish secondary areas in the course through which the student will have more chances of achieving a successful career.

This article provides a practical example of the application of two data mining techniques in the analysis of transcripts: cluster analysis and classification. For that, there were used an open source tool named WEKA and a sample of actual data regarding the Computer Sciences course of Universidade Presbiteriana Mackenzie.

Knowledge Discovery In Database (KDD)

In the 80's, managers of major organizations started to be concerned about the volume of stored data and its uselessness for their purposes [1]. In tandem, as the market became increasingly dynamic and competitive, companies had to quickly make strategic decisions, even though this could imply potential risks [2]. Given this scenario, studies on data exploitation became more intense and, in 1989, the term Knowledge Discovery in Databases (KDD) was created by [4] with reference to the entire process of extraction of useful information from large data volumes [4]. Ever since, the entire KDD process arouses the interest of people both within the academic and the corporate spheres.

As opposed to data mining itself, KDD is a complex methodology with focus on the production of consistent and construable knowledge, comprising issues such as storage and access, development of efficient algorithms, visualization and interpretation of results and the way the user is able to interact with them [4].

The definition of KDD as a process composed of sequential steps is quite renowned and, in its essence, it comes down to the preparation of data, and mining and interpretation of the results.

The preparation is an elemental step for the success of a KDD project, insomuch that 80% of the efforts for understanding the data are focused on its cleansing and preparation [7]. There are three subjects that provide grounds for the importance of data preparation:

- Real world data are impure like noise and manifested as errors which, when provided as an input to a mining algorithm, tend to spend the results acquired from reality, hiding relevant patterns or even compromising the credibility of ascertained patterns.
- High performance mining applications require quality data.
- Quality data produce quality patterns.

It is crucial for the representation of the results acquired in the search for patterns to be simple and construable. [8] base this need on the fact humans have the ability to analyze large amounts of information when such information is presented in a visually organized manner. Among the distinct data analysis methods, the majority derives from the statistics, and the graphic analysis is more specifically related to the branch of Exploratory Data Analysis (EDA), whose focus is precisely on visualization.

Data mining is the step where artificial intelligence concepts are employed to provide resources for understanding the information being handled.

Data Mining

Data mining represents the most important step of the process of Knowledge Discovery in Database (KDD), as in this step the patterns are effectively exploited and analyzed. Once the data have been organized and their quality assured, mining is applicable with a view to extract their relevant patterns.

Under a simplified perspective, the goals of data mining might be resumed as prediction or description. When the purpose is prediction, the database is analyzed so that unknown values can be found or future ones predicted based on previously existing data. The most common examples of predictive methods are Classification, Regression and Analysis of Time Series, which operate through supervised learning techniques. As to description, the goal is to find construable patterns that describe the input data, based on classic methods such as Cluster Analysis, Discovery of Association Rules and Sequential Pattern Analysis, which use non-supervised learning techniques [4], [3].

Classification, as the name suggests, consists basically of mapping a log back to existing predefined categories. A powerful but simple classification technique consists of building decision trees, which is accomplished through a series of questions carefully developed and focused on the training set [8].

When certain classes cannot be previously identified in a clear fashion, clusters provide a good alternative by identifying items that will naturally group together [6]. The analysis of clusters aims at analyzing a set of logs and dividing it so that the values of the elements within one group are more similar inwardly than as compared to other groups [9]. Clusters oftentimes indicate tacit data patterns.The WEKA (Waikato Environment for Knowledge Analysis) is a widespread open source tool that consists in the creation of a compilation with machine learning and data pre-processing algorithms, comprising interfaces for the operation of input data, statistical validation of learning schemes and visualization tools.

The WEKA environment gathers the most important data mining methods: Regression, Classification, Cluster Analysis, Association Rules and Attribute Selection.

Preparation Of The Data Used

The data used in this research were acquired through the General Office of Universidade Presbiteriana Mackenzie, São Paulo campus (where information secrecy has been ensured due to the fact that a numeric indication is used for each student without any reference to the records used for the control of their academic life). The pieces of information correspond to a sample of the historic records of grades and absences of Computer Science students, from the Computer Systems and Data Processing Faculty, in the period between the first semester of 1999 and the second semester of 2010.

The records were found in an Excel spreadsheet, in the xlsx format, comprising the fields Student, Discipline Code, Name of the Academic discipline, Final Average, Absences, Year/Semester and Final Situation, as shown by Figure 1. The attributes Academic discipline Code and Name of the Academic discipline represent the discipline to which the grades are related, and Year/Semester is the attribute that indicates the period in which the student attended the classes. The final averages, absences and final situation are the target information and the patterns are analyzed according to these measures.

In addition to the spreadsheet, it was possible to acquire the description of the curricula on the page of the university and a definition of the theme areas of the academic disciplines from the course coordination.

The description of the 1999 curriculum acquired from the website of the university is the most complete one, comprising code, name, stage and the total, theory and laboratory credit hours for each of the academic disciplines in the curriculum. For the 2004 curriculum, the same information is available, except for the code of the academic disciplines, which could be acquired through the Informative Academic Terminal (known as TIA). For the 2009 curriculum, the codes of the academic disciplines are also not present; however it shows all other attributes and the indication of the new academic disciplines regarding the previous curriculum. The most relevant data are the code of the academic discipline and the stage, through which it is possible to associate a discipline to a curriculum and the level that this discipline is taught; this information is not available in the original data.

Studente | Dicip. Code | Discipline Name | Final average | Absence % | year attended | Final situation

Aluno	cód. disciplina	nome da disciplina	média final	% faltas	ano/sem cursado	situação final
10677106	11351055	LABORATORIO DE PROGRAMACAO	0,10	34,28	20102	R
10677106	11353074	ORGAN E ARQUIT DE COMPUTADORES	5,10	0,00	20101	A
10677106	11353082	ESTRUTURA DE DADOS	6,30	0,00	20092	A
10677106	11353139	ANALISE NUMERICA I	7,00	0,00	20101	A
10677106	11354100	LINGUAGEM DE PROGRAMACAO II	5,40	0,00	20092	A
10677106	11355026	TEORIA DOS GRAFOS	8,80	0,00	20091	A
10677106	11355166	BANCO DE DADOS	6,50	0,00	20091	A
10677106	11356057	REDES DE COMPUTADORES	5,20	0,00	20101	A
10703394	07051042	FISICA PARA COMPUTACAO I	8,50	5,55	20102	A
10703394	09251022	INGLES INSTRUMENTAL	8,60	12,50	20102	A
10703394	09351019	ETICA E CIDADANIA I	8,00	5,00	20102	A
10703394	10011811	CALCULO DIF. E INTEGRAL I	9,70	2,56	20102	A
10703394	10051090	GEOMETRIA E VETORES	7,00	0,00	20102	A
10703394	11351039	AMBIENTES OPERACIONAIS	8,20	7,69	20102	A
10703394	11351047	INTROD ALGORITMOS E PROGRAMACAO	8,80	2,77	20102	A
10703394	11351055	LABORATORIO DE PROGRAMACAO	5,70	6,25	20102	A
10705936	07051042	FISICA PARA COMPUTACAO I	6,10	8,33	20102	A
10705936	09251022	INGLES INSTRUMENTAL	5,60	6,25	20102	A
10705936	09351019	ETICA E CIDADANIA I	9,20	0,00	20102	A
10705936	10011811	CALCULO DIF. E INTEGRAL I	7,30	2,56	20102	A
10705936	10051090	GEOMETRIA E VETORES	6,50	0,00	20102	A
10705936	11351039	AMBIENTES OPERACIONAIS	6,20	5,12	20102	A

FIGURE 1 Part of the data is provided in its original format

The document regarding the theme areas of the academic disciplines is based on the 2009 curriculum, but since its academic disciplines differ just little as compared to the previous one, it is possible to generalize it. The Academic disciplines are arranged in Programming, Humanistic and Complementary courses, Mathematics, Computer Models and Systems, Technology, Software Engineering and Graphics Processing. It is important to remark that an academic discipline might be part of more than one theme line, such as the Graph Theory. This content also exposes some information relevant for the interpretation of results and not for the mining algorithm.

Before the tests, the data went through a cleansing and transformation process in which the duplicate lines were removed and the uncommon values treated, as the case of averages 22.22 and 99.99 attributed to certain records. Furthermore, the data had their format changed to a new structure as shown in Figure 2.

ALUNO	ANO-SEM	id10054057-CALCULO DIF E INTEGRAL IV	id10054073-MATEMATICA DISCRETA
191333	20071	1.6	
191333	20072	5.5	
191333	20081		
191333	20082		
191333	20091		
191333	20092		
191333	20101		
191333	20102		
210060	20051		
210060	20052		
210060	20061		5.5
210060	20062		
210060	20071	2.9	
210060	20072	4	
210060	20081	5.5	
210060	20082		
210060	20091		
210060	20092		
210060	20101		
210060	20102		
210637	20051		
210637	20052		
210637	20061		7.3

FIGURE 2. Part of the data is provided in its original format

In this new structure, the absences and final situation were removed and the averages started to be arranged in columns, sorted by academic discipline. This data aggregation was loaded on a Sun MySQL database to be more efficiently exploited.

When attempting to run cluster analysis experiments with these data, by using WEKA's K-Means algorithm, it was noted that the field Year/Semester rendered the results unsatisfactory. The data were then clustered by the field 'Student' so that each student started to have only one data line and the grade of each academic discipline started to be the average of the acquired grades, where the historic representation of the data ceased to exist. For example, the student mentioned previously in the green rectangle of Figure 2, who has attended the course Differential and Integral Calculation IV twice, and had averages 1.5 and 5.5 respectively, starts to have a single record comprising the average 3 for this academic discipline. This aggregation bears the exploitation of patterns that reflect the profile of the students, because each student is represented by a single line, which, on the other hand, is considered an aspect to be analyzed by the K-Means algorithm.

Tests

As soon as the data were loaded on WEKA it was possible to observe certain statistical measures related to the data in the new aggregation. The set comprises 184 students, but the grades are mostly focused on the academic disciplines taught in the beginning of the course.

On WEKA's Cluster tab, the Simple K-Means algorithm has been selected, which offers the classic implementation of the K-Means technique, with the parameters preserved in their default value.

One of the parameters of the algorithm is the number of clusters in which the data must be segmented. This parameter, normally referenced as K, has been altered between the values 2 and 10 with a view to find out the number of clusters that best suits the data after the analysis of the results acquired for each of the K values.

When running the algorithm with the K value set to 2, an enhanced performance of one of the groups could be observed, which is shown as blue dots in Figure 3.

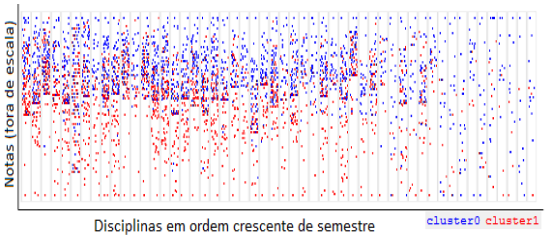


FIGURE 3. Part of the data is provided in its original format

Note that cluster 0 integrates most of the dots to the right of the horizontal axis of Figure 3, where the academic disciplines related to the end of the course are located. It can also be observed that, as semesters move forward, the students of both clusters show a tendency to standardize their performances, which is evidenced by the mixture of different colors at the end of the horizontal axis.

When the K value is gradually increased, these patterns are preserved and, beginning with K equals 5, the groups start to present extra peculiarities.

During the analysis in search of 7 clusters, two interesting points stood out, one regards the difficulties, and the other, the aptitudes. The students with difficulties, arranged in clusters 3 and 6, have demonstrated it especially in Mathematics and Programming, as shown in Figure 4.

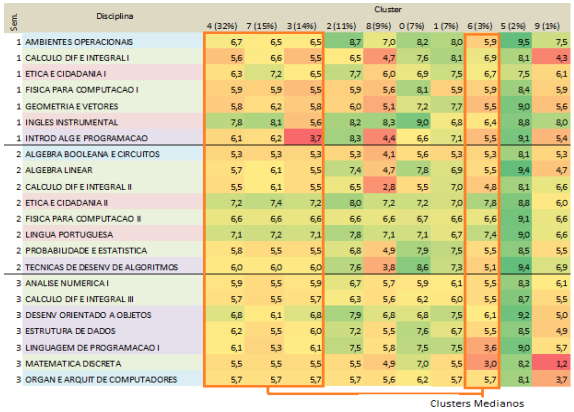
Sem.	Disciplina	Qtde	μ	σ	Cluster 4	3	2	6	1	0	5
1	AMBIENTES OPERACIONAIS	183	7,1	1,341	6,6	6,2	8,2	7,2	7,3	8,6	9,7
1	CALCULO DIF E INTEGRAL I	184	6,2	1,157	6,0	4,9	7,0	5,4	7,3	7,5	7,4
1	ETICA E CIDADANIA I	184	6,8	1,274	6,8	6,4	7,4	6,6	7,0	6,7	8,1
1	FISICA PARA COMPUTACAO I	106	5,9	1,953	5,8	4,7	6,1	5,8	6,2	8,1	8,3
1	GEOMETRIA E VETORES	138	5,8	1,802	6,0	4,2	6,6	4,8	6,6	7,3	8,0
1	INGLES INSTRUMENTAL	183	7,7	1,237	8,0	6,2	8,5	8,0	6,6	8,7	7,5
1	INTROD ALG E PROGRAMACAO	151	6,1	2,073	5,9	3,7	7,6	6,2	7,1	7,5	8,0
2	ALGEBRA BOOLEANA E CIRCUITOS	58	5,3	2,001	5,2	5,0	5,5	4,5	5,5	6,8	7,2
2	ALGEBRA LINEAR	181	6,0	1,612	6,1	4,5	6,9	5,0	6,4	8,3	8,6
2	CALCULO DIF E INTEGRAL II	143	5,5	1,975	5,6	4,0	6,5	4,5	6,3	6,8	7,3
2	ETICA E CIDADANIA II	128	7,2	1,408	6,9	6,8	7,6	7,2	7,4	7,7	8,6
2	FISICA PARA COMPUTACAO II	49	6,6	1,854	6,4	6,5	6,6	6,2	6,6	7,6	8,8
2	LINGUA PORTUGUESA	134	7,1	1,303	7,0	6,5	7,6	7,0	6,7	7,3	8,9
2	PROBABILIDADE E ESTATISTICA	179	6,0	1,624	5,9	4,6	6,8	4,4	6,8	8,0	7,7
2	TECNICAS DE DESENV DE ALGORITMOS	175	6,0	1,901	5,3	5,2	6,9	5,1	7,1	8,6	8,4
3	ANALISE NUMERICA I	131	5,9	1,393	5,8	5,6	6,7	4,5	5,9	6,3	7,6
3	CALCULO DIF E INTEGRAL III	134	5,7	1,379	5,6	5,4	6,5	4,1	5,9	6,8	7,7
3	DESENV ORIENTADO A OBJETOS	127	6,8	1,261	6,6	6,7	7,8	5,5	6,8	7,3	8,5
3	ESTRUTURA DE DADOS	159	6,2	1,536	6,1	5,3	7,2	4,7	6,4	7,7	7,7
3	LINGUAGEM DE PROGRAMACAO I	159	6,1	1,628	5,7	5,7	7,4	4,5	6,7	7,7	7,2
3	MATEMATICA DISCRETA	91	5,5	1,710	5,3	5,2	5,8	4,3	5,6	7,2	7,2
3	ORGAN E ARQUIT DE COMPUTADORES	92	5,7	1,553	5,6	5,8	5,9	4,9	5,7	6,9	7,2
4	ANALISE DE ALGORITMOS	57	5,5	1,754	5,3	5,4	5,5	4,6	5,7	6,7	7,0
4	ANALISE NUMERICA II	123	6,4	1,228	6,1	6,3	6,9	5,5	6,8	7,0	7,5
4	CALCULO DIF E INTEGRAL IV	129	5,3	1,258	5,2	5,1	6,0	4,2	5,6	6,1	7,0
4	INTROD A ENG DE SOFTWARE	141	6,4	1,155	6,3	6,0	6,9	5,5	6,7	6,9	8,0
4	INTRODUCAO A COMPUTACAO GRAFICA	53	6,1	1,845	6,0	5,8	6,1	5,5	6,3	7,0	8,7
4	LINGUAGEM DE PROGRAMACAO II	123	5,9	1,482	5,5	5,7	6,5	5,0	6,8	6,6	7,1
4	PROGRAMACAO MATEMATICA	49	5,8	2,011	5,7	5,8	5,9	5,1	5,9	6,5	7,7

Áreas Temáticas
Programação
Matemática
Tecnológica
Eng. de Software
Proc. Gráfico
Modelos e Sist. Comp.
Humanística e Compl.

FIGURE 4. Results acquired with the analysis of seven clusters.

As to aptitude, it is possible to note, for instance, that among groups 4, 2 and 1, interpreted as average performance, the best performance in instrumental English was acquired by clusters 4 and 2.

By replacing the distance function of the K-means algorithm from Euclidian to Manhattan, using the value 10 as parameter K, the results indicated differences that were less uniform and whose characteristics were more distinct among the clusters (Figure 5). In the average group, cluster 3 shows an outstanding difficulty in the academic discipline Algorithms and Programming Basics, while the logs of cluster 8, interpreted as having difficulties, show its aptitude in dealing with the English language.



Sem.	Disciplina	Cluster 4 (32%)	7 (15%)	3 (14%)	2 (11%)	8 (9%)	0 (7%)	1 (7%)	6 (3%)	5 (2%)	9 (1%)
1	AMBIENTES OPERACIONAIS	6,7	6,5	6,5	8,7	7,0	8,2	8,0	5,9	9,5	7,5
1	CALCULO DIF E INTEGRAL I	5,6	6,6	5,5	6,5	4,7	7,6	8,1	6,9	8,1	4,3
1	ETICA E CIDADANIA I	6,3	7,2	6,5	7,7	6,0	6,9	7,5	6,7	7,5	6,1
1	FISICA PARA COMPUTACAO I	5,9	5,9	5,5	5,9	5,6	8,1	5,9	5,9	8,4	5,9
1	GEOMETRIA E VETORES	5,8	6,2	5,8	6,0	5,1	7,2	7,7	5,5	9,0	5,6
1	INGLES INSTRUMENTAL	7,8	8,1	5,6	8,2	8,3	9,0	6,8	6,4	8,8	8,0
1	INTROD ALG E PROGRAMACAO	6,1	6,2	3,7	8,3	4,4	6,6	7,1	5,5	9,1	5,4
2	ALGEBRA BOOLEANA E CIRCUITOS	5,3	5,3	5,3	5,3	4,1	5,6	5,3	5,3	8,1	5,3
2	ALGEBRA LINEAR	5,7	6,1	5,5	7,4	4,7	7,8	6,9	5,5	9,4	4,7
2	CALCULO DIF E INTEGRAL II	5,5	6,1	5,5	6,5	2,8	5,5	7,0	4,8	8,1	6,6
2	ETICA E CIDADANIA II	7,2	7,4	7,2	8,0	7,2	7,2	7,0	7,8	8,8	6,0
2	FISICA PARA COMPUTACAO II	6,6	6,6	6,6	6,6	6,6	6,7	6,6	6,6	9,1	6,6
2	LINGUA PORTUGUESA	7,1	7,2	7,1	7,8	7,1	7,1	6,7	7,4	9,0	6,6
2	PROBABILIDADE E ESTATISTICA	5,8	5,5	5,5	6,8	4,9	7,9	7,5	5,5	8,5	5,5
2	TECNICAS DE DESENV DE ALGORITMOS	6,0	6,0	6,0	7,6	3,8	8,6	7,3	5,1	9,4	6,9
3	ANALISE NUMERICA I	5,9	5,5	5,9	6,7	5,7	5,9	6,1	5,5	8,3	6,1
3	CALCULO DIF E INTEGRAL III	5,7	5,5	5,7	6,3	5,6	6,2	6,0	5,5	8,7	5,5
3	DESENV ORIENTADO A OBJETOS	6,8	6,1	6,8	7,9	6,8	6,8	7,5	6,1	9,2	5,0
3	ESTRUTURA DE DADOS	6,2	5,5	6,0	7,2	5,5	7,6	6,7	5,5	8,5	4,9
3	LINGUAGEM DE PROGRAMACAO I	6,1	5,3	6,1	7,5	5,8	7,5	7,5	3,6	9,0	5,7
3	MATEMATICA DISCRETA	5,5	5,5	5,5	5,5	4,9	7,0	5,5	3,0	8,2	1,2
3	ORGAN E ARQUIT DE COMPUTADORES	5,7	5,7	5,7	5,7	5,6	6,2	5,7	5,7	8,1	3,7

Clusters Medianos

FIGURE 5. Results for the Manhattan distance from the analysis of ten clusters.

The instances exported by the analysis of clusters with Manhattan distance were used for classification, and gained an extra attribute in which the cluster associated to the instance is defined.

The instances that belong to clusters 4 and 7 were grouped under the label Average1, and cluster 3 was renamed to Average2. Clusters 0, 1 and 2 became WithAptitude1, WithAptitude2 and WithAptitude3. Lastly, cluster 8 was once again labeled WithDifficulty.

When running WEKA's J48 algorithm, which is an implementation of technique C4.5, the tree, which is textually shown below, was built: This tree presents a classification accuracy of 60%, however most of its inaccuracy is found under the labels WithAptitude.

```
Instrumental English <= 6.8
| Ethics and Citizenship I <= 8.1
| | Differential and Integral Calculation. I <= 6.2
| | | Instrumental English  <= 6.3: AVARAGE2
| | | Instrumental English > 6.3
| | | | Differential and Integral Calculation. I <= 5.8: AVARAGE1
| | | | Differential and Integral Calculation. I > 5.8: AVARAGE2
| | Differential and Integral Calculation. I > 6.2
| | | Ethics and Citizenship I <= 6.6:AVARAGE2
| | | Ethics and Citizenship I > 6.6: WITHAPTITUDE2
| Ethics and Citizenship I > 8.1: AVARAGE1
Instrumental English > 6.8
| Differential and Integral Calculation. I <= 6.1
| | Operational Environments <= 8.7
| | | Algorithms and Programming Basics <= 4.8: WITHDIFFICULTY
| | | Algorithms and Programming Basics > 4.8: AVERAGE1
| | Operational Environments > 8.7
| | | Ethics and Citizenship I <= 6.7
| | | | Differential and Integral Calculation. I <= 5.6: WITHDIFFICULTY
| | | | Differential and Integral Calculation. I > 5.6: AVARAGE1
| | | Ethics and Citizenship I > 6.7: WITHDIFFICULTY3
| Differential and Integral Calculation. I > 6.1
| | Programming Language I <= 5.75: AVERAGE1
| | Programming Language I > 5.75
| | | Probability and Statistics <= 6.9
| | | | Object-oriented Drawing <= 7.1
| | | | | Numerical Analysis <= 6.6: AVERAGE1
| | | | | Numerical Analysis > 6.6: WITHDIFFICULTY1
| | | | Object-oriented Drawing > 7.1: WITHDIFFICULTY3
| | | Probability and Statistics > 6.9
| | | | Linear Algebra <= 8.7
| | | | Linear Algebra <= 5.9: AVERAGE1
| | | | Linear Algebra > 5.9
| | | | | Instrumental English <= 8.3
| | | | | | Differential and Integral Calculation I <= 7.5: WITHDIFFICULTY1
| | | | | | Differential and Integral Calculation I > 7.5: WITHDIFFICULTY2
| | | | | Instrumental English > 8.3: WITHDIFFICULTY1
| | | Linear Algebra > 8.7: WITHDIFFICULTY3
```

Although its performance was not quite satisfactory for instances of the WithAptitude class, with 83% accuracy this tree has defined that Average2 is under the academic discipline "Instrumental English <= 6.8" and "Ethics and Citizenship I <= 8.1". Moreover, the instances of the class WithDifficulty have been placed in the academic discipline "Instrumental English > 6.8" and "Differential and Integral Calculation I <= 6.1".

Conclusion

This article highlights some data mining processes, from data procurement to the use of WECA, and the interpretation of patterns acquired through the use of algorithms.The test allowed the observation of stages such as the preparation of stored data and the application of the classification technique. These stages might be useful as a foundation for new tests and improvements, allowing its application in other courses and schools. The acquired results indicate the feasibility of data mining in transcripts, stressing that several patterns appear quantitatively during the day-by-day of courses, allowing the application of algorithms and the identification of trends, not always so evident, regarding the characteristics of the students and the courses. The method presented here might function as a base to the strategic development of new courses, if used as a tool to aid the learning process.

REFERENCES

1. Amo, S. D. (2004) "Técnicas de Mineração de Dados", In: Jornada de Atualização em Informática, Salvador: Sociedade Brasileira de Computação.
2. Bispo, C. A. F. (1998) "Uma Análise da Nova Geração de Sistemas de Apoio à Decisão", 174 f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia da USP São Carlos.
3. Cherkassky, V. and Mulier, F. M. (2007) "Learning from Data: Concepts, Theory, and Methods", Hoboken, NJ, USA: Wiley, 2th edition.
4. Fayyad, U. M. and Piatetsky-Shapiro, G. and Smyth, P. and Uthurusamy, R. (1996) "Advances in Knowledge Discovery and Data Mining", The MIT Press.
5. Pang-Ning, T. and Steinbach , M. and Kumar, V. (2005) "Classification: Basic Concepts, Decision Trees and Model Evaluation", In:_____. Introduction to Data Mining, cap. 4, p. 106-152, Addison Wesley, 1th edition.
6. Witten, I. H. and Frank, E. (2005) "Data Mining: Practical Machine Learning Tools and Techniques", San Francisco, CA, USA: Elsevier, 2th edition.
7. Zhang, S. and Zhang, C. and Yang, Q. (2003) "Data Preparation for Data Mining", Applied Artificial Intelligence, v. 17, p. 375-381.
8. Tan, P.-N.; Steinbach, M.; Kumar, V. Exploring data. In: . Introduction to Data Mining. 1. ed.

[S.l.]: Addison Wesley, 2005. cap. 3, p. 97_144. ISBN 0321321367.

9. Wu, X.; Kumar, V. The Top Ten Algorithms in Data Mining. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2009. ISBN 1420089641, 9781420089646.

Moving US SMB Customers to the "Sweetspot" using Predictive Analytics

Paromita Sen
Global Marketing Customer Intelligence, Hewlett Packard, Bagmane Tech Park
Bangalore, Karnataka, India

and

Sweta Agrawal
Global Marketing Customer Intelligence, Hewlett Packard, Bagmane Tech Park
Bangalore, Karnataka, India

and

Peter J. Jaumann
Global Marketing Customer Intelligence, Hewlett Packard, 10196 Flower Court
Westminster, Colorado, USA

ABSTRACT

The US SMB customer base shows wide variations in buying patterns and relationship with HP. There is a need to identify the most valuable customers, understand their profile, and develop strategies to influence "other" customers. HP is divided into 3 major business units – **PSG (Personal Systems Group - primarily laptops desktops and workstations), IPG (Ink and Printing Group – primarily printers/copier and ink) & EB (Enterprise Business – primarily servers).** Our SMB customers primarily purchase from only one or two of the business units. The ideal strategy for HP is to encourage customers to purchase across our full product portfolio – PSG, IPG & EB. The Customer Intelligence (CI) SMB COE team discovered that the most valuable customers are those who are purchasing from all three lines of business – what we define as the "sweetspot" and has developed a modeling framework using data mining and analytical techniques to predict the SMB accounts that can be converted to the "sweetspot" with the right marketing/sales activities. Recommendations for likely product purchase of the "sweetspots" are also identified in this paper.

Keywords:
Data Mining, Database Marketing, Logistic Regression, Cross – sell, Predictive Analytics

INTRODUCTION

Both proprietary and industry research indicates that there is a rising demand for products and services and increased spending by the SMB segment providing enough compelling evidence to target this segment for future business growth.
Figure 1 shows the current view of the SMB Customer Lifecycle framework:

Figure 1: SMB Customer Lifecycle Framework

Driving retention, acquisition and development simultaneously is always a challenge. Looking at retention and development by mining the existing base of customers is a fairly easy and cost effective method to drive top line growth, especially as new customer acquisition typically requires far more resources. However, as competition in IT sector becomes more intense, both the importance of and challenge in maintaining customer loyalty also increases. This challenge is more pronounced for small and medium business customers, who typically receive limited budget allocations for 1-1 targeted relationship programs. Without adequate knowledge on the customer's buying behavior, it then becomes very difficult to provide a customized solution for this segment.

In order to drive retention and development, we have used HP's large existing SMB customer base to explore ways to generate an increasingly larger share of revenues and product sales rather than relying on growth from new customers as was traditionally done in the past. The US SMB active existing base consists of ~358K accounts. These businesses have varying sizes, ranging from 10 – 999 employees, and varying IT needs.

CONCEPT OF "SWEETSPOT"

In order to develop a robust and scalable solution in the area of development and retention, we quantified a metric called "sweetspot" to gauge the relationship of a business with HP. A "sweetspot" is defined as a customer who has purchased products from all three business units of HP - EB (Enterprise Business – primarily servers), PSG (Personal Systems Group - primarily laptops desktops and workstations) and IPG (Ink and Printing Group – primarily printers/copier and ink) in the last 5 years (2006Q3 - 2011Q2). Figure 2 illustrates the profiles of these "sweetspot" customers.

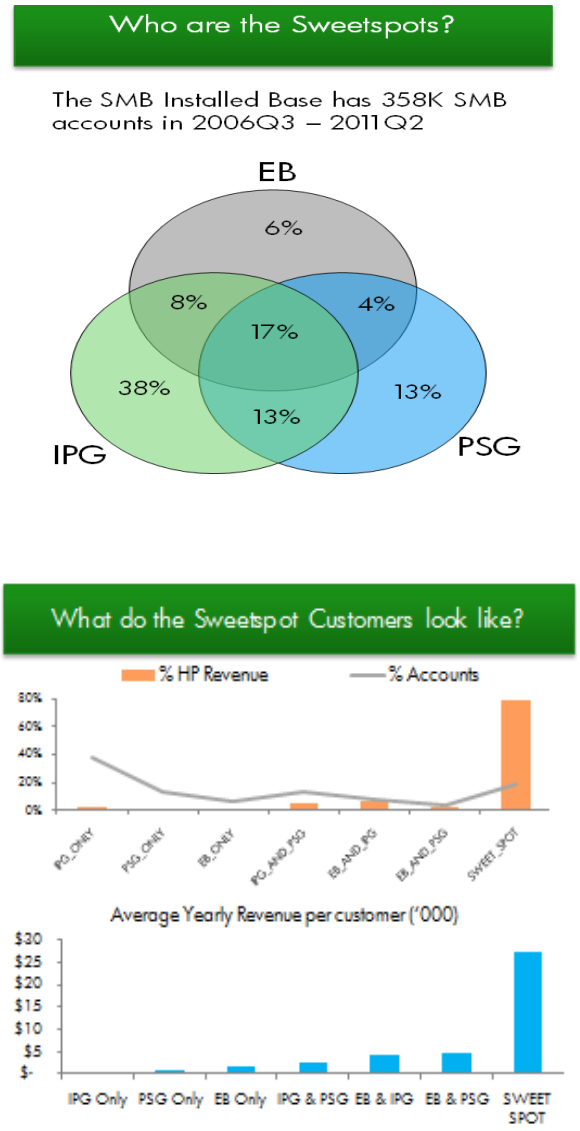


Figure 2: Profile of "sweetspot" customers (2006Q3 - 2011Q2)

As depicted in Figure 2, the "sweetspot" accounts comprise of 17% of the SMB existing customer base, but has generated as high as 79% of US SMB revenue in the last 5 years. The average revenue generated from the "sweetspot" customers is also much higher compared to the other customers. Hence conversion of customers (who buy from only 1 or 2 business units) into "sweetspots", can generate a considerable amount of incremental revenue.

Utilizing the concept of "sweetspot", we helped a specific IPG business to improve their relationship with HP through increased IPG purchases. We developed a statistical model to identify potential "IPG sweetspots". The IPG business wanted to identify and understand the profiles of the accounts, which have a high potential of improving their purchase portfolio with HP through increased IPG purchases. The idea was to understand their IT requirements and accordingly promote the right products, at the right time, with the right campaign. This would not only keep the business focused and engaged with the top customers, but also help the business augment sales by grabbing up-sell and cross-sell opportunities.

LITERATURE REVIEW

The traditional approach to identify the business potential in this case would be to profile the non-IPG SMB customers, based on their purchase history in terms of time and composition of purchase, revenue generated, and typical firmographic statistics, in order to understand their key characteristics using which marketing strategies can be devised. For example the approach utilized by Ron Kahan in "Using database marketing techniques to enhance your one-to-one marketing initiatives" - this paper describes the RFM Analysis framework (recency, frequency, and monetary value) in consumer behavioral analysis. "Strategic database marketing" by Arthur Middleton Hughes suggests the use of RFM when marketing a new or a different product to an already existing customer. However this provides a descriptive picture only and is not a forward looking approach.

As an enhancement, response models based on look-a-like approach are usually employed, which can be used to model the behavior of a typical IPG SMB customer, based on the assumption that accounts exhibiting a behavior similar to current IPG customers are also likely to become IPG customers in the future. Traditional segmentation techniques, like CART and CHAID, can also be used to create customer segments with a high IPG purchase rate. For example, "The use of the new ordinal algorithm in CHAID to target profitable segments" by Magidson, J suggests the use of CHAID to target pre-determined customer segments. However, both CART and CHAID are computationally very intensive and hence difficult to execute.

The next step after identifying the potential customers is to identify the products that they are likely to purchase. In order to recommend the next most likely product for any identified potential customers Sequential Market Basket Analysis is typically used for many businesses.

THE "SWEETSPOT" SOLUTION

Our solution improvises on the commonly used look-a-like response models by looking at converting accounts; we look at only those customers who have been buying EB &PSG products in the past but converted to "sweetspot" by buying IPG products in the last 6 months, and then we look for other customers that look similar to these converted customers using advanced statistical technique. This ensures that the identified IPG prospects have a higher likelihood of conversion when targeted with the appropriate marketing campaign. Logistic regression also differs from both RFM and CHAID in one important way. Logistic regression provides a response probability for individual members of the dataset rather than creating discreet groups of people as in RFM and CHAID. This response probability is important for the present purpose. Each person in the dataset will have a different response probability. Moreover, the RFM variables are incorporated in the logistic model. Use of logistic regression is also more robust and accurate, and easy to compute and validate.

In order to recommend the next most likely product for the "sweetspots" we have utilized a probabilistic approach instead of using the association based Market Basket Analysis. The technique is more robust, accurate, and provides a more synergistic approach, even though it might get somewhat challenging to compute and validate.

Our solution uses a three-step approach –
(1) Building the Conversion Model to identify key drivers of "IPG sweetspot"
(2) Identifying Likely "IPG Sweetspots"
(3) Recommend Most Likely IPG Product Purchase

(1) Building the Conversion Model to identify key drivers of "sweetspot"

The accounts which have purchased HP non-IPG products from only 1 or only 2 of the 3 business units in the observation period (2006Q3 – 2010Q4) are taken into consideration for building the model. Out of these accounts, the target variable is defined as follows:
"sweetspot" = 1 if the account moves towards a "sweetspot" with an IPG purchase in the conversion period of next 6 months, 0 otherwise.

The conversion rate is approximately 11% for our data. For statistical modeling purposes, a 10% random sample of the overall 5 years SMB space is used, split into Development (70%) and Validation (30%) samples. Variables considered for this analysis includes key customer behavioral characteristics such as total HP revenue, product portfolio, sophistication of IT needs, purchase channel, recency, frequency of purchase, price elasticity, click-stream behavior, IT spend, brand preference, and firmographics (e.g., D&B data like number of employees and industry segment). To ensure that the revenue amount does not directly influence "sweetspots", we created various derived variables which were used as predictors in the model. Some of these variables are as follows - Percentage of HP revenue from top 5 products, Maximum Share of Wallet among the 3 Business Units Percentage change in HP revenue from 1st to 5th year.

A stepwise logistic regression model is built on these accounts. We started with more than 70 variables at the customer level. After multiple iterations of data profiling, data completeness check, outlier detection, correlation analysis, and review of logistic regression outputs, we ended up with 9 variables which best describe the key characteristics that drive customers to convert to "sweetspots" with IPG purchases.

The model shows that mid-sized businesses from the financial industry are more likely to convert to "sweetspots" with IPG purchases. Customers with a long HP relationship and recent purchases also show promise of conversion. Other variables such as IPG TAM (total addressable market dollar value), PSG revenue in the current year, and percentage of EB revenue also emerge as important drivers. We classified the drivers into 4 types depending on their relative impact on the "sweetspots" and percent of customers who display the significant characteristics. Figure 3 illustrates this classification:

a. **Developmental Drivers:** Characteristics having high influence on "sweetspots" but fewer customers in the SMB space having these characteristics. For example, customers from the Financial Industry are likely to move to "sweetspots" with IPG purchases but only 4% of the non-IPG customers are currently from the Financial Industry.

b. **Maintain Drivers:** Most important characteristics lie in this block with high influence on "sweetspots" and higher number of customers in the SMB space having these

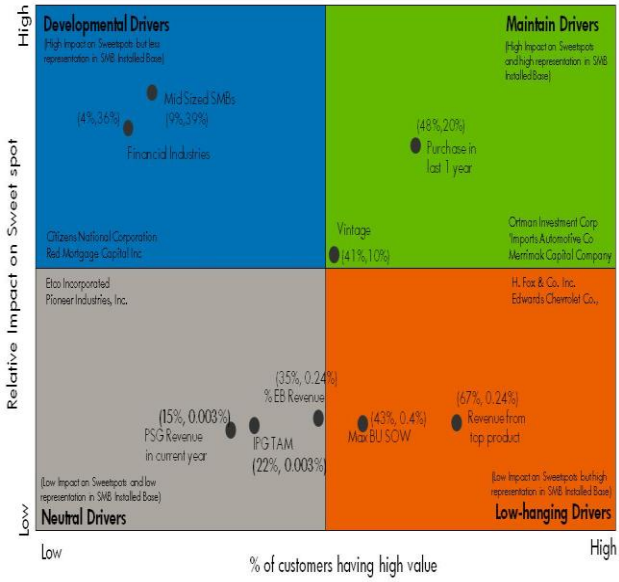


Figure 3: A four-quadrant classification of key "IPG-sweetspot"drivers

characteristics. For example, customers who purchased in the last 1 year are likely to move to "sweetspots" with IPG purchases and 48% of the non-IPG customers are currently in that category.

c. **Low Hanging Drivers:** Zone with characteristics having moderate influence on "sweetspots" but higher number of customers in the SMB space having these characteristics. For example, customers who have high revenue from top products are moderately likely to move to "sweetspots" with IPG purchases but a high percentage of 67% of the non-IPG customers are currently in this category.

d. **Neutral Drivers:** Low importance zone with characteristics having moderate influence on "sweetspots" as well as fewer customers in the SMB space having these characteristics. For example, customers who have high IPG TAM (total addressable market dollar value) are moderately likely to move to "sweetspots" with IPG purchases and lower percentage of 22% of the non-IPG customers are currently in that category.

This classification demonstrates that Maintain Drivers are the most important drivers which would drive customers to "IPG sweetspots" followed by Developmental and Low - hanging drivers.

(2) Identifying Likely "IPG sweetspots"

In the next step, we mapped the significant characteristics of potential "IPG sweetspots" onto the entire population to identify the potential IPG prospects. In other words, we scored the entire existing SMB customer base using the statistical model scores for each of the significant drivers through a logistic regression equation. We identified a list of 7K customers with top scores who are likely to move towards "sweetspot" with IPG purchases. A final prioritized list of approximately 5.6K was prepared comprising of top medium and large accounts, along with their last 3-year revenue figures and share of PSG, EB buys was also provided as a quick reference for the sales personnel.

(3) Recommend Most Likely IPG Purchase

Having identified the likely "IPG sweetspot" prospects which the US IPG business could target, the next logical step was to recommend which IPG products they are most likely to buy.

The "Next Most Likely Product (NMLP) Analysis" utilizes a conditional probability framework to recommend the top IPG products that US SMB customers are likely to buy next, based on product purchase history of the entire existing customer base.

To predict the product most likely to be purchased next, we considered the Markovian property – given the present, the future does not depend on the past. We say a sequence of random variables (X_t) is a Markov Chain if for all t, all x_0, x_1,…, x_t ,

$P\ (X_{t+1} = y/X_0 = x_0,\ X_1 = x_1,\ \dots,\ X_t = x_{t)} = P\ (X_{t+1} = y/X_t = x_t)$……………………………………………………………….(1)

Based on historical transactions, transition probabilities P (Y|X) or pij's were computed for every combination of (i,j). We built 2 transition matrices, one for one time buyer (zero order Markov Chain) and another for repeat buyers (first order Markov Chain). The transition matrix calculates the probability that a customer will buy product category (Y) given the customer has bought another product category (X) in the past.

According to the analysis, Mono laser printers emerged as the top selling next most likely product for PSG and EB buyers. Typically, PSG buyers don't buy supplies (ink/cartridges). Ink cross-selling opportunities are identified among EB buyers

VALIDATION

A pre-campaign validation exercise was undertaken by splitting the data into modeling and validation. A number of iterations were run on the data, across different time periods, to understand the robustness of the model. The model was created with an expectation to deliver ~60% correctly predicted customers. It also gave a lift of 26% from any randomly chosen list of customers. As seen in Figure 4, the top 4 deciles of the scored customers capture 64% of the total "sweetspots"

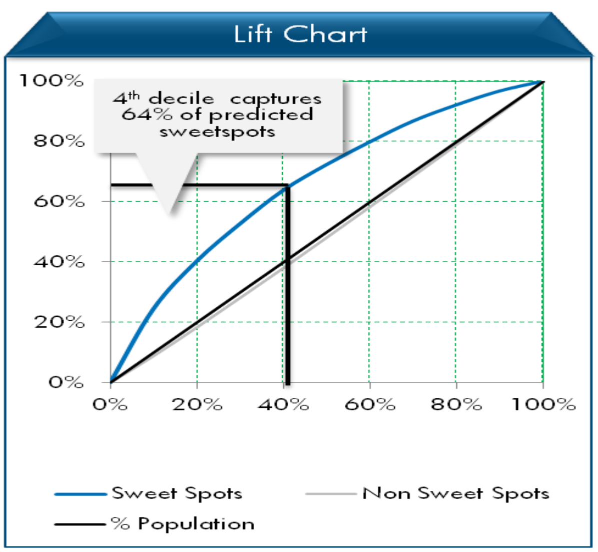


Figure 4: Lift Curve

CURRENT STATUS AND NEXT STEPS

Currently the IPG business is in the process of designing campaigns to implement this analytical approach. Targeting the right customers with the right products will push customers to the "sweetspot" and generate incremental revenue. The purchase history of each of these customers was provided as added information, for example, first order date, last order date, the kind of product purchased and the distribution of spend across business units and product categories. Demographic details of the industry, size segment, and telephone numbers were also added to the targeting list.

The "sweetspot" model is the first step to providing holistic IT solution to our SMB customers. Going forward, we will combine this approach with the estimate of the likely time for a next purchase, resulting in a complete go-to-market strategy.

CONCLUSION

We have presented a novel approach in the area of database marketing to identify top segment customers by utilizing a more evolved response modeling framework. We have used logistic regression to show how customers can be converted to a higher segment instead of looking at the usual look-a-like kind of a response model framework. This is particularly important for customer development campaigns (up-selling and cross-selling). The proposed methodology is simple and can be readily applied in any other cross selling activity. We hope that this paper will instigate new ideas and open up more research avenues in the database marketing industry.

REFERENCES

[1] Hosmer D, Lemeshow S. Applied logistic regression. New York: Wiley; 1989.

[2] Scott W. Menard . Applied logistic regression analysis, Volume 106; Volume 2002

[3] David G. Kleinbaum, Mitchel Klein, Erica Rihl Pryo. Logistic Regression: A Self-Learning Text

[4] Paul David Allison, Logistic regression using the SAS system: theory and application

[5] Victor S.Y. Lo, Fidelity Investments, Boston, MA. The true lift model: a novel data mining approach to response modeling in database marketing

[6] Ron Kahan, (1998). Using database marketing techniques to enhance your one-to-one marketing initiatives, Journal of Consumer Marketing, Vol. 15 Iss: 5, pp.491 – 493

[7] Markov Chain Monte Carlo in Practice (Chapman & Hall/CRC Interdisciplinary Statistics)

[8] John A. McCarty, Manoj Hastak. Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression

[9] Purba Rao. Database Marketing and CRM: A Case on Predictive Modeling for Ayurveda Product Offerings

Novel Image Representation and Description Technique using Density Histogram of Feature Points

Keneilwe ZUVA
Department of Computer Science, University of Botswana, P/Bag 00704 UB, Gaborone, Botswana

and

Tranos ZUVA
Department of Computer Systems Engineering, Tshwane University of Technology, Private Bag X680, Pretoria 0001, South Africa

and

Queen Miriam SELLO
Department of Computer Science, University of Botswana, P/Bag 00704 UB, Gaborone, Botswana

ABSTRACT

This paper introduces novel object shape representation using Density Histogram of Feature Points (DHFP). We use silhouette images where the image region ξ consists of only those pixels that correspond to points on the object and have a value one (1) indicating "*on*" pixels. We count the number of on pixels in a rectangle boundary around the centroid, in the event that there are no "on" pixels in a rectangle boundary then the value is zero and the rectangle boundaries that are outside the grid are represented by a dummy number. A similarity measure is used to calculate the probability of two image objects being similar. Depending on the value of the probability then a dissimilar can be calculated. This method showed improved retrieval rate due its selective way of calculating dissimilarity of object shapes. Analytic analysis was done to justify our method, experiments were conducted and we tabulated the results.

Keywords: Density Histogram, Similarity, Dissimilarity, Shape representation, and Silhouette images

1. INTRODUCTION

With vast collection of digital images on personal, institutional computers and on the Internet, the need to find a particular image or a collection of images of interest has increased tremendously. This has motivated the researchers to find efficient, effective and accurate algorithm that is domain independent for representation, description and retrieval of image(s) of interest. There have been many algorithms that have been developed to represent, describe and retrieve images using their visual features (shape, colour, texture) [5][6][8][10][11]. Visual feature representation and/or description play(s) a very important role in image classification, recognition and retrieval. A successful image representation and description is dependent on the following:

- the selection of suitable image feature(s) to encode
- the quantification of these features [11].

Shape representation and description has been dominant in research area of image processing because shape is considered to be the basis of human visual recognition [11]. The shape representation can be classified as Region based or Contour based representation.

The contour based techniques use the boundary of shape to describe an object. It is commonly believed that human beings can differentiate objects by their boundaries or contours [10]. Usually most objects form shapes with defined contours, making the use of these techniques most appealing. The techniques can generally be applied to different application areas with a considerable success. The techniques have a low computation complexity as compared to region based techniques and they are sensitive to noise. Some of the techniques in this group are as follows just to mention a few and are well described in [6] are: Compactness, Eccentricity, Shape signature, Hausdoff Distance, Fourier Descriptors and Wavelet Descriptor.

The region based shape representation uses the boundary pixels and the interior pixels of the shape. This group of shape representation algorithms are robust to noise, shape distortion and they are applicable to generic shapes [1]. Some of the techniques in this

group in [6] are: Geometric moments, Legendre moments, Zernike moments, Generic Fourier Descriptor and Object representation by the density of feature points.

In this paper we propose a novel image representation and description technique using **D**ensity **H**istogram of **F**eature **P**oints (DHFP) representation of an image object. This method imitates human visualization of image object shape and matching similar object shapes.

2. SHAPE REPRESENTATION BY THE ENHANCED DENSITY HISTOGRAM OF FEATURE POINTS (EDHFP)

This method describes the feature points within the rectangle boundary in an image grid. Assume we have a silhouette object shape segmented by some means such as Chan & Vese Active Contour without Edges and let the feature points set $P(x, y)$ (intensity function) of the object shape be defined as

$$P(x, y) = p_i(x, y)$$
$$such\ that\ i = 1, 2, ...n\ where\ n\ \varepsilon\ \mathrm{N} \quad \text{Eq. (1)}$$

We find the centroid of the object shape. The following formulae will be used to calculate the centroid [4][7]:

$$x_c = \frac{m_{1,0}}{m_{0,0}} \quad \text{Eq. (2)}$$

$$y_c = \frac{m_{0,1}}{m_{0,0}} \quad \text{Eq. (3)}$$

where $m_{1,0}, m_{0,1}, m_{0,0}$ are derived from the silhouette moments given by

$$m_{i,j} = \sum_x \sum_y x^i y^j P(x, y) \quad \text{Eq. (4)}$$

The following theorems will guarantee the uniqueness and existence of silhouette moments:

Uniqueness Theorem

Assuming that the intensity function $P(x, y)$ is a piece-wise continuous and bounded in the region ξ, the moment sequence $\{m_{i,j}\}$ is uniquely determined by the intensity function $P(x, y)$ and conversely.

Existence Theorem

Assuming that the intensity function $P(x, y)$ is a piece-wise continuous and bounded in the region ξ, the moments $m_{i,j}$ of all orders exist and finite.

Thus for silhouette image $P(x, y)$, $m_{0,0}$ the moment of zero order represents the geometrical area of the image region and $m_{1,0}, m_{0,1}$ moment of first order represents the intensity moment about the y-axis and x-axis of the image respectively. The centroid (x_c, y_c) gives the geometrical centre of the image region.

Suppose the size of the grid occupied by the object shape is NXN. The vector dimension to represent the density of object shape will be N-1. In reality we are going to have a vector dimension of N, the last element represents the number of vector elements that describe the object shape within the grid.

The centroid calculated by the two formulas above (2) and (3) is (x_c, y_c). From the centroid we count the number of "on" pixels in the rectangle boundaries in steps of one successively from the centroid. Some rectangle boundaries are incomplete but we count the "on" pixels in them in the grid. The number of "on" pixel in each and every rectangle boundary is denoted as $n_1, n_2, ...n_m$ where m is the number of rectangle boundaries from the centroid. If m is less than N-1 (m < N-1) it means the rectangle boundaries fall outside the grid where there is no image object. The positions of the vector that fall outside the grid, we represent them with dummy number for example "pi" an irrational number. In the event that there are no "on" pixels in a rectangle boundary within the grid we put zero.

Since all the pixels of the image contribute in calculating the centroid, it means the deviation of the point is minimal. The vector representation of the object shape of our method should have all or some of the following:

- Zeros-indicating no "on" pixels in a partial or complete rectangle boundary within the grid
- Natural numbers indicating number of "on" pixels in partial or complete rectangle boundary within the grid
- Dummy number > 4N or an irrational number for rectangle boundary outside the grid

Example (Representing an object shape)

Supposed we have the following object shape features on a grid given in table 1.

0,0	1,0	**2,0**	**3,0**	**4,0**
0,1	**1,1**	**2,1**	**3,1**	**4,1**
0,2	**1,2**	**2,2**	***3,2***	**4,2**
0,3	1,3	2,3	**3,3**	**4,3**

0,4	1,4	2,4	**3,4**	**4,4**

Figure 1. Segmented object shape

The bolded indicate the "on" pixels. The size of the grid occupied by the object shape is 5X5. It means the vector dimension to represent the density of object shape in grid will be four (4). The centroid calculated by the two formulas above (2) and (3) is (3, 2), the centroid pixel is in italics. The first rectangle boundary is made up of the following pixel

(2,1), (3,1), (4,1), (4,2), (4,3), (3,3), (2,3), (2,2)

and there are seven "on" pixels that constitute our first element of the vector. The vector that represents object shape above is

(7, 7, 1, Pi), using EDHFP

Since rectangle boundary 4 is outside the grid then a dummy number is used, in this case "pi" is the number used in EDHFP.

3. SIMILARITY OF SILHOUETTE IMAGES

Supposed we have two geometrical objects shapes defined as

$$P(x,y)\ and\ Q(x,y)$$

are called similar if they both have the same shape. The objects shapes are made up of pixels $p_i(x,y)\ where\ i = 1,2,.....n\ and\ n\ \varepsilon\ \mathrm{N}$ and $q_i(x,y)\ where\ i = 1,2,.....n\ and\ n\ \varepsilon\ \mathrm{N}$ respectively with intensity value one (1). Uniform scaling is done to all our objects shapes therefore in reality we are finding objects shapes that are congruent to each other. If the two objects shapes are congruent then the distribution of pixels and the area are the same in the two objects shapes. We also regard that congruent shapes are similar shapes with a scale factor of one making similarity of shapes more generic terminology to use.

Similarity is defined [9] as:

Having two subsets $P(x,y)\ and\ Q(x,y)$ of Euclidean space R^n are called similar if

$$f : P \to Q$$

such that for any two points x and y that belong to P we have

$$d(f(x), f(y)) = rd(x,y), \qquad \text{Eq. (5)}$$

where $d(x,y)$ is the Euclidean distance from x to y. $P(x,y)\ and\ Q(x,y)$ are called similar if P(x, y) is the image of Q(x, y) under such a similarity. In our case r=1 thus there exist isometry f. The following condition will be fulfilled

$$d(P,Q) = 0 \qquad \text{Eq. (6)}$$

d is invariant under a chosen group of transformations G if for all $f \in G,\ d(f(P), f(Q)) = d(P,Q).$ This is a requirement for object shape recognition under affine transformation [9].

The acquisition of the image using different camera enabled devices and the segmentation technique used makes the objects shapes not to be exactly the same causing problems in image retrieval. Any small change in the object shape causes changes in the similarity distance. One point that should not deviate much due to some minor distortion in the object shape is the centroid due to the fact that every pixel of the object shape contributes in calculating it.

So if $P(x,y)\ and\ Q(x,y)$ are similar then

- The centroid is approximately the same in relation to pixels that make up the object shape due to the fact that similar objects shapes have the same shape.
- The area of two objects shapes are approximately the same because they are mapped in the same grid of NXN.

The centroid is the point of reference when generating the rectangles. It means the number of rectangles in each object shape is the same in the grid in the event of objects shapes with the same centroid. The distribution of pixels in both objects shapes that are similar is the same thus we can establish a pattern of number of pixels within each and every rectangle generated from the centroid of the object shape. Knowing the pattern for $P(x,y)$ then we know approximately the pattern for $Q(x,y)$. Then the following hold for similarity of objects shapes

- Number of pixels ($p_i(x,y)$) in $P(x,y)$ is equal to number of pixels ($q_i(x,y)$) in $Q(x,y)$
- Centroid of $P(x,y)$ in relation to $p_i(x,y)$ is also the centroid of $Q(x,y)$ in relation to $q_i(x,y)$
- The number of rectangles generated in $P(x,y)$ is also the same number generated in $Q(x,y)$
- The number of pixels of the object shape in each rectangle in $P(x,y)$ is also the number of pixels in each rectangle in $Q(x,y)$

In measuring the dissimilarity on objects shapes, the fundamental concept is to compare corresponding rectangles in each object shape to find out the difference in the corresponding rectangles. The following dissimilarity measurement techniques in [2][3] were experimented with in our method

- L_p *Minkowski family* and
- L_1 *family* similarity distance measures.

Dissimilarity Measurement

We use the Euclidean L_2. When we have two vectors representing two object shapes and the vectors are

$V_1 = x_i$ *and* $V_2 = y_i$ where i=1, 2,.....n, $n \varepsilon N$

The dissimilarity between the two object shapes is given by [3][2] as:

$$d_s(V_1, V_2) = \| V_1, V_2 \|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \ .$$

Eq. (7)

Applying the Euclidean similarity distance to our method, the dummy value is used in calculating the distance and is taken as zero.

Similarity Measurement

$$s_{s1} = \frac{p + s}{p + q + r + s}$$ Eq. (10)

Where p = number of rectangles occupied by both objects

s = number of rectangles not occupied by both objects

q = number of rectangles occupied by object 1 and not by object 2

r = number of rectangles occupied by object 2 and not by object 1

When $s_{s1} \geq 0.5$ then there is a higher probability of the two objects to be similar, if not then it is very unlikely for the objects to similar to each other.

Example (Calculating Similarity and Dissimilarity)

Suppose the two object shapes are represented as follows:

The object shape being queried is V_1

EDHFP

$$V_1 = (4, 13, 14, 2, \text{pi}, \text{pi})$$
$$V_2 = (4, 12, 15, 3, 5, pi)$$

We calculate similarity measure s_{s1} as follows:

p = 4, s = 1, q = 0, r =1

using the formula above we have

$$s_{s1} = 0.833$$

indicating a high probability of similarity between the two objects.

We can now calculate the dissimilarity of the two objects using the Euclidean distance formula. $d_{s2} = \sqrt{(4-4)^2 + (13-12)^2 + (14-15)^2 + (2-3)^2 + 5^2 + 0}$

$$d_{s2} = \sqrt{28}$$

So that is the dissimilarity of the two object shapes.

4. EXPERIMENTATION

Our main objective is to find the effectiveness of the representation algorithms DHFP in retrieval of image objects. We used the Euclidean Dissimilarity in retrieving similar image objects after calculating their likelihood of being similar. We created image database of shoes image shapes. Some of the image objects were not rotated lossless at 90, 180 and 270 degrees that means degradation of the image object shapes occurred during rotation. The query images were captured using different camera enabled devices. The images objects were of different dimensions MXN or NXN where M and N belong to natural numbers.

The images that we used were only having one image object with a homogeneous background. We then segmented the image object shape by a 45X45 grid. All images were converted to gray scale images. After segmentation the output was a binary image object (silhouette). They were then represented using our method the novel DHFP.

We measured the accuracy of our system by calculating the recall, the precision and effectiveness. The following formulas were used [8]

$$recall = \frac{A}{N}$$ Eq. (11)

$$precision = \frac{A}{A + C}$$ Eq. (12)

$$N = A + B \qquad \text{Eq. (13)}$$

Where A is the number of relevant image objects retrieved, B is the number of relevant image objects not retrieved and C is the number of not relevant image objects retrieved.

5. RESULTS

The results obtain from comparing the two methods are as follows:

Grid size 45X45	
Average Recall %	Average Precision %
20	100
33	100
50	100
67	100
75	100
80	100
100	85

Table 1: The average precision of the method with the grid size of 45X45

6. SUMMARY AND CONCLUSION

From our results we can conclude that EDHFP method of image object representation was able to retrieve human-beings perceived similar image object shapes. The combination of similarity and dissimilarity measures made our system to have a high precision values. The researchers found an efficient, effective and accurate algorithm that are domain independent for representation, description and retrieval of image(s) of interest. This was deduced from the retrieval results.

7. REFERENCES

[1] E. M. CELEBI, & A. Y. ASLANDOGAN, A comparative Study of Three Moment-Based Shape Descriptors. **Proceedings of the International Conference on Information Technology**: Coding and Computing, 2005.

[2] S. H. CHA, Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. **International Journal of Mathematical Models and Methods in Applied Sciences**, 1(4), 2007, pp. 300-307.

[3] C. C. CHEN & H.-T. CHU, Similarity Measurement Between Images. **Paper presented at the Proceedings of the 29th Annual International Computer Software and Applications Conference**, 2005.

[4] J. FLUSSER, T. SUK, & B. ZITOVA, **Moments and moment invariants in pattern recognition.** West Sussex: John Wiley & Sons Ltd 2009.

[5] Y. LI,. & L. GUAN, An effective shape descriptor for the retrieval of natural image collections. **Paper presented at the. Proceedings of the IEEE CCECE/CCGEI**, Ottawa 2006.

[6] Y. MINGQIANG, K. KIDIYO, & R. JOSEPH, A survey of shape feature extraction techniques. **Paper presented at the Pattern Recognition**, 2008.

[7] R. MUKUNDAN, & K. R. RAMAKRISHNAN, **Moment functions in image analysis: theory and applications.** Singapore: World Scientic Publishing Co. Pte. Ltd. 1998.

[8] M. X. RIBEIRO, J. MARQUES, A. J. M. TRAINA, & C. T. JR, Statistical Association Rules and Relevance Feedback: Power Allies to Improve the Retrieval of Medical Images. **Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems**, 2006.

[9] R. C. VELTKAMP & L. J. LATECKI, Properties and Performance of Shape Similarity Measures. **Lecture notes in computer science: 2006**, pp. 1-9.

[10] D. ZHANG, & G. LU, Review of shape representation and description techniques. **Pattern Recognition Society**, 37: 2004, pp. 1-19.

[11] X. ZHENG, S. A. SHERRILL-MIX, & Q. GAO, Perceptual shape-based natural image representation and retrieval. **Paper presented at the Proceedings of the IEEE International Conference on Semantic Computing**, 2007.

YouDemo: Capturing Live Data from Videos

ICT applications in Education and Training
For ICTA 2011

Mike BOROWCZAK
School of Electronics and Computing Systems, University of Cincinnati
Cincinnati, OH 45221, USA

and

Andrea BURROWS
Secondary Science Education, University of Wyoming
Laramie, WY 82071, USA

ABSTRACT

YouDemo is a technology concept, created by the primary author, which allows video viewers to rate two metrics of the video at the same time. The metrics can be adjusted according to the needs of the data collection. This technology can be adapted for use in the K-20 classroom and anyplace where the effectiveness of a video is important for learning or understanding. YouDemo's three main visible interfaces are explained. Two data sets using YouDemo were collected. One data set is from a K-12 audience and the other is from a university audience. Initial findings show that use of YouDemo helps guide the viewer to focus on certain aspects of the video. Additionally, if the viewers are also video producers, YouDemo indicates the areas of strength and weakness in the original videos. These reviewed videos can then be reproduced using the feedback from the audience for an enhanced new version of the video product.

Keywords – Computer-Based Training, Web-Based Training, Internet-Based Teaching, Video, Video critique, YouDemo

INTRODUCTION

With the ever-increasing use of technology in today's K-20 classrooms, and a new generation of teachers who have grown up in the digitally connected age, it is only fitting that the training and continued education of teachers should involve both formal instructional theory and the practical usage of technology.

The authors want to provide K-20 teachers with a tool, *YouDemo*, to assess their skills at creating engaging, content-rich online and in-class demonstrations and laboratories for their students. The focus here is on the use and continuous assessment of self-created video demonstrations as a method to enhance live classroom demonstrations with a particular emphasis on student engagement. The goal is to provide a mechanism to share and most importantly assess online demonstrations. This is accomplished, by providing a social-driven tool integrating the common YouTube interface, with a novel, continuous assessment feedback interface.

This evaluation product consists of three main visible interfaces, two private to the video creator (primary author) and a third that is publicly available and shared via various social media sites. The first interface available to the video's creator allows the posting of both video content and desired assessment metrics, while the second interface allows the reviewing and analysis of assessment data. The third interface, to the public, allows for live, continuous assessment of the video demo as the video progresses.

The technology to critique video during "real time" is a vital innovation. YouDemo is an important advancement in using videos to enhance learning and understanding. As Hodson [1] explains, "Any discussion of technological literacy inevitably raises important issues relating to computer technology." Computer technology, or computer literacy, "extends well beyond the acquisition of basic computer skills to encompass "…the capacity to evaluate information for accuracy, relevance, and appropriateness, and the ability to detect implied meaning, bias and vested interest."

YOUDEMO AT A GLANCE

YouDemo allows a teacher, or anyone, to upload video(s), select a question(s) they wish to have answered about their video, and then share the audience assessment with students, peers, instructors or even their entire social circle (via sites such as Google+, Twitter, and Facebook). Those viewing the video, and using *YouDemo*, will be able to provide continuous assessment throughout the duration of the video, allowing the teacher to gain valuable, authentic feedback data on their presentations (e.g. demonstration videos). The teacher will be able to view and share the aggregation of their assessment results to further improve their videos.

YouDemo is a needed concept since validity and reliability in the classroom are important assessment aspects. As stated by Mertler [2]

> Evidence must be continually gathered and examined in order to determine the degree of validity possessed by decisions. Three formal sources of evidence that support the existence of validity include content, criterion, and construct evidence. Content evidence relies on professional judgment; whereas, criterion and construct evidence rely on statistical analyses. Content evidence of validity is the most important source of evidence for classroom assessments. …As with validity, reliability addresses assessment scores and their ensuing use.

YouDemo will allow K-20 teachers a means of assessing the validity and reliability of the presentations, or videos, that they show to a student audience.

Finally, the showcasing product tool, *YouDemo*, is built upon free and readily available software development kits (SDKs) and application programming interfaces (APIs), and is itself open source allowing future modification and improvement.

BEHIND THE SCREEN WITH YOUDEMO

As mentioned earlier, *YouDemo* is an evaluation platform that consists of three main visible interfaces. Two of the interfaces are private to the video creator (or evaluation requestor) and a third is publicly available and shared via various social media sites using one of the private interfaces.

Of the two private interfaces, the first (Figure I) allows YouTube videos to be registered and assigned two custom evaluation metrics. Both actions are combined on one simple, unified screen allowing interested and potentially novice users easy access to YouDemo. After the preliminary release of the product all previous selected metrics will be catalog and based on those results the top metrics will be provided as suggestions to future users.

The *YouDemo* registration process has been streamlined to allow users of any ability simplified access. Since *YouDemo* simply references existing YouTube videos, two crucial points arise. First, only embeddable videos (a YouTube feature) are accessible via *YouDemo*. Secondly, a YouTube video can be associated with multiple users, or the same user and multiple metric pairs. Future iterations of YouDemo will allow inline searches of videos which fit this criterion as well allow direct uploads to YouTube.

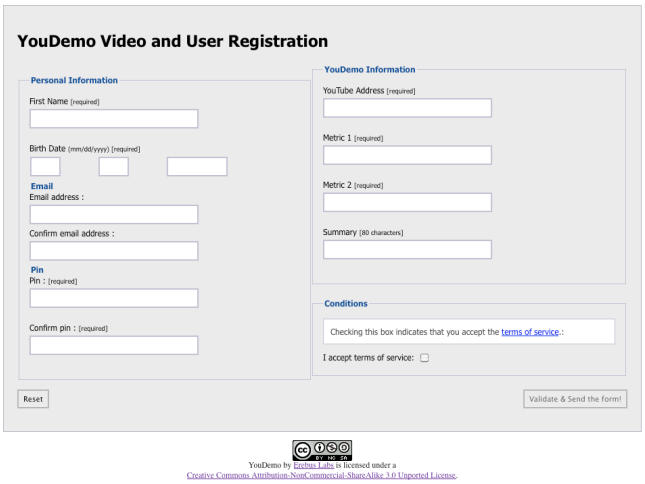


Figure I: The YouDemo Unified Video and User Registration allows quick access to users of all abilities.

The second private interface, the Rating Dashboard (Figure II), allows access to social media distribution of the unique, public, Video Rating Request Page and the collected statistics data. Currently, the collected statistics exist only in a comma separated value (csv) form. This common format is easily imported to any spreadsheet tool such as Microsoft Excel, LibreOffice Calc or Google Docs Spreadsheet. Currently, a unique analytics engine is being developed to quickly render and display key data trends allowing the entire *YouDemo* experience to become self-contained. Based on preliminary feedback, the dashboard may evolve to allow sharing of collected data between users, e.g. between a student and professor, between colleagues, or perhaps to the public.

Figure II: The YouDemo Rating Dashboard highlights a user's rating requests along with sharing and data download options.

The final, public, interface is the one seen by most users. As mentioned before, it is unique every video-rating pair. A video "XYZ" with metrics *A* and *B* uploaded by Alice has a page different than those exact parameters uploaded by Bob. Similarly, Alice could also upload video "XYZ" with metrics *A* and *Z* and that page would also be unique (and valid). With this in mind, Figure III shows an early version of the YouDemo rating system in which the two specified metrics are rating using the left/right and up/down arrow key pairs. Data is collected live and after the video a sequence of alignment questions are displayed.

Figure III: An early version of the YouDemo Rating Engine showing both metrics and the associated YouTube video.

YouDemo utilizes standard back-end web technologies to support the three interfaces. The standard, freely available, LAMP solution stack is used exclusively, providing access to MySQL and PHP. We further supplement the existing toolsets with the addition of freely available, Open Source, graphing package. While outside the scope of this paper, we also mention that due to real-time data collection, we have optimized our MySQL data tables to record approximately 4000 minutes of rating data per MB of storage. Considering the average target demo video is under 5 minutes this allows for 800 ratings of a single video in 1 MB of data.

THEORETICAL FRAMEWORK

In using YouDemo we embrace a socio-cultural theory as the framework of this research. We look to the process of learning while the videos are being critiqued, not the product produced (the video itself). Socio-cultural theory has grounding in the work of Vygotsky[3] and Bandura[4]. There is an emphasis on the interaction between learners and learning tasks. Since STEM education is currently in the spotlight, gaining insights into STEM video production and critique using a socio-cultural perspective is important. Current K-20 work directly impacts youth in low socioeconomic conditions. Roth and Lee [5] state that "a researcher… does not separate the poverty or culture of urban students' home lives from conditions of schooling, consideration of the curriculum, problems of learning, or learning to teach under difficult settings." When working with urban youth, the socio-cultural perspective becomes fundamentally important to use. Wertsch [6] also showcases the process of learning through all aspects of the relationship between the mind and setting. Thus, socio-cultural theory is about the "whole scene" of learning, or the process. It is not the individual parts in isolation that create the scene. Using the "whole scene" approach will sharpen our understanding of how STEM videos, and their peer and instructor critiques, can impact learning and understanding for the K-20 student audience.

METHODOLOGY

YouDemo was created by the primary author, gently shaped by suggestions for use by the secondary author, and then used by K-20 students as they viewed videos. There were two sets of data collected. The first set of data came from secondary students in NSF GK-12 Fellows' classrooms near Cincinnati, Ohio. These Fellows showed STEM content videos, created by former Fellows, and asked the secondary students to critique the videos. The second set of data came from university students critiquing videos created by their peers. The university students were studying to become science teachers, and created videos for a class project. Both the secondary students and the university students used YouDemo to critique the videos that they were presented. All of the K-20 students were asked three questions after using YouDemo. The questions were: 1) What specifically do you remember about the video you just watched?, 2) How did YouDemo affect your viewing of this video? 3) How do you think YouDemo could or should be used? Using YouDemo required the video viewer to press the up/down for one metric and right/left for the other metric. Pressing either up or right represented "more" and either down or left represented "less" of the metric selected. The data generated from both metirics, on each video, by both groups was captured and stored by YouDemo.

BACKGROUND ON NSF GK-12 FELLOWS AND SECONDARY STUDENTS

To put the contents of this paper in proper context, it will be appropriate to first briefly describe the background of the Fellows, or graduate engineering students, who implemented the STEM lessons, and showed the STEM videos in secondary classes. These STEM videos were entitled "Science in a Minute." The Fellows were prepared to take their engineering content expertise into the secondary classrooms. Preparation for the Fellows occurred in "Instructional Planning," a formal three credit hour course offered by the College of Education. The course addressed a wide range of topics: STEM achievement standards, lesson and unit planning, instructional models of teaching, instructional management, the nature of students, skills of or connecting with students at a personal level, understanding student cultures and responding appropriately, and assessment or evaluation of student learning and instructional efforts. The course was taken by the Fellows during the summer before they entered the classrooms, but while they were engaged with secondary students in an enrichment summer academy. When the school year began, the

Fellows enrolled in another Education course, "Field Practicum," which was a one credit hour course taken in the fall, winter, and spring quarters. This course supported Fellows as they encountered unfamiliar territory as a teacher upon their entry into the secondary schools. Fellows are required to focus on important aspects of the teaching-learning situation and the culture of the school and students as well as their relationships with the teachers. These experiences led the Fellows to create the "Science in a Minute" videos (http://www.eng.uc.edu/step/) for the K-20 audience. These videos highlight STEM content and were critiqued by the secondary students, most of whom are urban students, in the Fellows' classes.

VIDEO REVIEWS FROM SECONDARY STUDENTS

The secondary students watched the "Science in a Minute" videos, which were created by NSF GK-12 Fellows, and critiqued those videos. Based on preliminary findings of the pilot data set (data collection will continue through June 2012), the secondary students exhibited three characteristics. The first characteristic was that the secondary students were more likely to push an arrow key for a metric when the video changed scenes. The second characteristic was that the "more" button was used more frequently in the beginning of the video and the "less" button was used more frequently at the end of the video. The third characteristic was that the secondary students reported that they "paid more attention" to the videos when they had to critique them.

BACKGROUND ON UNIVERSITY STUDENTS

The university students were undergraduates and graduates obtaining majors in both a pure science and science education. As part of their degree requirements they took a course on how to teach science, called methods. The methods course required an assignment to create a video, eventually posted to YouTube, that showcased a laboratory or demonstration directed at a K-12 audience. There were specific guidelines that the university students were asked to follow. The instructor provided guidelines that the video should be between 3 and 6 minutes long and highlight specific science content. Additionally, the video should relate to a real world STEM application in an engaging manner. The university students were able to recall more specific details from the videos, but this finding cannot be directly compared to the secondary students.

VIDEO REVIEWS FROM UNIVERSITY STUDENTS

The university students watched the peer created videos of laboratories and demonstrations. They critiqued the videos using YouDemo. Based on preliminary findings of the pilot data set (data collection will continue through June 2012), the university students exhibited two characteristics. As with the secondary students, the first characteristic was that the university students were more likely to push an arrow key for a metric when the video changed scenes. The second characteristic was that the university students reported that used in moderation, YouDemo would focus their attention and allow them to recall more details of the videos.

SUCCESSES AND CHALLENGES

The successes of YouDemo are many. First, it was created and worked to assess two metrics of a video at the same time. This technology, until now, could not be found free to the public. Second, a video viewer has a collection of video history ratings that can be accessed at any time by the user. Third, YouDemo showed promise in promoting student focus and learning during video sessions. Fourth, it allowed teachers and K-20 students the opportunity for real time feedback that enabled them to revise their videos for more impact on the viewer.

There were several challenges of creating and using YouDemo as well. There were technical constraints, such as providing a clean, professional, polished interface that a novice would be comfortable using. Figuring out how to

represent the data that was collected is still a challenge. Due to real-time data collection, data aggregation requires extra background computation and analysis. Finally, the age and video game experience of the user, impacted the ease of pressing up/down and left/right for two independent metrics.

CONCLUSION

YouDemo is an emerging platform to evaluate existing content. The platform has shown promise in K-20 settings. Based on the successes (mentioned earlier) and the pilot data sets, one set from the K-12 audience and one set from the university students, we conclude that the use of YouDemo can focus viewers on specific video content. Students can concentrate on two particular metrics, either chosen or provided, when using YouDemo. Overall, students believe it helped them to recall more details from a video viewing than they would be able to recall without using YouDemo.

SUMMARY AND IMPLICATIONS

The K-20 audience showed an appreciation for the metirics in the videos that were being analyzed. Based on self reporting, they were able to recall more details. Even if future evidence shows that students recalling more details are not reliable or valid, student beliefs, or their self-efficacy, could influence their learning and understanding. In other words, just believing that they are learning more with YouDemo could create an environment for the K-20 audience to increase their content knowledge from the videos critiqued. Better understanding and increased STEM learning in the K-20 audience was the ultimate goal, and YouDemo is one technology that can both improve video quality and increase student learning. A future research direction for this technology would be more in depth studies of YouDemo use and expansion of YouDemo use into other education and business realms.

REFERENCES

[1] Hodson, D. **Teaching and learning about science,** Boston: Sense Publishers, 2009.

[2] Mertler, C. **Classroom assessment: A practical guide for educators**, Los Angeles: Pyrczak Publishing, 2003.

[3] Vygotsky, L, **Mind in society: The development of higher psychological processes**, Boston, Harvard University Press, 1978.

[4] Bandura, A, **Social learning theory**, New York: General Learning Press, 1977.

[5] Roth, W. & Lee, Y., "**Vygotsky's neglected legacy**," Educational Research, 77, 1998, 186-232.

[6] Wertsch, J, **Mind as action**, New York: Oxford University Press, 1998.

The Four-rotor Helicopter used for Real-life Introduction to Multivariable Control Problems

Dag A. H. Samuelsen
Department for technology, Buskerud University College
3611 Kongsberg, Norway
Dag.Samuelsen@hibu.no

And

Olaf H. Graven
Department for technology, Buskerud University College
3611 Kongsberg, Norway
Olaf.Hallan.Graven@hibu.no

ABSTRACT

Giving students laboratory exercises that are both theoretically challenging and inspiring, while having the elements of real-life challenges like sub-optimal models, limited processing time and large degree of uncertainty, is a challenging task, partly due to the need of adapting the level of complexity to the student or group of students doing the exercise in order to keep them engaged throughout the exercise, and in part due to the university's need to reduce expenses related to the administration, supervision, and execution of laboratory exercises.. The use of the four rotor helicopter presented in this paper allows for the desired adaptation in complexity through the type of assignment given to the students and the students' option of choosing between different models . The eager student might be tempted by the better performing, but more complex models, while the struggling student can find satisfaction in stabilising the aircraft using the less complex models. The laboratory setup presented uses low-cost components, which allows for low investment and maintenance costs.

Keywords: DSP, helicopter, multivariable feedback control

INTRODUCTION

Over last few decades the four-rotor or quadrotor helicopter[1], has been in existence as a full scale vehicle. The quadrotor helicopter as a small scale vehicle is also used in some universities as a training platform for students, in particular master and PhD students. The use of a quadrotor helicopter as a small vehicle is an attractive approach for several reasons: The low cost of the building materials, high reliability, and the use of a simple mechanical construction that is easy to continuously alter and adjust to fit the specific needs of the user. The low cost stems from the use of low cost components, such as simple standard available mechanical parts, standard power converters used in hobby R/C devices, low cost processors, and simple motors without gears. The latter part also introduces high reliability in the sense that there are few moving parts which introduces wear and tear on the device. This is in stark contrast to the standard helicopter, with a main rotor and a tail rotor. A standard helicopter has a highly complex mechanical system to interconnect the rotors. A quadrotor helicopter uses highly complex control structures to control the speed of the four rotors independently in order to give the aircraft balance and controlled movement. This control structures are implemented in digital hardware giving added reliability and the reduced cost.

The implementation of the control structure and algorithms is even today an unresolved matter, making it a research topic of several institutions [2-5]. Sub-optimal solutions to this problem do exist, and today's implementations of the four-rotor helicopter use a simplified model [6] which is possible to balance out using the known control theory and the processing power available in the aircraft. The processing power is limited first and foremost by the power consumption, as the power consumed in the processor(s) will effectively limit the flying time of the aircraft. Low power DSPs can have a power consumption around 1W [7], but the processing power is then limited and only allows for simple models and low order controllers. Medium power DSPs have

more processing power, but power consumption is then increased both in the DSP itself and in the external memory banks which needs to be introduced [8]. This may also require cooling fans. For very complex models and controllers, several DSPs is necessary to handle the amount of data processing and gives an unrealistic solution in terms of both power consumption and size. Cost is normally not a limiting factor in this context.

The known control algorithms for the four-rotor helicopter are simple enough to make a realistic implementation in an advanced control module of an engineering bachelor degree[9]. The laboratory installation is a low cost device, affordable for many institutions, and the maintenance cost (as students tend to break things they use) will also be low due to low component cost and simple mechanical construction. It is with such a setup possible to give the students all the different challenges associated with this type of control problem: limited knowledge of the true models of the system that are to be controlled, limited time and processing power to run the control loop algorithm, and the student's limited experience with problem solving of this type. When students starts working on problems for which they have limited experience with, it affects their ability to reach a working solution, and any solution is greatly affected by choices made in early stages of the controller synthesis. When success fail to come after completing an iteration, it is not always immediately obvious that an earlier bad choice might be the main reason for the control system failing. It is then important for the tutor to intervene and set the students on the right track. This may involve everything from simple steps like tuning parameters, to the need to change the control structure or even designing a completely new model. The students need to understand that each iteration actually must contain an evaluation and assessment of the model and a potential redesign. These types of challenges are typical for control system development for the not-so-many-years-of-experience engineer, and are therefore in the authors' opinion an important lesson to experience for the students.

At the authors' institution, the final year bachelor students in electrical engineering complete a module in multivariable control theory. The learning objectives of the exercise is that the students are be able to set up models of basic multivariable processes, set up different controller structures and find the controller parameters for the same processes, and find the properties of the combined systems consisting of process, perturbation block and controller regarding both nominal and robust stability and performance. Doing this on a theoretical basis by simulations will give the students only parts of the real-life physical challenges multivariable control in the presence of uncertainties poses to engineers. Thus, there is a definite need for a hands-on physical laboratory for the students to work on. It is also desired that this

Figure 1: Basic design of the helicopter

laboratory also should give the students experience in the previously mentioned challenges:

- How limited knowledge of the true models of the system poses a challenge as to how to model the uncertainty covering for both the neglected dynamics and the unknown dynamics, which are both represented in some way by perturbations of the nominal system. Failure to cover all possible perturbations from the nominal model might cause the system to be unstable for a set of states. And the understanding that this is clearly not desirable and why.
- How limited time and processing power will effectively limit the number of calculations or processor cycles that can be used to calculate the next command signal. How a control loop is typically performed by using a timed signal to do measurements of the process output at fixed periodic intervals. Then how to do calculations based on measurements and previous controller states and then sending the command signal to the actuators to "push" the process in the desired direction. In order for the controller to be able to stabilise the process, the control loop has to be run at a specific rate, giving the sampling rate of the system. When the speed of the processor is limited (as it always will be), this gives the limit on the combination of sampling rate and number of calculations done in the control loop. The students should be able to maximise the utilisation of this limit but still keep within the limit.
- The knowledge of the students varies, and each student have to decide at what level the controller design should be laid, so that the control problem can be solved within the given time period set aside for completing the laboratory work within the module.

When the students start on this module, they have theoretical as well as practical skills in a

number of areas from previous completed modules. The prerequisite knowledge needed for mastering this module is:

- Mathematics with linear system theory where they are supposed to develop basic mathematical models for physical systems and processes, mainly limited to amplification, time delay, 1st, and 2nd order linear, time invariant systems. This part also includes mathematical tools such as complex numbers.
- Control theory for single input - single output (SISO) systems. These modules cover the standard feedback and feed forward control loop with both continuous and discreet controllers, as well as stability and performance analysis.
- Basic programming of microcontrollers in C, with emphasis on the control of peripheral units and other basic tasks.

The purpose of the module for which the exercise program is described in this article, is to further extend this knowledge and skills to master the control of multivariable processes. The learning objectives of the module are [10]:

- Weighted sensitivity for SISO systems.
- Linear system theory: Coprime factorization, State controllability and observability, Stability, Zeros, Internal stability, Nyquist stability, Norms. H2-norm, H-infinity-norm, Hankel norm.
- Limitations on performance in SISO and multiple input multiple output (MIMO) systems.
- Uncertainty and robustness for SISO systems.
- Robust stability and performance analysis
- Controller design and controller structure design, including LQG and H-infinity methods

LABORATORY EXERCISE

The purpose of the laboratory exercise is to aid theoretical understanding by creating a link between the theoretical part and the physical world that is to be controlled, and to develop practical skills needed to perform the task of creating a control system. Thus, the laboratory exercise includes the creation of mathematical models for the process to be controlled, synthesis of a controller, simulations and uncertainty analysis to verify the controller operations on the model and all of its perturbations, discretisation and application of the controller on the physical system for a final verification that the controller actually is apt for controlling the physical system. After the final operation, it is usually necessary to tune the parameters of the controller in order to optimise the performance and stability properties.

The aircraft used in the laboratory exercise consists of a stiff frame with four arms perpendicular to each other for mounting of the motors driving the propellers. The power source is a high-capacity, lightweight battery pack, with low loss voltage converters to supply the microcontrollers/DSP processors, and power converters for controlling the speed of the motors. There is an option for several processors in the aircraft, each with a specific list of tasks. The main processor has significantly more processing power than the others, and is used for running the control loop algorithms. The other processors can be set aside to do other tasks not running in the main processor, such as communication, sensor interfacing, simple signal processing, and housekeeping in general. Some of these are mandatory, while the presence of others is up to the students to decide on.

The helicopter is given ready assembled to the students in order for them to have their full focus on development of the control algorithms. Which choices are then presented for the students? In the aircraft given to the students, the main processor will be a digital signal processor (DSP) [11], as the processing power needed to stabilise the aircraft is considered too much for a normal microcontroller. Installed is also a battery pack with fixed capacity, normal housekeeping circuits for power supervision, security, charging, and receiver for the remote control. The sensors available to the students are a solid state gyro and an accelerometer. The readings from these sensors are not very accurate and filtering through a Kalman filter/observer is highly recommended. This filter will have to be integrated within the control loop, and it is therefore natural to run the filter algorithm in the main processor. The helicopter can receive radio signals from a normal remote control for model aircrafts. In order for the helicopter to be controlled by a user via the remote control, whenever the receiver mounted in the aircraft receives control signal this must be read by the main processor or possibly buffered by any other microcontroller the students decide to put into the aircraft.

The helicopter and the motors must also be modelled by the students, and the choice of model will highly influence how well the control system will perform, or even if it is possible to synthesise a controller for the process model. The full model itself is not challenging to achieve, the main challenge lies in the highly non-linear characteristics of the model [12]. The mathematical tools for verification of controllers for non-linear models are limited, and hence the ability to develop stabile, robust controllers is likewise limited. In this context, the full helicopter model has to be reduced [13] to a sub-model for which the mathematical tools exist. This model reduction is one of the main issues of the laboratory exercise, and is left for the students to handle.

When a suitable simplified model is found, the controller has to be synthesised and discretisised so that it can be programmed as part of the software running in the main processor of the aircraft. As part of the controller discretisation, a specific sampling rate for the control loop is chosen. The first limiting factor is that a low sampling rate means that the command signal is updated too slowly to counteract unstable, low or mid frequency dynamics of the aircraft. So the sampling rate has to be fast enough. As a consequence of the limits set on the sampling frequency the limited processing power effectively dictates how many processor cycles that are available to calculate the next command signal to the motors. Due to the requirement for low power consumption, the DSP [11] in the aircraft has limited operating frequency. However the processor included has floating point capabilities. The processing limitations has the impact that the students will have to write efficient code, based on models that are optimal in the crossing point between model complexity and processing time of the controller for that specific model. This challenge is further complicated by the other limiting factors: limited knowledge of creating controllers for non-linear models, and the limited knowledge and skills the student or student group exhibit.

An average student group is expected to implement a controller which is able to stabilise the aircraft in the presence of small, possibly ramped changes in reference. For better performing groups, large, rapid steps are necessary to handle while keeping the aircraft stable, and also requirements on performance in the presence of model deviations may be set for the control system. For student groups with achievements below average, the aircraft must be possible to stabilise using the remote control for the aircraft.

LABORATORY TRIALS

The four-rotor helicopter model has been tried by a student group of 6 students in their final year project where the task given was to develop control algorithms to stabilize the helicopter so that is possible for a person with just basic training to fly the helicopter. The assignment included design and construction of the hardware, which is not a part of the assignment that will be given to future students within the multivariable control theory module. The main reason is that the hardware construction would take up to much time and that hardware construction is not a part of the learning objectives in the module. The project group managed to synthesise and tune a fairly good controller for the aircraft, meaning that the aircraft as difficult to control, but manageable. The students were considered to be above average in theoretical understanding and skills. When extracting the time resources spent on modelling the aircraft and the design of the control system, the students in the project group used on average less than 100 hours each, but stated that more hours spent on the design of the controller structure and tuning of parameters would likely result in a better performing

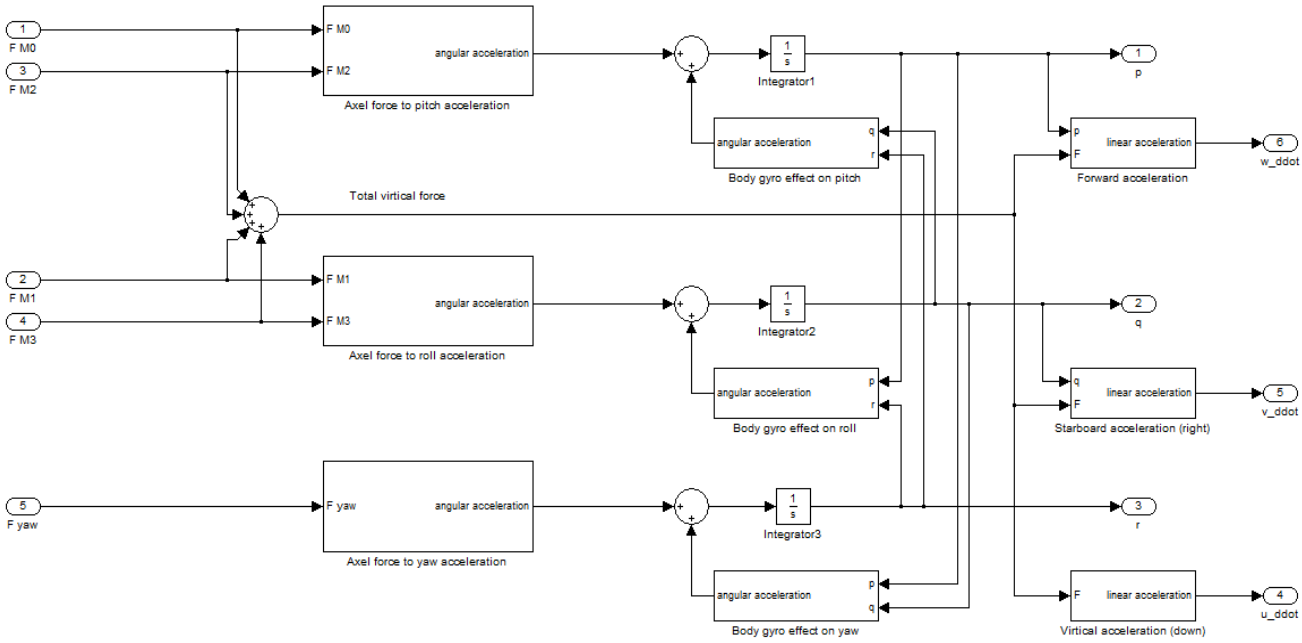


Figure 2: Flight dynamics for the model developed by the student group

aircraft. The students were not specifically prepared for this task, and were distracted in the project period by other side-activities such as group administration; meaning that 100 hours per person for a 3 person group is deemed sufficient for the laboratory exercise is more focused on the specific tasks.

In a 10 ECTS module, a student is expected to put a total of about 300 hours effort, including lectures, exercises, self-study, laboratory work preparing for exam and exam. It is the authors' plan that in this module in multivariable control theory a student would typically spend 60-70 hours on lectures and exercises, and equally much on self-study. Adding the exam preparation and exam, this leaves between 130 and 150 hours for the laboratory exercise project. An average student would then typically need more hours to complete the project at an acceptable level then did the student group in the trial. It is the authors' opinion that the laboratory is within the student's ability if groups are formed with about 3 students in each. In order to give all students a feeling of satisfaction the assignment will be divided into different levels, where the first assignment would be to stabilise the aircraft around one axis at a time, before moving on to the multivariable problem. This will allow the students to gradually approach the multivariable problem, while acquiring an intuitive understanding of the behaviour of the aircraft in relation to the models developed.

Figure 2 shows the Simulink diagram of the flight dynamics as developed by the student group doing the trial case with the four rotor helicopter. For controlling the process, the student group decided mainly to use a set of PI/PID controllers with decoupling, or decentralised control.

CONCLUSION

In this paper a laboratory exercise setup allowing for hands-on training for students in multivariable feedback control has been presented. Using the four-rotor helicopter, the exercise can be adapted to be challenging enough for any student by letting the students develop models on their own, in search of a better performing control loop, while at the same time the less skilled students might fall back on simpler and well known models in order to stabilise the aircraft. The exercise can be done evolutionary, in the sense that the students are given basic tasks at the start, like parameter tuning, before engaging in more complex elements of the control system design. In this way the students can be given tasks that are manageable at their own level, and the exercise setup is therefore adaptable to each student's needs and level of skills.

The laboratory has also been shown to give students challenges similar to what they might experience in their career as control engineers, like model errors, parameter uncertainty and limited processing time and power for the control loop. Another important aspect is how students handle their own limited experience with regard to the handling of unsuccessful control loop implementation and the strategies for solving these issues, e.g. the decision between of more parameter tuning or redesign of the controller or model.

The laboratory setup follows the syllabus of a control-theory module, and the steps necessary to complete the exercise are described with an estimate of 130 and 150 hours needed to complete the assignment. The test of the laboratory setup with a group of students seems to confirm the estimates given.

In addition to the shown advantages in learning outcome and engagement for the students of the described setup, the complete system is created with the use of low cost components making this setup an attractive alternative for institutions in need of training students in multivariable control theory.

REFERENCES

[1] J. Detore and S.Martin. Multi Rotor Options For Heavy Lift [Online]. Available: http://papers.sae.org/791089

[2] A. Das, *et al.*, "Dynamic inversion of quadrotor with zero-dynamics stabilization," in *Control Applications, 2008. CCA 2008. IEEE International Conference on*, 2008, pp. 1189-1194.

[3] A. Zul Azfar and D. Hazry, "A simple approach on implementing IMU sensor fusion in PID controller for stabilizing quadrotor flight control," in *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*, 2011, pp. 28-32.

[4] H. Voos and H. Bou-Ammar, "Nonlinear tracking and landing controller for quadrotor aerial robots," in *Control Applications (CCA), 2010 IEEE International Conference on*, 2010, pp. 2136-2141.

[5] T. Dierks and S. Jagannathan, "Output Feedback Control of a Quadrotor UAV Using Neural Networks," *Neural Networks, IEEE Transactions on,* vol. 21, pp. 50-66, 2010.

[6] W. Jun, *et al.*, "RBF-ARX model-based modeling and control of quadrotor," in *Control Applications (CCA), 2010 IEEE International Conference on*, 2010, pp. 1731-1736.

[7] TI. (2011, 7.7). *C6747/45/43 Power Consumption Summary*. Available: http://processors.wiki.ti.com/index.php/C6747/45/43_Power_Consumption_Summary

[8] dSPACE. (2011, 7.7). *DS1104 R&D Controller Board* Available:

http://www.dspace.de/en/pub/home/products/hw/singbord/ds1104.cfm

[9] Gonza*, et al.*, "A New Nonlinear PI/PID Controller for Quadrotor Posture Regulation," in *Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2010*, 2010, pp. 642-647.

[10] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design, 2nd Edition*, 2nd ed.: Wiley, 2005.

[11] TI. *TMS320C6742 Fixed/Floating-Point DSP*. Available: http://focus.ti.com/lit/ds/symlink/tms320c6742.pdf

[12] M. Y. Amir and V. Abbass, "Modeling of Quadrotor Helicopter Dynamics," in *Smart Manufacturing Application, 2008. ICSMA 2008. International Conference on*, 2008, pp. 100-105.

[13] A. A. Mian and D. Wang, "Nonlinear Flight Control Strategy for an Underactuated Quadrotor Aerial Robot," in *Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on*, 2008, pp. 938-942.

Variable Time Step Dynamics With Choice

Lev Kapitanski
Department of Mathematics, University of Miami
Coral Gables, FL 33124, USA

and

Sanja Gonzalez Živanović
Department of Mathematics and Computer Science, Barry University
Miami Shores, FL 33161, USA

ABSTRACT

We develop a simple and general approach to study long term behavior of deterministic systems that switch regimes and have dwell times of variable length. We investigate the results of all possible as well as restricted, and/or controlled, switchings. To analyze all these situations, we introduce the notions of variable time step dynamics with choice and variable time step iterated function systems. We establish general sufficient conditions for the existence of global compact attractors of such systems and describe how they are related.

Keywords
Global Attractors, Dynamical Systems, Control, Dwell Time, and Iterated functions system.

1. INTRODUCTION

We are interested in time evolution of systems that can and do switch their modes (regimes) of operation at discrete moments of time. The intervals between switching may, in general, vary. The number of modes (regimes) may be finite or infinite. Such systems are very common in life. In control theory the systems we talk about are often called hybrid, see [5, 6]. The switching times and the switching of the regimes may be deterministic or random. Here we will discuss the deterministic case only. A few examples may be helpful.

The first example is a switched system, i.e., a special case of a hybrid system governed by a finite set of systems of ordinary differential equations,

$$\dot{x} = f_{w(j)}(x) \quad \text{on the interval } [t_j, t_{j+1}), \qquad (1)$$

where $t_0 < t_1 < \dots$ are the switching times, and each $f_{w(j)}$ is taken from a finite set of functions $\{f_0, \dots, f_{N-1}\}$. Here $w(\cdot)$ is a regime switching function, it maps the non-negative integers into the label set of the available regimes, $\{0, 1, \dots, N-1\}$. Denote the states at the regime switching times by $x_j = x(t_j)$ and the time intervals between the switching (dwell time) by $h(j) = t_{j+1} - t_j$. We can write the transition from x_j to x_{j+1} symbolically as $x_{j+1} = S^{h(j)}_{w(j)}(x_j)$. Here $S^{h(j)}_{w(j)}$ is a transformation that solves system (1), i.e., for every y, $S^{h(j)}_{w(j)}(y)$ is the solution $x(t)$ of the system $\dot{x} = f_{w(j)}(x)$ with the initial condition $x(0) = y$, evaluated at time $t = h(j)$.

The second example is a discrete version of (1),

$$x_{j+1} = x_j + h(j)\, f_{c(j)}(x_j)\,, \qquad (2)$$

where $h(j)$ is a variable, in general, time step. (In fact, (2) does not have to be related to (1) and could have a totally independent origin.) Again, we can write the transformation from the state x_j to the next state, x_{j+1}, as $x_{j+1} = S^{h(j)}_{w(j)}(x_j)$, but now $S^{h(j)}_{w(j)}$ is given explicitly by the right hand side of (2). In both examples, x takes on values in some region of a finite dimensional space $\mathbb{R}^d$. However, it is not hard to come up with meaningful examples of infinite dimensional systems with continuous or discrete time where parameters switch during the evolution.

The switching times and the regime switching function can be given in advance, or can be generated step-by-step depending on external or internal information. In the latter case, $h(j+1)$ and $w(j+1)$ may depend on the state x_j, on the regime $w(j)$, or even on x_{j-1}, $w(j-1)$, etc. The rule generating $h(j+1)$ and $w(j+1)$ can be viewed as a control law.

In the present report we address the long term behavior of systems of the type just described. In

particular, we are interested whether or not a system possesses global attractor. A global attractor is a compact subset of the state space that attracts all bounded sets (and not just individual points). This is an important notion because in practice the states of the system are known only approximately. If the global attractor consists of one point, this point (state) is automatically asymptotically stable. The global attractors of nonlinear dissipative systems may have a very complicated geometry (as, e.g., the Lorenz attractor).

There is a large literature on asymptotic stability (of one state, the origin, $x = 0$) of both linear and nonlinear switched systems, see e.g. [17, 19] and references there. Outside of engineering applications, we should mention an interesting paper [18] on using switching controls in epidemic models. A number of papers are devoted to attractors of switched systems, mostly of the form (1) in the finite-dimensional setting, see e.g. [7, 8, 14].

In [11, 12, 22] we developed a simple and general approach to dynamics of fixed time step switching systems. We use the term *dynamics with choice* because we handle the systems more general than (1) and (2). In this paper we extend our approach to deal with variable time step. We give sufficient conditions for the existence of global compact attractors. One of those conditions may be hard to check in practice as we show with an example.

2. DYNAMICS WITH CHOICE

Consider a set X (the state space) and a collection of maps $S_j^\tau : X \to X$. The lower index, j, indicates the chosen regime. All available regimes are labeled by the points of some set $\mathcal{J}$. The upper index, τ, is interpreted as the dwell time. The values of τ are taken from some interval $\mathcal{I} = [a, b]$, $0 < a \le b$. Thus, $S_j^\tau(x)$ is interpreted as the result of the regime j acting on the state x over the time period τ. Switching regimes and dwell times leads to the dynamics

$$x_{n+1} = S_{j_n}^{\tau_n}(x_n)\,. \tag{3}$$

We would like to be able to work with all the trajectories $x_0, x_1, x_2, \dots$ generated this way with arbitrary $j_n \in \mathcal{J}$ and $\tau_n \in \mathcal{I}$. One can think of several different mathematical formalizations of this dynamics. We will use two. Denote by $\Sigma_\mathcal{J}$ (respectively, $\Sigma_\mathcal{I}$) the space of maps from the non-negative integers $\mathbb{Z}_{\ge 0}$ into $\mathcal{J}$ (respectively, $\mathcal{I}$). Sometimes it is convenient to view a map, $w \in \Sigma_\mathcal{J}$, as a one-sided infinite string (word) of symbols $w(0)$, $w(1)$, $w(2)$, ... from $\mathcal{J}$; we write $w = w(0)w(1)w(2)\dots$. Sometimes we refer to w as a regime switching function. In a similar fashion, $h \in \Sigma_\mathcal{I}$ can be viewed as $h(0)h(1)h(2)\dots$, and we call h a dwell time function, or a time step function. The sets $\Sigma_\mathcal{J}$ and $\Sigma_\mathcal{I}$ are equipped with a shift operator, σ, that acts as follows: $\sigma(w)(n) = w(n+1)$. [In the language of symbolic dynamics, $\Sigma_\mathcal{J}$ is a full one-sided shift over the alphabet $\mathcal{J}$, see [13].] Now, we view equation (3) as a "non-autonomous system" and apply the well-known skew-product construction, [20], to obtain an "autonomous system."

Definition 1. *The variable time step dynamics with choice on X associated with the maps S_j^τ, $j \in \mathcal{J}$, $\tau \in \mathcal{I}$, is the discrete time dynamics generated on the product space $\mathfrak{X} = X \times \Sigma_\mathcal{J} \times \Sigma_\mathcal{I}$ by the iterations of the map*

$$\mathfrak{S} : (x, w, h) \mapsto (S_{w(0)}^{h(0)}(x), \sigma(w), \sigma(h))\,. \tag{4}$$

The second formalization of the deterministic dynamics (3) is motivated by iterated function systems (IFS), [2, 3, 9]. Here one works not with points of X, but with its subsets. Define the maps F^τ, $\tau \in \mathcal{I}$, acting on the subsets of X according to the rule

$$F^\tau(A) = \bigcup_{j \in \mathcal{J}} S_j^\tau(A)\,. \tag{5}$$

Definition 2. *The variable time step iterated function system on X associated with the maps S_j^τ, $j \in \mathcal{J}$, $\tau \in \mathcal{I}$, is the discrete time dynamics generated on the product space $\mathfrak{Y} = 2^X \times \Sigma_\mathcal{I}$ by the iterations of the map*

$$\mathfrak{F} : (A, h) \mapsto (F^{h(0)}(A), \sigma(h))\,. \tag{6}$$

This notion of a variable time step iterated function system seems to be new. In addition, we propose another useful variant of a variable time step iterated function system *inside* the state space.

Definition 3. *The variable time step iterated function system* **inside** *X associated with the maps S_j^τ, $j \in \mathcal{J}$, $\tau \in \mathcal{I}$, is the discrete time dynamics generated on 2^X by the iterations of the map*

$$\mathcal{F} : A \mapsto \bigcup_{\tau \in \mathcal{I}} F^\tau(A)\,. \tag{7}$$

3. GLOBAL ATTRACTORS

Consider a discrete time dynamics generated on a metric space Y by iterations of a map $\Phi : Y \to Y$. The global compact attractor of such semi-dynamical system is the smallest compact set $\mathfrak{A} \subset Y$ that attracts all bounded sets. The

latter means that for every $A \subset Y$ of finite diameter, and for any open neighborhood U of $\mathfrak{A}$, there exists N such that $\Phi^n(A) \subset U$ for all $n \geq N$.

Both definitions in the previous section present the variable time step dynamics (3) in the form of a discrete semi-dynamical system. The theory of attractors for continuous time as well as for discrete semi-dynamical systems is well developed, see e.g. [16, 21] and a brief account in [10]. We will apply this theory in our situation. To this end we first need to outfit the state space X and the maps S_j^τ with certain properties. The following will be our basic assumptions. We assume that X is a complete metric space with metric d_X, that the set $\mathcal{J}$ is a metric compact with metric $d_\mathcal{J}$, and that each map S_j^τ is continuous and bounded. Also, we assume that there is a measure of noncompactness, ψ, so that each map S_j^τ is ψ-condensing. Recall that a function $\psi : 2^X \to [0, +\infty]$ is a measure of noncompactness on X iff (see [1])

(i) $\psi(A) = 0$ iff A is relatively compact;

(ii) If $A_1 \subset A_2$, then $\psi(A_1) \leq \psi(A_2)$;

(iii) $\psi(A_1 \cup A_2) = \max\{\psi(A_1), \psi(A_2)\}$;

(iv) There exists a constant $c(\psi) > 0$ such that

$$|\psi(A_1) - \psi(A_2)| \leq c(\psi)\,\mathrm{dist}(A_1, A_2),$$

where $\mathrm{dist}(A_1, A_2)$ is the Hausdorff distance between A_1 and A_2.

For example, the Kuratowski measure of noncompactness of a set A (denoted $\alpha(A)$) is the infimum of the numbers $\epsilon > 0$ such that A admits a finite cover by sets of diameter less than ϵ.

Recall that a map $S : X \to X$ is ψ-condensing (condensing with respect to ψ) iff $\psi(S(A)) \leq \psi(A)$ for any bounded A, and $\psi(S(A)) < \psi(A)$ if $\psi(A) > 0$ (i.e., if $\overline{A}$ is not compact). The compact maps $X \to X$ are ψ-condensing. Also, in a Banach space X, any map of the form *contraction + compact* is ψ-condensing for any ψ.

Now we introduce two additional assumptions. The first assumption is always true if the number of regimes (the set $\mathcal{J}$) is finite and all dwell times are the same.

Assumption 1. *For any closed, bounded set $A \subset X$, the maps S_j^τ, restricted to A, depend uniformly continuously on j and τ. More precisely, given a closed, bounded A, for every $\epsilon > 0$ there exist $\delta_\mathcal{J} > 0$ and $\delta_\mathcal{I} > 0$ such that*

$$\sup_{x \in A} d_X(S_{j_1}^{\tau_1}(x), S_{j_2}^{\tau_2}(x)) \leq \epsilon$$

provided $d_\mathcal{J}(j_1, j_2) \leq \delta_\mathcal{J}$ and $|\tau_1 - \tau_2| = d_\mathcal{I}(\tau_1, \tau_2) \leq \delta_\mathcal{I}$.

In the second assumption we use the following notation. Given a one-sided infinite string $h \in \Sigma_\mathcal{I}$, $h[n]$ denotes the finite word made by the first n symbols in h, i.e., $h[n] = h(0)h(1)\cdots h(n-1)$, and $\|h[n]\|$ stands for $h(0) + h(1) + \cdots + h(n-1)$. Also, we denote by $S_{w[n]}^{h[n]}$ the composition

$$S_{w[n]}^{h[n]} = S_{w(n-1)}^{h(n-1)} \circ \cdots \circ S_{w(1)}^{h(1)} \circ S_{w(0)}^{h(0)}\,.$$

Assumption 2. *Assume there is a closed, bounded set $\mathbf{B} \subset X$ such that for every bounded $A \subset X$ there exists $T(A) > 0$ such that $S_{w[n]}^{h[n]}(A) \subset \mathbf{B}$, for any word $h \in \Sigma_\mathcal{I}$ such that $\|h[n]\| > T(A)$ and any word $w \in \Sigma_\mathcal{J}$.*

This assumption requiring the existence of an absorbing set is necessary for the existence of a global attractor.

We are ready to state our main result, but first make a small adjustment to the definition of the variable time step iterated function system. We modify the map F^τ in (5) by taking the closure of the union as follows

$$F^\tau(A) = \overline{\bigcup_{j \in \mathcal{J}} S_j^\tau(A)}\,. \tag{8}$$

Similarly, we adjust the map $\mathcal{F}$ in (7):

$$\mathcal{F}(A) = \overline{\bigcup_{\tau \in \mathcal{I}} F^\tau(A)}\,. \tag{9}$$

Also, in Definition 2, we replace the space 2^X by the space of all *closed nonempty* subsets of X, which we denote by $\mathfrak{C}(X)$. The Hausdorff distance furnishes a complete metric to this space, [15].

Theorem 4. *Under the above assumptions on X, $\mathcal{J}$, $\mathcal{I}$, and S_j^τ, we consider the corresponding variable time step dynamics with choice and iterated function system.*

1. *The variable time step dynamics with choice possesses a global compact attractor. This attractor, $\mathfrak{A}$, is a subset of the product space $X \times \Sigma_\mathcal{J} \times \Sigma_\mathcal{I}$, and is invariant under the map $\mathfrak{S}$ defined in (4), $\mathfrak{S}(\mathfrak{A}) = \mathfrak{A}$.*

2. *The variable time step iterated function system possesses a global compact attractor. This attractor, $\mathfrak{B}$, is a subset of the product space $\mathfrak{C}(X) \times \Sigma_\mathcal{I}$, and is invariant under the map $\mathfrak{F}$ defined in (6), $\mathfrak{F}(\mathfrak{B}) = \mathfrak{B}$.*

3. *The variable time step iterated function system inside X possesses a global compact attractor. This attractor, $\mathcal{K}$, is a subset of $\mathfrak{C}(X)$ and is invariant under the map $\mathcal{F}$.*

4. *Define the set $K \subset X$ as the union of all the points (which are closed subsets of X) of the attractor $\mathcal{K}$. This is a compact set. The attractors $\mathfrak{A}$ and $\mathfrak{B}$ are related to the fractal K in the following sense:*

$$\mathfrak{B} = K \times \Sigma_{\mathcal{I}}\,, \quad \mathfrak{A} = K \times \Sigma_{\mathcal{J}} \times \Sigma_{\mathcal{I}}\,. \quad (10)$$

The proof of this theorem relies on the following lemma.

Lemma 5. *Under the above assumptions, the maps $\mathfrak{S}$, $\mathfrak{F}$, and $\mathcal{F}$ are condensing with respect to the appropriate measures of noncompactness on the spaces $X \times \Sigma_{\mathcal{J}} \times \Sigma_{\mathcal{I}}$, $\mathfrak{C}(X) \times \Sigma_{\mathcal{I}}$, and $\mathfrak{C}(X)$ respectively.*

The measures of noncompactness in the lemma are defined as follows. If Y and Z are two complete metric spaces, with some measures of noncompactness ψ_Y and ψ_Z, we construct a measure of noncompactness on the product space $Y \times Z$ as follows:

$$\psi_{Y\times Z}(A) = \max\{\psi_Y(A_Y), \psi_Z(A_Z)\},$$

where A_Y and A_Z are the projections of $A \subset Y \times Z$ on the spaces Y and Z. On the space $\mathfrak{C}(Y)$ we define a measure of noncompactness as follows:

$$\psi_{\mathfrak{C}(Y)}(\mathcal{N}) = \psi_Y\left(\bigcup_{A\in\mathcal{N}} A\right).$$

Our approach allows us to easily handle *dynamics with restricted choice in the variable time step setting*. By restricted choice we mean that certain sequences of regime and/or dwell times switches may be forbidden. Mathematically this means that the full shift $\Sigma_{\mathcal{J}}$ may be replaced by a subshift (i.e., closed and shift-invariant subset of $\Sigma_{\mathcal{J}}$) $\Lambda_{\mathcal{J}}$. Similarly, $\Sigma_{\mathcal{I}}$ may be replaced by a subshift $\Lambda_{\mathcal{I}}$. The Definitions 1 and 2 with these replacements give rise to the variable time step dynamics with restricted choice and restricted iterated function systems. Theorem 4 does hold for such systems.

4. CONTROL IN DYNAMICS WITH CHOICE

Within our general framework of Sec. 2 we see many opportunities to introduce controls. For example, depending on you goals, choose a map $u : X \times \mathcal{J} \times \mathcal{I} \to \mathcal{J} \times \mathcal{I}$ that will determine the next regime and the next dwell time based on the current state, regime, and dwell time. Denoting the $\mathcal{J}$ and $\mathcal{I}$ components of u by $u_{\mathcal{J}}$ and $u_{\mathcal{I}}$, we have

$$\begin{aligned} j_{n+1} &= u_{\mathcal{J}}(x_n, j_n, \tau_n)\,, \\ \tau_{n+1} &= u_{\mathcal{I}}(x_n, j_n, \tau_n)\,. \end{aligned} \quad (11)$$

Note that the system of equations (3) and (11) defines a discrete time dynamics on the space $X \times \mathcal{J} \times \mathcal{I}$.

Theorem 6. *If the assumptions of Sec. 3 are satisfied, then for any continuous control map u, the system governed by equations (3) and (11) possesses a global compact attractor, $\mathcal{M}$, in $X \times \mathcal{J} \times \mathcal{I}$. If $(x, j, \tau) \in \mathcal{M}$, then, setting $x_0 = x$, $w(0) = j$, and $h(0) = \tau$, and defining recursively*

$$\begin{aligned} x_{n+1} &= S^{h(n)}_{w(n)}(x_n), \\ w(n+1) &= u_{\mathcal{J}}(x_n, w(n), h(n)), \\ h(n+1) &= u_{\mathcal{I}}(x_n, w(n), h(n))\,, \end{aligned}$$

we obtain the infinite strings $w \in \Sigma_{\mathcal{J}}$ and $h \in \Sigma_{\mathcal{I}}$. It turns out that (x, w, h) belongs to the attractor $\mathfrak{A}$ of the corresponding variable time step dynamics with choice.

5. EXAMPLE

In this section we give an example of a two-dimensional switched system that shows that, in practice, checking Assumption 2 of Sec. 3 may be quite tricky. We build the example by first patching a couple of systems of ODEs on the plane and then transplanting the result to an infinite cylinder.

The first system is this:

$$\begin{aligned} \dot{x} &= -\frac{y}{x^2+y^2} + x\,\left(1 - x^2 - y^2\right) \\ \dot{y} &= \frac{x}{x^2+y^2} + y\,\left(1 - x^2 - y^2\right) \end{aligned} \quad (12)$$

The origin is its unstable focus and $x^2 + y^2 = 1$ is the stable limit-circle. In the neighborhood of the origin, we shall glue in two linear systems between which the switching will occur. Those systems are

$$\begin{aligned} \dot{x} &= \epsilon x - 4y \\ \dot{y} &= x + \epsilon y \end{aligned} \quad (13) \qquad \begin{aligned} \dot{x} &= \epsilon x - y \\ \dot{y} &= 4x + \epsilon y, \end{aligned} \quad (14)$$

where ϵ is a positive parameter. This type of systems is well known in the stability theory of switched systems, [17]. For each (13) and (14), the origin is an unstable fixed point, but with the right switching between (13) and (14) (e.g., using (13) if $xy > 0$ and using (14) if $xy < 0$), the origin becomes globally asymptotically stable, see

[4]. Transfer systems (12), (13), and (14) to a cylinder by changing the coordinates x and y to s and θ, where $x = e^s \cos\theta$, $y = e^s \sin\theta$. In the (s, θ) coordinates system (12) reads

$$\dot{s} = 1 - e^{2s}\,, \quad \dot{\theta} = e^{-2s}\,, \tag{15}$$

system (13) reads

$$\dot{s} = \epsilon - \frac{3}{2}\,\sin(2\theta), \quad \dot{\theta} = \frac{1}{2}\,(5 - 3\cos(2\theta)), \tag{16}$$

and system (14) reads

$$\dot{s} = \epsilon + \frac{3}{2}\,\sin(2\theta), \quad \dot{\theta} = \frac{1}{2}\,(5 + 3\cos(2\theta)). \tag{17}$$

The half of the cylinder with $s < 0$ corresponds to the interior of the unit circle, and the half with $s > 0$ - to its exterior. We want system (15) to operate in the region $s \geq -1$ and systems (16) and (17) to operate where $s < -4$. This will be achieved by adding the right sides of (15) with coefficient $\zeta(s)$ to the corresponding right sides of systems (16) or (17) with coefficient $(1 - \zeta(s))$, where $\zeta(s)$ is a smooth, monotone increasing function such that $\zeta(s) = 0$ for $s \leq -4$ and $\zeta(s) = 1$ for $s \geq -1$. We write the resulting systems symbolically as

$$\frac{d}{dt}\begin{bmatrix} s \\ \theta \end{bmatrix} = f_{w(j)}(s, \theta)\,, \tag{18}$$

where $w(j) = 1$ corresponds to (16) combined with (15), and $w(j) = 2$ corresponds to (17). Note, that on the cylinder, system (15) and each of the two systems (18) have the circle $s = 0$ as the global compact attractor. However, depending on the strategy of switching between system number 1 and system number 2 of (18), the attractor may survive, or there may be no global attractor. To understand this we only need to understand how switching works for systems (13) and (14). Indeed, consider the corresponding switched system

$$\frac{d}{dt}\begin{bmatrix} x \\ y \end{bmatrix} = \epsilon \begin{bmatrix} x \\ y \end{bmatrix} + A_{w(j)} \begin{bmatrix} x \\ y \end{bmatrix} \tag{19}$$

where $w(j) = 1$ or 2 and

$$A_1 = \begin{bmatrix} 0 & -4 \\ 1 & 0 \end{bmatrix}\,,\; A_2 = \begin{bmatrix} 0 & -1 \\ 4 & 0 \end{bmatrix} \tag{20}$$

We will exhibit an interval $[a, b] \subset (0, +\infty)$ such that for any choice of time steps τ_j from this interval and for any sequence of regime switching $w(j)$, all trajectories starting not at the origin go to infinity. This will imply that on the cylinder, all trajectories of system (18) starting at the points with $s < 0$ (in fact, any disc in this region) will move in the direction of the circle $s = 0$. The trajectories starting at the points with $s > 0$ will move to the same circle because in the xy-coordinates $s = 0$ is the limit-circle of system (12). Thus, the variable time step system (18) with $\mathcal{J} = \{1, 2\}$ and $\mathcal{I} = [a, b]$ does have a global attractor. In the setting of Sec. 2, the global attractor of the variable time step dynamics with choice is $\mathfrak{A} = \{s = 0\} \times \Sigma_{\mathcal{J}} \times \Sigma_{\mathcal{I}}$ and the fractal K is the circle $s = 0$.

Choosing the wrong interval for the time steps will destroy this picture. Some trajectories of system (19) will converge to the origin. On the cylinder this means $s \to -\infty$, and there is no global attractor for (18). We will give an example of a bad interval.

To find a "good" interval, it suffices to find those τ for which the solutions

$$e^{\epsilon\tau}\, e^{\tau A_1} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \text{ and } e^{\epsilon\tau}\, e^{\tau A_2} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

of (13) and (14) are both farther from the origin than the initial condition $[x_0, y_0]^T$ by some factor. In other words,

$$e^{\epsilon\tau}\, \|e^{\tau A_{1,2}} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}\| \geq \gamma\, \| \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}\| \tag{21}$$

for some $\gamma > 1$ and all $[x_0, y_0]^T$ on the unit circle. Just for an example of some "good" interval, we find that

$$\|e^{\tau A_{1,2}} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}\|^2 \geq \frac{1}{2}\, \| \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}\|^2$$

when $5/9 < \cos(4\tau) \leq 1$, and take those τ, which in addition guarantee that $e^{2\epsilon\tau}/2 \geq \gamma^2 > 1$, for some $\gamma > 1$. If $\epsilon = 0.3$, we may take $\tau \in \mathcal{I} = [1.326, 1.810]$.

A different argument is needed to find the "bad" intervals. Note, that the eigenvalues of the matrices $e^{\tau A_j}$ are $e^{\pm 2\tau i}$, and the trajectories corresponding to sequences without switching do not go to the origin. However, both products $e^{\tau A_1} e^{\tau A_2}$ and $e^{\tau A_2} e^{\tau A_1}$ have real, negative eigenvalues, and the largest eigenvalue, $\lambda(\tau)$, is such that $|\lambda(\tau)| < 1$, provided τ satisfies $|\cos(2\tau)| < 3/5$. Thus, given an $\epsilon > 0$, we find intervals of τ such that $e^{2\epsilon\tau}|\lambda(\tau)| < \delta < 1$, for some δ. Those intervals are bad. Denote $M_1^\tau = e^{2\epsilon\tau} e^{\tau A_1} e^{\tau A_2}$ and $M_2^\tau = e^{2\epsilon\tau} e^{\tau A_2} e^{\tau A_1}$. We can choose any sequence τ_j from any of the bad intervals and look at the trajectories $[x_{j+1}, y_{j+1}]^T = M_{w(j)}^{\tau_j} [x_j, y_j]^T$, for any infinite string w of 1's and 2's. The distance from the origin to the consecutive images of any circle $(x_0)^2 + (y_0)^2 = r^2$ will go to 0 exponentially fast (although some points in the images

will go to infinity). If $\epsilon = 0.3$, and $\mathcal{I} = [0.5, 1.06]$, the dynamics with choice does not have a global compact attractor.

6. CONCLUSION

To investigate the long term behavior of deterministic systems that switch regimes at discrete moments of time and have dwell times of variable length, and to investigate the results of all possible as well as restricted, and/or controlled, switchings, we have introduced the notions of a variable time step dynamics with choice and a variable time step iterated function system. Using these notions as a rigorous foundation, we establish general sufficient conditions for the existence of global compact attractors of such systems and describe how they are related. Unlike in the case of systems without switching, one of the conditions (Assumption 2) may be hard to check in practice. We give a two-dimensional example that illustrates this difficulty.

7. REFERENCES

[1] R. R. Akhmerov, M. I. Kamenskiĭ, A. S. Potapov, A. E. Rodkina, B. N. Sadovskiĭ, **Measures of noncompactness and condensing operators**, Operator Theory: Advances and Applications, 55. Birkhuser Verlag, Basel, 1992.

[2] M. F. Barnsley, **Fractals everywhere**, Second edition. Academic Press Professional, Boston, MA, 1993.

[3] M. F. Barnsley, **Superfractals**, Cambridge University Press, Cambridge, 2006.

[4] M. Branicky, "Stability of switched and hybrid systems", **Proc. of the 33rd Conference on Design and Control**, Lake Buena Vista, FL, 1994, pp. 3498-3503.

[5] M. Branicky, **Studies in Hybrid Systems: Modeling, Analysis, and Control**, Sc.D. thesis, MIT, 1995

[6] M. Branicky, **Introduction to hybrid systems**, in D. Hristu-Varsakelis and W.S. Levine (eds.), Handbook of Networked and Embedded Control Systems, pp. 91-116. Birkhauser, 2005.

[7] D. N. Cheban, "Compact Global Attractors of Control Systems", **Journal of Dynamical and Control Systems**, vol.16 (2010), no.1, pp. 23–44.

[8] D. N. Cheban, **Global Attractors of Set-Valued Dynamical and Control Systems.** Nova Science Publishers Inc, New York, 2010.

[9] J. E. Hutchinson, "Fractals and self-similarity", **Indiana Univ. Math. J. 30**, 1981, no. 5, pp. 713–747.

[10] L. Kapitanski, I. N. Kostin, "Attractors of nonlinear evolution equations and their approximations", (Russian) **Algebra i Analiz 2**, 1990, no. 1, pp. 114–140; translation in **Leningrad Math. J. 2**, 1991, no. 1, pp. 97–117

[11] L. Kapitanski, S. Živanović,: "Dynamics with choice", **Nonlinearity 22**, 2009, pp. 163–186.

[12] L. Kapitanski, ; S. Živanović, : "Dynamics with a range of choice", **Reliable Computing**, vol. 15, no. 4, 201, pp. 290-299.

[13] B. P. Kitchens, **Symbolic dynamics. One-sided, two-sided and countable state Markov shifts**, Universitext: Springer-Verlag, Berlin, 1998.

[14] P. E. Kloeden, "Nonautonomous attractors of switching systems", **Dynamical Systems**, vol. 21, no. 2, 2006, pp. 209–230.

[15] K. Kuratowski, **Topology**, Volume I. New York - London - Warszawa, Academic Press: Polish Scientific Publishers, 1966.

[16] O. A. Ladyzhenskaya, **Attractors for semigroups and evolution equations**, Lezioni Lincee. [Lincei Lectures] Cambridge University Press, Cambridge, 1991.

[17] D. Liberzon, **Switching in systems and control**, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 2003.

[18] X. Liu, P. Stechlinski, "Pulse and constant control schemes for epidemic models with seasonality", **Nonlinear Analysis: Real World Applications**, 12, 2011, pp. 931–946.

[19] M. Margaliot, "Stability analysis of switched systems using variational principles: an introduction", **Automatica J. IFAC**, 42, 2006, no. 12, pp. 2059–2077.

[20] G. R. Sell, "Nonautonomous differential equations and topological dynamics", **I, II, Trans. Amer. Math. Soc.**, 127, 1967, pp. 241–262, pp. 263–283.

[21] G. R. Sell, Y. You, **Dynamics of evolutionary equations**, Applied Mathematical Sciences, 143. Springer-Verlag, New York, 2002.

[22] S. Živanović, **Attractors in Dynamics with Choice**, PhD Thesis, University of Miami, Coral Gables, Fl, USA, 2009.

Extended LZCode Algorithm for the fast Binary Code Decompression in Mobile Devices

Hyunchul Lee, Kangseok Kim, Okkyung Choi, Taeshik Shon, Hongjin Yeh, Manpyo Hong
Dept. of Knowledge Information Security, Graduate School of Ajou University, San 5, Woncheon-dong, Yeongtong-Gu, Suwon, Keonggi-do, Korea 443-749
E-Mail: {deletenim, kangskim, okchoi, hjyeh, tsshon, mphong}@ajou.ac.kr

Abstract

Embedded devices such as mobile phones use NAND Flash Memory for cost reduction. Space saving and booting time reduction can also be expected by compressing program code and storing it in NAND Flash Memory. Before Program codes are executed essentially it is loaded on main memory. At that point, loading time is to be sum of time about compression and decompression data read from NAND Flash Memory.

Binary Code means general system data and such Binary Code is sometimes loaded on main memory and executed or utilized as important data.

When an embedded device that uses compression algorithm reads data from NAND Flash Memory into main memory, a compressor is triggered to compress or decompress data in reading or writing from or onto NAND Flash Memory.

Using a compressor allows more efficient utilization of NAND Flash Memory space, and the booting time will be the sum of the time to read in compressed data from NAND Flash Memory and the time to decompress those data. In conclusion, a more efficient algorithm can be used by comparing the time to read original data as it is and the sum of time to read in compressed data and decompress it. This will allow selecting and using an optimized compression algorithm for fast booting speed and efficient memory space saving as well.

Therefore faster compression and decompression speed is to be important factor on the embedded device. Generally, in case of mobile device, in contrast with desk top because of fewer battery capacity, limited processor and NAND Flash Memory size saving program, it didn't show optimized performance.

In this paper, we progress our research about lossless compression algorithm and present one half improved algorithm for decompression speed comparing with LZCode suitable for mobile system. We increased compression speed by eliminating the table and relational operators utilized by LZCode and replacing with an algorithm corresponding to such eliminated data.

Key Words: NAND Flash Memory, Compression, Decompression, Compressor, Embedded devices. Mobile devices

References

[1] Hyojin Kim, Youjip Won, Yohwan Kim. “MNFS : Design of Mobile Multimedia File System based on NAND FLASH MEMORY Memory”, KOREA INFORMATION SCIENCE SOCIETY, Journal of KIISE : Computer Systems and Theory, Vol.35 No.11·12, pp.497-508, Dec. 2008.

[2] Y. Kim, Y. Wee, "A Program Code Compression Method with Very Fast Decoding for Mobile Devices," Journal of KIISE : Software and Applications, Vol.37, No.10, pp.851-858, Nov. 2010.

[3]P.Deutsch, “DEFLATE Compressed Data Format Specification version 1.3”, In http://www.ietf.org.IETF

RFC1951, Mar.1996.

[4]Taehwa Kim, Youngcheul Wee, "A New Code Compression Method for Compressed Firmware Over the Air on Mobile Devices", IEEE Transactions on Consumer Electronics, Vol. 56, No. 4, Nov.2010.

[5]A. Wolfe and A. Chanin, "Executing compressed progr ams on an embedded RISC architecture", Proc. 25th Ann. International Symposium on Micro architecture, pp.81-91, Dec.1992.

ENGINEERING THE CLOTHING INDUSTRY TOWARDS COMPETITIVE ADVANTAGE: A MANAGERIAL DILEMMA

Kem Ramdass

Senior Lecturer, Faculty of Arts, Design and Architecture, University of Johannesburg,

Auckland Park Bunting Road Campus, Auckland Park, Johannesburg, 2092

ABSTRACT

The global economy which is enhanced through changing technologies of all types is pressurizing organisations to improve productivity of their business processes. Competition is forcing organisations to focus their energy on "core competencies." Like many industries, the clothing industry is witnessing changes in technology, diversification of labour, managerial implications while competing on the global market. The South African clothing and textile industry has the potential to create jobs, but this potential has been steadily diminishing over the last ten years before 2007 [7]. In this context the performance of the clothing industry, whether in terms of efficiency, working conditions or degree of social protection, is unstable. The industry's ability to generate sustainable and productive employment varies according to geographical locations.

This paper explores the experiences of employees at a clothing manufacturer in South Africa through empirical data that was gathered through a series of focus group and individual interviews and analysed in terms of the idyllic relationship between management commitment and process improvement implementation in the workplace. In the development of these insights, the study aims to inform the process of the implementation of business process improvement particularly for the clothing industry in South Africa [1].

Keywords: clothing industry, business process improvement, management commitment

LITERATURE REVIEW

The current economic distress faced by many manufacturing companies in South Africa both large and small has forced the leadership to review business performance and implement measures to reduce costs across all levels [4]. The competency of leaders comes into question when a business faces difficult times. According to the findings of a report by [10] assessing the management status in South Africa, South Africa lacks competent managers. Nienaber [10] concludes that proper and relevant education and training is critical in mastering "management" both in theory and practice.

According to [2] many companies are in the process of radical transformation aimed at achieving the ability to respond simultaneously and efficiently to meet heightened customer requirements in quality, service, innovation, speed, and price. In a global business environment, organisations are seeking ways to maintain a competitive edge. From past studies as quoted by [6] it is widely accepted that organisations must build an effective management strategy by implementing managerial skills that embrace improvement which produces high quality products or services.

Leaders provide the driving force to create the values, expectations, goals and the systems in order to guide and sustain the pursuit of quality excellence in satisfying customer requirements and performance improvement. However, according to [10], South African managers cannot create and maintain competitive advantage and therefore neglect customers.

According to [8], the problem of poor organisational performance is rife in many organisations in South Africa. They state that this problem is further highlighted by the annual research conducted and published by the World Economic Forum. The World Competitive Reports of 1997, 1998 & 1999 indicate that South African business organisations fare exceptionally poorly when compared to other developed and developing nations. Furthermore, a few of the more disconcerting facts are that the capacity of management to identify and implement competitive practices falls in the bottom 25% for all developed and developing nations and South African organisations fall in the bottom 10% for productivity when compared with other developing nations. [8] states that this situation requires leadership of organisations in South Africa particularly to take responsibility for developing new management skills and applying these skills sensitively to their specific workforce situation [13].

LEADERSHIP COMPETENCIES

Leadership competencies can be defined as the ability to adapt effective interpersonal communication, and good decision-making skills in order to be an effective teamplayer [2]. Leadership competencies are considered important for several reasons, including the fact that they guide direction, they are measurable, and competencies can be learned [2].

According to [8]. from their study of leadership competencies in a manufacturing environment in South Africa, they recommended that the main focus of SA manufacturing companies should be instilling the following competencies; leadership with credibility, having a sense of mission and purpose, ability to communicate a vision, ability to inspire others, emotional intelligence, ability to participate fully with people on all levels, ability to detect positive qualities in others, and the willingness to share responsibility in a measure appropriate to those qualities and willingness to learn, adapt and grow since change is often a step into the unknown.

SURVEY EVIDENCE THROUGH CASE STUDY APPLICATION

A qualitative approach using a case study is used in the implementation of line balancing methodology. This production facility manufactures men's and ladies fashion wear and operates in a small town in Kwa-Zulu Natal. Currently, approximately 300 people work in the plant. The factory opened in 1970 and did not implement modern technology due to financial constraints. The facility had 16 supervisors and a plant manager.

The plant manager agreed to perform a pilot project on line balancing to determine its effectiveness. The sewing department used to the bundle system of manufacture. Work is passed to sewing machines in bundles of cut pieces. The number of cut parts in the bundle may vary according to weight or the complexity of operations required, but the principle remains the same: the operator unties the bundles, sews the cut parts together, re-ties the bundle, processes the work ticket and places the bundle into a bin or on a transporter system (a U shaped manual conveyor).The bundle then goes to another machinist who repeats a similar sequence; a bundle may be tied and untied several times before it completes its lengthy journey. Units move from operator to operator for completion of the respective operation. The bundle production system is a prominent production system used in the clothing industry. Manufacturers use it as a "buffer feeder" and fail to implement process improvement techniques to enhance production flow. Bundles of work-in-progress are found at workstations and sub-assemblies [3].

Before the year 2000 the production facility was accustomed to lot sizes of between 2000 and 10000 units per order. Currently, there are lot sizes of approximately 100 units per order. The garments were not as complex in construction as the ones received currently. The factory was "flooded" with high lots of work-in-process throughout the plant. Employees who were loyal and employed for the last 30 years said that the environment in which they worked was hostile and they did the same operation for several

years. It is important to note that there was no process improvement methodologies implemented at this organisation.

RESULTS AND DISCUSSION

This section contains a qualitative discussion of the experiences of the people involved in the implementation of line balancing in the organisation. Employees felt that management commitment and education/training is the most important aspect of any initiative in an organisation[6].

LEADERSHIP QUALITIES

The workforce of the organisation complained that management did not treat them as "assets" of the organisation. They claimed that they are often treated poorly and management would not consider their views on issues. Labour relations are considered "sub-standard" as management regard workers as another "input" for production. Workers mentioned that all management is concerned about is production, and didn't care how it is achieved. The portrayal of an authoritative management style is common in the clothing industry due to its labour intensity. But the ability to improve the morale through the philosophy of total management could have a positive impact on the output performance of the industry [12].

The organisation realised the benefits of work-study principles, but complained that they did not have the capacity to apply the process improvement principles. The implementation of innovative practices with regards to production techniques, design and development of production, manufacturing processes, supply chain management and labour relations should enable clothing manufacturers to maintain and grow within the industry. The multitude and magnitude of challenges facing the SA clothing and textile industry are clear from the information presented. Both the domestic and international markets are demanding and require a new operating framework that could assist in the survival of the industry.

Management commitment

Any change in the organisation stems from top management. Commitment from management drives the process of change and nothing can be achieved if management does not support the initiative. Once management gives their approval any change is possible, but employees need to understand and support the changes for it to be successful. Management realised that in order to counteract the competitive pressures of the industry they would try out the line balancing methodology. Employees were delighted that the plant manager supported the initiative and frequently visited to find out how they were performing. An employee of the team briefly summarized how he felt.

Any project has to have the "blessings" of management and the acceptance from employees for it to be successful. The managing director of the organisation initiated the process of change in terms of funding labour for the project. Support from management, especially in terms of funding is important for a project of this nature.

The planning, organising, leading and controlling of the project are important as it would benefit the organisation over a period of time. The clothing industry is in need of radical change that would be able to counteract the competition faced. Employees were thankful that they had commitment and the necessary expertise from the management team.

Education and training

A number of training sessions were held with the team of employees to provide orientation with the objectives of the project. Employees held discussions regarding their concerns so that everybody understood their role in the project. The researcher explained that this was a pilot project for the purpose of adding value to the organisation and if it did not work, they would revert back to the old system.

The organisation invested in training and development of employees on an ongoing basis. It was mentioned that training of employees in the latest developments would enhance

employee skills and workers would embrace changes in future. Another employee's experiences was that people would be willing to change if they knew what the change was all about and how it would impact on their work. Mention was made that employee involvement from the very outset would clear any negativity that may be spread through the grapevine within the organisation. It was said that management discussions behind closed doors regarding changes are unhealthy for an organisation. Open communication and the building of trust among the people are extremely important.

An employee mentioned that learning can only take place by change in attitude and behaviour. She also mentioned that training makes employees aware of what is happening and what to expect and it removes barriers between people and is also a great motivator for the workforce.

Another employee mentioned that the concept would be ineffective and that government intervention was the only way that the industry could be saved. The researcher interacted with the individual and convinced him of the way forward. The employee admitted that he was sceptical and did not want change, but since there was communication with management and training of workers, he would "go with the flow." The comments suggest that a project such as this needs education, training, communication and management support.

Open communication is important in a project. The sharing of information between management and employees enhances the success of the project [5]. It was mentioned that the dissemination of too much information and the interpretation of the information could cause problems within the work environment. The "grapevine" misinterprets information and employees become despondent. It was mentioned that 15 years ago operators were not allowed to speak and at present communication is encouraged.

An employee mentioned that "this was quite a change for them." It was mentioned that approximately 15 years ago the floor manager had an elevated office at a centralised point on the machine floor where there was a clear view of all employees. "Management by walk about" (MBWA) had become a prominent feature in the clothing industry. It was mentioned that the manager should be a part of the team on the production floor, know the employees by name and understand the problems experienced. Much could be achieved if team-work is implemented throughout the organisation and all employees strive to achieve the mission and vision of the organisation. Human assets need to be appreciated to enhance their motivational level. Working together could "change a mountain into a molehill," mentioned an employee.

It was mentioned that employees were often ignored and management made all the decisions. Issues such as product quality, customer expectations, productivity were never disclosed to employees. A motivated workforce can achieve labour efficiency without the pressure from management. It was explained that communication among the employees and management improved quality of production and an empowered employee could definitely add value to the organisation, no matter what problems were faced.

The implementation process outcome elucidates that active employee participation with knowledge sharing could improve the performance of the organisation. Sharing information about the costs that go into production and the financial position of the organisation makes employees understand the importance of "right the first time, every time." With work-study officers involved in the process, all work measurement and method study evaluations were done with the team that shared ideas on methods and ergonomics. With the adoption of transparency in all activities employees understood their situation and that of the organisation.

RECOMMENDATION

Strategic focus for manufacturing excellence

The objective of this strategy is on the improvement of quality production, cost and delivery through the application of seven

elements. [9] defines "quality as the development of customer closeness where the workforce understands customer requirements and aims to fulfil these requirements. The researcher concurs with Ng and Hung and considers their approach valuable for the development of the strategies applicable to the clothing industry.

Management approach

- Development of an organisational culture that practices an open and participative management style that supports innovation
- Set achievable goals for the organisation and measure against set standards.
- Understand the production processes and capabilities thoroughly.
- Remove barriers between departments so that processes are seamless to achieve optimal customer satisfaction.
- Manage processes across functional boundaries.
- Managers are to be seen regularly on the production line, engineers in the proximity of the process and there should be regular face to face communication

Manufacturing strategy

- Institute a clear vision and mission of the organization with a long term plan that is understood by everyone;
- Ensure continuous improvement of manufacturing operations.
- Understand globalisation and the impact on the organisation. Develop an understanding of competitive forces.
- Create a plan of action through the involvement of stakeholders in the decision making process.
- All employees should participate in understanding and sharing the strategic intent of the organisation.
- Examine strategies on a regular basis to maintain its applicability.
- Keep abreast with the latest developments that may affect the organisation.

Organisation

- Flatter structures enable effective communication.
- Eliminate "silos" and encourage team-work between departments.
- Create relationships with strategic stakeholders, suppliers and customers (and even competitors).

Manufacturing capabilities

- Adopt process improvement principles in product, delivery and service in all operations.
- Create operations that are adaptable to customer needs.
- Engineer operations towards the elimination of non-conformances.
- Eliminate harm to the environment by determining the impact of processes.

Performance measurement

- Measure customer satisfaction.
- Create measurement systems that enhance productivity.
- Apply business management principles.
- Align the performance measurement system to the organisation's strategic objectives.

Human assets

- Empower employees to strive for the accomplishment of the organizations the goals.
- Supervision should be removed and coaching and mentoring should be implemented.
- Coachers should promote team development, team problem solving and team performance rewards.
- Create an enabling environment where change is embraced.
- Initiate comprehensive programmes of learning and development for continuous improvement.
- Treat the workforce as assets of the organisation and encourage loyalty among employees.

Technology

- Strategize towards technological advancement.
- Understand the competitive status and implement technology accordingly
- Align upgrades with infrastructure
- Implement software solutions that provide on time information [11].

References

[1] D.R. Cooper, and P. Schindler, Business **Research Methods**. McGraw-Hill. New York, 2006.
[2] A. Das, V. Kumar, & U.Kumar. The role of leadership competencies for implementing TQM. **International Journal of Quality & Reliability Management**. Vol.28 No.2, 2010, pp.195-219.
[3] L. Edwards, and S. Golub, South Africa's International Cost Competitiveness and Productivity: A Sectoral Analysis. **Report prepared for the South African National Treasury under a USAID/Nathan Associates SEGA Project**.2002.
[4] C.Forza, and A.Vinelli, 2000. Time compression in production and distribution within the textile-apparel chain. **Integrated manufacturing systems**. Vol. 11, No.2. 2000.
[5] P.Kilduff, 2000. Evolving strategies, structures and relationships in complex and turbulent business environment: The textile and apparel industries of the new millennium. **Journal of textile and apparel, technology and management**. Vol.1, No.1.
[6] D. Kim, V.Kumar, and U. Kumar, U. 2008. A performance realisation framework for implementing ISO 9000. **Unpublished research paper**, University of North Florida, USA and Carleton University, Canada.2008.
[7] R.Mamoepa, R. Minister Dlamini Zuma to hold discussions with Chinese counterpart.2006. **http://www.dfa.gov.za.**
[8] S.M.Mollo, K. Stanz, and T. Groenewald, T. 2005. Leadership competencies in a manufacturing environment. **SA Journal of Human Resource Management**. Vol.3, No.1, 2005, pp33-42.
[9] K.C.Ng, and I.W. Hung, 2001. A model for global manufacturing excellence. Vol.50, No.2, 2001. **MCB Press**.
[10] H.Nienaber, Assessing the management status of South Africa. **European Business Review**, Vol 19, No.1,2007, pp 72-88.
[11] K.Ramdass, An engineering management framework for the clothing industry in SA with a focus on Kwa-Zulu Natal. **Thesis, University of Johannesburg**.2009.
[12] W.A.Taylor, 1994. Senior executives and ISO 9000: attitudes, behaviours and commitment. **International Journal of Quality & Reliability Management**. Vol.12 ,No.4,1994,pp 40-57.
[13] J. Van Wyk, The utilisation of a 360º leadership assessment questionnaire as part of a leadership development model and process. **Unpublished doctoral thesis**, University of Pretoria, Pretoria.2007.

KPE (Knowledge Practices Environment) Supporting Knowledge Creation Practices

Merja BAUTERS
Media Engineering, Helsinki Metropolia University of Applied Sciences
PL 4070, 00079, Finland

Minna LAKKALA
Technology in Education Research Group, University of Helsinki
PO Box 9, 00014, Finland

Sami PAAVOLA
CRADLE, Center for Research on Activity, Development, and Learning
Faculty of Behavioural Sciences, University of Helsinki
PO Box 9, 00014, Finland

Kari KOSONEN
CRADLE, Center for Research on Activity, Development, and Learning
Faculty of Behavioural Sciences, University of Helsinki
PO Box 9, 00014, Finland

and

Hannu MARKKANEN
Media Engineering, Helsinki University of Applied Sciences
PL 4070, 00079, Finland

ABSTRACT

The paper introduces the Knowledge Practices Environment (KPE), and a learning approach called trialogical learning. The virtual environment (KPE) is aimed at providing some solutions to the challenges arising from the constant flow of new tools, use trends, services and terminologies, which question the "oldish" learning practices, and create confusion amongst teachers and students. The trialogical learning suggest practices that support learning in the new and emerging environments by emphasizing students learning activities, which are organized around shared "objects". KPE is designed to support trialogical learning and collaborative knowledge creation processes. Both KPE and trialogical learning have been developed in a large EU-funded KP-Lab project (2006-2011). KPE has been created to provide an integrated system and tools for supporting collaborative knowledge creation; i.e., emphasis is placed on collaborative, iterative and sustained efforts of creating artifacts and/or knowledge practices and processes together, and the role of the tool is to mediate the process smoothly and flexibly. Knowledge creation processes refer to a broader class of purposive and situated activities of a learning community (underlining e.g., object-orientedness) aiming at developing knowledge artifacts and the trialogical approach. This means that KPE is designed to support versatile ways of working with shared "objects".

Keywords: Knowledge creation, Collaboration, Tool Ecology and Trialogical Learning.

INTRODUCTION

Many activities have moved to the Web, offering a medium for numerous everyday tasks related to home, community, office, education, etc. The landscape of tools changes constantly and the tools are complemented with a new generation of open source and access tools, social media tools, services, and enhancements. This includes tools, for example, for social bookmarking and note taking (e.g. Diigo), community-building environments (e.g. LinkedIn and Facebook) and different collaborative working tools build on wiki engines as well as photo-, music-, and video-sharing tools (e.g. Flickr, Vimeo and YouTube) [1]. The challenge of combining an appropriate solution to work, study and various other forms of practices is constant.

The learners are faced with the fact that they have to select, combine and use various materials, online tools and services [2]. It means that learners need to be guided and supported in their choices of learning trajectory including tools and resources (i.e., the learning environment) as well as provided with examples of tool ecologies and collaborative work practices with the tools. The set of tools and practices that these new possibilities allow influences also learners studying practices within the environment [3] and [4]. Still, most tools used for collaborative work and practices are based on approaches that do not support reflection, holistic perspective, or the changing of perspectives (see [5]). To tackle the above challenges, this present article introduces the Knowledge Practices Environment (KPE), a virtual environment aimed at providing some solutions to the challenges. KPE has been created to provide an integrated system and tools for supporting collaborative knowledge creation; i.e., emphasis is placed on collaborative, iterative and sustained efforts of creating artefacts and/or knowledge practices and processes together, and the role of the tool is to mediate the process smoothly and flexibly.

THEORETICAL BACKGROUND OF KPE

KPE is a web-based application designed to provide specific affordances for working with shared objects, that is, joint development of concrete knowledge artifacts as well as for planning, organizing and reflecting on related tasks and user networks (see [6] and [7]). The features, design and interaction potentials of KPE were derived using the co-design processes with several cycles where the theoretical perspectives, research-based pedagogical ideas, and technological development were

integrated. The trialogical approach is a kind of metatheory of knowledge practices. It owes to knowledge building theory because both have as their central focus the interest on understanding knowledge processes and how the technology can support these kinds of processes. The trialogical approach emphases the conceptual artifacts as having two aspects: the conceptual realm but have also the material characteristics, and being in close connection to the practices where they are used. Charles Peirce semiotic and pragmatistic theory provides the theoretical and ontological basis for the trialogical approach than Popper theory of cultural artefacts used within knowledge building [8]. All kinds of texts, project plans, models, sketches are conceptual artifacts, meaning conceptual artifact is not just the ideas that they inhere. The trialogical approach aims at finding ways of supporting people to organize their work and learning around the artifacts. The important point is that ideas are fully embodied and merged within material practices in knowledge work. Other central background for the trialogical approach to learning is provided by the cultural-historical activity theory. The cultural-historical approach builds on the idea that human activities are mediated by artifacts (often just referred by tools), used and modified in iterative everyday activities [9: 108-110]. Practices, and cultural artifacts are developed in interaction with each other. This intertwined process historically situated and evolving [10]. The activity theory has provided some basic orientations to the trialogical approach, such as: attempt to understand the intertwined system of material, social and practical components of learning. The trialogical approach is closely connected to the role of new technology, and especially how it transforms practices. As mentioned, the trialogical approach concentrates on processes and forms of mediation where a group of people are developing some concrete things (like texts, products, ways of working) for some purpose. 'Shared - Trialogical objects' are concrete objects, which people develop collaboratively (see also [11] for boundary objects). This focus creates a clear difference to the pedagogical approaches where the focus is more on participatory or dialogical meaning making processes. Thus for trialogical learning the focus is on an interaction between subject(s), other subjects, and "objects", not, for example, only between subjects. The temporal dimension is also important in trialogues, meaning the shared objects are developed and modified iteratively while in the same time using and developing existing practices. The objects that learners develop are meant for some subsequent use and/or potentially to be modified later on. As said, the object is something concrete – even ideas and conceptions must be tangible artifacts – still these also material objects keep evolving within the process. Emphasis is on developing something new collaboratively, not repeating existing knowledge. To support the change of the prevailing pedagogical practices into more trialogical ones: six aspects that aim at defining general characteristics of trialogical learning have been drafted. The aspects guide also the evaluating process of the practices. These aspects are called Design Principles:

1. Organizing activities around shared "objects" (artifacts, practices)
2. Supporting interaction between personal and social levels, and eliciting individual and collective agency
3. Fostering long-term processes of knowledge advancement
4. Emphasizing development through transformation and reflection between various forms of knowledge and practices
5. Cross fertilization of various knowledge practices across communities and institutions
6. Providing flexible tool mediation

The implementation of these functional requirements called for open, modular and loosely coupled technical design, which was decided to be pursued with he service-oriented architecture (SOA). The project carried out state-of-art studies on existing software, comparing the functional and technical requirements with various groups of collaborative learning and working environments, such as knowledge building environments (FLE, Knowledge Forum, CMap Tools), web collaboration environments (BSCW, Google Apps, ZoHo), collaboration and learning environment (SAKAI), as well as on-line classroom and eLearning platforms (Moodle, Claronline). Although the different environments provided similar features and functionalities as KP-Lab project was aiming at, none of them provided the solid software base to build on. Major prohibiting factors were that the software was not open or the architecture did not support extending of the functionality as required by the KP-Lab pedagogical scenarios.

The above- mentioned Knowledge Forum has inspired the development of KPE because it provides a knowledge space with functionalities to create, link and build on shared multimedia objects. FLE3, was developed for progressive inquiry practices [12] and [13]. KPE aims at providing a holistic and more integrated perspective into the work in contrast to environments, which separate processes and different aspects of work more clearly (such as LAMS and Sky Lab). Combining the web 2.0 tool provide personal and collaborative tool ecologies (see e.g., [14], [15] and [16]. These combination include such tools as: file sharing system such as DropBox, combined social media tools e.g., Facebook, Google's applications, Zoho, ad hoc tools, such as Piratepad, Typewith.me, Zotero, including Confluence wiki, which though is commercial, just to mention few well known available tools. These tools provide a start for collaborative elaboration of shared knowledge artefact, but the tools do not provide further affordances for systematic and sustained creation and formation of collaborative practices and knowledge (see [17], [18] and [19: 24].

FEATURES IN KPE TO PROVIDE AFFORDANCES FOR COLLABORATIVE KNOWLEDGE CREATION

In this section, Knowledge Practices Environment (KPE) is described in more detail. With KPE, users are able to build collaboration environments by creating and configuring the means of the common practice, as opposed to operating with predefined structures. KPE is a virtual environment that includes a set of basic, integrated tools such as: real-time and

history-based awareness, wiki, note editor, commenting, chat, semantic tagging, linking, process organization, filtering and search. KPE is based on strong visual and spatial ways of organizing the work, building on a kind of a desktop metaphor. It means that the spaces do not have folder structures, but items can be filtered using structural and semantic tagging, spatially organized. This approach provides a novel perspective to relations between knowledge and practices as will be described below. KPE enables object-oriented and threaded commenting on all items as well as viewing of knowledge artefacts and their relations from several perspectives. Three basic perspectives provided are: Content, Process and Community Views.

Sharing and co-construction of knowledge artefacts with free visual arrangement and linking in Content View of KPE. DP 1, and 2 are supported in KPE by functionalities that enable users to create, modify, build on and organise various knowledge artifacts as well as their relations. Below, some central characteristics related to the work with knowledge artefacts are briefly described. The DP6: Providing flexible tool mediation, belongs to all of the described issues below.

In KPE, user groups can create 'shared spaces' through which various knowledge artifacts can be shared and co-constructed. The basic features include uploading any type of files or web-links into the shared spaces. But instead of providing only a space to store or manage versions and the synchronisation of vast number of documents, KPE enables the users to organize knowledge artefacts (represented by graphical icons) through visual representations. A central view in KPE for working on knowledge artefacts is the *Content View* that allows free visual arrangement and linking of its content (see Figure 1). The organisation of a shared space reminds the organisation of files on desktop, except KPE allows better tools for spatial arrangement and linking of items, filtering of items based on metadata and tags. These features and functions also allow reflecting on the artifacts, their relations and organisation. KPE is not based on folder structures or hierarchical presentation of the content; it does not hide the content into folders, which detach items from their relations. One of the most interesting ideas in KPE is this strong approach on integrating visual and spatial organisation, filtering, categorizing, prioritizing, semantic meaning creation and process visualisations.

With *Note editor,* users can directly write their ideas and thoughts as content items in the Content View, without the labour of creating and uploading an external text file [20]. All members of a space can open and edit the created notes and view their previous versions. It is an important functionality since it enables fast access to previous thoughts and arrangements of ideas and knowledge, which is needed for further developing, pondering and reflecting on the joint procedures, goals and achievements. The Content View includes a *Sketch pad* tool that is based on the same idea as Note editor, but enabling the creation, co-editing and versioning of simple drawings and visual sketches.

The ability to write collaboratively in a sustained manner – an essential feature of knowledge work – is supported through integrated wiki offering the possibility to access the same wiki document from a shared space. The actual use (observed during four years and in six different courses) showed that the wiki was usually taken to be for more thoughtful writing and for producing more finished texts. The students intuitively used (meaning here without guidance) the combination of the tools. Note editor was used for idea generation, sketching and drafting. After the sketching and drafting phase were over and the subject matter was felt to be better understood, the students moved on writing wiki, where the goal was to polish and structure previous writings.

Object-oriented interaction around knowledge artifacts (DP 4) is possible by using commenting functionality, which means that asynchronous, threaded discussions are attached directly to knowledge artifacts. One object can have many comment threads, thus enabling users to discuss various aspects of the objects, directly, in the context. This object-oriented aspect places KPE beyond isolated discussion forums, threaded notes or argumentative discussion supports, which concentrate only on dialogical aspects of collaboration with threaded discussions and lose easily the context and the object. The KPE answers to the need to have individual contributions attached in collaborative work that is organized around shared knowledge artefacts embedded and embodied in a shared space. Similarly, object-oriented chat enables synchronous interchange attached directly in the items at hand. Chat log is saved and linked to the targeted item, therefore keeping the log attached to its object for possible re-use and continuation. The object-oriented features and functions are further supported by the visual metaphor in keeping everything in sight, allowing different spatial arrangements that can be flexibly changed according to different phases of the work. No other tool so clearly allows contextualised work, which keeps all objects visible still allowing their filtering after the phase or work is done. The products and processes do not disappear and get lost in folders, sub-pages, tabs or separate forums.

Flexible use of tags is one of the aspects of the KPE related to DPs 1, 2, and 3. It makes KPE go beyond current learning environments and especially combinations of social media tools and tool economies, is the use of metadata and semantic features to support the usage and integration of knowledge artefacts in various ways. Tags, tag clouds and tag vocabularies can be created and edited by participants. In the Content View and Alternative Process View, all items can be tagged. This provides additional affordances for various types of knowledge practices in education, as compared to existing tools. For example, in typical research seminars, semantic tagging can be used to help students find common areas of interest and related materials, or to analyse the elements and concepts of existing and produced research papers. The tag cloud generated automatically from the tags assigned by users enables easy filtering of the items according to the subject matter, categories, or other user defined taxonomies. The tags users define, are implemented in the underlying technology in a way that allows search through the semantics or relations between tags; e.g., semantic information can be reused across various integrated tools. Such functionalities allow the users to create their own cognitive and conceptual tools and instruments based on the potentialities of the semantic web. The filtering using the tag cloud also allows emphasizing different

knowledge artifacts and practices depending on what kind of issues or phases the group or individual is going through. This supports the use of the same Content View for longer time periods, enabling sustained work, reuse of items and the reflection of previous work and practices without separating the phases or distributing the items across tools and time.

Organizing processes (pragmatic mediation)
For planning, monitoring, and regulating joint activities and working processes (DPs from 1-4). These functionalities enable users to define tasks as well as draft visual, spatial and semantic representations of processes. They also provide users with 'awareness features' of the activities in the spaces.

The process planning can be executed through defining tasks and drafting visuo-spatial and semantic process representations. In KPE users can explicitly define, modify and arrange *task items and areas* to represent the process and domain elements of activities. Task items may include, e.g., title descriptors, responsible users, start and end dates and status. Areas attached with semantic meanings can be created to represent a phase, an action, or a category depending on the manners and needs of the users to organise their knowledge artefacts. These features allow users to explicate their process elements and promote responsibility and ownership over the decisions and actions. The *Alternative Process View* includes the spatial representation of user-defined areas for organizing knowledge artefacts and processes, which enable users to illustrate processes, phases, groups and categories according to shape, colour and place of the areas in question. It emphases relationships between task and content items and their meaning, since the areas can be tagged, and the tags are inherited to all items placed into the particular area. The tags are also presented in the Content View in the tag could, from which users can filter the items according to the meaning of the area specified in the Alternative Process View (see Figure 2 and 3). The area–tagging feature makes the tagging process easier than it is with most other tools using tags (e.g., Google mail, Diigo, Delicious). It lowers the threshold to use tags and to think of the meanings knowledge artefacts and their relations have. This is important since experience has shown that for students it is often a challenge to see the benefits of laboriously explicating the semantic meaning and relations of knowledge artefacts [21].The features of Alternative Process View are especially useful in those educational settings, where the chronology of the work is not essential, but there is a requirement to see connections, associations and causal relations between the various elements of the process.

Awareness features for supporting process planning and coordination of collaborative working processes be they asynchronous or synchronous are not often consciously noticed or paid attention to but they play an essential role in tool-mediated collaboration, keeping track of on-going and past actions. Without such information, the work is severely hindered. Most of the awareness features in KPE that are meant to support synchronous work; for example: visual clues and on-line notifications about who is online, who is working with whom, or who is working on what object (a lock or a glove is displayed on the item with the name of the users) and doing what. Historical perspective is provided, e.g., by a list about modifications of knowledge artefacts and tasks or by e-mail or mobile device notifications about the events in a shared.

Social relations around shared objects and processes
The organizing of social structures, responsibilities and roles for a smooth coordination of collaborative work, it is crucial to explicitly define social structures among the participants (reflecting DPs of 2 and 5). For each content or task item visible in the Content or Alternative Process Views, it is possible to define persons responsible for that item. In addition, a third basic view of KPE, called the *Community View* (see Figure 4), is especially meant to support the formation of groups (e.g., by visually displaying the groups/teams formed with the visual information of the users, and their roles, the same members can have more than one role) as well as coordination of tasks and responsibilities between participants. The users are presented as items in the Community View but they are also presented as a list in the Network View on the right hand tab. Both displaying manners present also the information of the users' online status. Detailed user information includes a list of all tasks and knowledge artifacts that have been created and modified by or assigned to a particular member.

CONCLUSION

Summarising the experiences and results of the scientific research from five years, it can be concluded that KPE captures the essence of the trialogical perspective, that is, gives means for working with shared objects and processes from multiple perspectives in an integrated way:

- It allows commenting, collaboration as well as organizing and sharing of work in a holistic and visuo-spatial manner stressing the process besides the outcomes. The KPE desktop metaphor provides multiple perspectives into the knowledge artifacts and practices;
- It supports the reflection of practices in context, not separating activities into fragmented reflection parts. The KPE's object-oriented interaction enhances possibilities for reflecting on individual and collaborative products and practices;
- It enables flexible group formation;
- It supports information display of online statuses, social relations, roles information, etc., and use as well as multiple perspectives to the work by various filtering methods (e.g., with tags, visuo-spatial organization, linking etc.).

The managing of collaborative and/or sustained knowledge creation processes in a versatile multimediational way is one evident strength of KPE. KPE appears especially to support early phases of the knowledge creation process and the integration of different activities. In addition, in the examined courses, the possibility to get visual overviews of things, to organize processes flexibly and visuo-spatially and to tag items through placing them in particular areas were especially appreciated, special features of KPE (related to "a virtual desktop" metaphor).

However, there exist challenges that need to be taken into account and corrected when developing KPE further. Such challenges are, for example, the following:

- KPE is too complex and needs serious reduction of

features and functions.

- KPE is competing with other tools, which users already know and which are continuously emerging in the Internet. These tools are easy to use and do not require registration. KPE needs to be opened up so that these kinds of tools can be added and used in collaboration with KPE.
- The previous point relates to the requirement of integrating individual self-reflections with group activities and to offer awareness information about the social system in which individual activities are embedded. New distributed social tools and services e.g. pushing feeds for the group, mashing and filtering group feeds that enable people to interact in the group environment from within personal learning environments, would help to provide scaffolding both for an individual learning process and for collaborative activities.

REFERENCES

[1] Väljataga, T., Pata K. and Tammets K. (2010). Considering Students' Perspectives on Personal and Distributed Learning Environments in Course Design. In Web 2.0-Based E-Learning: Applying Social Informatics for Tertiary Teaching. Mark J.W. Lee (Ed.) Catherine McLoughlin.

[2] S. Fiedler and K. Pata, "Distributed learning environments and social software: In search for a framework of design", in S. Hatzipanagos, and S. Warburton, (Eds.), **Handbook of research on social software and developing community ontologies**, Hershey, PA: Information Science Reference , 2009, pp. 151–164.

[3] K. D. Könings, S. Brand-Gruwel, J. J. G. van Merriënboer, and Broers, "Does a new learning environment come up to students' expectations?", A longitudinal study. **Journal of Educational Psychology**, Vol., 100, No. 3, 2006, pp. 535–548.

[4] N. Entwistle, and H. Tait, "Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments", **Higher Education**, Vol. *19*, No. 2, 1990, pp. 169–194.

[5] G. Conole, "Stepping over the Edge: The Implications of New Technologies for Education The Open University", in **Web 2.0-Based Learning**, in Web 2.0-Based E-Learning: Applying Social Informatics for Tertiary Teaching. Mark J.W. Lee (Ed.) Catherine McLoughlin, 2010), 380-393.

[6] H. Markkanen, M. Holi, L. Benmergui, M. Bauters and C. Richter, "The Knowledge Practices Environment: a Virtual Environment for Collaborative Knowledge Creation and Work around Shared Artefacts", in **Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications**, Chesapeake, VA: AACE, 2008, pp. 5035-5040.

[7] M. Lakkala, S. Paavola, K. Kosonen, H. Muukkonen, M. Bauters and H. Markkanen, "Main functionalities of the Knowledge Practices Environment (KPE) affording knowledge creation practices in education", in C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), **Computer supported collaborative learning practices: CSCL2009 conference proceedings (297-306)**. Rhodes, Creek: International Society of the Learning Sciences (ISLS), 2009, Retrieved July 22 2011 from: http://www.helsinki.fi/science/networkedlearning/texts/Lakkala_et_al_2009_KPE_cscl09.pdf

[8] S. Paavola, and K. Hakkarainen, "From meaning making to joint construction of knowledge practices and artefacts – A trialogical approach to CSCL", in C. O'Malley, D. Suthers, P. Reimann, and A. Dimitracopoulou (Eds.), **Computer Supported Collaborative Learning Practices: CSCL2009 Conference Proceedings**, Rhodes, Creek: International Society of the Learning Sciences (ISLS), 2009, pp. 83-92.

[9] M. Cole, **Cultural Psychology. A Once and Future Discipline. Cambridge**, MA: The Belknap Press of Harvard University Press, 1996.

[10] R. Miettinen and J. Virkkunen, "Epistemic Objects, Artefacts and Organizational Change", **Organization**, Vol. 12, No. 3, 2005, pp. 437-456.

[11] S. L. Star and J. R. Griesemer, "Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals", in Berkeley's Museum of Vertebrate Zoology, 1907-39, **Social Studies of Science**, Vol. 19, No. 3, 1989, pp. 387-420.

[12] H. Muukkonen, K. Hakkarainen and M. Lakkala, "Collaborative technology for facilitating progressive inquiry: Future Learning Environment Tools", in C. Hoadley, and J. Roschelle (Eds.), **Designing new media for a new millennium: Collaborative technology for learning, education, and training**, Mahwah, NJ: Erlbaum, 1999, pp. 406–415

[13] T. Leinonen, G. Kligyte, T. Toikkanen, J. Pietarila and P. Dean, **Learning with collaborative software – A guide to FLE3**, Helsinki: University of Art and Design 2003, Retrieved July 22 2011 fromhttp://fle3.uiah.fi/papers/fle3_guide.pdf)

[14] E. Arenas, "Personal learning environments: Implications and challenges". In D. Orr, P. A. Danaher, G. Danaher, & R. E. Harreveld (Eds.), **Lifelong learning: Reflecting on successes and framing futures**. Keynote and refereed papers from the Fifth International Lifelong Learning Conference. Rockhampton, Australia: Central Queensland University Press, 2008, pp. 54–59.

[15] M. Crosslin, "When the Future Finally Arrives: Web 2.0 Becomes Web 3.0." in **Web 2.0-Based Learning**, in Web 2.0-Based E-Learning: Applying Social Informatics for Tertiary Teaching. Mark J.W. Lee (Ed.) Catherine McLoughlin, 2010, 394-415.

[16] H. Huijser and M. Sankey, "You Can Lead the Horse to Water, but … : Aligning Learning and Teaching in a Web 2.0 Context and Beyond", in **Web 2.0-Based Learning**, in Web 2.0-Based E-Learning: Applying Social Informatics for Tertiary Teaching, Mark J.W. Lee (Ed.) Catherine McLoughlin, 2010, 267-283.

[17] M. E. Cigognini, M. C. Pettenati and P. Edirisingha, "Personal Knowledge Management Skills". In **Web 2.0-Based Learning**, in Web 2.0-Based E-Learning: Applying Social Informatics for Tertiary Teaching. Mark J.W. Lee (Ed.) Catherine McLoughlin, 2010, 109-127.

[18] S. Downes, "E-learning 2.0", **eLearn Magazine.** Retrieved June 22, 2011, from http://

www.elearnmag.org/subpage.cfm?section=articles&article=29-1.

[19] T. Bates, "Understanding Web 2.0 and its Implications for E-Learning". In **Web 2.0-Based Learning**, in Web 2.0-Based E-Learning: Applying Social Informatics for Tertiary Teaching. Mark J.W. Lee (Ed.) Catherine McLoughlin, 2010, pp. 21-42.

[20] I. H. Furnadziev, V. P. Tchoumatchenko, T. K. Vasileva, M. Lakkala, M. Bauters, "Tools For Synchronous Communications in Collaborative Knowledge Practices Environment (KPE)", **V International Conference on Multimedia and Information & Communication Technologies in Education m-ICTE2009**, Lisbon, Portugal, 22–24 April, 2009, pp. 588–592.

[21] S. Jalonen, M. Bauters, and K. Kosonen, "Verkkoteknologia tietokäytäntöjen kehittämisessä yliopiston semiotiikan metodologia-kurssilla", Presentation in congress of **Blended Learnign 2010, 11.3. – 12.3.2010**, Helsinki. Retrieved March 24 2010, http://blogs.helsinki.fi/sulautuvaopetus/verkkoteknologia-tietokaytantojen-kehittamisessa-yliopiston-semiotiikan-metodologia-kurssilla/

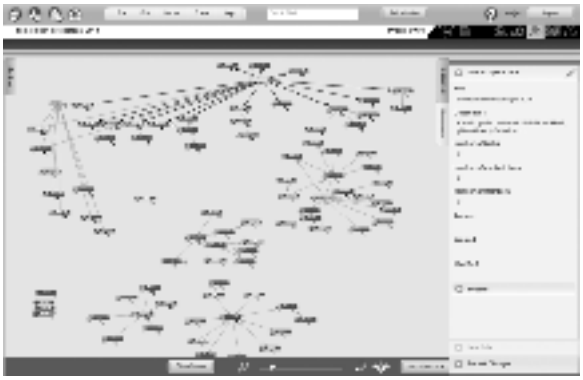


Figure 1. Content View: visual arrangement of content items in the course of semiotic methodology at University of Helsinki. The blackish items are content items, such as up-loadable files, links to Internet, notes and chats. The links display a particular kind of relation between the items; or represent hierarchical relations ship. The greyish items are tasks.

Figure 2. Alternative Process View (APV): a student team shared space from a project course where lean programming methods were used. The figure present the 'Kanban' table of the tasks, issues to be done and the state the items are in.

Figure 3. The figure presents the Content View related to the Alternative Process View of the figure 2 above. As can be seen from the left side tab's tag cloud, the same tags that were in there are in the Content View. On the right side image: the items have been filtered using one tag ('Backlog').

Figure 4. Community View: the groups of a project course have been formed, on the right hand tab is presented the items created by one, selected user (displayed with a halo) in this shared space.

Making it Real: Faculty Collaboration to Create Video Content

Claudia Jennifer DOLD, MLIS
Louis de la Parte Florida Mental Health Institute Research Library
USF Libraries
University of South Florida - Tampa
Tampa, Florida 33620

and

Gary DUDELL, PhD
Department of Rehabilitation and Mental Health Counseling
College of Behavioral & Community Sciences
University of South Florida - Tampa
Tampa, Florida 33620

ABSTRACT

Interest in integrative health care is a growing area of health practice, combining conventional medical treatments with safe and effective complementary and alternative medicine. These modalities relate to both improving physical and psychological well-being, and enhancing conventional talk therapy. In an interdisciplinary collaboration, teaching and library faculty have created a series of sixteen on-line video interviews that introduce practitioner-relevant experiences to students as supplemental course material. These videos are available through the department web-pages to students in other related disciplines as well, including Social Work, Counselor Education, Psychology, and the Colleges of Public Health, Nursing, and Medicine. The video series was undertaken as part of the educational mission of the library, bringing to the classroom new material that is essential to the professional development of future counselors.

Keywords: video interviews, content development, collaboration, integrative health techniques, mind-body techniques, holistic approach, mindfulness, well-being, counseling

INTRODUCTION

While there is growing interest in exploring complementary medicine in counseling [1], there is no text book on cutting edge practice, by practitioner, by location. Not all students can afford to go for a session or two with local therapists to experience a technique firsthand. Practitioners don't have time to field student emails over the semester, and the content would probably be repetitive. A one-time guest visit to a class benefits that class only. The professor can't be expected to ask the same practitioners again and again to his classes, year in and year out.

At the College of Behavioral & Community Sciences in the University of South Florida-Tampa, Dr. Dudell discussed the problem of practitioner-relevant resources with librarians familiar with his coursework. The professor wanted video content specific to his class. When nothing appeared to be available commercially, they decided to engage in a joint project to create the target content. One of the librarians, Claudia Dold, had extensive experience in filming, editing, and producing academic videos using in-house equipment and software.

The videos would be available for other teaching staff with similar interests, as well as to the community at large. Specifically, the professor would ask practicing colleagues, many in the local area, to engage in

videotaped interviews to discuss their philosophy, practice, and outcomes using mind-body techniques. The interviews would be structured in a similar manner to make possible the comparing and contrasting of various integrative health care techniques. The professor would conduct the interviews in order to guide the discussion to the important aspects of each technique and also to maintain a consistent structure, and the librarian would address the videotaping and technical aspects of the work.

This paper discusses the details of the collaboration, with the intention of shedding light on the collaborative process so other faculty may engage in a similarly rewarding and productive experience for themselves, for their students, and for their academic institutions when published material is not available to enrich a course.

VIDEO IN HEALTH EDUCATION

Video is currently used as an education tool in the health field in numerous facets: in preventive health measures [2], [3], [4], [5]; to teach medical students and practitioners new procedures [6], [7], [8], [9]; to demonstrate interventions [10], [11]; to improve counseling skills [12], [13], [14]; and to deliver health information to the patient [15], [16]. A review of the literature identified one article that addressed the influence of demonstration videos on the students' perspective on the counseling profession [17]. The Keats article discusses the trainee's reflections on viewing expert counselors in session with patients and recommends research into how students select modalities to incorporate into their own counseling style. The author notes that video gives students a view of the practicing therapist at work, offering them a window on professional demeanor and performance. The literature did not reveal studies concerning the use of video to introduce emerging alternate therapies to students.

The College of Behavioral & Community Sciences at the University of South Florida (USF) prides itself on translating theory to practice. The video interviews produced in the collaboration are viewed as a means of bridging that gap and making the variety of mind-body techniques real to future professionals in the counseling field.

VIDEO APPLICATION IN COUNSELING EDUCATION

In the case study discussed in this paper, video interviews were used to introduce future professional counselors to the application of successful therapies practiced by counselors usually within driving range of their university. The geographical area is significant, since students could follow up a particularly appealing modality with the practitioner they had seen. Furthermore, if they established a practice in the Tampa area, they could also maintain contact through regional conferences and potentially make referrals to a known practitioner.

Choice of Modalities: Defining the students' need concerned decisions about content selection. The initial series was entitled, "Conversations with Mind/Body Practitioners" and was defined as a series of eight interviews with integrative health professionals. It was designed for students in the Department of Rehabilitation and Mental Health Counseling at USF. The definition of integrative medicine was taken from the National Center for Complementary and Alternative Medicine (NCCAM), a branch of the National Institutes of Health: "an approach to medicine that combines mainstream medical therapies and CAM (complementary and alternative medicine) therapies for which there is high-quality scientific evidence of safety and efficacy"[1, p. 65]. Suggested topics included yoga, tai-chi, nutrition, fitness and stress-management, all of which engage in a relation between maintaining positive lifestyle habits and increasing emotional well-being.

RESULTS

The first series was well received by the students in Dr. DuDell's counseling class and by the department chair. Students commented that they were impressed by the sincerity and commitment of many of the interviewees. They could see ways to enrich their future practice by suggesting some of these alternative therapies when appropriate to meet particular needs of their clients. However, no formal assessment was made at the end of the first course.

A second series of eight videos was planned and is now complete. A student in the Masters of Library and Information Science program at USF was hired to record,

transcribe, and edit the raw video. The librarian provided technical assistance and performed administrative tasks; the professor again selected the therapists he wanted to interview and conducted the interviews.

The two series of eight videos were merged into one series of sixteen videos, prefaced by a brief video interview with the professor. In the brief introductory unit, he discusses the recurring themes in the videos: connection, personal responsibility, and therapist passion for improving the lives of others.

DISCUSSION

Grant Application: The project was funded by the Center for 21st Century Teaching Excellence, a unit within USF that encourages innovative teaching techniques. The grant enumerated the advantages for the students and listed the courses for which this series would enhance the curriculum. Other venues were also mentioned that might be interested in posting the videos in their outreach programs. To further strengthen the grant proposal, the application listed the alignment of the project with the department, the college, and the university strategic goals.

A cost projection was submitted, detailing the estimated hours of labor per video and the cost of essential equipment: videocassettes, headphones, and a one-terabyte hard disk for external storage. The library already owned the video camera, tripod, video-processing computers, and software. The grant was awarded for just under $2000 with a deadline of six months to complete the project.

The Working Collaboration: The professor and the librarian brought a variety of useful skills to the project. The former had experience conducting interviews on radio; he also had been in private counseling practice for years. He knew integrative health workers in the local area and across the country from attending professional conferences. The librarian had been working in the mental health library for several years. She was well acquainted with the library's video equipment and Camtasia, the software product used to process the raw film into captioned, edited units. With practice, the team improved its filming technique. Experience pointed out the importance of planning the interview space so that the background was unremarkable, the light was ambient, and the recorded sound was crisp. No interviews had to be repeated due to faulty performance of equipment, personnel, or planning, and the viewer's experience improves in subtle ways over time as modifications in the process were applied.

The typical interview lasted an hour. The professor and the guest were seated and the microphone was placed centrally to capture their conversation. The interview was recorded using the videocamera affixed to a tripod. After the interview, the taped session was downloaded from the videocassette to a desktop computer. The professor and videographer/librarian would look at the film together and decide what pieces to keep, to zoom in on, and to amplify. The goal was a finished video interview of not more than thirty minutes. A standard title and credit page were created to bring a sense of unity to the series. Fourteen hours were allotted per interview for the videographic work, which spanned the initial interview set up to the posting of the final product

The Unique Role of the Librarian: The academic librarian is poised in the college structure to assist both faculty and students. Teaching faculty ask for literature reviews and syllabus updates, and students ask for help refining their search topics and finding information. As faculty, librarians are both colleagues to teaching faculty and teachers to students, not only in bibliographic skills, but also in subject matter within their own sphere of academic background and interests. Librarians may collaborate with faculty before the semester begins to ensure that the resources are on hand for use in a course. Librarians then assist students who come to the library during the semester to work those assignments. From their unique position, librarians are situated to observe what works in the academic setting and to notice what could work better for both faculty and students.

Librarians have a long history of partnering in health education, teaching research skills that complement nursing faculty curriculum and prepare nursing students to keep up with the latest in professional literature concerning treatment protocols and patient care [18], [19]. For example, the University of Arizona in Tucson placed librarians on site in the colleges of medicine, pharmacy, nursing, and public health to readily serve their patron groups [20]. "Embedded librarians" collaborate with teaching faculty, gain specialized

knowledge of the field, and become familiar with the course material that students will cover during the semester and over the course of their larger program of academic study [21].

One of the unexpected benefits of the video project collaboration at USF was the insight gained by the librarian as the interviews were recorded, the audio track was transcribed, and the film was edited. The content brought the librarian up-to-date with complementary therapies, informed her searches in the literature as part of her consulting activities with the students, and situated complementary and alternative medicine within the larger field of contemporary health care. For the information professional, the experience of interdisciplinary collaboration broadens one's awareness of expanding fields of knowledge and of the challenge of accessing reliable research [22].

CONCLUSION

Creating content is a challenge at many levels. The video project collaboration succeeded for several reasons. The principal investigators had a very clear idea of what each one would do, and each was fully capable of performing the tasks assigned. The goals and purpose of the project were clear from the start: the video series filled a gap in the education of future professionals. The grants were well-written and showed the potential impact on the university and the community. The collaborators respected each other's talents, personality, values, and time. Each gained an understanding of the other's expertise and mission within the university. In the end, they created a professional relationship, which in itself is a valuable resource. They created original content that will serve students for several semesters and become part of the teaching repository. They also demonstrated the advantages of interdisciplinary collaboration in advancing education.

REFERENCES

[1] National Center for Complementary and Alternative Medicine, **Expanding Horizons of Health Care: Strategic Plan 2005-2009**, U.S. Department of Health and Human Services, National Institutes of Health, NIH Publication Number 04-5568, 2004. http://nccam.nih.gov/about/plans/2005/strategicplan.pdf

[2] Y. Calderon, J. Leider, S. Hailpe, M. Haughey, R. Ghosh, P. Lombardi, P. Bijur, and L. Bauman. "A Randomized Control Trial Evaluating the Educational Effectiveness of a Rapid HIV Posttest Counseling Video", **Sexually Transmitted Diseases**, Vol. 36, No. 4, 2009, pp. 207-210.

[3] Y. F. Y. Chan, R. Nagurka, L. D. Richardson, S. B. Zaets, M. B. Brimacombe, and S. R. Levine. "Effectiveness of Stroke Education in the Emergency Department Waiting Room", **Journal of Stroke & Cerebrovascular Diseases**, Vol. 19, No. 3, 2010, pp. 209-215.

[4] M. A. Chiasson, S. Hirshfield, and C. Rietmeijer. "HIV Prevention and Care in the Digital Age", **JAIDS Journal of Acquired Immune Deficiency Syndromes**, No. 55, 2010, pp. S94-S97.

[5] D. M. N. Paperny and V. A. Hedberg, "Computer-Assisted Health Counselor Visits – A Low-Cost Model for Comprehensive Adolescent Preventive Services", **Archives of Pediatrics & Adolescent Medicine**, Vol. 153, No. 1, 1999, pp. 63-67.

[6] J. W. Y. Chung, T. K. S. Wong, K. K. P. Chang, C. B. Chow, B. P. M. Chung, G. Chung, S. Ho, J. S. C. Ho, C. K. Y. Lai, A. Lai, V. S. F. Lam, J. Lau, J. Liu, E. Mok, and D. Wong. "Rapid Assessment of a Helpdesk Service Supporting Severe Acute Respiratory Syndrome Patients and their Relatives", **Journal of Clinical Nursing**, Vol. 13, No. 6, 2004, pp. 748-755.

[7] C. Nicolaidis, M. Curry, and M. Gerrity, "Measuring the Impact of the Voices of the Survivors Program on Health Care Workers' Attitudes Toward Survivors of Intimate Partner Violence", **Journal of General Internal Medicine**, Vol. 20, No. 8, 2005, pp. 731-737.

[8] A. Scardovi, P. Rucci, L. Gask, D. Berardi, G. Leggieri, G. B. Ceroni, and G. Ferrari. "Improving Psychiatric Interview Skills of Established GPs:

Evaluation of a Group Training Course in Italy", **Family Practice**, Vol. 20, No. 4, 2003, pp. 363-369.

[9] J. J. Price, A. W. Bedell, S A. Everett, and L. Oden, "Training In Firearm Safety Counseling in Family Practice Residency Programs", **Journal of Community Health**, Vol. 22, No. 2, 1997, pp. 91-99.

[10] J. Dale, H. Sandhu, R. Lall, and E. Glucksman, "The Patient, the Doctor and the Emergency Department: A Cross-Sectional Study of Patient-Centeredness in 1990 and 2005", **Patient Education and Counseling**, Vol. 72, No. 2, 2008, pp. 320-329.

[11] J. R. Soble, L. B. Spanierman, and H. Y. Liao. "Effects of a Brief Video Intervention on White University Students' Racial Attitudes", **Journal of Counseling Psychology**, Vol. 58, No. 1, 2011, pp. 151-157.

[12] P. Marita, L. Leena, and K. Tarja. "Nurses' Self-Reflection Via Videotaping to Improve Communication Skills in Health Counseling", **Patient Education and Counseling**, Vol. 36, No. 1, 1999, pp. 3-11.

[13] M. S. Shafer, R. Rhode, and J. Chong. "Using Distance Education to Promote the Transfer of Motivational Interviewing Skills Among Behavioral Health Professionals", **Journal of Substance Abuse Treatment**, Vol. 26, No. 2, 2004, pp. 141-148.

[14] V. Verhoeven, D. Avonts, E. Vermeire, L. Debaene, and P. Van Royen. "A Short Educational Intervention on Communication Skills Improves the Quality of Screening for Chlamydia in GPs in Belgium: A Cluster Randomised Controlled Trial", **Patient Education and Counseling**, Vol. 57, No. 1, 2005, pp. 101-105.

[15] I. L. Beale, P. M. Kato, V.M. Marin-Bowling, N. Guthrie, and S. W. Cole, "Improvement in Cancer-Related Knowledge Following Use of a Psychoeducational Video Game For Adolescents and Young Adult with Cancer", **Journal of Adolescent Health**, Vol. 41, No. 3, 2007, pp. 263-270.

[16] M. R. Weaver, M. Myaya, K. Disasi, M. Regoeng, H. N. Matumo, M. Madisa, N. Puttkammer, F. Speilberg, P. H. Kilmarx, and J. M. Marrazzo, "Routine HIV Testing in the Context of Syndromic Management of Sexually Transmitted Infections: Outcomes of the First Phase of a Training Programme in Botswana", **Sexually Transmitted Infections**, Vol. 84, No. 4, 2008, pp. 259-264.

[17] P. A. Keats, "Buying into the Profession: Looking at the Impact on Students of Expert Videotape Demonstrations in Counsellor Education", **British Journal of Guidance & Counselling**, Vol. 36, No. 3, 2008, pp. 219-235.

[18] B. Layton and K. Hahn, "The Librarian as a Partner in Nursing Education", **Bulletin of the Medical Library Association**, Vol. 83, No. 4, 1995, pp. 499-502.

[19] A. Hallyburton and B. St. John, "Partnering with Your Library to Strengthen Nursing Research", **Journal of Nursing Education**, Vol. 49, No. 3, 2010, pp. 164-167.

[20] G. Freiburger and S. Kramer, "Embedded Librarians: One's Library's Model for Decentralized Service", **Journal of the Medical Library Association**, Vol. 97, No. 2, 2009, pp. 139-142.

[21] D. Shumaker, "Who Let the Librarians Out? Embedded Librarianship and the Library Manager", **Reference & User Services Quarterly**, Vol. 48, No. 3, 2009, pp. 239-242, 257.

[22] E. T. Crumley, "Exploring the Roles of Librarians and Health Care Professionals Involved with Complementary and Alternative Medicine", **Journal of the Medical Library Association**, Vol. 94, No. 1, 2006, pp. 81-89.

Literary rewriting through information and communication technologies: an educational exercise

Alexandre GUIMARÃES
Centro de Comunicação e Letras, Universidade Presbiteriana Mackenzie
Rua Piauí, 143, 2º andar, 01241-001
São Paulo, São Paulo, Brasil

Valéria MARTINS
Centro de Comunicação e Letras, Universidade Presbiteriana Mackenzie
Rua Piauí, 143, 2º andar, 01241-001
São Paulo, São Paulo, Brasil

ABSTRACT

The Brazilian educational universe is still far from the complete usage of the resources of information and communication technologies in Fundamental Education. In the overall scenario of the country, generally speaking, few schools even the private ones have the necessary tools of pedagogical-didactic technologies of information and communication. It's worthwhile to point that many teachers and even students are not digitally and mediatically literate to put into practice projects that evolve such technologies. However, it doesn´t exclude the possibility to accomplish productions such as this one having as main goal the stimulation of reading, unfortunately unsuccessful, developing, though, through information and communication technologies abilities of text comprehension in verbal, imagetic, and sonorous levels – linguistic thinking, of effective reading, of rewriting, of reflection, ludic, of rewriting, of intertextuality, of intermediality, of interdisciplinarity, seeking for a significant learning as a result of new effective ways to evaluate the reading process of universal literature icons among students ranging from thirteen to fifteen years old.

Keywords: Evaluation, Intermediality, Reading production, Text production, Information and Communication Technologies.

1. SCHOOL AND READING

Nowadays, in Brazil, the education is structured in: Children´s Education, Fundamental Education and Higher Education.

The Fundamental Education is composed by the Elementary Education I (1st to 5th grades) Elementary Education II (6th to 9th grades) and High School (1st to 3rd grades). When one thinks about the progress of these students it´s right to say that they start the 1st grade when they are about 6 years old and finish the 3rd grade when they are about 17 years old.

Officially, it is on Children´s Education that they start teaching how to read and write. However this is not the, constantly, found reality because once started, several students get to the Elementary Education II without being able to develop a dynamic, efficient and reflective reading that results in a fragile text production.

Poorly prepared, the students end up not developing a strong attachment to the reading exercise which is, indeed, a citizenship exercise.

Besides the indifference due to the lack of sensitization of contextualization and incentives in reading proposals on the side of the teachers, students many times develop antipathy in

relation to the reading practice, mainly, when it´s about the literature icons, a priori when poorly studied, far from the universe of the Brazilian teenager.

> The problem is that the initiation rituals proposed to the beginners don´t seem to please: the literary text, object of a not always discrete but always disturbing, lack of interest and boredom of the faithful – unfaithful – worth pointing, that did not ask to be there. (LAJOLO, 2008, p. 12)

2. INFORMATION AND COMMUNICATION TECHNOLOGIES: SCHOOL AND UNIVERSITY

In Brazil, Fundamental Education schools are maintained by the government, the Municipality, the States and the Union or by private institutions. Generally speaking, but not as a rule, the private schools are the ones that show greater efficiency in the teaching-learning process.

Several questions, such as financial ones involve the acquisition and maintenance of computer equipments. So, there is not the possibility to think, with exceptions, of schools with a broad technological infrastructure.

It´s a fact that teaching is not only through technological tools but it´s also obvious that such tools are extremely relevant in and for the teaching-learning process of youngsters that today live in daily computerized and virtual worlds.

The process of formation of teachers in the Bachelor degrees courses that allow, in Brazil the teaching practice is still old-fashioned, rare are the university courses that really offer in its pedagogical projects subjects connected to comprehension, usage, information and communication technologies reflection.

So, in Higher Education few professionals are formally able to use technologies in his/her teaching routine, also because the future teacher´s molder has not been literated in technologies that, nowadays, serve education.

Besides that scenario that the older college teachers find shelter in questions that capture the technological scenario and his/her lack of inabilities towards these instruments, many professionals and also new teachers do not show interest in working with these tools.

3. THE READING EXERCISE

It´s a habit to imagine that the reading exercise starts in the beginning of the individual schooling phase. Reading is a process that starts out of the school environment. Besides the reading of verbal texts, people are, according to Paulo Freire (2009a) readers of the world.

When they get to school, children have already tried several reading forms and as a consequence, discoveries.

Nevertheless, at school, the process plastering, the not effective reading methods the obligation of reading itself, the lack of interdisciplinary tasks based on reading proposals, the lack of dialogue between the literary text and other text forms and the demands on books reading, lead students to a displeasure when they face reading so important to the creation of a citizen engaged to his/her chronotop.

Choosing the aspects to be criticized is an easier task than building knowledge. The observation of obstacles is not an impeditive for the

proposition of projects of a size that has as a goal the stimulation of a broader reading.

Information and communication technologies, in this stage, are instruments of great importance to the didactic-pedagogical procedures, even with the awareness of the lack of technological apparatus in Brazil.

Some years ago, Freire facing the characteristics of education in Brazil dictated:

> I have no doubt of the great potential of the incentives, challenges and the curiosity that technology put in service of children and teenagers of the so called less favored groups. (FREIRE, 2009b, p. 877)

The profile of these schools, unfortunately, hasn´t changed yet. However, the work here portrayed happened in the city of São Paulo, the largest educational ad financial center of the country.

The school where the present project was created is over a hundred years old, it is located downtown, with a surrounding neighborhood of population from classes A, B, C and D.

The students from this institution, that ranges from children education to strictu sensu post-graduation, also belong to different social status due to government and institution scholarships. Among the Fundamental education students, main characters of the process, many do not have a personal computer at home, and as a consequence internet access.

4. THE LITERARY REWRITING THROUGH TECHNOLOGICAL INSTRUMENTS

The project, developed by the Elementary School II, started with the choice of the books that would better feed the educational needs of the 9th grade. The book selection was done by the group of professionals that taught the classes that would put the proposal into practice.

The first book chosen was *Dr. Jekyll and Mr. Hyde*, by Robert Louis Stevenson. Besides being a classic novel, the masterpiece made possible the work between the subjects of Portuguese, Science and Ethics. In this case, besides reading and literary rewriting, it promoted insights over Ethics in Medicine, so present nowadays, through the means of communication, due to discoveries that evolve bioethics.

The second book of the selection was *Um Certo Capitão Rodrigo*, by Erico Verissimo, an exponent of the second generation of the Brazilian Modernism. The proposal embraced the subjects of Portuguese, History, as the book portraits the Farroupilha Revolution, a bloody civil war that split the south of Brazil, in special the state of Rio Grande do Sul, dividing families with controversies brother versus brother.

The next step was the presentation of the books to the students. In the beginning, the comments about the authors and over the context in how the books were written. Then, the first chapter of each book was read and the teacher added the comprehension out loud in the classroom.

After this reading the students exposed their feelings and when questioned the work ad its contents, the showed their questions. This sensitization stage was an utmost.

In the end, the students were invited to continue the reading at home observing that, during all

the reading process they had the support of the teachers.

The checking of the reading of these books happened in the computer lab of the institution. They were all told about this question that would lead them to the animated narration rewriting of the story.

Analyzing, the goal of the project it is important to remember that:

> Through reading, I´m opening a door between my world and the other´s world. The meaning of the text only completes itself when this exchange happens, when the interchange of meanings one to the other is completed. If I believe that the world is absolutely complete and nothing else has to be said, reading makes no sense for me… I need to be opened to the world´s variety and the ability of the word to say that the reading activity is meaningful. (COSSON, 2009, p. 27)

Ten days later, the students were taken to the computer lab and the rewriting proposal started. In the first class, the teachers explained the resources of the technological and communicational tools that would be used.

In the case of the book *Dr. Jekyll and Mr. Hyde*, the PowerPoint software from the Office package of the Microsoft Corporation was selected. As a bunch of the students did not know how to use the computer thoroughly this tool was used because it is easy to deal with.

For the accomplishment of the proposal of the book *Um Certo Capitão Rodrigo*, the free software Hagáquê, from the Universidade Estadual de Campinas (UNICAMP) a reference in the educational context, available for free in the site of the institution. The Hagáquê is a software that makes possible the creation of cartoons, already with previews of design structures, allowing the student to create new structures and importing them to the software.

Weekly the students went to the lab and, through a previous reading at home, they rewrote the stories using extracts of the texts and of texts in direct speech. The back image of the cartoons, done by the students, were taken from sites and the characters were created by the students. In the end of the project, they added the animation effects.

From the initial incentive and with the possibility of rewriting books through easy technological tools, and also, the freedom of creation, the students felt challenged to read the books and then create their own stories.

> There is an important subject. If the students didn´t read, how would they produce their rewritings? It´s also important to add that even students with great problems in text creation executed the work in good will and the results were more satisfactory than the reading demands of traditional tests. So, reading is not an automatic process of capturing a text as a piece of photosensible paper captures light, but an ordinary, puzzled, rebuilding process and, however, personal (MAGUEL, 1997, p. 54)

The possibility of recreation, with the information technology instruments and also unrestricted access to medias that are interesting to the students such as sites, movies and games is a guarantee of an effective and meaningful reading of the masterpieces.

> The artistic creation started by didactic procedures creates a tension that establishes or breaks boundaries, making possible to the subject to produce knowledge about the object.

> Creating, the individual puts in the limelight structure of values and meanings parallel to the processes developed in class. […]. There is, in the artistic creation of the student, an attempt to answer what was required, but also to reveal him/herself. There is a personal brand in the creation. (PEREIRA, 2010, p. 12)

5. CLOSING

Any teaching activity demands a real, not naïve commitment from the teacher with the pedagogical practice, the contents, the institution, the society and most of all the students.

The teacher must be aware of the cultural and social changes and also, make use of them while working. If this teacher, for instance, faced problems on his/her academic rising, it is up to him/her to rebuild and improve his/ her knowledge.

The technological tools, the spreading of information and the appearing of updated medias must be part of the scholar routine, eve because the students are nearer these languages, especially in the future work environment.

We don´t question the old, also because of the fact that the school is an entity that preserves the knowledge from the past. What is questioned is the non-appropriation of the technological and communicational instruments offered to the pedagogical development, once checked the lack of efficiency of the methods in reading processes and, as a consequence, of its examination.

> There is a broader dialogue between different languages powered by the technological development, and an intense intertextual variety. This rich can be much better understood by us, in its complexity, through knowledge of the original texts in which the media seeks reference. And the school is a favored space where questions about relations between literature, communication and education can happen stimulating the birth of more competent readers and of various texts (HIGUCHI, 2008, p.15)

The teachers were put face to face to the fast access to information through the internet, the contact with different forms of speech and culture, the need of reflex for the selection of coherent information through different sources, development of ability to work with responsible autonomy of the creation of media work (PowerPoint and Hagáquê) in which the verbal, visual and sound languages interchange.

As observed through the works presented we do not announce the use of technological and communicational last generation tools, but demands from the teacher a fundamental knowledge over the information and communication technology that guarantees an acceptable and a more substantial feedback from the students.

BIBLIOGRAPHIC REFERENCES

A. Manguel, **Uma história da leitura.** São Paulo: Companhia das Letras, 1997.

K. K. Higuchi, **Literatura, Comunicação e Educação:** um romance em diálogo com a mídia. São Paulo: Cortez, 2008.

K. Pereira, **Como usar artes visuais na sala de aula.** São Paulo: Contexto, 2010.

M. Lajolo, **Do mundo da leitura para a leitura do mundo.** São Paulo: Ática, 2008.

P. Freire, **A importância do ato de ler:** em três artigos que se completam. São Paulo: Cortez, 2009a.

_____. **Pedagogia da autonomia:** saberes necessários à prática educativa. São Paulo: Paz e Terra, 2009b.

R. Cosson, **Letramento literário:** teoria e prática**.** São Paulo: Contexto, 2009.

The usage of the computer as a tool in the development of the teenager´s reading and writing process

Alexandre GUIMARÃES
Centro de Comunicação e Letras, Universidade Presbiteriana Mackenzie
Rua Piauí, 143, 2º andar, 01241-001
São Paulo, São Paulo, Brasil

Anne PERES
Centro de Comunicação e Letras, Universidade Presbiteriana Mackenzie
Rua Piauí, 143, 2º andar, 01241-001
São Paulo, São Paulo, Brasil

The usage of computers has changed the way to conceive new ideas and also to communicate them to our pairs in the XXI Century. It can be called a "New Rebirth" or a new species of spreading and culture acquisition of culture.

Up to this point, we intend to elucidate how it is possible to make the use of information and communication technologies a successful instrument, in a real teaching-learning tool in relation to the pedagogical works, facing the acquisition and maintenance of the reading habit and the written production, in the Brazilian fundamental education.

The hypermedia is a way of communication in the digital system that enables the creation of a text that aggregates and mixes diverse Medias and languages.

The adolescent, the majority of the students enrolled in Brazilian High Schools, ranging between 15 to 17 years old, is a user of these ways of communication, providing greater creativity, besides linguistic abilities.

There is through the hypermedia the possibility to express in a way to give conditions of providing the group what they understand as information, allowing, thus, their mere receivers' condition, offering them the role of contents generators.

Facing this reality, we present a proposal to evolve the student in the reading of a book written five centuries ago in Portugal. *Auto da Barca do Inferno*, by Gil Vicente, pictures a reflex of the change of time and the change from the Middle Ages to the Rebirth.

The masterpiece written in an age of transition presents a period that the hierarchy and the social order were conducted by inflexible rules, the Middle Age and the appearing of a new society started to revolutionize the established order when questioning it.

The reading of the previously mentioned book is demanded in High Scholl because besides all cultural questionings, is part of the obligatory readings of one of the greatest processes to enter the University in Brazil, the FUVEST.

The work proposed evolved Portuguese, History and Arts teachers, that received as a feedback from the reading production in the computer lab, of a hypermediatic product that, as published on the internet afterwards. For this production the students used information obtained during the Arts, Literature and History classes. They chose a character from the Vicentine book proposed and described the organization of the Portuguese and Brazilian societies of the 16th and 21st centuries.

Besides converting students in authors, the productions weren´t done only for the teachers, a very common practice, unfortunately, in the school environment, the project worked with the real production development and the textual structure in virtual environment.

Developing the Humanness of the Art of Nursing: Using Technologic Tools to Support Distance Nursing Curricula

Linda L. Strong
Nursing Department, Sacred Heart University
Fairfield, Connecticut

and

Debbie L. Shadd Simmons
Nursing Department, Sacred Heart University
Fairfield, Connecticut

ABSTRACT

The explosive application of information technology can be seen in all aspects of daily and professional life. Technology has revolutionized the way we live, learn, do business, communicate and play. In nursing and nursing education this technology has had significant impact on skill training and clinical thinking. On-line and distance-learning formats permit individuals the ability to learn from a variety of locations and the ability to learn twenty-four hours per day. The use of technology is pervasive in nursing education and practice. But is that all that is can contribute? While technology has blossomed, the humanistic segment of practice has been minimized. The basis for humanistic practices is the development of aesthetic knowledge. Aesthetic knowing is associated with the integration of the arts and humanities into the nursing curriculum and nursing practice. In traditional curriculum planning, this is easily accomplished but a challenge in online classrooms. This presentation describes the use of information –technology in a distance-learning/on-line undergraduate and graduate nursing program. Discussion of the teaching strategies used in the virtual classroom to integrate and encourage aesthetic knowledge development. Recommendations for educational research into the meaning of technology and development of aesthetic knowing, and other methodological inquiry is explored.

Key words: Aesthetic knowledge, Humanness, Online learning, Technology

INTRODUCTION

The explosive application of information technology can be seen in all aspects of daily and professional life. Technology has revolutionized the way we live, learn, do business, communicate and play. What once seemed an oddity telecommuting employees are now common. Businesses and universities may no longer be situated solely in a specific location; rather they may have a dual existence, one tied to a parcel of land and the other traced to a virtual on-line existence. Faculty and other employees share in this duality as well, many academics may never visit the university where they are faculty members, but instead commute to their university and communicate with fellow faculty members through electronic media. They use information technology in the same means by which they teach and interact with students. Technology is one of several factors which influence nursing and nursing education in the new millennia . Western approaches to medicine, nursing and healthcare have caused a fracturing of the arts and sciences of the caring professions. As well, the segregation of the liberal arts and humanities in professional academic settings has eroded the ability to develop competence in the aesthetic knowledge which supports clinical reasoning and affective learning outcomes as a whole. There is a great deal of evidence which supports re-integration of liberal arts and humanities into the curricula of the helping professions such as nursing but advancing trends in technology for education create certain barriers. The following discourse presents one University Nursing Departments attempt at this re-integration while attending to the need for meeting consumer demands for flexible classroom settings.

FACTORS INFLUENCING NURSING AND NURSING EDUCATION

Heller, Oros, & Durney-Crowley (nd) suggested that in the new millennium nursing education would need to respond to ten emerging trends:

1. Changing demographics and diversity
2. Explosion of technology
3. Globalization of economy and society
4. Era of the educated consumer requiring practitioners knowledgeable of alternative therapies, genomics and palliative care
5. Increased complexity of care and a shift to population-based care
6. Challenge of managed care to answer the increasing costs of health care
7. Implementation of new federal and state policies and regulations to counter costs and to define and measure quality
8. Need for Interdisciplinary education and collaborative practice
9. Implementation of new federal and state policies and regulations to counter costs and to define and measure quality
10. Need for Interdisciplinary education and collaborative practice
11. Persistence of a nursing shortage, need for life-long learning and work-force development

12. Expanding body of nursing research and advances in nursing science. (p.1-7)

Each of these trends can be explored in isolation. Each of these trends suggests changes in nursing curriculum and practice. For instance, the United States alone, 22% of the population will be 65 by 2030 (Vincent et al. 2010), as the population ages the prevalence of chronic illness increases, almost 75% of those age 65 and older have at least once chronic illness, and about 50% have at least two chronic illnesses (AHRQ, 2002) and despite increasing frailty and loss of physical and mental ability studies show that 95% of seniors would prefer to age in their own homes and live independently (He et al. , 2005). In this one statement can be seen five out of the ten trends listed. But when this statement is examined as a macro-picture an interrelated phenomenon appears and all ten trends emerge. The existing cadre of nurses and other health care providers are not prepared for these changes and must receive both preparatory and on-going education and skills. Responding to these predicted trends raises the question; is there one or several best way(s) to address these challenges?

Educational technology would seem to offer one of the most efficient means to provide the knowledge and skills needed to meet such changes and prepare the health profession workforce. Nursing education has used technology to significantly impact skill training, clinical thinking and clinical practice. Neuman (2006) commented that nursing education has embraced the innovations offered by technology. The rapid growth of health related technology provides nursing and other health profession students with virtual reality via games and others simulation experiences thus enhancing the delivery of education "through immersion experiences that allow users to absorb knowledge through all their senses" (Neuman, 2006,p.2). Nursing students of today, and other students of the millennial generation, are expecting to engage the technology as a routine part of their educational experience. Skiba & Barton (2006) stress that students in the Net generation prefer to be taught with teaching strategies which promote digital literacy allow for interactivity and immediacy and finally, supports experiential learning.

Nursing education is no longer tied to a specific location. On-line and distance-learning formats permit individuals to learn from a variety of locations and the ability to learn twenty-four hours per day-which in most instances can promote efficiency with learning. And while there are mixed opinions to support that e learning is more efficient in terms of time engaged in learning, these technology based learning environments produce enhanced learner outcomes (Cook, Levinson & Garside, 2010). Just as the reign of the dinosaur has become an era of history so too have restrictions to learning. Educational technology has unearthed a new creative horizon for the acquisition of skills, clinical knowledge and simulated practice in nursing education and nursing practice The acquisition of these skills and the ability to manipulate this technology prepares the student and the current practitioner to operate in a highly technologic environment, whether it is in acute care or extended care facilities or in the home. (Benner, Sutphen, Leonard, & Day, 2010).

But is efficiency all that technology can contribute? Is there more to practice than the knowing how to perform skills, to knowing the rationale for why health care providers intervene through ordered treatments, tests, medications, and diets? If so, what is the missing element of this education, preparation and formation into a professional nurse? Can educational technology help to resolve the missing link?

THE MISSING ELEMENT

Westernized medicine and the education of physicians, nurses, physical therapists and other helping professions have followed the biomedical model of practice, a model punctuated by four characteristics;

- illness perceived as a biological process,
- interventions [that] are guided by scientific principles [and] that are supported by evidence evolved from extensive research,
- a dichotomization of body and mind, and
- an emphasis on cure rather than prevention.

A consequence of this model has been the segmentation of knowledge and skills into distinct schools of practice and education that distanced the knowing and doing of health care and curing from the feeling and arts of caring. However this compartmentalization is not limited to the health sciences. The university environment has existed in knowledge silos since the 1800s, and this academic specialization has resulted in a lack of connectedness between knowledge and ideas and majors. Liberal arts and the humanities have suffered as have students and faculty. (Warch, 1990)

In the mid to late 20th century the biomedical model began to be questioned as the singular focus on knowledge for doing alienation of the knowledge of caring. The void left by the segregation of humanities and the liberal arts from the curriculum of health professions saw the emergence of various specialties that incorporated the social sciences, ethics and law into medical, nursing and other education arose.(Greaves, 2001). However, these new disciplines made a moderate impact on the resolution of the compartmentalization of knowledge, courses and technology have blossomed, but the humanistic segment of practice has remained minimized. Greaves recommendations calls for the re-conceptualization of the methods used to prepare health care practitioners while specifying the need for integration of the liberal and social arts along with the humanities.

CRISIS IN NURSING EDUCATION AND PRACTICE

The recognized need for nursing education to be closely aligned with the liberal arts and the humanities has existed sense its inception as profession, the degree to which this association has varied, with a distance between them occurring in the early twentieth century. Over the span of that century it became clear that this segregation was harmful to the formation of the practitioner and to the practice as a whole. (Hermann, 2004).

Sullivan (2005) noted in an interview on his text, *Work and Integrity: The Crisis and Promise of Professionalism in America* that nursing and other health professions are challenged by the current social climate that calls for a focus on thinking and doing that is primarily based in the utility and instrumentality of the knowledge and associated skills. This focus limits the development of the values that undergird the caring of these professions and the responsibility of and for these professions by their practitioners. (Sullivan & Benner, 2005). It inhibits the health professions' from investing in the development and sustainability of the "quality of their craft, the inventiveness of

their practice…a contribution of public value and as a source for motivation and deep personal satisfaction" (p.78). This statement suggests the potential relevance of educational technology in promoting the cognitive and affective domains of learning.

In this interview Sullivan asks this question as he commented that while professional education has done an exceptional job in preparing practitioners in the analytical and scientific portions of practice they have been less than adequate in "teaching skillful practice and wise judgment" in complex situations. (Sullivan & Benner, 2005, p.78; Benner, Sutphen, Leonard, & Day, 2010, p. 8-9). This void underscores the lack of success that education in integrating thinking and practice skills with the development of social responsibility and social contracts.

Benner et al. (2010) concurs that nursing programs in the United States are not effective in teaching "the nursing science, natural sciences, social sciences, technology, and the humanities" (p.12) Concentrating on analytical and scientific thinking unwittingly perpetuates the long-held view of the mind-body dichotomy, where disease is studied without reference to the meaning attached to the disease. Students amass discrete facts and skills. It produces a proficient technological professional who acts in concert with the current practice environment that is focused on efficiency, measurable competencies, and cost-savings. (Benner, 2010; Sullivan & Benner, 2005).

While we have prepared graduates skilled and competent in applying nursing knowledge and performing the psychomotor aspects of care, what has been missing has been the connection of this work to addressing the needs of the society. The understanding of the relationship of the profession to the greater good has slipped from the profession's grasp (Benner, 2011; Sullivan & Benner, 2005). We have lost the understanding and valuing of liberal arts and the humanities. We have lost the ability to see and understand that the individual lives in a "historical and life-world" that has been invaded, attacked, or otherwise penetrated by disease and that this has meaning to individuals, families and societies (Benner, 2011).

THE CASE FOR LIBERAL ARTS AND THE HUMANITIES

Hutchins (1952) noted that liberal education seeks to "clarify the basic problems and to understand the way in which one problem bears upon the other" (p.3). Liberal arts prepares students to appreciate the differences and similarities between different fields, to be able to use the ways of knowing of these fields to address problems, to discern appropriate interventions, and gives the student the ability to "read, write, speak, listen, understand and think" (Ibid. pg. 4).

Mantzorou & Mastrogiannis (2011) and Casey (2009) speak to the art of nursing and the ways of knowing. Of the four patterns of knowing as described by Carper (1978) the liberal arts and the humanities contribute most to the aesthetic ways of knowing in nursing practice. Aesthetic knowing is the understood to the "direct feeling of experience" (Carper, 1978). It is through this feeling that the nurse connects to the person, with a genuineness of spirit and regard, exuding caring and translating this into nursing care that values the person, their point of view, and their differing opinions. How this aesthetic knowing is acquired is often equated with schooling in the liberal arts and the humanities. Casey (2009) writes that the arts and literature are often used to foster creative thinking and inquiry, as the story that is told is a metaphor for the feelings, doubts, problems, fears, and joy associated with health and illness. It is through the use of the liberal arts and the humanities the practitioner moves from their own world to that of the other, it allows the practitioner to notice the importance of the world lived by the other, and in so doing provides a way of knowing that is unlike those of empirical, ethical, and clinical knowing. Yet when all four types of knowing are synthesized the understanding of the clinical picture provides for a richer, broader field from which to inform clinical reasoning and decision making. The importance of the liberal arts and humanities for nursing education and practice can be best summarized in an essay by Link (2009) titled *Why we need the humanities* answers the question stating that we "need the humanities because humanity itself is something we all share" (p.3).

The American Association of Colleges of Nursing (AACN) articulates the need, importance, and significance of the liberal and social arts and humanities for undergraduate and graduate nursing. These courses provide the foundation for viewing and understanding the diversity of cultures, languages, religions and life-ways. They offer alternative ways of knowing and thinking about issues and problems that may be missed by focusing solely on the thinking taught in nursing education. They promote "an understanding of self and others and contributes to safe quality care" (Essential of Baccalaureate Nursing, 2008, p.11). These courses promote the development of "skills of inquiry, analysis, critical thinking and communication" (p.11) which can be used in collaboration with other health care providers. Together with nursing courses they co-create and form the graduate into a practitioner that can articulate their values, engage in a practice that is ethical, altruistic and unbiased, one who can advocate for the individual and the community, and can be involved in making the society into a better place (Essentials, 2008; Benner et al., 2010; Benner, 2011; Hermann, 2005; Sullivan & Benner, 2005).

Education must prepare both the former as well as prepare and develop a nurse who understands and values the connections among nursing science, natural sciences, social science, technology and the humanities. This encompassing approach will result in a skillful artful practitioner with strength in the science as well one who has a depth and richness in practice that is founded on the humanness of care and caring. An education that builds on the integration of these factors will be a factor in the revolutionizing nursing education as it will be one of the building blocks of curriculum that promote salience and situated cognition and clinical reasoning (Benner et al., 2010).

NURSING CURRICULUM: LIBERAL ARTS AND NURSING EDUCATION

While it is clear that the liberal arts and humanities are crucial to nursing education, what is not so clear is how to best deliver this content (Hermann, 2005). Development of the humanistic segment of practice has been addressed by numerous authors (Carper, 1978; Chinn & Watson, 1994; Darbyshire, 1994; Casey, 2009) and is perceived as the way of aesthetic knowing. Aesthetic knowing is grounded in perception, interpretation, and expression (McEwen & Wills, 2011). Aesthetic knowing is associated with the integration of the arts and humanities into the nursing curriculum and nursing practice.

The use of the arts in nursing education has been reported in the literature (Casey, 2009; Leight, 2001; Mareno, 2006; Smith, Bailey, Hydo, Lepp, Mews, Timm & Zorn, 2004; Wainwright, 2005). Roberts (2009) used poetry to promote emotional intelligence, Mareno (2006) described the use of the art of the Great Masters to promote critical thinking, and the purpose of Leight's (2001) work was to enhance the awareness and understanding of women's health. Common to these authors and others is that all of the courses were taught in traditional on-ground nursing courses.

In the era on on-line/distance learning where course faculty, students and oil and canvas, dance, sculpture, museums and all the other forms of art are distant to both the teacher and the learner how can nursing faculty develop aesthetic knowing?

EDUCATIONAL TECHNOLOGY AND AESTHETIC KNOWING: DEVELOPING THE HUMANNESS OF NURSING

The challenge facing nursing curriculums that are taught online is how to use educational technology to teach in not only the cognitive and psychomotor domains of learning but how do they do so for the affective domain. To be considered are: what course designs are most amenable to achieving the outcomes associated with affective/aesthetic learning, what technology is available to help meet these objectives, and how to evaluate affective/aesthetic learning and achievement of course outcomes? The following section will describe how one department of nursing answered the questions and challenge of integrating aesthetic knowing and the liberal arts into a distance-learning curriculum using educational technology.

The decision to engage in a full evaluation and re-design of a long-standing and well proven RN-BSN/MSN curriculum was prompted by several factors: a new core-curriculum; The Human Journey instituted at the university level, an Association of American Colleges and Universities Core Commitments initiative; a reaccreditation visit by the Commission on Collegiate Nursing Education (CCNE) and new curriculums at the first professional degree and graduate levels; changes in the Association of American Colleges of Nursing Essentials (AACN) document, and reports by the Institute of Medicine (IOM) Reports on Safety and Quality and PEW reports on the Future of Nursing. In turn they also informed the direction of the curriculum redesign. Existing program characteristics that were not open for discussion were the commitment to on-line learning and a commitment to the established reputation of the RN-BSN/MSN curriculums for rigor and quality.

In order to develop course curricula supportive to the liberal arts core curricula, it was necessary to articulate the concepts with professional nursing concepts and theoretical knowledge. Aesthetic knowing is the basis of the art of nursing; clinical practice techniques performed properly represent aesthetic knowledge. Likewise is clinical reasoning and the ability to transfer and translate empiric knowledge to the clinical scenario an example of aesthetic knowledge. Most nursing curricula separate these two ways of knowing and use technology that is fairly unique to each situation to support knowledge development. Clinical reasoning strategy is usually focused on development of empiric knowledge but it is also important for affective learning. Nursing students must learn to use clinical reasoning relative to basic human values when engaged in clinical practice. The use of technology for education of nursing students has provided nursing faculty with new tools for both development and evaluation of learning but also the potential for analysis of specific technology to support this knowledge development. In order to further understand how technology is employed in nursing educational environments where the learner must engage technology for learning to occur, we must discuss knowledge development relative to technology.

Pickstone (2001) discusses that technoscience now involves our everyday human existence and while we may not fully comprehend the intricacies of each technology, we understand its value. And through our understanding of its value, we necessarily explore its meaning. These authors would go further to assert that nursing education curricula which engages the idea of technoscience as a way of knowing, and uses technology for the benefit of the learning process, facilitates learner ability to extrapolate the value and meanings of their world. Technology for classroom learning involves the technology used in support of a teaching strategy classroom but also includes the technology which supports or creates the classroom environment. This paper discusses the use of technology to create deeper meanings for students relative to specific course curricula and further asserts that the asynchronous e-environment of the online classroom can allow for a reflective dialogue of shared meanings to occur. In this way aesthetic knowledge (know how) is expanded to include knowing how to translate conceptual knowledge within the framework of a discipline and beyond and knowing how to use technology to support this learning process.

The nature of the online environment can facilitate greater acquisition of aesthetic knowledge. In synchronous (or real-time) learning environments the student must have the ability for instant recall and has little time for reflective thought. Often times this classroom environment does not foster student reflection as they are challenged to attend to didactic lecture simultaneously. The nature of the asynchronous online classroom allows the student time to engage the classroom readings and other media before offering reflective response to discussion questions, written assignments etc. It is important to remember here that development of aesthetic knowledge depends greatly on the time engaged in reflection relative to the learner values and experiences that can be measured against this new classroom content.

E-learner environments use many different technologies and approaches to support acquisition of aesthetic knowledge. Many are delivered in similar ways to those used in traditional classrooms but the learning environment itself is web based. These strategies for nursing education can include visual and auditory media in static or streaming format (Schermer, 1988; Smith-Stoner, M. & Willer, A., 2003) written reflection (Epps, S: 2008; Cohen & Welch, 2002) and even kinesthetic type (Meehan-Andrews, T., 2008). In synchronous online learning environments the use of telecommunications is often used to support acquiring nursing knowledge. This includes web-based chat, Skype and other real time technologies. While many of these strategies are used in online learning and the traditional classroom, most are not directed with the primary intention to gain aesthetic knowledge. There is a relative dearth of literature describing the primary use of any of these strategies in other than the clinical setting of learning (Northington, L., Wilkerson, R, Fisher, W. & Schenk, L., 2005) to support development of aesthetic knowledge. The following information describes the experience of these authors with development and

implementation of a nursing course specifically designed to ensure development of aesthetic knowledge.

THE HUMAN JOURNEY OF NURSING: THE INTEGRATION OF LIBERAL ARTS AND HUMANITIES

The University's core curriculum; the Human Journey consists of four essential questions that are threaded through four required courses for all undergraduate students. Taken in four disciplines; History, English, Social or Biological Science and Religious studies students are asked to explore the disciplines through the lens of these questions: what does it mean to be human, what does it mean to lead a life of meaning and purpose, what does it mean to understand and appreciate the natural world and what does it mean to forge a more just society for the common good?

This exploration uses various forms of literature; poetry, short story, fiction, novels, historical records, and performing arts. For those students who choose the biological sciences over sociological study they explore the questions using works from science regarding topics such as the environment and sustainability, genetics, health and disease such as HIV AIDS. Students studying the social sciences read works by psychologists, sociologists, and political scientists and consider questions such as what makes people prejudiced, or issues of poverty, or the tension between individual rights and the common good. These strategies are the primary methods and approaches used in the liberal arts and sciences. However, two to three years after implementation of these courses on-ground they needed to be re-mastered into 8-week sessions taught solely via-distant learning technology.

The Human Journey in Nursing is a vertical development of these courses, builds upon the four essential questions and asks four correlated questions. Nursing in its theoretical discussions has used literature from humanities and liberal arts to support these questions. The Nursing Code of Ethics, discusses of the need for evidence in nursing practice, and for social justice in public health courses. They have been explored using poetry (Roberts, 2009), and art (Mareno, 2006).

EDUCATIONAL TECHNOLOGY: SCULPTING A COURSE

Technology and Design

However, when grouped together these question are situated differently, the questions require an alternative way to know the answers, the answers cannot be found solely through empirical, ethical, and clinical knowing, aesthetic knowing is essential to finding the answer to these questions. Educational technology is the tool that permits the integration of aesthetic knowing with the knowing of empiricism, ethics and praxis. In building the course the designer was explicit in what was integral to the course, use of visual and audio technology beyond the simple taping of a lecture or use of formats for synchronous participation any methods could not jeopardize the asynchronous design of the course. Web-sites must be used that would allow students to visit museums and collections that were distant to their location of study. Assigned and supplemental readings would be diverse, sometimes controversial, and promote integration of each of the four university and nursing core questions. The end product *The Human Journey of Nursing* uses audio recordings of poetry, the technology of Voice Thread, and visual access to paintings and virtual trips to museums. Not yet available is the technology of touch or smell but when this technology becomes available it will be quickly incorporated.

The nursing course was specifically designed by one of the authors for the purpose of transferring foundational knowledge from the liberal arts core curriculum which is based on four key concepts:

- Being human
- Living with meaning and purpose
- Understanding and appreciating the natural world
- Forging a just society for the common good

and translating this for the discipline of nursing to:

- Protecting and promoting patient health and well-being
- Embracing core beliefs, values and standards of the profession
- Examining advancing science and technology in nursing practice
- Modeling professional behaviors which contribute to the greater good of society

The course was author, in consultation with both an instructional designer and online teacher expert, matched teaching materials and strategy to ensure that aesthetic knowledge was gained as a primary objective and clinical reasoning was advanced via the application of this knowledge to clinical scenario's and discussion. This course design required no clinical practicum for the student. Student's clinical knowledge to engage the clinical scenarios and other course activities was however supported by the fact that they were practicing nurses enrolled in a degree completion program.

This course utilized both embedded and web-based technologies to support the objectives in addition to traditional written assignments. See Table 1 for an excerpt of primary assignments and matched teaching strategies for the four translated areas of the core curriculum. Voice thread is a collaborative multimedia site that can hold visual media in any form, and allow multiple participants to comment on the held media. Directed web-surfing allowed students to explore sites that promote the core values in various ways. Reflective journaling (technology mediated preferred) allowed the students to document initial thoughts regarding comparing assigned readings to their own experiences and/or examining their own ideas about a sub-concept presented in the course. The platform supported discussion board provided a means to share collaboratively and evaluate student understanding of course content via weekly discussion questions. Streaming media in general offers the learner the opportunity to tap into their own ability to learn through sensory information. And finally, students were encouraged to use technology for completion of the interview assignment – synchronous strategies were desired.

Media streaming was used in both audio and video formats within the course. Audio streaming was used frequently within the course with the poetry presentations primarily allowing for the capture of student initial impressions and commentary. These comments were posted in audio or video format and shared amongst the group. Using this tool allowed the students

to develop shared meanings regarding specific media designed to promote the essence of the core value. Video streaming was used to explore paintings and sculpture while taking virtual tours in museums. Streaming media such as video and audio can help learners understand complex concepts and procedures that are difficult to explain with simply text and graphics (Klass, 2003). "By using visual and auditory messages, students can process the information quicker, which in turn, helps foster their learning acquisition of the material" (Hartsell & Yuen, 2006, pg.32).This media stream strategy was important because we are human beings who base many of our initial impressions and understandings of our world on our visual and auditory representations.

Table 1. Excerpt of course teaching plan

Assignments	Technology based teaching strategy	Technology tools
Group Analysis of 4 Books - thematically selected	View the painting: *The cycle of terror*, Parrish, G. (2006)	Voice Thread
Selected literature for review – includes humanities based literature	Listen to this poem: Dickinson, E. (nd), *Hope.* Audio version: http://poetryoutloud.org/poems poem.html?id=171719	Audio Streaming
Discussion of national resources	Web based exploration of sites: AHRQ	Web-surfing
Reflective journal	Online journal	Learning Management System (LMS) tool
Interview on Service learning	Conduct technology mediated interviews & archive then reviewing for thematic analysis	Virtual Video conference, Chat or audio recording device

Directed web-surfing encouraged exploration of sites that were data repositories for nursing information on quality, healthcare, and cultural diversity. This technology also encourages guided exploration of the internet for the purpose of gaining skill with identifying and evaluation professional resources on the web. In a recent study Schullo et al (2005), found that half of the students surveyed in an online course found this approach useful to their learning.

Reflective journaling allowed the students to document initial thoughts regarding comparing assigned readings to their own experiences and/or examining their own ideas about a sub-concept presented in the course. Students were encouraged to include all forms of expression within the journals to respond to the question posed. "Journals occupy a unique space in the array of reflective practices by giving students a safe place to withdraw temporarily and create an ongoing and informal record of meaningful aspects of their own learning processes" (Mills, 2001, pg.27). These individual journals were submitted for faculty feedback. Exposure to a web-based journal was an option which allows for beginning skills with other types of journaling like blogs.

Markel (2001) asserts that the discussion board is a key vehicle in asynchronous online education. The discussion board was utilized on a weekly basis and this technology services as a shared format for individual journaling which was monitored frequently each week throughout the course. It was primarily used for guided comparative reading activities similar to those described by Alstete (2007).

Chats and Skype type technology had the ability to connect the student across great distances to conduct the interview as directed. This format allowed for archival retrieval of the interaction and analysis of themes associated with the communication. Students were also introduced to a web based technology that could be used for real-time communication in future professional endeavors. Avery, Cohen and Walker (2008) discuss the importance of interaction within online courses to ensure quality in online nursing curricula. Synchronous learning strategies support this interaction. While the bulk of the interaction in the course should be amongst the learner group, along with the faculty member, the opportunity to interact with a professional nurse outside of the classroom offered students the ability to make comparisons relative to the core values.

Technology and Course Content

The first module *Pathway: Past, present and future* begins with exploration of the first question; what does it mean to be human- how does the profession protect and promote the health and welfare of patients. The concepts explored include vulnerability, resilience, spirituality, safety and diversity. Using educational technology the first assignment is to view painting by Grayden Parrish (2006) *The cycle of terror,* using Voice Threads student's record their first five impressions of this painting. Each student participant must listen to the other classmates and then conduct a discussion through written response. Other portions of required readings/assignments include listening and responding to the audio version of Emily Dickinson's *Hope* and written version of Edna St. Vincent-Millay *To the wife of a sick friend*. A course assignment begins in this module, student's form groups to select one book; T. Kidder (2004) *Mountains beyond Mountains*; A. Fadiman, (1997) *The Spirit Catches You and You Fall Down; J.* Barry, (2005) *The Great Influenza; or R.M. Zaner (2005) Conversations on the Edge* to read, analyze and lead the class in discussion of how the book addresses the four essential and nursing core questions.

The second module *Do no harm* explores the second question what does it mean to lead a life of meaning and purpose- what are the core beliefs, values, and principles of the profession. The concepts explored include codes of ethics, caring, care giving and the experience of caring. Directed to read Robert Wadsworth Longfellow's (1893) Santa Filomena; Milton Mayeroff (1971) *Major ingredients of caring and* P. Wicker (1988) *When caring doesn't mean cure* students must debate in written response the similarities and differences of the meanings of caring that are stated in these passages.

Module three *Who we are or who are we* explores question three what does it mean to understand and appreciate the natural world- what is the effect of emerging scientific technology. The concepts explored include evidence, alternative care and

stewardship. Directed to read required content students must also explore the National Center for Alternative and Complimentary Medicine and the Richard and Hinda Rosenthal Center for Complementary and Alternative Medicine and relate the findings of this exploration.

The final module Citizen Nurse explores the fourth question what does it mean to forge a more just society for the common good-what is the responsibility of the profession in contributing to the greater good for society. The concepts explored include the social policy statement of nursing, historical and social influences, and race, gender and class. Directed to read required content students must read or listen to John Milton (1608-1674). On His Blindness, read Robert Penn Warren (1905-1989) *Timeless, Twined* and *Truth*, and revisit Grayden Parrish (2006) *The cycle of terror, and* using Voice Thread to record their impressions of this work and how the course has impacted their understanding of this work.

DISCUSSION

Educational Technology: Lessons learned

The sculpting of a course that is committed to integrating aesthetic knowing and incorporates the liberal arts and humanities requires partnership with colleagues in the disciplines of English, Art, History, Philosophy and Religious studies, as this partnership extends past the sharing of references. It requires, as is consistent with most interdisciplinary initiatives, the ability to see the content and meaning through their lenses. It requires the time needed for discussion and exploration of the concept and content in order to speak to these concepts as an ally.

Secondly, the partnership with librarians is invaluable, and this is not just the librarian for the medical arts. The use and integration of the liberal arts required the knowledge of these works from the librarians that are the caretakers of this work.

Thirdly, the partnership with the instructional design technologist is invaluable, they can take the vision of what is wanted and make it real. Streaming media and the other tools used within the course facilitated student visualization and listening of expressive thoughts in literature and performing arts. It also allowed for the capture and sharing of student reactions to visual and performing arts of student reactions to a controversial paintings and writings, so that shared meaning and reflection could occur . In the future this partnership will permit inclusion of dance and music in the next version of the course and hopefully a greater opportunity for kinesthetic interactions via planned and recorded, non-virtual travel experiences.

Implications for further study

A course such as the Human Journey of Nursing opens the avenue for research into as to the impact of this course on the student who takes this course using distance learning technology. While there are many avenues this research might take there are some key areas which might be explored directly as a result of this foundational work.

Quantitative work might begin with examining:

- Student's satisfaction with this technology and efficiency of use within the online classroom ;
- Educational outcomes research relative to impact on affective and cognitive learning goals with nurses who receive technology mediated curricula in online environments for the purpose of translating the liberal arts core curricula to the professional nursing values.

Qualitative endeavors might include:

- Student perceptions of value of liberal arts in translating nursing professional values.

CONCLUSION

The challenge of re-integration of the liberal arts and humanities into professional nursing curricula remains for most colleges and university faculty. Developing curriculum which translates this foundational knowledge as a basis for humanness of nursing supports current healthcare issues and trends relative to quality and coordination within communities. Teaching strategies within the classroom should reflect affective learning activities which support aesthetic knowledge development used for clinical reasoning and caring. Technology presents certain barriers and constraints within the online classroom and an even greater challenge exists when combined with a distance format for offering. Faculty need to be open to and competent with using newer synchronous and asynchronous technologies and e-tools to support these learning goals.

REFERENCES

AHRQ Publication No. 02-0018. (2002, April). *Preventing disability in the elderly with chronic disease.* Research in Action, Issue 3. Agency for Healthcare Research and Quality. Rockville, MD. Retrieved 12/17/10 http: //www.ahrq.gov/research/elderis.htm

Alstete, J (2007). Using comparative reading discussions in online distance learning courses. *International Journal of Instructional Technology and Distance Learning 4*(10) 51-61

American Association of College of Nursing, (2008). The Essentials of Baccalaureate Education for Professional Practice. Washington, D.C.: American Association of Colleges of Nursing.

Avery, M., Cohen, B. and Walker, J. (2008*).* Evaluation of an online graduate nursing curriculum: examining standards of quality. *International Journal of Nursing Education Scholarship 5*(1): art. 44

Benner, P. (2011). Formation in Professional education: An examination of the relationship between theories of meaning and theories of self. *Journal of Medicine and Philosopy,* 1-12.

Benner, P., Sutphen, M., Leonard, V. & Day, L. (2010). Stanford, CA: The Carnegie Foundation for the Advancement of Teaching.

Carper, B.A. (1978). Foundational patterns of knowing in nursing. *Advanced of Nursing Science,* 1(1), 13-23.

Casey, B. (2009). Art-based inquiry in nursing education. *Contemporary Nurse,* 32(1-2), 69-82.

Cohen, J. and Welch, L. (2002). Web journaling: using information technology to teach reflective practice. *Nursing Leadership Forum 6* (4): 108-12

Cook, D., Levinson A and Garside, S. (2010). Time and learning efficiency in internet based learning: a systematic review. *Advances in hearth sciences education: theory and practice. 15*(5): 755-70

Epps, S. (2008). The value of reflective journaling in nursing education: a literature review. *International journal of nursing studies 45*(9): 1379-88

Greaves. D. (2001). Two conceptions of medical humanities. *Nursing Philosophy*, 2, 270-271.

Hartsell, T., & Yuen, S. (2006). Video streaming in online learning. *AACE Journal, 14*(1), 31-43.

He, W., Sengupta, V.A., Velkoff, V.A. & DeBarnes, K.A. (2005). *65+ in the United States*. U.S. Census Bureau, Washington, D.C. Pp.23-209.

Heller, B.R., Oros, M.T. & Durney-Crowley (n.d.) The future of Nursing Education: Ten trends to watch. *National League for Nursing.* Accessed 10/09/11. Available: www.nln.org/nlnjournal/infotrends.htm .

Herman, M. (2004). Linking liberal & professional learning in nursing education. *Liberal Education,* Fall, 42-47.

Hutchins, R.M. (1952). The Tradition of the West. *The Great Conversation: The substance of Liberal Education. Great Books of the Western World. (Vol.1). Chicago, Encyclopedia Britannica, Inc.*

Klass, B. (2003). Streaming media in higher education: Possibilities and pitfalls. *Syllabus*, *16*(11). Retrieved Oct 144, 2011, from http://www.syllabus.com/article.asp?id=7769

Leight, S.B. (2002). Starry Night: Using story to inform aesthetic knowing in women's health nursing. *Journal of Advanced Nursing, 37*, 108-114.

Link, P. (2009). *Why we need the humanities.* College of Humanities, Arts and Social Sciences, University of California, Riverside.

Mantzorou, M. & Mastrogiannis, D. (2011) The value and significance of knowing the patient for professional practice, according to the Carper's patterns of knowing. *Health Sciences Journal, 5* (4): 251-261.

Mareno, N., A. (2006). A nursing course with the Great Masters. *Nursing Education Perspectives,* 27(4): 182-183.

Markel, S. (2001). Technology and education online discussion forums: It's in the response. *Online Journal of Distance Learning Administration, 4*(2). Retrieved October 14, 2011.

Meehan-Andrews, T. (2008). Teaching mode efficiency and learning preferences in first year nursing students. *Nursing Education Today 29* (1): 24-32

Mills, S. (2001). Electronic journaling: using the web-based group journal for service learning reflection. *Michigan Journal of Community Service Learning 8*(1): 27-35

Neuman, L.H. (2006). Creating new futures in nursing education: envisioning the evolution of e-nursing education. *Nursing Education Perspectives,* January-February. Retrieved 10/09/11. Available http://fidarticles.com/p/articles/mi_hb3317/is_1_27/ai_n29248908

Northington, L, Wilkerson, R., Fisher, W. & Schenk, L. (2005). Enhancing nursing students' clinical experiences using aesthetics. *Journal of Professional Nursing 21* (1): 66-71

Pickstone, J. (2001). *Ways of knowing: A new history of science technology and medicine.* The University of Chicago Press: Chicago, IL.

Roberts, M. (2010). Emotional intelligence, empathy and the educative power of poetry: A Deleuzo-Guatarian perspective. *Journal of Psychiatric and Mental Health Nursing, 17, 236-241.*

Schermer, J. (1988). Visual media, attitude formation and attitude change in nursing education. *Education Communication and Technology. 36*(4): 197-210

Schullo, S., Barron, A., Kromrey, J., Venable, M., Hilbelink, A., Hohlfeld, T. and Hogarty, K. (2005). Enhancing online courses with synchronous software: An analysis of strategies and interactions. *Annual Meeting of the American Educational Research Association, Montreal, CA April 11-15*: 1-31 Retrieved 10-14-11 from: http://www.coedu.usf.edu/cream/papers/AERA__STARS_final_paper_v5.pdf

Skiba, D. and Barton A. (2006). Adapting your teaching to accommodate the Net generation of learners. *Online Journal of Issues in Nursing. 11* (2):5

Smith, R.L., Bailey, M., Hydo, S.K., Lepp, M., Mews, S., Timm, S. & Zorn, C. (2004). *Nursing Education Perspectives,* 25(6), 278-283.

Smith-Stoner, M. and Willer, A. (2003). Video streaming in nursing education: bringing life to online education. *Nursing Education 28*(2): 66-70

Sullivan W. *Work and Integrity: The Crisis and Promise of Professionalism in America.* 2nd ed. San Francisco, Calif: Jossey-Bass; 2005.

Vincent, G. & Velkoff, V. (2010). The next 4 decades: the older population in the United States: 2010-2050. *Current Population Reports, P25-1138.* U.S Census Bureau, Washington, DC.

Wainwright, S.P. (2005). Culture and ageing: Reflection on the arts and nursing. Journal of Advanced Nursing, 52(5), 518-525.

Warch, R. (1990). Communities of mind and spirit. *Liberal Education*, 76(5). November-December, 6-13.

Simulation Study on Performance Evaluation of MF-TDMA-based Military Satellite Communication Systems

Geunkyung Choi[1,2], Bosung Kim[3], Ki-yeol Ryu[3], Young-Bae Ko[3], Byeong-hee Roh[2,3]

[1] LIG Nex1, Seoul, 135-982, Korea
2 Dept. of NCW, Graduate School, Ajou University, Suwon 443-749, Korea
3 Dept. of Computer Engineering, Graduate School, Ajou University, Suwon 443-749, Korea

ABSTRACT - The multi-frequency time-division multiple access (MF-TDMA) has become general in satellite communications. In this paper, we developed a simulation model for MF-TDMA-based satellite networks using OPNET simulator. With the simulator, we evaluated the efficient use of DAMA schemes using MF-TDMA compared to PAMA based on FDMA.[1]

Key Words: Satellite Communications, MF-TDMA, OPNET, Simulation Study, DAMA, PAMA

1. INTRODUCTION

Evaluating performances of a satellite communication system is often costly, even impossible for systems in development phase. M&S (Modeling and Simulation) is most important tools at this phase to inspect a suitable requirements and solutions.

The multi-frequency time-division multiple access (MF-TDMA) has become general in satellite communications. The access method used for the NCW (Network Centric Warfare) is MF-TDMA which is an excellent choice when a mixed network with different terminal capabilities are to be supported[1]. It has great promises in multimedia networks such as commercial broadcasting, long-range military communication and emergency rescue. MF-TDMA allows a large number of users to dynamically share satellite resources efficiently.

In DVB-RCS standard[2][3], a return channel satellite terminal (RCST) sends a message, capacity request (CR), to the Network Control Center (NCC), after it has scheduled how to request timeslots needed. An RCST may explicitly requests the needed capacity to the NCC in a Demand Assignment Multiple Access (DAMA) scheme. The NCC then allocates return channel timeslots based on each CRs. This is the service as a scheduling and CR evaluation phase as described in Fig. 1. The NCC will allocate the CRs in a two-dimensional resource in the resource allocation packing phase. This space is called a Terminal Burst Time Plan (TBTP) that informs to terminals when (timeslots) and where (carrier) to transmit their traffics. The RCSTs,

This research was partially supported by the MKE, Korea, under the ITRC support program supervised by the NIPA (NIPA-2011-(C1090-1121-0011)), and also supported by Basic Science Research Program through the NRF funded by the MEST (2010-0016938).

terminals or users, are capable of buffering the traffics arriving from the user applications, before transmitting CRs to the satellite link. In a RCST, there can be heterogeneous services, and these are requested to be supported the Service Level Agreements (SLAs).

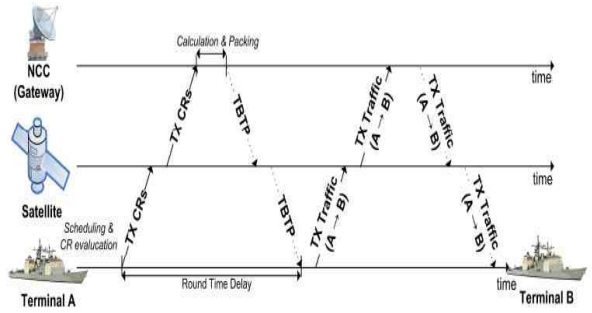


Fig. 1. MF-TDMA procedure

In this paper, we develop a military satellite network simulator based on MF-TDMA, and evaluate its performances compared with FDMA-based PAMA networks. For the simulator, OPNET simulation software is used.

2. IMPLEMENTATION OF MF-TDMA SIMULATION MODELS

To evaluate the performances of the MF-TDMA operation, we implemented simulation models using OPNET [4] as shown in Fig. 2. The model consists of three OPNET node models such as SAT (Satellite), NC (Network Controller) as NCC, and NM (Network Member) as RCST and terminals as shown in Fig.3.

TBTP and C2P are all implemented on NC_MAC, user traffic scheduling mechanism is implemented on NM_MAC, and user traffic is delivered through NM_Eth_Intf. Satellite propagation link (SAT_Link) is modeled as a simple pipe to covert from uplink signal to downlink signal. We set the propagation delay as 250ms with varying BER (Bit Error Rate) characteristics.

For traffic generations, self-similar traffic model is considered according to the self-similarity phenomenon study on the aggregate IP traffic in [5].

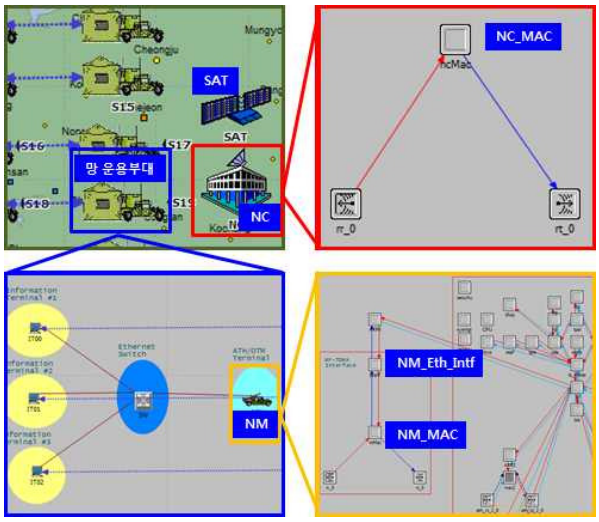


Fig. 2. OPNET Simulation Model for MF-TDMA Satellite Networks

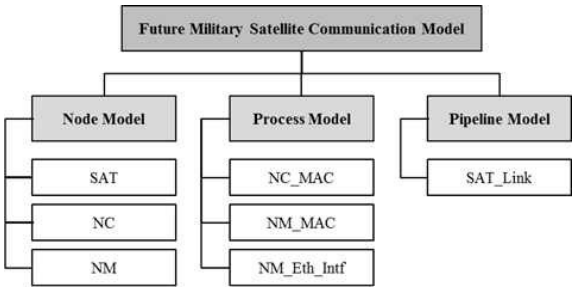


Fig. 3. Components of Simulation Models

3. EXPERIMENTAL RESULTS

The delay for different traffic classes are compared through the MF-TDMA-based satellite networks. For the traffic differentiation, DiffServ (differentiated services) mechanism is utilized.

As shown in Fig. 4, with the DiffServ operation, EF traffic delay can be guaranteed below a specified

delay requirement even the high traffic intensity environments. On the other hand, the delay of BE class traffic increases as the traffic intensity increases.

The efficiency of MF-TDMA is compared to that of FDMA-based PAMA (Pre-Assigned Multiple Access) scheme. Resources are constantly assigned to each user in PAMA scheme, while MF-TDMA assigns the resources on demand by considering the condition of the resources' usages.

In Fig. 5, number of acceptable terminals when traffic burst ratio varies. The traffic burst ratio is defined as the ratio of peak rate to average rate. As shown in Fig. 5, as the burst ratio increases, i.e. the traffic variation becomes larger, the number of acceptable terminals significantly improved compared to PAMA.

4. CONCLUSION

Modeling and Simulation (M&S) work is a very cost effective and efficient method to perform performances evaluation and to validate network protocols. In this paper, we developed a simulation model for MF-TDMA-based satellite networks using OPNET simulator. With the high utilization feature of MF-TDMA scheme, MF-TDMA scheme is expected to be one of the promising resource management schemes for next generation satellite systems. Our study on developing the MF-TDMA-based satellite simulation model can contribute to the construction of those next generation satellite systems.

REFERENCES

[1] J.Wiss, R. Gupta, "The WIN-T MF-TDMA Mesh Network Centric Waveform," IEEE MILCOM'2007, Nov. 2007

[2] ETSI. "EN 301 790 V1.5.1, Digital Video Broadcasting (DVB) Interaction Channel for Satellite Distribution Systems," 2009.

[3] ETSI. "TR 101 790 V1.4.1, Digital Video Broadcasting (DVB) Interaction channel for satellite distribution systems Guidelines for the use of EN 301 790," 2009.

[4] OPNET, http://www.opnet.com

[5] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (extended ver-sion)," IEEE/ACM Tr. on Networking, Vol. 2, No. 1, February 1994, pp.1-15

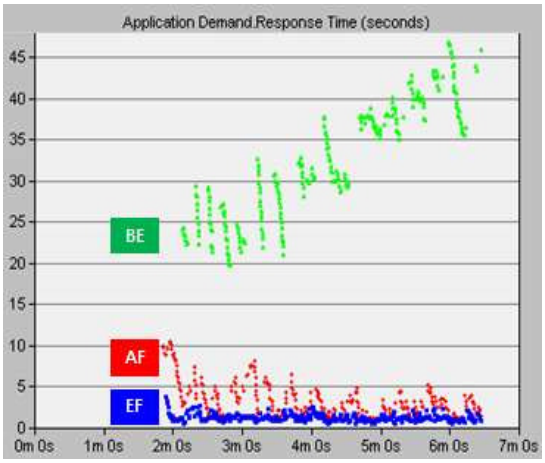


Fig. 4. Delay Characteristics

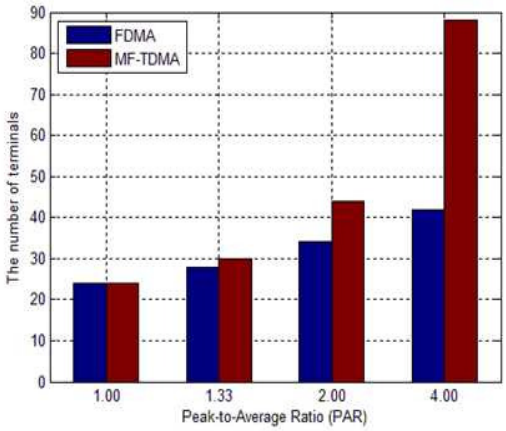


Fig. 5. Number of Acceptable Terminals varying the Burst Ratio

Computer Aided Engineering of Cyber-Physical Information Gathering and Utilizing Systems

Alfred P. DeFonzo, Anthony P. Hopf
Department of Electrical and Computer Engineering
University of Massachusetts Amherst
Amherst, MA USA
email: defonzo@engin.umass.edu, anthony.hopf@gmail.com

***Abstract*—Engineering Cyber-physical information gathering and utilizing systems(CIGUS) presents the systems engineer with a difficult, multi-criterion, multi-objective decision problem. Research, development and design is done over many disciplines, across many domains, each with their specific models. Systems engineers are expected to provide a common level of communication amongst the domains to promote convergence to a design. We present novel information measures that enable combination of the underlying domain specific subsystems parameters in a way that makes the information yield of the system intelligible to decision makers and domain experts. These measures enable, for the first time, the application of multi-objective evolutionary algorithms and end-to-end computer aided engineering of CIGUS.**

Our novel approach is validated and verified through the application and direct comparison of simulated and experimental results of state-of-the-art weather radar network test bed designs. The approach resulted in Pareto optimal point within an average of 10% of the actual case study design parameters and within 25% of the Pareto ideal point. No additional parameters beyond the underlying domain parameters were introduced. This demonstrates that the computationally aided engineering approach presented in this work facilitates engineering feasibility decisions and the subsequent evolution of the engineered systems in way that reduces cost and effort.

***Index Terms*—information gathering and utilizing systems, cyber-physical, network sensors, multi-objective problem, optimization.**

I. Introduction

Interest in the engineering of cyber-physical information gathering and utilizing systems (CIGUS) has burgeoned in part due to the proliferation of wireless technology [1] and in part due to the growing demand for intelligible information. Such systems are complicated, with hierarchies of interfaces containing underlying complexity. They often involve distributed network sensors. The configuration can be dynamic, static and adaptive. Increasingly they involve real time collaboration among agents of varying degrees of autonomy. The interface of high yield systems often hides underlying subsystem complexity which pose new challenges to systems engineering[2]. Systems engineers are expected to provide a common level of communication amongst the domains of expertise that enable research, development and design of the system to converge. As the domains become highly optimized, the language and models become so specialized that it becomes extremely difficult to communicate across the domains. Prior to this work there was no practical and well founded way to combine the parameters of the underlying subsystems in order to represent the overall intelligible information yield. Moreover, in order for systems engineers to make the multicriteria tradeoffs and optimizations required for such systems, it is necessary to introduce new sets of objective functions without which existing multi-objective evolutionary algorithms[3], [4], [5], [6] can not be applied to CIGUS.

In the case of CIGUS, specific domain experts do the component subsystem design and subsequent modeling. Each of these domain specific subsystem models are developed in their particular domain language. Signal processing and communication models are essential to these systems. Weather Radar networks are a classic example. The sub-domains models involved in the systems engineering include; models of the component radars and their subsystems[7], network[8], signal processing[9], [10], and control[11]. What they have lacked is a systematic approach to overall optimization supporting the decision making process. The obstacle is combining parameters from different domains of expertise. The systems engineers ability to provide a level of abstraction that captures the entire system design problem at all levels will determine how quickly, or slowly, the design will converge to meet the requirements and how rapidly the systems will evolve. Clearly, for CIGUS, the underlying parameters and measures should resolve themselves in terms of the essential product: intelligible and useful information.

Moreover, CIGUS may be system of systems with uncertain and evolving requirements. Decisions made at multiple levels present a difficult multi-criteria, or multi-objective, decision problem. The systems engineer is presented with a difficult task of providing the decision makers with the information needed to support investment into further system evolution and development. By introducing information measures we are able to express the quality of the system in terms of more generally understood notions such as accuracy, precision, and bit rates as objective functions. We show that these objective functions, which encapsulate underlying domain specific parameters without introducing additional parameters. These can be combined with cost and throughput functions in a way that enables the application of state-of-the-art multi-objective evolutionary algorithms and automated decision support tools. Moreover, the predictions of this analysis can be directly compared with experimental data from test beds. One recent state-of-the-art weather network test bed, the Collaborative Adaptive Sensing of the Atmosphere (CASA) Integrated Project 1 (IP1), enables the comparison of simulations and experimental results presented in this paper and in more detail elsewhere.

II. Approach

To capture the salience of the engineered system, the systems engineer must separate the domain experts concerns, which are pursuant to providing objective content from the decision makers concerns, which are pursuant to ensuring that higher-level requirements are satisfied. While not conceived as such, a non-obvious example, rich in engineering challenges is the recently deployed the CASA IP1[12] experimental network of weather radars. The development is directed toward demonstration of the engineering feasibility of an end-to-end (TRL 6) [13] hierarchical emergency response and real time numerical weather forecast system. Its primary purpose is to improve tornado and severe weather warnings and to assist

emergency management response to such events[12]. As a case study for demonstrating the need and effectiveness of extending multi objective analysis to the computer aided engineering of CIGUS and to improve the quality of high consequence technology transition decisions associated with their design and development, "IP1", has the unique advantage of being intensively and extensively reported in public documents and the open literature[12]. The present study thus provides a foundation for extending computer engineering aids to support and evaluate technical readiness decisions to cases where such information is not so readily available (e.g. SBInet[14]).

The design of complex sensor systems, such as weather radars and weather radar networks, was accomplished over years of exploration and iteration[15], [16] by multiple uncoordinated efforts. While this traditional process, which involves both trial and error and systematic design, has provided the sensor community with a new means of weather sensing and prediction[12], it cannot solve the present communication problem. One limitation of this approach is that it only allows for a temporary solution to a particular systems engineering problem that will need to be revisited as future requirements are introduced case by case. Here we present for the first time, the Pareto optimal multi-objective analysis of CIGUS. As we discuss elsewhere[17] this enables us to capture the evolution of a particular species of CIGUS over many generations. Various benefits such as: evolutionary context, reuse, accelerate development, and reduced risk.

While the primary and essential quality that is demanded of CIGUS is informativeness, *un*informativeness provides the principled way to construct quality loss functions. The theory underlying the present formulation is developed elsewhere[17], in this paper we present the salience of a specific application. Information produced by such systems is uninformative to the extent that it is already known, that is to say the *prior* or to the extent that it is *uncertain*. Up to now, genetic and evolutionary algorithms have offered or developed neither effective nor principled approaches to incorporating such priors and uncertainty. (Un)informativeness is key and well suited to the engineering of such adaptive intelligence oriented systems and systems of systems because it is directly related to the principle of maximum entropy[18] as pioneered by Jaynes[19] and subsequently developed[20], making the form of the engineering problem presented here intelligible in a way that enables the application of multi-objective evolutionary algorithms. Weather radar networks are particularly suited to our innovative approach because, although implicit, maximum entropy principle is embedded in the core signal processing formulation[21]. (Un)Informativeness provides a natural level of abstraction which fully respects and consistently subsumes lower levels such as those associated with traditional approaches to sensing, signaling and communication [9], [22], [23]. In this paper, we make use of the connection between maximum entropy and Shannon information theory to cast objective functions in terms familiar to the engineering community. This has the added benefit of separating the concerns of channel provider and content provider.

As shown in figure 1, sets of information oriented measures of the performance of sensor systems may be represented in components of an overall objective vector for purposes of evaluation and optimization. Work completed in [17] show how these measures abstract the sensor system estimators of the underlying parameters of the overall system in terms of virtual sensors. By extracting the relevant information from the underlying parametric signal models, expressed in terms of the language of the subdomain, experts enable a reduced set of information metrics that are most relevant to CIGUS. The complexity of the sensor networks considered here results in vectors with high dimensions that make it difficult for the decision

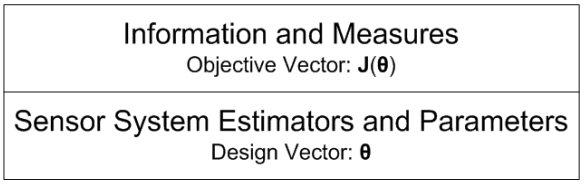


Fig. 1: The informative measures are abstraction over the sensor system estimators and parameters allowing integration over, and characterization of, a single or network of sensors. Objective functions formulated with informative measures capture the impact of varying parameters, design vector, on systems and networks of systems.

makers to comprehend. Here we explore the effectiveness of using multi-objective genetic algorithms(MOGA) in concert with recent visualization advances for computer aided engineering to facilitate the decision making process that goes into the evolution of complex information gathering and utilizing systems, such as weather radar networks and particularly prospective adaptive networks.

A. Information Oriented Objective Functions for Atmospheric Sensors

Information based objective functions enable channelization sensor information flows in accordance with the value and impact of the information. A virtual sensor is comprised of an element, called a test charge, that interacts with the environment that provides a measure of the stimulus, an element that receives the signal corresponding to this measure and a mediating element. In general, a phenomenological field, such as the weather, is sampled by sets of virtual sensors, each corresponding to a different measure and having its own characteristic channel.

The information oriented measures are built on the principles of maximum entropy and the concept of adaptive channel models that capture the scenes multiple spatial and temporal distributions. Adaptive channels model the interaction between the radar and test pattern, including propagation effects. The measures can be aggregated and stored in a data structure that consolidates all collaborative viewpoints on a common grid of vectors, containing all the utilizable information gathered from the scene[17]. The sensors may be mixed or fused at the channel level of abstraction enabling design and intensive optimization of diverse sensor networks.

A particular distribution of the phenomenological field salience and sensing instrumentation is modeled by a test pattern which represents a scenario from a set of viewpoints in support of requirements engineering

Scanning of test patterns by the simulated sensing system in space and time can be modeled as a graph traversal problem with the nodes representing subspaces to be sampled and the arcs weighed by the time cost. Each *subspace*, node on the graph, is a region defined by the beam solid angle, $\Delta\theta_s \times \Delta\phi_s$, and range extent, ΔR_s.

The objective functions used in the present work are chosen to explore the trade-offs between the conflicting objectives of information capacity, gathered information, quality of information, cost, and scan time.

The objective functions, $J_i(\theta)$, where the subscript i is the ith objective function, and θ is the design vector, are constructed for a typical weather scene as follows:

$$J_1(\theta) = \sum_{s=1}^{S} \left(I^s_{r_{cap}} + I^s_{v_{cap}} \right) \tag{1}$$

$$J_2(\theta) = \sum_{s=1}^{S} \left(I_{r_{cap}}^{HDs} \right) \tag{2}$$

$$J_3(\theta) = \sum_{s=1}^{S} \left(I_r^s \right) \tag{3}$$

$$J_4(\theta) = \sum_{s=1}^{S} \left(I_v^s \right) \tag{4}$$

$$J_5(\theta) = \sum_{s=1}^{S} \left(I_r^{HDs} \right) \tag{5}$$

$$J_6(\theta) = \frac{\sum_{s=1}^{S} \left(BER_r^s + BER_{\hat{v}_r}^s + BER_{\hat{\sigma}_{vr}}^s \right)}{3S} \tag{6}$$

$$J_7(\theta) = \sum_{s=1}^{S} \left(T_{subspace}^s \right) + \sum_{s=1}^{S-1} \left(T_{trans}^s \right) \tag{7}$$

$$J_8(\theta) = \text{cost}_{Rbase} + \text{cost}_{power} + \text{cost}_{agility} + \text{cost}_{antenna} \tag{8}$$

There are two classes of targets, six weather subspaces and six hard target subspaces. The information oriented measures of information capacity ($I_{r_{cap}}^s$ and $I_{v_{cap}}^s$), information (I_r^s and I_v^s), and Bit Error Rate ($BER_r^s, BER_{\hat{v}_r}^s$, and $BER_{\hat{\sigma}_{vr}}^s$) are captured in equations (1)-(6). The superscript HD indicates hard target information oriented measures and the subscript s is used to identify the s^{th} subspace.

The first two objective functions, (1) and (2), sum the information channel capacity for weather and hard targets over the subspaces, respectively. Three types of information capacity, reflectivity($I_{r_{cap}}$), velocity($I_{v_{cap}}$), and hard target($I_{r_{cap}}^{HD}$), are defined instantaneously as the maximum bit rate that can be sustained by channel models of a gaussian white noise channel, and noiseless gaussian channel, and a Swerling 1 model channel, respectively[17]. The hard target velocity capacity is not calculated. In the present analysis the objective functions of channel capacity are minimized to ensure maximum capacity utilization.

Objective functions J_3, J_4, and J_4 are comprised of the aggregated information gathered over the individual reflectivity, velocity, and hard target reflectivity channels, which are then summed over the subspaces, respectively. Hard target velocity information is not calculated. These functions are maximized.

The bit error rates are a measure of the quality of the information extracted and are a function of the errors in the underlying estimators. Objective function J_6 used in this analysis consolidated the BER associated with the various channels to provide an overall quality of information measure. The summation is over S, the subspaces, of the individual terms of each subspace referring to the reflectivity, (BER_r) the velocity ($BER_{\hat{v}_r}$), and the spectrum width ($BER_{\hat{\sigma}_{vr}}$). Hard target reflectivity or velocity bit error rate is not calculated. Minimizing the BER, maximizes accuracy and precision of the information[17].

Objective function J_7 is a measure of the total time it takes to acquire the information in the scene. It is a measure of the information gathering throughput of the system, the amount of information collected for the time to complete the test pattern scan. The time objective function is split into two summation, the first is the time to scan each subspace, the second is the time taken to scan between each subspace. The time to scan each subspace, $T_{subspace}^s$, is given by the dwell time of the radar, DT, and the number of positions in azimuth, $B_{az} = \frac{\Delta\theta_s}{\theta_{az_{3dB}}}$, and elevation, $B_{el} = \frac{\Delta\phi_s}{\phi_{el_{3dB}}}$, necessary to scan the entire subspace and the time to transition from beam to beam within the subspace. The time to move from subspace to subspace, T_{trans}^s, is given by rotating the sensor. Equations (9) and (10) define the subspace time and transition time.

$$T_{subspace}^s = B_{az}B_{el}DT + B_{el}\left[(B_{az}-1)az_{tB2B}\right] + (B_{el}-1)el_{tB2B} \tag{9}$$

$$T_{subspace}^s = az_{tS2S}^s + el_{tS2S}^s \tag{10}$$

where az_{tB2B}, and $el_{tB2B}$are the times to transition from beam position to beam position. In the case of the transition from subspace to subspace, az_{tS2S} and el_{tS2S}, the time is given by the angular difference in azimuth and elevation multiplied by the angular velocity in that direction. Minimizing J_7, maximizes the throughput.

Objective function J_8 is a measure of the cost of the system. The cost objective function, $J_8(\theta)$, is made up of four factors; *base radar cost, excess power cost, excess agility cost*, and *excess antenna cost*. Our initial objective cost function is a first approximation to the true cost function to be created and is referenced to the cost values for the IP1 weather radars[24], [12]. Cost is minimized.

In this study we chose the following decision variables: *maximum transmit power, half power beam width in azimuth and elevation, and maximum angular velocity of the pedestal*, given in table I to make up the decision vector, $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$. These variables were chosen because the object functions are most sensitive to them and are sufficient for validating the approach.

In the present case of computer aided engineering of a single radar we have reduced our objective vector, $J(\theta)$, to eight dimensions, corresponding to the six aspects of the scene about which we seek to gather information, the time interval over which we seek it, and the cost of the deployed system.

B. Multi-Objective Genetic Algorithms

Multi-objective optimization seeks to optimize problems that require the simultaneous optimization of multiple, often competing objectives [3]. Genetic Algorithms were originally developed to imitate the process by which living organisms evolve [4]. They have since been applied to multi-objective optimization problems as algorithms to supply reasonable approximations to the Pareto front and set [25]. Here they are used in a computer aided engineering approach to simulate the evolution of complex engineered systems. The technical analysis supports the decision makers in making a selection of a particular design out of the set of Pareto optimal designs. Each of the solutions returned by the analysis, see Figure 2, is a valid optimal design resulting from tradeoffs among the conflicting objectives reaching mutually non-dominated solutions referred to as the Pareto front. The discrete set of optimum points can then be used by the various decision makers to drive the evolution of the complex system being optimized, in this case a cyber-physical information gathering and utilizing system. The use of genetical algorithms to calculate the Pareto front and set of a multi-objective optimization problem is referred to as MOGA. Within the present approach we will demonstrate how MOGA can be used to calculate the Pareto front and set for low order models of a single weather radar. Higher order models can be incorporated into MOGA through the use of a more sophisticated simulation[17].

TABLE I: MOGA Settings

Decision Variables	**Parameter (Unit)**			
		lower	initial	upper
θ_1	Peak Power (W)	5e3	12.5e3	20e3
θ_2	$\theta_{az3dB}(deg)$	1	2	4
θ_3	$\theta_{el3dB}(deg)$	1	2	4
θ_4	Agility (deg/sec)	10	40	80
Cost Variables			Value	
R_{base}			\$220e3	
λ_1			245	
γ_1			2	
κ_1			8e3	
λ_2			1463	
γ_2			1.5	
κ_2			20	
λ_3			736	
γ_3			1.5	
κ_3			4	

For these higher dimensional multi-objective problems, the present approach is an 8 dimension problem, a visualization technique called *Level Diagrams*[5] will be used to enable an improved analysis of the Pareto front and will provide an excellent tool for the decision makers. The Level Diagrams classify each Pareto front by the distance of the Pareto front from the ideal point, accounting for all the objectives simultaneously. It is extremely unlikely for an optimized solution to the Pareto front to achieve the ideal point[6], but we define the Pareto optimal point as the point with the shortest 1-norm distance from the ideal point. Every objective $(J_i(\theta), i = 1, \ldots, m)$ is normalized and classified with respect to its minimum and maximum values on the Pareto front, $J_i^{norm}(\theta), i = 1, \ldots, m$ [5]:

$$J_i^{max} = \max_{\theta \in \mathbf{\Theta}_P^*} J_i(\theta), J_i^{min} = \min_{\theta \in \boldsymbol{\theta}_P^*} J_i(\theta), i = 1, \ldots, m \tag{11}$$

$$J_i^{norm}(\theta) = \frac{J_i(\theta) - J_i^{min}}{J_i^{max} - J_i^{min}} \tag{12}$$

such that,

$$0 \leq J_i^{norm}(\theta) \leq 1 \tag{13}$$

The Y-axis on all the Level Diagram graphs, figure 2, corresponds to the value of the normalized objective function, and this means that all graphs are synchronized with respect to this axis. The X-axis corresponds to values of the objective, or decision variables, in physical units. Using this representation, all plots are synchronized with respect to the y-axis, meaning a single level on the y-axis returns all the information for a single point on any of the objective function or decision variables plots[5].

III. MOGA Analysis: Case Study

A. Scanning Analysis

The MOGA analysis is done with an agile mechanical pedestal using the decision variables and cost variables listed in table I.

The Level Diagrams of the Pareto front and set for the MOGA analysis of the agile mechanical X-band radar is given in figure 2. The Pareto optimal point is the light green square referenced by the arrow. Black vertical lines in plots of $J_7, J_8, \theta_1, \theta_2$ and θ_3 represent the specifications given in [24], [26] for the IP1 weather radars. Given the complexity of the multi-objective problem, it is surprising to see the Pareto optimal point coming in close comparison to the documented values of the IP1 weather sensing radar. The Pareto optimal point returns $\theta_{az3dB} = 1.6^o, \theta_{el3dB} = 1.9^o, P_t = 9.4kW$, cost $= \$459k$ and time $= 53sec$, compared to the IP1 values of $\theta_{az3dB} = 1.8^o, \theta_{el3dB} = 1.8^o, P_t = 8kW$, cost $= \$459k$ and heart beat time $= 60sec$.

TABLE II: MOGA Analysis Summary

	Power (P_t) (W)	θ_{az3dB} (deg)	θ_{el3dB} (deg)	Scan Time (sec)	Cost (k\$)
Simulated	9359	1.6	1.9	53	458.6
IP1	8000	1.8	1.8	60	459.0

IV. Discussion

The present computer aided engineered approach applied to the given weather radar sensor results in a well formed high dimension Pareto front yielding the Pareto optimal point close to the ideal point. The 1-norm Level Diagrams, shown in figure 2, have smooth objectives with well defined minima where no single objective dominates, suggesting convexity of the Pareto front. Combined with location of the 1-norm Pareto optimal point to within 25% of the ideal point, we can characterize the Pareto front as well formed. Therefore, the Level Diagrams are providing insight into high dimension Pareto fronts when based on information oriented measures and test patterns.

The resulting Pareto optimal design vector yielded values, on average, in excellent correspondence with the actual IP1 design. An agreement between the optimal design vector and IP1 design of within 10% for the scan time is evidence that the current test pattern is a good representation of a multitask scene. Further indication is the similarity, within 10%, of the optimal azimuth and elevation beam width to the IP1 design. The Pareto optimal peak transmit power, a relatively outlier at 18% greater than the IP1 design, is a result of the magnetron transmitter in the IP1 radar operating below its maximum rated peak power. The present computer aided engineered approach accurately models the evolution of IP1 system.

Although the results exhibit excellent convergence, extending the objective vector to include a reliability/availability component would likely result in further convergence between the Pareto design and real case. However, a valid and verified reliability/availability model for the present case under study has not appeared in the literature. As the models become available, they can be incorporated into the multi-objective optimization aiding in the engineering of the system.

The computer aided engineering approach provides isolation from the other objective functions allowing higher level models for cost, reliability, maintainability, volume manufacturing, industrial learning curves, and other potential non-functional and functional requirements to be readily incorporated or modified. MOGA simultaneously evaluates each of the objective function individually. This allows the objective functions to be individually modified without the need to update subjective weights. The additional abstraction of the informative objective functions allows the inclusion of uncertainty and priors into the MOGA analysis and encourages the use of other multi-objective evolutionary algorithms(MOEA) that may be better for other applications.

The method presents an approach allowing for the acceleration of the evolution of complex, multi-criterion information gathering and utilizing systems. Extension to higher order models of signal estimators and test patterns in the presence of multiple weather sensors is of interest to provide insight into design trades over changing weather conditions and different venues. Specifically, creation of higher order models of the sensor system and test pattern will facilitate exploration into the trade space of polarimetric weather radar networks and waveform design for network multifunction radars. Moreover, the

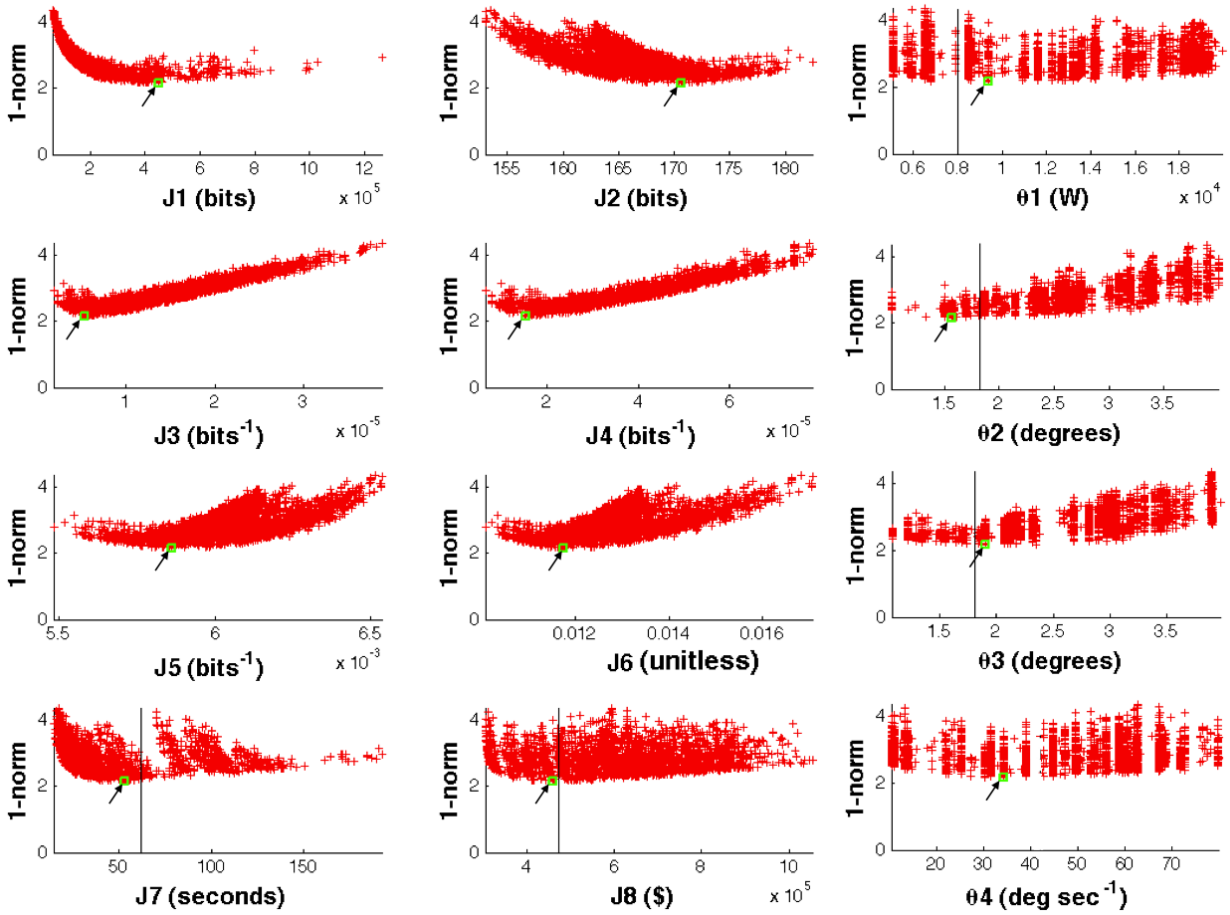


Fig. 2: 1-norm Level Diagram of the Pareto front and set the eight objective functions comprising the objective vector, $J_i(\theta)$, where the subscript i is the ith objective function, and θ is the design vector used in the MOGA analysis of the case study X-band weather radar described in section III-A. The Pareto optimal point is the light green square referenced by the arrow. Black vertical lines in plots of $J_7, J_8, \theta_1, \theta_2$ and θ_3 represent the specifications given in [24], [26] for the IP1 weather radars.

method can be extended to incorporate further decision support for more complex trade-off analysis that may be required to assess the evolution at higher levels to support business modeling and planning.

V. Conclusion

We have shown that by introducing integrative objective information oriented measures, we can define a level of abstraction which captures the underlying sensor estimators and parameters that solves the communication problem between the systems engineers, domain experts and decision makers. Not only will the obstacle be eliminated, the design of these complex sensor systems will converge much more rapidly, allowing for an acceleration in the evolution of the systems, with the inclusion of the preferences of decision makers a posteriori to the objective analysis, hence acknowledging subjective influences.

The analysis is applied to weather radar designs providing complex multi-objective design problems with evolving specifications and requirements. Without any adjustable parameters, any subjective weighting, and in such a complex design space where a multiplicity of results could have occurred, the informative methodology of systems engineering resulted in decision parameters very close to that of the IP1 system. The results of the MOGA analysis case study, show that the approach is successful in modeling the complex system by producing a Pareto optimal point within an average of 10% of the case study's design specifications and providing an objective basis for evaluating the engineering feasibility of the end-to-end system and its transition into operational environments for further development.

The foregoing capabilities facilitate the demonstration of engineering feasibility and subsequent development and evolution of the CIGUS. We develop objective functions, combining measures of cost and throughput with the underlying domain specific parameters, enabling the application of state-of-the-art multi-objective evolutionary algorithms and automated decision support tools. The novel systems engineering approach is further validated and verified by the agreement of the predictions of the analysis and the experimental data from the IP1 test bed. Clearly, in the case of weather radars had the present approach been available, considerable time and money could have been saved[17].

References

[1] Z. Song, Y. Chen, C. R. Sastry, and N. C. Tas, *Optimal Observation for Cyber-physical Systems: A Fisher-information-matrix-based Approach*, 1st ed. Springer, Aug. 2009.

[2] J. M. Carlson and J. Doyle, "Highly Optimized Tolerance: Robustness and Design in Complex Systems," *Physical Review Letters*, vol. 84, no. 11, pp. 2529–2532, Mar. 2000. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.84.2529

[3] C. M. Fonseca and P. J. Fleming, "Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization," in *Genetic Algorithms: Proceedings of the Fifth International Conference*, San Mateo, CA, July 1993.

[4] J. Holland, *Adaptation in Natural and Artificial Systems*. Cambridge, Ma.: First MIT Press Edition, 1992.

[5] X. Blasco, J. Herrero, J. Sanchis, and M. MartÌnez, "A new graphical visualization of n-dimensional pareto front for decision-making in multiobjective optimization," *Information Sciences*, vol. 178, no. 20, pp. 3908 – 3924, 2008, special Issue on Industrial Applications of Neural Networks, 10th Engineering Applications of Neural Networks 2007. [Online]. Available: http://www.sciencedirect.com/science/article/B6V0C-4STYV2P-1/2/98278c13175ff0eb13b8c87c5cce61da

[6] S. Azar, B. J. Reynolds, and S. Narayanan, "Comparison of two multiobjective optimization techniques with and within genetic algorithms," in *Design Engineering Technical Conferences*. ASME, September 1999.

[7] S. Frasier and P. Siqueira, "Ground-based atmospheric research radar systems at the university of massachusetts." Amer. Meteor. Soc., 2009.

[8] M. Zink, E. Lyons, D. Westbrook, D. Pepyne, B. Pilips, J. Kurose, and V. Chandrasekar, "Meteorological command & control: Architecture and performance evaluation," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, vol. 5, july 2008, pp. V –152 –V –155.

[9] R. J. Doviak and D. S. Zrnić, *Doppler Radar and Weather Observations*, 2nd ed. Academic Press, 1993.

[10] V. N. Bringi and V. Chandrasekar, *Polarimetric Doppler Weather Radar*. New York, NY: Cambridge University Press, 2001.

[11] D. Pepyne, D. Westbrook, B. Philips, E. Lyons, M. Zink, and J. Kurose, "Distributed collaborative adaptive sensor networks for remote sensing applications," in *American Control Conference, 2008*, june 2008, pp. 4167 –4172.

[12] (2011, May). [Online]. Available: http://www.casa.umass.edu/

[13] E. NASA. (2011, May) Definition of technology readiness levels. [Online]. Available: http://esto.nasa.gov/files/TRL_definitions.pdf

[14] (2011, May). [Online]. Available: http://en.wikipedia.org/wiki/SBInet

[15] R. C. Whiton, P. L. Smith, S. G. Bigler, K. E. Wilk, and A. C. Harbuck, "History of operational use of weather radar by u.s. weather services. part i: The pre-nexrad era," *Weather and Forecasting*, vol. 13, no. 2, pp. 219–243, 2011/08/21 1998. [Online]. Available: http://dx.doi.org/10.1175/1520-0434(1998)013⟨0219:HOOUOW⟩2.0.CO;2

[16] ——, "History of operational use of weather radar by u.s. weather services. part ii: Development of operational doppler weather radars," *Weather and Forecasting*, vol. 13, no. 2, pp. 244–252, 2011/08/21 1998. [Online]. Available: http://dx.doi.org/10.1175/1520-0434(1998)013⟨0244:HOOUOW⟩2.0.CO;2

[17] A. P. Hopf, "Informativeness and the computational metrology of collaborative adaptive sensor systems," Ph.D. dissertation, University of Massachusetts Amherst, Amherst, MA, May 2011.

[18] E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 939 – 952, 1982.

[19] ——, "Information theory and statistical mechanics," *The Physical Review*, vol. 106, no. 4, pp. 620–630, May 1957.

[20] A. Caticha, "Entropic inference," in *MaxEnt 2010: The 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Chamonix, France, July 2010.

[21] R. Doviak, D. Zrnic, and D. Sirmans, "Doppler weather radar," *Proceedings of the IEEE*, vol. 67, no. 11, pp. 1534 line 4–9, nov. 1979.

[22] A. V. Oppenheim, A. S. Willsky, and S. Hamid, *Signals and Systems*, 2nd ed. Prentice Hall, August 1996.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: John Wiley & Sons, 2006.

[24] J. Brotzge, R. Contreras, B. Philips, and K. Brewster, "Radar feasibility study," NOAA, Tech. Report, 2009.

[25] C. Fonseca and P. Fleming, "Multiobjective genetic algorithms made easy: selection sharing and mating restriction," in *Genetic Algorithms in Engineering Systems: Innovations and Applications, 1995. GALESIA. First International Conference on (Conf. Publ. No. 414)*, Sep. 1995, pp. 45 –52.

[26] F. Junyent, V. Chandrasekar, D. McLaughlin, E. Insanic, and N. Bharadwaj, "The casa integrated project 1 networked radar system." *Journal of Atmospheric & Oceanic Technology*, vol. 27, no. 1, pp. 61 – 78, 2010.

Safety Lead Curve and Entertainment in Games

Hiroyuki Iida, Takeo Nakagawa, Shogo Sone, Apimuk Muangkasem, Taichi Ishitobi
School of Information Science, Japan Advanced Institute of Science and Technology
Nomi-city, Ishikawa 923-1292, Japan

Abstract

This paper is concerned with the safety lead curve and entertainment in games, where the safety lead curve is that once information of the game outcome goes above it, the advantageous team will win the game with 100% certainty. The safety lead curves have been derived by using a series of approximate solutions of the flow between two parallel flat walls, one of which is at rest, the other is suddenly accelerated from rest to a constant velocity. The safety lead curve (SLC) is critical for neutral observer(s) to assess entertainment in games, for when the information of game outcome is below it, the game is interesting, while when the information is above it, the game is boring. The power balance between the two teams (players) causes the information of game outcome to follow along the safety lead curve so as not to lose entertainment during the game. The safety lead curve can be a critical edge of game, at which two different things, viz. interesting game and boring game, happen, and thus we feel a particular emotion while the game proceeds along it. The four Japan games in 2010 FIFA World Cup are under the safety lead curve, and thus are interesting for neutral observer(s).
Key Words: Safety Lead Curve, Entertainment, Game Advantage, Information of Game Outcome, Soccer, Fluid Mechanics

I. Introduction

Nothing perhaps is more intriguing than to know game progress patterns or how information of game outcome varies with the game length or time, where information of game outcome is the data that is the certainty of game outcome. Information of game outcome and its development are therefore topics that have attracted many researchers (e.g. Iida et al 2011a, 2011b), but much remains to be done before a satisfactory understanding is obtained and real prediction is possible.

Browne(2008) has determined whether the quality criteria of a game is precisely defined and automatically measured through self-play in order to estimate the likelihood that a given game will be of interest to human players, and whether this information is used to direct an automated search for new games. Yannakakis & Maragoudakis(2005) have introduced an effective mechanism for obtaining computer games of the player's satisfaction. The proposed approach is based on the interaction of a player modeling tool and a successful on-line mechanism.

Game information dynamic models (Iida et al 2011a) make it possible to treat and identify game progress patterns. The two models are expressed, respectively, by

$$\text{Model1:} \qquad \xi = \eta^n \tag{1}$$

and

$$\text{Model2:} \qquad \xi = [sin(\frac{\pi}{2}\eta)]^n \tag{2}$$

where ξ is the non-dimensional information of game outcome, η the non-dimensional game length or time, and n a positive real number parameter depending on the fairness of the game, strength of the two teams(players), and strength difference between the two teams (players).

It has been confirmed that game information dynamic models are quite useful for understanding and explaining game progress patterns in Base Ball(Iida et al 2011a), Soccer, Chess, Shogi and others. However, the effect of the safety lead on game progress patterns has not been taken into account in these models, where the safety lead is such that once the lead exceeds its value, the leading team will win the game with 100% certainty. The safety lead is sometimes a critical factor in game entertainment, for if one team (player) gets the safety lead against the other team (player) the game becomes immediately boring, but if not it is kept to be interesting. Thus, the safety lead curve, which is that once the information of game outcome goes above it, the team having advantage will win the game with 100% certainty, plays a part as a game information dynamic model (see Appendix), along which the game proceeds under certain conditions, as to be shown.

It is evident that winner(s), loser(s) and neutral observer(s) have different feeling, or emotion during the game from each other, where winner(s) is winning player(s) and winner-sided supporter(s), and loser(s) is losing player(s) and loser-sided supporter(s). Thus, in this study concerning entertainment for the sake of clarity we will only inquire whether neutral observer(s) feels interested or bored during a game. However, how one feels emotion during the game essentially belongs to each person, so that the present discussion on entertainment is based on authors subjective views.

The main purpose of the present study is to propose two novel information dynamic models representing safety lead or uncertainty of game outcome, and to

TABLE I: Time history of goals during three artificial Soccer games between team A and team B.

Game	Result*	Goal time**(minutes)
balanced game	0 - 0	
seesaw game	5 - 4	9(A), 18(B), 27(B), 36(A), 45(A), 54(B), 63(B), 72(A), 81(A)
one-side game	50 - 0	From 1 to 50 minutes, one goal in every 1 minute

*In the column ˝Result˝, the left value is the goal sum for team A after the game, while the right value is the goal sum for team B.
**In the column ˝Goal time˝, characters A and B in brackets denote team A and team B, respectively.

clarify the role of the safety lead curve and its relation to entertainment in games.

II. Safety Lead Curve and Entertainment in Games

In any game, it is realized that there exists a safety lead curve, which is defined in such a way that once the information of game outcome goes above it, the advantageous team will win the game with 100% certainty. The safety lead curves have been derived as a series of approximate solutions of the flow between two parallel flat walls, one of which is at rest, the other is suddenly accelerated from rest to a constant velocity (see Appendix), and it is expressed by

$$\xi = (1-\eta)^q \quad \text{for } 0 \le \eta \le 1 \tag{3}$$

where ξ is the non-dimensional current safety lead of game, η the non-dimensional current game length, and q a positive real number parameter. The safety lead curve depends on characteristics of the game, strength of the two teams (players) and strength difference between the two teams (players). To know the relation between the safety lead curve and entertainment in game, the three artificial elemental game progress patterns are introduced, viz., ˝balanced game˝,˝seesaw game˝and ˝one-sided game˝, as listed in Table I(Iida et al 2011c), where in a ˝balanced game ˝, both of the teams have no goal through the game, in a ˝seesaw game ˝, one team leads, then the other team leads, and this may repeatedly alternate, and in a ˝one-sided game ˝, the current goal sum of one team (winner) is always greater than that of the other team (loser), so that the goal difference between the two teams is always positive.

The non-dimensional information ξ_s in Soccer is here defined as follows: When the total goal(s) of the two teams at the end of game $G_T \neq 0$,

$$\xi_s = \begin{cases} \frac{|G_A(\eta)-G_B(\eta)|}{G_T} & \text{for } 0 \le \eta < 1 \ , \\ 1 & \text{for } \eta = 1 \ , \end{cases} \tag{4}$$

where $G_A(\eta)$ is the current goal sum for the team A (winner), and $G_B(\eta)$ is the current goal sum for the team B (loser). The measure of non-dimensional information of game outcome ξ_s has been derived by considering only goals scores in this study. It is, however, evident that number of shoots and corner kicks, ball possession time, strength difference between the two teams, player's morale or stamina and so on, affect the value of the measure, it is not straightforward to quantify these factors as a measurement and also their contribution to the measure is considered to be insignificant. This is the reason why they have not been taken into account for evaluating ξ_s. At η=1, ξ_s is assigned the value of 1, for at the end of game the information must reach the total value. On the other hand, when G_T=0,

$$\xi_s = \begin{cases} 0 & \text{for } 0 \le \eta < 1 \ , \\ 1 & \text{for } \eta = 1 \ , \end{cases} \tag{5}$$

Note that in a draw case ξ_s may also take the value of 0 other than 1 at η=1, depending on the game rules. In the case of a tournament match, ξ_s=1 at η=1, while in the case of a league match, ξ_s=0 at η=1.

The game length is defined as the current time (minutes), and it is normalized by the total game length or the total time to obtain the non-dimensional value η. The total game length of Soccer is normally 90 minutes, but in the case of extended games it becomes 120 minutes.

Figure 1 shows the relation between non-dimensional information ξ and non-dimensional game length η for three artificial Soccer games, viz. ˝balanced game˝, ˝seesaw game˝, and ˝one-sided game˝. In this figure, two safety lead curves are concurrently plotted for reference: The safety lead curve 1 is ξ=$(1-\eta)^2$, while safety lead curve 2 is ξ=$(1-\eta)^{0.4}$.

Firstly, let us discuss the entertainment of the game, by assuming the safety lead curve 1. In the case of a ˝balanced game˝, the information is always under the safety lead curve 1 through the game except for the value at η=1. In the case of a ˝seesaw game˝, the information exceeds the safety lead at $\eta \simeq 0.69$, so that the game is solved at this game length. In the case of a ˝one-sided game˝, the information is under the safety lead curve 1 until crossing each other, but after that the information is above the safety lead curve 1. This means that before the cross point this game is interesting, but after the cross point it becomes boring for neutral observers.

Secondly, let us discuss the entertainment of the game, by assuming the safety lead curve 2. In both a ˝balanced game˝and ˝seesaw game˝, the information is below the safety lead curve 2 through out the game except for the value at η=1. This means that the entertainment in these games is maintained through the total game length except at the end. In the case of a ˝one-sided game˝, the information is under the safety lead curve 2 until crossing it, but after that the information is above the safety lead curve. This means that before the cross point this game is interesting, but after the cross point it becomes boring. Note that when

the safety lead curve changes from the curve 1 to 2, the interesting game length becomes longer or boring game length becomes shorter. It may be instructive to consider the two extreme cases, viz. the parameter q=0 and ∞. When q=0, the safety lead curve becomes ξ=1 for $0 \leq \eta < 1$, but ξ=0 for η=1. In this case, every game is interesting. On the other hand, when q=∞, the safety lead curve $\xi \simeq 0$ for $0 < \eta < 1$, and ξ=1 for η=0 and ξ=0 for η=1 and approximately coincides with the "balanced game" except for the value at η=0. Thus, in this case, only a "balanced game" is interesting, and the games are boring.

It is suggested that we normally try to design a game in such a way that it proceeds so as to keep the information under the safety lead curve. This is because when the information is under the curve, the game is interesting, while when the information is above the curve it is boring. The safety lead curve is, therefore, critical to assess entertainment in games. In Soccer, while one team is losing, the players make their efforts to avoid the safety lead of the other team. On the other hand, while one team leads, the players try to secure their safety lead against the other team. This power balance between the two teams may result in that the information of game outcome follows along the safety lead curve. In another words, Soccer players in one team severely struggle with those in the other team to avoid the safety lead of the other team, for as far as the opponent's lead is within the limit, it is quite possible to revert the game later. The situation is similar with a Marathon race. A strong runner often is willing to take rear position in keeping the distance within the safety lead of the front runner, for this provides her or him the highest probability to win the race. This also results in that a Marathon proceeds in such a way that the distance between the two competing runners is kept to be the safety lead of the front runner approximately, so that the information of race outcome follows along the safety lead curve. Thus, the safety lead curve can be a critical edge of the game, at which two different things happen (Iida 2007). This is because we feel a particular emotion at the edge, or the safety lead curve.

III. Data Analysis

Four Japan games in 2010 FIFA World Cup South Africa have been considered and analysed, where Soccer is a form of football in which the use of the hands and arms either for playing the ball or for interfering with an opponent is prohibited. Some of the relevant information is summarized in Table II.

A. Group E 1st game: Japan vs. Cameroon

Figure 2 indicates that ξ_s is kept to 0 until $\eta \simeq 0.433$, but it jumps to 1 at η=0.433 and is kept to be the same value until the end of the game. In this game, there are two intervals, where ξ_s is constant; In the earlier interval, ξ_s=0, and in the later interval ξ_s=1. Thus, although this game represents a typical information curve

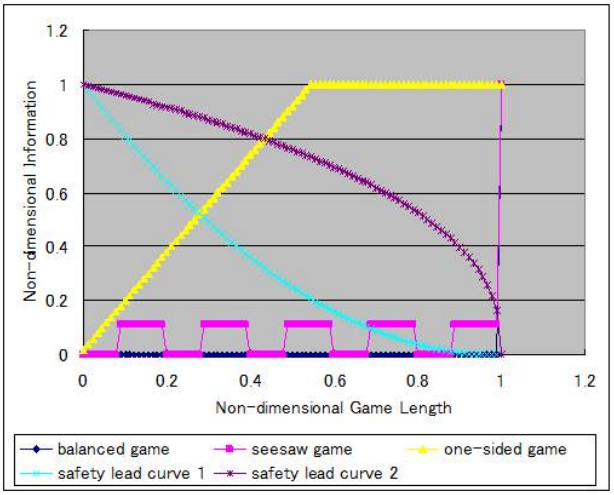


Figure 1:Non-dimensional information ξ against non-dimensional game length η for artificial Soccer games, viz. "balanced game, "seesaw game, and "one-sided game.

TABLE II: Four Japan Games in 2010 FIFA World Cup South Africa

Game	Result	Goal(minute)
Group E	1 - 0 (45 min)	39th (Japan)
1^{st} game	0 - 0 (45 min)	
June 14,	- - - - - - - - - - -	
Bloemfontein	Japan 1 - 0 Cameroon	
Group E	0 - 0 (45 min)	
2^{nd} game	1 - 0 (45 min)	53th (Holland)
June 19,	- - - - - - - - - - -	
Durban	Holland 1 - 0 Japan	
Group E	2 - 0 (45 min)	17th (Japan)
3^{rd} game		30th (Japan)
June 24,	1 - 1 (45 min)	81th (Denmark)
Rustenburg		87th (Japan)
	- - - - - - - - - - -	
	Japan 3 - 1 Denmark	
Round of 16	0 - 0 (45 min)	
June 29,	0 - 0 (45 min)	
Pretoria	0ex0 (15 min)	
	0ex0 (15 min)	
	- - - - - - - - - - -	
	Japan 3PK5 Paraguay	

as one-sided game, in fact, it is interesting. During the earlier interval ξ_s=0, the game may proceed, experiencing alternate changes from offense to defense by the two teams many times. On the other hand, during the later interval ξ_s=1, in addition to the alternate changes from offense to defense by the two teams many times, the game is still pending state, for if Cameroon gets one goal during this interval, the game immediately reverts back to a balanced state again.

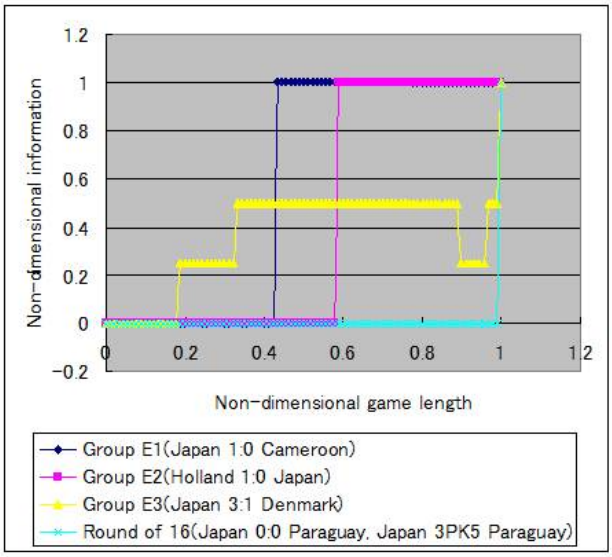


Figure 2: The relation between non-dimensional game information ξ_s and the non-dimensional game length η for the four Japan games in 2010 FIFA World Cup South Africa.

B. Group E 2nd game: Holland vs. Japan

Figure 2 indicates that ξ_s is kept to 0 until $\eta \simeq 0.589$, but it jumps to 1 at $\eta \simeq 0.589$ and is kept to be the same value until the end of the game. In this game, there are two intervals, where ξ_s is constant, In the earlier interval ξ_s=0, while in the later interval ξ_s=1. Thus, similarly to Group E 1st game although this game represents a typical information curve as "one-sided game", in fact it is interesting. During the earlier interval ξ_s=0, the game may proceed experiencing alternate changes from offense to defense by the two teams many times. On the other hand, during the later interval ξ_s=1, in addition to the alternate changes from offense to defense by the two teams many times, the game is still in a pending state, for if Japan gets one goal during this interval, the game will immediately revert back to the balanced state again.

C. Group E 3rd game: Japan vs. Denmark

Figure 2 shows that ξ_s is kept to be 0 until $\eta \simeq 0.189$, but it jumps to 0.25 at $\eta \simeq 0.189$ due to Japan's first goal and is kept to be the same value until $\eta \simeq 0.333$. At $\eta \simeq 0.333$, ξ_s jumps to 0.5 due to Japan's second goal and is kept to be the same value until $\eta \simeq 0.9$. At $\eta \simeq 0.9$, ξ_s decreases suddenly to 0.25 due to Denmark's first goal and is kept to be the same value until $\eta \simeq 0.967$. However, at $\eta \simeq 0.967$, ξ_s jumps to 0.5 due to Japan's third goal and is kept to be the same value until $\eta \simeq 0.989$. Then, ξ_s becomes 1 at the end of the game.

D. Round of 16: Japan vs. Paraguay

Figure 2 shows that ξ_s is kept to be 0 until $\eta \simeq 0.992$, but it jumps to 1 at η=1. This game is a draw case, so that the winner is determined by a penalty match, where five kickers of each team participate. During the penalty match, Paraguay gets 5 goals, while Japan gets 3 goals, so that Paraguay wins the game. This game is a typical "balanced-game", in which both teams cannot get any goal but they repeat from offense to defense by the two teams many times, and thus it is interesting. This is because the strength of both teams is quite high and the strength difference between the two teams is very small. During the game offensive and defensive battles between the two teams are so severe that no team can get any goal for 120 minutes.

It may be worth noting that all of the balanced-games are not always interesting. For example, when the strength of the both teams is very low, both teams may not get any goal due to lack of skill. Moreover, when one team intentionally tries to make a game draw against the other team, the game may not be interesting.

IV. Discussion

This section discusses the relation between game progress patterns, which are how the non-dimensional information ξ varies with non-dimensional game length η, and information dynamic models.

Figure 3 shows the relation between the non-dimensional information ξ and non-dimensional game length η for Group E1, Group E2 and Model 2. Group E1 is roughly accounted for by Model 2 at n=1.5, while Group E2 is roughly accounted for by Model 2 at n=3. This denotes that the maximum information velocity (increase rate) of Group E2 is greater than that of Group E1. In another words, Group E2 is more interesting and exciting than Group E1.

In Group E1 and Group E2, the strength difference between the two teams is extremely small. In these games, before the winning goal is scored by either Japan or Holland, they are balanced, and after that they follow the safety lead curve $\xi \simeq 1$. Hence, it is considered that these games are exciting through the total game length, even though they may look like one-sided games at first glance.

Figure 4 shows the relation between the non-dimensional information ξ and non-dimensional game length η for Group E3, Round 16 and Model 1. Group E3 cannot be accounted for by Model 1 with any value of n, while Round 16 is roughly accounted for by Model 1 at n=50. In Round of 16, the strength difference between the two teams is also extremely small, so that the safety lead curve becomes $\xi \simeq 1$ for $0 \leq \eta < 1$, but ξ=0 for η=1 as in Group E1 and Group E2. Hence, it is considered that this game is interesting through the total game length.

Figure 5 illustrates a possible interpretation of Group E3. This game follows Model 1 $\xi=\eta$ until crossing the safety lead curve $\xi=(1-\eta)^{0.25}$, and then it bifurcates into two branches at the cross point (referred to as bifurcation point, here after). One branch is Model 1 $\xi=\eta$ and the other branch is the safety lead curve

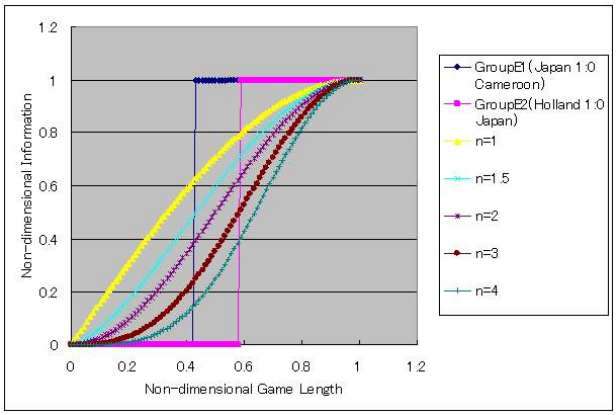


Figure 3: Non-dimensional information ξ against non-dimensional game length η for Group E1, Group E2 and Model 2.

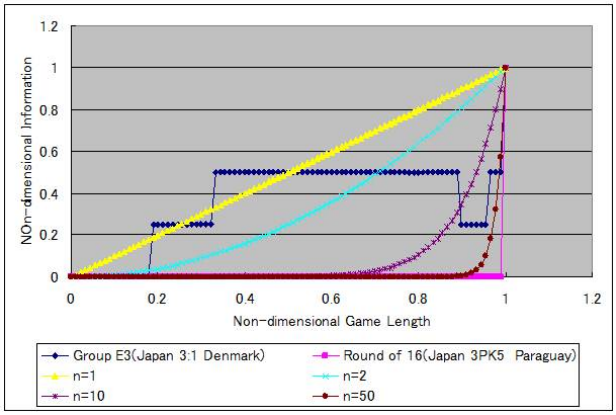


Figure 4: Non-dimensional information ξ against non-dimensional game length η for Group E3, Round of 16 and Model 1.

ξ=$(1-\eta)^{0.25}$. Information of Group E3 follows Model 1 ξ=η until the bifurcation point, then it switches to the safety lead curve ξ=$(1-\eta)^{0.25}$ and follows the curve until just before the end. Then, information of Group E3 jumps and joins to Model 1 ξ=η at the end η=1.
This game progress pattern can be expressed analytically, by introducing the unit step function,

$$u(\eta) = \begin{cases} 0 & \text{for } \eta < 0 \\ 1 & \text{for } \eta > 0 \end{cases}$$

Then, the analytical expression of this game progress pattern becomes

$$u(\eta) = \begin{cases} \eta[u(\eta) - u(\eta - a)] + \\ (1-\eta)^{0.25} \times \\ [u(\eta - a) - u(\eta - 1)] & \text{for } 0 \le \eta < 1 \\ 1 & \text{for } \eta = 1 \end{cases} \quad (6)$$

where a≃0.72. The coordinate (ξ, η) at the bifurcation point is (0.71, 0.71) approximately. This finding is critical in view of entertainment in games, for if information of Group E3 follows Model 1 $\xi = \eta$ after the bifurcation point, the game becomes boring because the information is above the current safety lead curve $\xi = (1-\eta)^{0.25}$

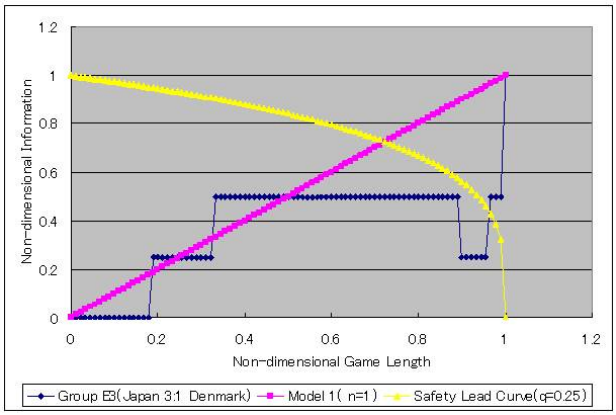


Figure 5: Non-dimensional information ξ against non-dimensional game length η for Group E3, Model 1 and Safety lead curve.

or Eq. (6). On the other hand, as far as information of Group E3 is kept under the safety lead curve $\xi = (1-\eta)^{0.25}$ or Eq. (6), Group E3 is interesting, for the winner and loser are still uncertain. This result reflects the nature of a game when the strength difference between the two teams is fairly small. In Group E3, Japan gets two consecutive goals first, so that Denmark fights very severely against Japan to avoid Japan's third goal, for this may result in providing Japan a safety lead.

V. CONCLUSION

The new knowledge and insights obtained through the present investigation are summarized as follows. Safety lead curves have been proposed. The safety lead curves have been derived by using a series of approximate solutions of the flow between two parallel flat walls, one of which is at rest, the other is suddenly accelerated from the rest to a constant velocity, and are expressed by

$$\xi = (1-\eta)^q \text{ for } 0 \le \eta \le 1$$

where ξ is the non-dimensional safety lead of game, η the non-dimensional game length, and q the positive real number parameter. It is realized that the proposed model of safety lead also represents the uncertainty of game outcome.
The safety lead curve is critical for neutral observers to assess entertainment in game, for when the certainty of game outcome is below it, the game is exciting, while when the certainty is above it, the game is boring, where excitement is a feeling that we perceive when our mind is stirred emotionally. It is suggested that any game is designed in such a way that it proceeds so as to keep the certainty of game outcome under the safety lead curve.
The power balance between the two teams (players) results in that the certainty of game outcome follows along the safety lead curve so as not to lose entertainment in game.

The safety lead curve can be a critical edge of game, at which two different things, viz. interesting game and boring game, happen, and thus one feels a particular emotion while the game proceeds along it.
The four Japan games in 2010 FIFA World Cup are under the safety lead curve, and thus they were interesting for neutral observers. In Group E1, Group E2, and Round of 16, the safety lead curve is common, and the safety lead ξ takes approximately value of 1 through the game except at the end, where ξ=0. In Group E3, the certainty of game outcome follows the information dynamic model ξ=η until crossing the safety lead curve $\xi = (1-\eta)^{0.25}$, and then it bifurcates into the two branches, follows the safety lead curve, and takes the value of 1 at the end.

VI. Recommendation for Future Work

It is realized that the analytical function $\xi = (1-\eta)^q$ represents the safety lead of game and/or the uncertainty of game outcome, which is data that is uncertainty of game outcome (see Appendix). This function represents how uncertainty of game outcome depends on the game length. Májek & Iida(2004) have calculated how uncertainty of game outcome for Chess or Soccer changes with increasing the game length during the game. Thus, a direct comparison between the information dynamic model $\xi = (1-\eta)^q$ and the calculated uncertainty of game outcome by Májek & Iida(2004) is strongly encouraged and recommended, for this provides us some clue to discover the relation between the present approach(Iida et al 2011a, 2011b) and Shannons approach(Shannon 1948, 1951) to information.

References

[1] H. Iida. Backward-game puzzle and the principle of the edge. IPSJ SIG Technical Report, 57-63, 2007-GI-17(8), 2007
[2] H. Iida, T. Nakagawa, and K. Spoerer. A novel game information dynamic model based on fluid mechanics: case study using base ball data in world series 2010. In Proc. of the 2nd International Multi-Conference on Complexity Informatics and Cybernetics, pages 134-139, 2011a.
[3] H. Iida, and T. Nakagawa, Game information dynamics. In: 10th International Conference on Entertainment Computing ICEC 2011, J. Anacleto et al.(Eds.) LNCS, 403-406, 2011b
[4] H. Iida, T. Nakagawa, K. Spoerer, and S. Sone. Three elemental game progress patterns. IScIDE 2011c(submitted for publication)
[5] P. Májek, H. Iida. Uncertainty of game outcome. In Proc. Int. Academia, Hungary, 171-180, 2004
[6] T. Nakagawa, and H. Chanson. Fluid Mechanics for Ecologists: Fundamental Edition, IPC, Tokyo, pp.336, 2002.
[7] T. Nakagawa, and H. Chanson. Fluid Mechanics for Ecologists: Applied Edition, IPC, Tokyo, pp.394, 2006.
[8] T. Nakagawa. Dr. Shunichi Tsugé's Statistical Theory of Turbulence-Microscopic View of Motion of Fluids, White Mountain Academia Press, Hakusan, pp.125, 2008
[9] C. Shannon. Prediction and entropy of printed English. The Bell System Technical Journal, 30, 50-64, 1951.
[10] C. Shannon. A mathematical theory of communication. The Bell System Technical Journal, 27:379-423, 623-656, 1948.
[11] G. Stokes. On the effect of the internal friction of fluids on the motion of pendulums. Cambridge Philosophical Transactions, IX, 8, 1851; Mathematical and Physical Papers, III, 1-141, Cambridge 1901.
[12] S. Tsugé, Approach to the origin of turbulence on the basis of two-point kinetic theory. Physics of Fluids, 17, 22-33, 1974
[13] M. Hansen, Die Geschwindigkeitsverteilung in der Grenzschicht ab der längsangeströmten ebenen Platte. ZAMM 8, 185-199, 1928
[14] R. Solso, Cognition and the visual arts, pp.294, MIT Press, Cambridge, 1994
[15] C.Y.Wang, Exact solutions of the steady-state Navier-Stokes equations. Annu. Rev. Fluid Mech. 23:159-77, 1991
[16] C. Browne. Automatic generation and evaluation of recombination games. Ph.D Thesis,Queensland University of Technology, 2008.
[17] G. Yannakekis, M.Maragoudakis. Player modeling impact on player's entertainment in computer games. L. Ardissono et al (Eds.): UM2005, LNAI3538, 74-78, 2005

Appendix: Derivation of SLC

The modeling procedure of information dynamics based on fluid mechanics has been established by Iida et al(2011a). An information dynamics model for a series of approximate solutions of the flow between two parallel flat walls, one of which is at rest, the other is suddenly accelerated from rest to a constant velocity U_0, Figure A1, will be constructed by following the procedure step by step.

As a similar flow near a flat plate which is suddenly accelerated from rest and moves in its own plane with a constant velocity is solved by Stokes(1851, 1901). For a brief sketch of the solution, see Schlichting(1968).
(a) Let us assume the flow between two parallel flat walls, one of which is at rest, the other is suddenly accelerated from rest to a constant velocity U_0 as shown in Figure A1. Note that the walls are two-dimensional, horizontal and infinitely long.

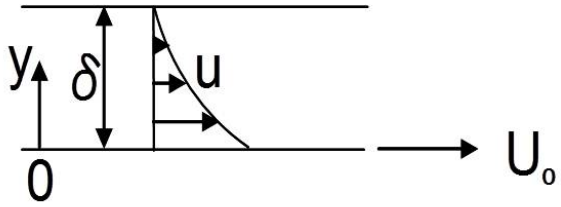


Figure A1: A definition sketch of flow between two parallel flat walls, one of which is at rest, the other is suddenly accelerated from rest to a constant velocity U_0.

Since the system under consideration has no preferred length in the horizontal direction, it is reasonable to suppose that the velocity profiles are independent of the horizontal x-direction, which means that the velocity profile u(y) for varying distance x can be made identical by selecting suitable scale factors for u and y. The scale factors for u and y appear quite naturally as the lower wall velocity U_0 and gap between the two walls δ. Hence, the velocity profile after the time $t > 0$ can be written as a function in the following way.

$$\frac{u}{U_0} = \mathrm{f}\left(\frac{y}{\delta}\right) \qquad \text{(A-1)}$$

(b) Get the solutions.
The velocity profile is here accounted for by assuming that the function f depends on $\frac{y}{\delta}$ only, and contains no additional free parameter. Since the fluid particles

are fixed on the surface of two walls due to the viscous effect, the function must take the value of 1 on the lower wall (y=0) and the value of 0 on the upper wall (y=δ), because owing to the viscous effect the fluid particles are fixed on the walls. The boundary conditions are:

$$t \leq 0 : \frac{u}{U_0} = 0 \text{ for } 0 \leq \frac{y}{\delta} \leq 1$$

$$t > 0 : \frac{u}{U_0} = 1 \text{ for } \frac{y}{\delta} = 0;\ \frac{u}{U_0} = 0 \text{ for } \frac{y}{\delta} = 1.$$

When writing down an approximate solution of the present flow, it is necessary to satisfy the above boundary conditions for $\frac{u}{U_0}$. It is evident that the following velocity profiles satisfy all of the boundary conditions.

$$\frac{u}{U_0} = (1 - \frac{y}{\delta})^q \qquad \text{(A-2)}$$

in the range $0 \leq \frac{y}{\delta} \leq 1$, where q is a positive real number parameter. Eq. (A-2) is considered as the approximate solutions on the flow between two parallel flat walls, one of which is at rest, the other is suddenly accelerated from rest to a constant velocity U_0, where each solution takes a unique value of q. The value of q must be determined by the boundary conditions and the Reynolds number $R_e = U_0 \cdot \frac{\delta}{v}$, where v is the kinematic viscosity of the fluid.

It is known that the transition from laminar to turbulent flow in the boundary layer is governed by the Reynolds number $R_e = U_\infty \cdot \frac{d}{v}$, where U_∞ is the free stream velocity, d the boundary layer thickness. The critical Reynolds number $(Re)_{crit.}$, at which the transition is initiated, is of 2,800 approximately(e.g. Hansen 1928, Schlichting 1968).

In the case of the present flow, as shown in Figure A1, at 1 atmospheric pressure and temperature at 20^o C, water has the kinematic viscosity ν=1.004×10^{-2} cm^2/s. When water is chosen as the fluid, and the constant velocity U_0 = 10 cm/s and the gap between the two walls δ=10 cm are set, we obtain the Reynolds number $R_e \simeq 10^4$. The result of this calculation clearly illustrates how the flow is liable to be turbulent under an ordinary situation.

The solution (A-2) is smooth analytical functions and thus this is only valid for laminar flow. The fundamental equations for fluid mechanics are the Navier-Stokes equation. This inherently nonlinear set of partial differential equations has no general solution, only several exact solutions, which are trivial in practice, have been found(Wang 1991). All of these exact solutions are for laminar flows, and no turbulent flow solution is available yet. However, it is considered that each of the laminar solutions in (A-2) represents an approximate turbulent solution.

(c) Let us examine whether this solution is game information or not.

The non-dimensional velocity $\frac{u}{U_0}$ varies from 1 to 0 with increasing non-dimensional distance $\frac{y}{\delta}$ in many ways with changing the parameter q. It can be considered that $\frac{u}{U_0}$ represents the safety lead curve or uncertainty of game outcome.

TABLE AI: Correspondences between flow and game information

Physical world(flow) Black	Informatical world(game) White
u : flow velocity	I : current uncertainty of game outcome
U_0: plate velocity	I_0: initial uncertainty of game outcome
y : vertical distance	L : current game length
δ : gap between two walls	L_0: total game length

(d) Visualize the assumed flow with some means. Imagine that the assumed flow is visualized with neutral buoyant particles. Motion of the visualized particles is detected by the eye almost instantaneously through light and is mapped on our retina(Solso 1994), so that during these processes, motion of the "fluid particles"is transformed into that of the "information particles"by light carrying the images of fluid particles. This is why motion of the fluid particles is intact in the physical space, but only the reflected lights, or electromagnetic waves consisting of photons can reach the retina. Photons are then converted to electrochemical particles and are passed along the visual cortex for further processing in parts of the cerebral cortex(Solso 1994). Photons and/or electrochemical particles are considered to be information particles. It is, therefore, natural to expect that flow in the physical world is faithfully transformed to that in the informatical world, or brain including eye, which is referred to as "informatical world"hereafter. During this transformation, the flow solution in the physical world changes into the information in the informatical world.

(e) Proposed are correspondences between the flow and game information, which are listed in Table AI.

(f) Obtain the mathematical expression of the information dynamic model. Considering the correspondences in Table AI and Eq. (A-2), it can be rewritten as

$$\frac{I}{I_0} = (1 - \frac{L}{L_0})^q \qquad \text{(A-3)}$$

Introducing the following non-dimensional variables in Eq. (A-3),

$$\xi = \frac{I}{I_0} \text{ and } \eta = \frac{L}{L_0}$$

we finally obtain the mathematical expression of the uncertainty of game outcome ξ as

$$\xi = (1 - \eta)^q \text{ for } 0 \leq \eta \leq 1 \qquad \text{(A-4)}$$

where η is the non-dimensional current game length, and q a positive real number parameter.

Storage Frameworks for Large Models within Model-Driven Data Warehouse Metadata Management Systems: Criteria and Evaluation

Frieder JACOBI, Robert KRAWATZECK, Marcus HOFMANN and André MÜLLER
Faculty of Economics and Business Administration, Chemnitz University of Technology
Chemnitz, D-09126, Germany

ABSTRACT

Many metadata arise during the process of data warehouse engineering (DWE). In order to achieve a maintainable data warehouse (DW), this metadata should be organized in a metadata management system (MDMS). Modern software development technology suggests a model driven approach. Following this approach, the emerging metadata are stored in form of models and metamodels. The DW domain usually comprises large models with strong dependencies between each other. This characteristic has to be considered when building performant MDMSs. This paper defines a set of criteria, which storage frameworks for large models within model-driven MDMSs should meet to ensure a performant and secure business intelligence (BI) system. Furthermore, it presents an evaluation, on the basis of the defined criteria set, of latest storage frameworks based on EMF/Ecore technology.

Keywords: Data Warehouse Engineering; Ecore; Large Models; Metadata Management System; Model-Driven Architecture; Model Repository

1. INTRODUCTION

During the last decade, data warehouses (DW) emerged into significant components of contemporary decision support systems [1]. The DW management process includes the maintenance of a lot of different information, lasting from physical system description to usage patterns and business objectives [1]. This information is called metadata, which describes other data. Metadata originates in various software systems enclosed in the DW system, like the staging area, data storage and data processing. As the number of connected systems rises, a need for managing the metadata arises. Following this line of thought, [2] proposes a reference architecture for metadata management systems (MDMS) in the context of data warehousing, as shown in figure 1.

The reference architecture focuses on the collection and preparation of metadata and does not explicitly mention the possibility for integrated data warehouse engineering (DWE). The Computer Aided Warehouse Engineering (CAWE) research group at Chemnitz University of Technology focuses on developing a process model for DWE and metadata management, including a software prototype [3]. This prototype's architecture is based on the reference architecture proposed by [2].

Following latest changes in software engineering methods, CAWE uses a model-driven approach for DWE [3]. Model-driven architecture (MDA) [4] and architecture-driven modernization (ADM) [5] are recent approaches covering system engineering and modernization mainly based on conceptual models describing the systems' attributes on an abstract level. Those models are defined platform-independent, thus do not contain technical specifications. Platform models describing concrete systems' technological specifications complete the description of a certain software product. Using transformations and platform models, conceptual models can be automatically transformed into a technical specification. Subsequently, this specification is used to configure or create a certain software artifact, such as a database schema.

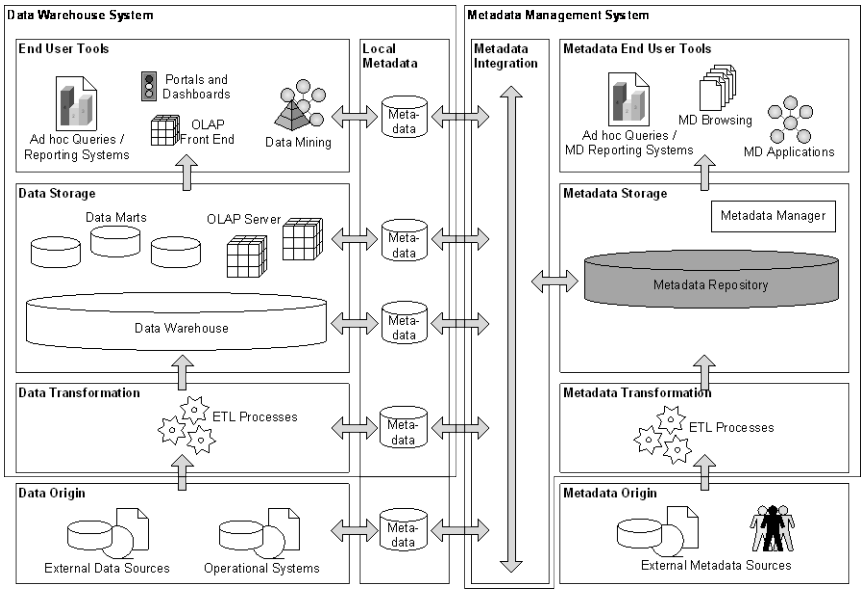


Fig. 1: Layer-oriented architecture of an integrated MDMS (based on [2], [7]).

The model-driven approach naturally leads to the existence of a certain number of software artifacts in form of models. Those models tend to contain a number of references amongst each other, since metadata used by components in DW systems strongly depend on each other [2]. For being able to define multiple models (M^1) using the same syntax, a common definition is required. Therefore, metamodels (M^2) are introduced. Any model must conform to a certain metamodel. This means the metamodel defines the model's syntax. For supporting a type-safe, non-ambiguous navigation through connected models, the metamodels in turn have to conform to a common meta-metamodel (M^3). The Object Management Group (OMG) proposes the Meta Object Facility (MOF): a four-layer architecture for models in model-driven development, defining four kinds of model types, each one describing the one below [6]. M^3 defines itself and thus forms the topmost level in this architecture. Figure 2a shows the concept of the MOF architecture. A modeling framework based on MOF subsequently provides all required characteristics. Metamodels, models as well as elements, attributes and references contained in models are henceforth referred to as modeling artifacts.

In the field of DWE models are highly connected [8], tend to be rather large and numerous [9]. Currently, all artifacts used by the CAWE prototype are stored in the file system. In the context of MDA and MDMS, this has certain deficiencies. First, a file system is not able to support automatic validation of models with regard to the metamodels they conform to. Second, the high degree of dependency between models is not taken into account. Moving files containing models to a different location leads to invalid references when not specially treated, since the referred model usually is identified by a path [10]. Third, the performance when working with large models is hardly satisfying. In many situations, the user is only interested in a rather small part of the model [2], for example a certain cube or dimension. Using a file system as storage technology requires a complete loading of the containing file, which leads to performance issues with large models.

Problem Statement: According to the reference architecture shown in figure 1, the file system implements the metadata repository. But since using a file system as a repository does have significant deficiencies, as shown above, an alternative solution has to be found. A repository is a software component whose goal is to store models and contents of software artifacts [11], and may be realized by using a client server architecture. A database server is an example of such architecture. With CAWE utilizing the model-driven approach for DWE and MDMS, the repository must at best be model-aware. For this reason, the mentioned database system is not sufficient by default. The objective of this paper is to define the special set of criteria which a metadata repository in the domain of model-driven DWH should meet to ensure a performant and secure BI system. Further, the paper at hand determines if a storage framework exists which meets the special requirements.

Paper Structure: The remainder of the paper is structured as follows: First, we outline a scenario where support for large models is desirable. Second, we establish criteria on a model-aware metadata storage framework, taking into account the three different perspectives DWE, MDA and classic software engineering. Third, we present storage frameworks to be evaluated. Since CAWE is based on EMF/Ecore technology [12], which is basically aligned on MOF (see figure 2). We constrained the pool of possible candidates to those natively supporting Ecore-based metamodels. Afterwards, we present the evaluation results, discussing interesting aspects of each individual framework. Finally, we show that a promising framework exists and give an outlook to further research questions.

2. SCENARIO

During the processes of model driven DW engineering, maintenance and evolution of the multidimensional data model plays an important role. Within a large company, a company-wide multidimensional model may consist of several hundred data cubes as well as dimensions. Additional business metadata may be attached to each element. Usually, users who work with the model do not need all provided information at once [2]. Therefore, for performance, usability and security reasons, there exist many scenarios where support for large models is desirable—for developers as well as for business users.

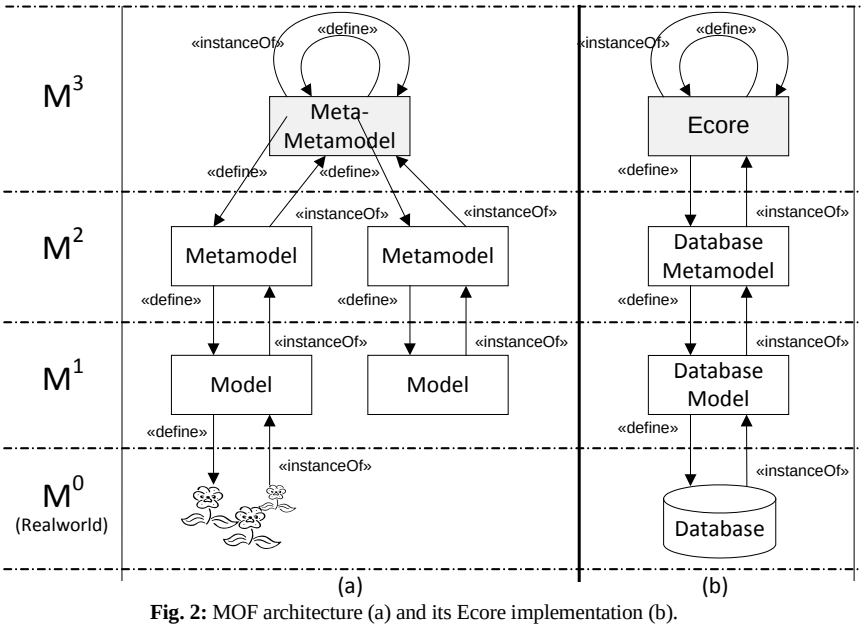


Fig. 2: MOF architecture (a) and its Ecore implementation (b).

A concrete scenario where the special requirements of large DW models should be considered is the applying of semantic concepts on multidimensional data structures (for further information about semantic multidimensional models, see [13], [9]). Conditioned by a semantic multidimensional model, precise requests are possible and typical lineage and impact analysis questions like 'Which measures are available for the calculation of product turnover?' or 'Which cubes are affected if we change our customer dimension?' can be answered [10]. These precise requests often serve only a subset of the whole multidimensional model as result. Thereby, the response can be given in two manners: (1) pure textual—the corresponding identifiers and related metadata will be presented (no user interaction possible) or (2) the corresponding model objects will be returned—which allows users to interact with. In the first case, the server which handles the semantic request can build and deliver the textual response easily without any overhead. In the second case, the connection of the particular objects to the whole multidimensional model—caused by the possibility of interaction—must be retained. As a trivial solution the whole model could be returned and only the requested objects are displayed.

For *performance*, *usability* and *security* reasons, this trivial solution should be avoided. As mentioned above, a company-wide multidimensional model can be large in size. Consequently, returning the whole model leads to high network latency by every request. Since multidimensional model objects are highly interconnected, potential subsequent requests initiated by users raise the amount of traffic. Besides the network performance, the long latency time affects the usability of the system. The long idle time during each request and subsequent interaction hinders fluent work. To avoid these problems, only relevant model objects—including their connections to other model elements—should be returned. Further model objects have to be provided on demand.

Additionally, there are situations where the whole multidimensional model is undesirable as response. For security reasons the view of model objects has to be limited appropriately [1]. Therefore, different types of user privileges are required—especially in large distributed teams [14].

3. CRITERIA FOR EVALUATION OF MODEL REPOSITORIES

To ensure the development of performant and secure BI systems, we present and discuss a set of relevant criteria for model repositories in the domain of DWE. The criteria are classified into three different perspectives according to the model driven data warehouse domain: criteria in the domain of data warehousing, criteria in the domain of MDA, and requirements in the context of classic software engineering. Following, the motivation for each chosen criterion is described separately.

Data Warehousing Perspective

Support for Large Models: According to [9], large DWs may consist of several hundred cubes and several hundred key performance indicators. Including the dimensions' definitions, the resulting model is likely to contain some thousand elements. Each of those elements is in turn described by a set of attributes [1], which leads to a large number of artifacts being stored in one model. As long as this model is stored in a model repository which is able to handle its size, this does not necessarily lead to a problem.

However, transferring the whole model to a user's personal computer leads to performance issues, like long network latency, especially if the model is bigger than some megabyte. But users who work with DW models usually do not need all of the information at once [2]. Hence, the repository should support a method for delivering only those artifacts which are currently of interest for the user.

The criterion is: Does the framework support the selection of model fragment, and which is the smallest supported level of granularity of model artifacts?

The Transaction Orientation: It is understood that DW models contain sensitive information, for example internal company structure or other inside knowledge. Furthermore, changes on those models may be complex and consequently may consist of a high number of atomic actions.

In case of a failure, the repository must allow rolling back those actions for ensuring the model's integrity. Thus it is necessary that sets of associated atomic actions are treated as transactions. For example a developer adopts the current model, which describes the extract, transform and load process (ETL) of data, according to new business user requirements. Assumed he has to replace a flat file by a more complex data source. As first action he checked-out the current ETL-model from model repository. Secondly, he adapts to model to the new requirements: he would probably delete the old data source (1), before he added the new one (2) and all necessary dependencies (3). Afterwards he commits the local changes of the model to the model repository. Thereby, it is important that all three modeling actions taken would be recognized to ensure the model integrity. Otherwise inconsistencies appear, like multiple data sources (the delete actions (1) was not recognized) or missing data sources (the add action (2) was not noted).

The criterion is: Does the framework support transaction orientation?

Multiuser Capability and Access Rights: In field of DW management, typically more than one person is involved in the development process [1] [15]. Each of them may take a different role in this process.

In the context of multiuser capability it is required that multiple users may work concurrently on the same artifacts. To satisfy the need for distinct roles it is furthermore necessary to provide means for defining access rights.

Additionally, according to [2] „In most cases today, DW/BI system security is largely about recognize legitimate users and giving them very specific rights to look at some but not all of the data". So it is necessary "to ensure that only authorized users can access the DW/BI system, and limit everyone's view of data as appropriate" [2] to avoid that untitled users get access to sensitive data. The same holds for modeling artifacts containing sensitive information.

The criterion is: Does the framework support multiple users, and which is the smallest level of granularity allowed for defining access rights on artifacts?

Transparent Replication: In large concerns with head offices at different locations network latency becomes relevant. From a large distance between client and server results a long latency.

For performance reasons, it is therefore required to minimize this distance. This is done by installing additional servers located

near to the end user. For transparency reasons, this should be taken care of the framework without user interaction.

The criterion is: Does the framework support automatic replication?

Model-Driven Architecture Perspective

Unique Identifiers for Modeling Artifacts: Working with models requires the possibility to uniquely identify modeling artifacts [16]. For example, if two models are named similarly, there is no way to figure out which is the searched-for, if the name is the only request parameter. As for other criteria we establish, the level of granularity of artifacts that may be identified plays an important role. Furthermore, the unique identifier should be persistent throughout the whole artifact's lifetime [16].

The criterion is: Does the framework provide unique artifact identifiers, and what is the smallest level of granularity for identifying artifacts, and are those identifiers persistent?

Inter-model Dependency Support: Following the approach for the design of domain specific languages, using small metamodels (M2) for representing specific views on the modeled domain is preferred [17]. Consequently, complex situation are expressed using different metamodel instances (models, M1). Models hence may contain strong dependencies to other models conforming to different metamodels [8]. To keep track of those dependencies the repository is required to support and maintain inter-model dependencies.

This is best realized by regarding the model's meta-models, which implicitly define the dependencies' syntax [17]. According to the MOF architecture, meta-models must conform to the same meta-meta-model (M3) to be used for dependency tracking [6]. Since all models of the CAWE prototype are based on the Ecore meta-metamodel, the repository must allow this M3 to be used as meta-metamodel.

The criterion is: Does the framework provide support for Ecore based metamodels?

Physical Transparency: In model-driven software engineering, models are core piece of information. The physical implementation of the storage mechanism is not relevant and thus should be hidden from the modeling tool.

Physical transparency is achieved when the framework's internal storage mechanism does not have any effect on the framework's utilization [18].

The criterion is: Does the framework provide physical transparency regarding its storage mechanism?

Software Engineering Perspective

Versioning Support: Artifacts in software engineering processes are typically subject to changes in form of an evolution. When incompatibilities occur it must be possible to move back in time for analyzing changes which have been made to the artifacts. Thus, versioning support is required [16].

The criterion is: Does the framework support versioning, and what is the smallest level of granularity of the artifacts underlying version control?

Locking Support: Since multiple users may work concurrently on models stored in the repository, conflicts may occur and must be handled.

One reliable approach is pessimistic blocking, giving users the possibility to lock artifacts [16]. Locked artifacts may not be changed by other users until the lock is released.

The criterion is: Does the framework provide locking, and what is the smallest level of granularity for locking artifacts?

Branching and Merging: In the software engineering process it is likely to happen that, while the system is in use, a new version is being developed.

Separation of deployed models and those currently under development is usually implemented by a branch and merge capability [16]. For that, a copy of the models in their current status is moved in a separate part of the repository called branch. In a branch, changes do not have any effect on the original models. After all changes have been implemented and tested, support for a reintegration of the changed models into the original location is required. This reintegration step is called merge.

The criterion is: Does the framework support branching and merging?

4. TOOL PRESENTATION AND SELECTION

The repository products to be explored can be classified in two main categories: client/server applications and service oriented applications. The main difference is that clients have to establish an explicit connection to a server in the first category, whereas more flexible tools can be developed by using a model-aware service registry mechanism in the latter tool category. According to our research initiative that is based on the Ecore meta-metamodel, we identified the following projects that will be outlined and examined in this chapter: Connected Data Objects (CDO) [19] and EMF¬Store [20] in the first category as well as ModelBus [21] and Morse [22] in the second category.

There are other model repository approaches based on the EMF/Ecore technology which do not explicitly focus on model storage. Those projects focus on conflict detection and resolution like Adaptive Model Versioning (AMOR) [23] as well as the semantic cross-tool integration like ModelCVS [24], [25].

CDO: CDO has been developed to foster modeling within an enterprise environment. The main features are scalability, distribution, persistency and transaction support. Its focus is on distributed and shared Ecore models [26]. The basic idea behind CDO is the permanent synchronization of modeling objects between client and server. Access to single resources is provided in a way similar to file systems organized via directories and resources.

EMFStore: EMFStore is part of the UNICASE project and was developed to provide a repository of models within the software engineering process [27]. The aim is to remedy deficiencies of other model repositories, especially to provide a continuous offline operation. Models are checked out locally and are synchronized upon upload in the repository.

ModelBus. ModelBus was developed within the MODELPLEX project funded by the European Union. On completion of the project, ModelBus has been transferred to the Fraunhofer FOKUS research group and is a closed-source project by now. It is a framework for integration and communication between different systems. It offers a service registry as well as a model repository based on the well-known Subversion (SVN) [28].

Morse: Morse (Model-Aware Repository and Service Environment) is developed by a cooperation of the University of Trento and Vienna University of Technology. It offers "a service-based environment for the storage and retrieval of models and model-instances at both design- and runtime" [22].

5. EVALUATION

The presented candidates were evaluated according to the established set of criteria for model-driven DWE. Table 1 gives a contrasting juxtaposition over the result of the evaluated frameworks. Furthermore, Each of the frameworks will be discussed in detail.

The mechanism behind *EMFStore* is comparable to SVN or CVS, where the user works on the models in an offline mode [27]. That means the models have to be checked out locally before the user can work on them and have to be committed after work is done. EMFStore records the user interactions as a sequence of operations and the commitment is realized as a transformation on the model. Further, EMFStore organizes models in projects. References between elements are expressed by a unique identifier, but only work within the same project. So, always the whole project has to be checked out, which results in a high footprint.

ModelBus uses SVN as a persistence tier, which originally comes with multi-user support. ModelBus also works offline; it maintains a mechanism to work with smaller models for saving resources while working on large models. Large models are supported by a fragmenting mechanism, which allows the user to partially check out a model [8]. The peculiarity here is that the user has to know all elements he wants to receive before he can start the request. As a notable feature, ModelBus provides a service registry mechanism which enables users to extend the functionality by adding individual components. This allows, for example, registering a service for automatic model transformations, which ensures the consistency of all models depending on the one being altered.

Morse follows a service-oriented approach, too. Compared to ModelBus, the user is able to transparently fetch any kind of model elements without specifically defining the required part before [29]. In contrast to the frameworks mentioned above, Morse is designed for an online operation. In the current version of Morse, no locking mechanism is supported. This hinders collaborative work substantially.

CDO is also designed for online operation [30]. Due to a specific storage solution, it is possible to manage very large models in a performant manner. CDO fetches only those model elements which are currently needed. Hence, model elements are requested on demand. Collaborative work is ensured by a locking mechanism on element level.

Table 1: Results of the evaluation of Ecore-based model repositories.

Criteria	CDO	EMFStore	ModelBus	Morse
Support for large models / Granularity	Yes / Element	No / Only the whole model	Yes (so-called "fragments")	Yes / Element
Transaction orientation	Yes	Yes	Yes (based on SVN)	Yes
Multi-user capability / Access rights / Granularity	Yes / No / Not supported	Yes / Yes / Model	Yes / Yes / Model	Yes / No / Not supported
Transparent Replication	Yes	Yes	Yes	Yes
Artifact UUID / Granularity / Persistence	Yes / Element / Yes	Yes / Element / Yes	Yes / Element / Yes	Yes / Element / Yes
Inter-model dependency support	Yes	Yes (only project-wide)	Yes	Yes
Physical transparency	Yes	Yes	Yes	Yes
Versioning support / Granularity	Yes / Element	No	Yes / Model	Yes / Element
Locking support / Granularity	Yes / Element	No / Not supported	Yes / Model[1]	No / Not supported
Branching / Merging	Yes / Yes	No / Yes	Yes / Yes	Yes / Yes

6. CONCLUSION AND FURTHER WORK

As result of the presented evaluation, none of the evaluated frameworks fulfills all of the defined criteria for the domain of model driven DWH. Nevertheless, CDO seems to be the most promising framework for the outlined domain. Conditioned by the missing support for large models, EMFStore is not yet qualified for the development of performant BI Systems. The issue with Morse is the missing locking support which detains a fluent cooperative work. The Framework ModelBus seems promising too, but the need to predefine model fragments to support large models does not suit in the dynamic BI domain, as outlined in the semantic modeling scenario. To obtain a performant system, it is not possible to predefine all possible answers to dynamically user-generated requests. In contrast, CDO supports large models occurring in DWE by design. Weak spots like missing support for access right definitions are unfavorable, but by the fact of CDO's active community [19] it is expectable that this will be fixed. The CAWE project will adopt this technology and integrate it into its prototype for model-driven DW lifecycle management. To improve the usability of CDO, the enhancement of the framework by a service registry mechanism – like presented by Morse and ModelBus – seems attractive. Therefore, service integration into the CAWE prototype is subject of further work.

Acknowledgements: This research has been partly funded by the European Social Fund and the Federal State of Saxony, Germany.

[1] ModelBus supports locking on element level, if the ModelBus adapter for the component *Eclipse Papyrus* is installed [28].

7. REFERENCES

[1] Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., Becker, B.: The Data Warehouse Lifecycle Toolkit (2nd Edition). Wiley, Indianapolis (2008).

[2] Melchert, F.: Integriertes Metadatenmanagement: Methode zur Konzeption von Metadatenmanagementsystemen für das Data Warehousing. Logos Verlag Berlin, Berlin (2006).

[3] Kurze, C., Gluchowski, P.: Computer-Aided Warehouse Engineering (CAWE): Leveraging MDA and ADM for the Development of Data Warehouses. AMCIS 2010 Proceedings, paper 282 (2010).

[4] MDA Guide Version 1.0.1. Miller, J., Mukerji, J. (eds.). Object Management Group (2003).

[5] Khusidman, V.: ADM Transformation White Paper - Part II. Object Management Group (2008).

[6] Meta Object Facility (MOF) Specification - Version 1.4. Object Management Group (2002).

[7] Auth, G.: Prozessorientierte Organisation des Metadatenmanagements für Data-Warehouse-Systeme. Dissertation, University of St. Gallen, Difo-Druck GmbH, Bamberg (2003).

[8] Sriplakich, P., Blanc, X., Gervais, M.-pierre. Collaborative Software Engineering on Large-scale models: Requirements and Experience in ModelBus. In: Proceedings of the 2008 ACM symposium on Applied computing - SAC'08, pp. 674–681. ACM, New York (2008).

[9] Kurze, C., Gluchowski, P.: Towards Principles for Structuring and Managing Very Large Semantic Multidimensional Data Models. AMCIS 2009 Proceedings, paper 315 (2009).

[10] Kurze, C., Gluchowski, P.: Reporting Repository: Using Standard Office Software to Manage Semantic Multidimensional Data Models. In: Proceedings of the Joint Workshop on Advanced Technologies and Techniques for Enterprise Information Systems, pp. 67–76. INSTICC Press, Setúbal (2009).

[11] Bernstein, P.A.: Repositories and object oriented databases. SIGMOD Rec. 27(1), 88–96 (1998).

[12] Steinberg, D., Budinsky, F., Paternostro, M., Merks, E.: EMF: Eclipse Modeling Framework (2nd Edition). Addison-Wesley Professional (2008).

[13] Böhnlein, M., Ulbrich-vom Ende, A.: Semantisches Data Warehouse-Modell (SDWM) - ein konzeptuelles Modell für die Erstellung multidimensionaler Datenstrukturen. In: Informationssystem-Architekturen. Rundbrief des GI-Fachausschusses 5.10, No. 1, (2001).

[14] Marco, D.: Building and Managing the Meta Data Repository: A Full Lifecycle Guide. Wiley, New York (2000).

[15] Dolk, D.R.: Model Management and Structured Modeling: The Role of an Information Resource Dictionary System. Communications of the ACM 31(6), 704–718 (1988).

[16] ISO/IEC 10027:1990, Information technology - Information Resource Dictionary System (IRDS) framework. ISO/IEC (1999).

[17] Stahl, T., Voelter, M., Czarnecki, K: Model-Driven Software Development: Technology, Engineering, Management. Wiley, Indianapolis (2006).

[18] Karagiannis, D., Kühn, H.: Metamodelling Platforms. In: Bauknecht, K., Tjoa, A.M., Quirchmayr, G. (eds.) EC-Web 2002. LNCS, vol. 2455, pp. 182–195. Springer, Heidelberg (2002).

[19] CDO Model Repository, http://www.eclipse.org/cdo/.

[20] EMFStore project home. http://www.eclipse.org/emf-store/.

[21] ModelBus project home, http://www.modelbus.org/modelbus/.

[22] MORSE, Model-Aware Service Environment, http://www.infosys.tuwien.ac.at/prototypes/morse/.

[23] AMOR - Adaptable Model Versioning, http://www.modelversioning.org/.

[24] Kramler, G., Kappel, G., Reiter, T., Kapsammer, E., Retschitzegger, W., Schwinger, W.: Towards a semantic infrastructure supporting model-based tool integration". In: GaMMa '06: Proceedings of the 2006 international workshop on Global integrated model management. pp. 43–46, ACM New York, NY (2006).

[25] ModelCVS: A Semantic Infrastructure for Model-based Tool Integration, http://www.modelcvs.org/.

[26] Stepper, E.: Modeling goes Enterprise. eclipse magazin 3, 38–42 (2009).

[27] Koegel, M., Helming, J.: EMFStore - a Model Repository for EMF models. In: ICSE'10 Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 2, pp. 307–308. ACM, New York (2010).

[28] ModelBus Users Guide, http://www.modelbus.org/modelbus/downloads/current/ ModelBus_UserGuide_0961_v1.9.6.pdf.

[29] Holmes, T., Zdun, U., Dustdar, S.: MORSE: A Model-Aware Service Environment. In: 2009 IEEE Asia-Pacific Services Computing Conference, pp.470–477, IEEE (2009).

[30] Stepper, E.: Modeling goes Enterprise II. eclipse magazin 4, 31–38 (2009).

Semi-Updating the Correctness of Point of Interest Information by Multi-Level Collective Intelligent

Chun-Hung Lu, Wen-Nan Wang, and Yi-Hsung Li
{enricoghlu, wennen, sbmk}@iii.org.tw
Innovative Digitech-Enabled Applications & Services Institute, Institute for Information Industry

The adoption of mobile devices (laptops, smart phones, and tablets) has been increasing rapidly. Many of them use the LBS (Location-Based Service) recommendation system [1, 5, 7, 9]. These recommendation systems are used to find several kinds of places, including viewpoints, restaurants, shops, etc. The recommendation system provides accurate and specific information about the locations of POI (Point of Interest) [8, 12]. However, the data on which the recommendations are based can be provided by the service provider or may be publicly available data that is collected from sites like Google Places [4]. Regardless of the data source, failure to check the POI information can result in the pushing of incorrect information to users, giving them a negative experience of the service. Therefore, a means of ensuring the correctness of POI information is essential.

This work proposes an information checking approach to maintain the correctness of POI information, as shown in Fig. 1. POI is assumed to be any location that is popularly discussed on the Internet. Nowadays, several forums [10] and websites [11] discuss POI, and some even rank them [6]. The goal is to identify changes in publicly available information about POI. The proposed system can semi-update the correctness information that is pushed to users.

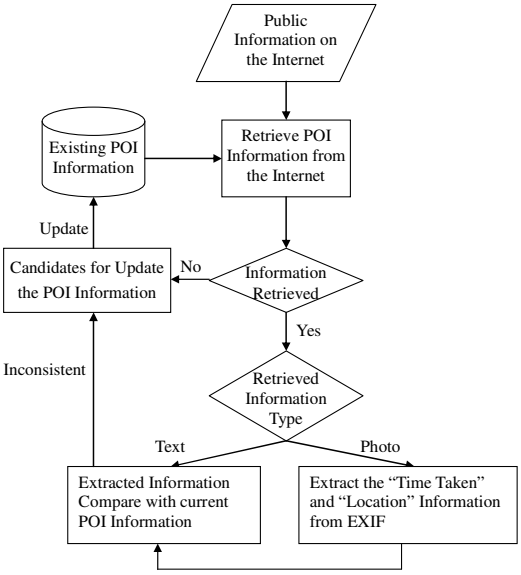


Fig. 1 POI Information Updating

POI information is retrieved via the API (Application Programming Interface) which provided by the website service provider [3, 4]. It is also collected using a web clawer, which developed by us, from the websites [6, 10, 11] that do not provide the API. The source website can be a forum, social network, blog, or others. Types of retrieved data include text and photographs. Data is extracted from the EXIF (Exchangeable Image File Format) of photographs, including the "time

taken" and "location" information. The "time taken" can be used to ensure that the POI actually existed at the time the photograph was taken. If the camera consists with GPS module, the EXIF information content with "location" information (latitude and longitude) which can be used to check whether the POI has moved. Combining CKIP (Chinese Knowledge and Information Processing) [2] word segmentation with Chinese Named Entity Recognition (Chinese NER) enables text data to be analyzed to obtain current POI information. After the relevant information has been extracted from the text and photographs data, the existing POI is compared with the extracted information to check the address, name, ranking and other data, to determine which, if any, POI should be updated or removed from the existing POI database. The POI thus identified is added as candidates for update.

Without the proposed approach, the recommendation system providers need to check the correctness of the POI information one by one. More POI information the POI providers have, more manpower cost they need to spend. Thus, this approach greatly reduces the number of POI data that have to be checked, and reduces the manpower required to do so. The proposed approach automatically detects incorrect POI information. Moreover, the ranking information that is retrieved from the Internet can be provided to users along with relevant comments, to help to prevent them from making a bad decision.

Acknowledge

This study is conducted under the "III Innovative and Prospective Technologies Project" of the Institute for In-formation Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

References

1. R. Bader, E. Neufeld, W. Woerndl, and V. Prinz, "Context-Aware POI Recommendations in an Automotive Scenario using Multi-Criteria Decision Making Methods," Proceedings of *the 2011 Workshop on Context-awareness in Retrieval and Recommendation*, pp. 23-30, February 2011.
2. CKIP, http://ckipsvr.iis.sinica.edu.tw/
3. Flickr, http://www.flickr.com
4. Google Places, http://code.google.com/intl/en/apis/maps/documentation/places/
5. T. Horozov, N. Narasimhan, and V. Vasudevan, "Using Location for Personalized POI Recommendations in Mobile Environments," Proceedings of the International Symposium on Applications on Internet, pp. 123-128, January 2006.
6. iPeen, http://www.ipeen.com.tw/

7. Y. Kawai, J. Zhang, and H. Kawasaki, "Tour Recommendation System Based on Web Information and GIS," Proceedings of the 2009 IEEE international conference on Multimedia and Expo, pp. 990-993, April 2009.
8. Kingwaytek Technology Co., Ltd., http://www.kingwaytek.com/
9. M. H. Kuo, L. C. Chen, and C. W. Liang, "Building and Evaluating a Location-Based Service Recommendation System with a Preference Adjustment Mechanism," Expert System with Applications, Vol. 39, Issue 2, pp. 3543-3554, March 2009.
10. Mobile01, http://www.mobile01.com
11. Yahoo Knowledge, http://tw.knowledge.yahoo.com/
12. K. Zhai and W. K. Chan, "Point-of-Interest Aware Test Case Prioritization: Methods and Experiments," Proceedings of the 10th International Conference on Quality Software, pp. 449-456, July 2010.

Information and communication technologies (ICT) as drivers in the globalisation process of small and medium-sized firms (SME) from Asia/Pacific

Dr Thomas Borghoff

School of Marketing and International Management, PO Box 600, Victoria University of Wellington, New Zealand

Email: Thomas.Borghoff@vuw.ac.nz

Abstract
This paper focuses on the role of ICT as a facilitator of the globalisation process of firms in Asia/Pacific as firms in this region develop in the most dynamic context. Histographic case studies are used in order capture the globalisation process right from inception. The results show that ICT can substantially drive the globalisation process of firm depending on the mindset and international experience or contacts of the entrepreneur and the management.

Introduction

The disciplines of information management and international management have so far developed in quite separate trajectories and are not well integrated despite the high relevance of their interplay in practice. This article therefore aims to provide an interdisciplinary and processual approach to the influence of ICT on the globalisation of SMEs in Asia/Pacific as the currently most vibrant region.

Literature review

From a management perspective, ICTs were first viewed as a potential source of competitive advantages that even culminated in a 'magic bullet theory" assuming *"that the gun fires itself*" (Markus/Benjamin 1997: 58). A similarly performance-oriented perspective developed suggesting that ICTs increase the efficiency of organisations first in a one-dimensional and later in a multidimensional approach. A third perspective linked the ICT to strategy and organisation. In this context, SMEs have limitations such as limited resources and size, preventing the adoption of sophisticated ICT. Most studies suggest that SMEs are rather followers in the adoption of ICT due to these constraints but also suggest a "*levelling of the playing field*" in international business (Shneor 2009). The extended connectivity of SMEs allows for expanding boundaries of activities and networks.

A lot of attention has been directed towards the potential of e-business as a new and distinct way of internationalisation. Most studies indicate that particularly firms in B2B businesses profit from ICT applications (Nieto/Fernandez 2006) and indicate a high correlation of the degree of internationalisation and B2B (Jaw/Chen 2006). A similar effect is generated by collaborative efforts and network leverage (Samiee 1998). There are different views how ICT impact the choice of entry modes but literature is basically consistent in assuming that ICTs increase the speed of market entry and of internationalisation in general (Berry/Brooke 2004). Another influence of ICT is supposed to be the strengthening of intra- and interorganisational ties. For example, Chen (2002) suggests that ICT promote cooperation along the international supply chain and enhance international network building (McMahon 2002). Berry/Brooke (2004) even suggest that networks accelerate internationalisation. A positive influence on the speed of internationalisation is proposed by several authors (e.g., Jaw/Chen 2006).

Methodology

This study builds on social systems theory (e.g., Luhmann 1995, Borghoff 2005) in order to provide a holistic perspective and to allow for the integration of more specific perspectives under this umbrella. From this perspective, globalisation is an increasing level of complexity within and across social systems on geographic and informational level and also on the level of organisations such as SMEs. Globalisation thus causes an increase in environmental complexity for the individual organisation. The consequence is a necessary increase in its own complexity to provide the 'requisite variety' (Ashby 1956) in the global context. In order to conceptualise the globalisation process for the study, it is broken down into four sub-processes (Borghoff 2005):

1. Global foundation (e.g. "born globals or "international new ventures")
2. Internationalisation: changes in the level and dispersion of activities in different national markets;
3. Global networking: development and management of internal and external network structures in the global context;
4. Evolutionary dynamics: motors of change driving the differentiation and integration of social systems on global scale. Firms develop in a co-evolutionary process with their environment.

This paper serves to explore the influence of ICT on the globalisation of firms. The focus is on SMEs in order to capture the whole development right from the beginning of this process. A histographic research design covers the whole process from inception to the latest steps in globalisation.

Sampling

The study follows a qualitative sampling strategy (Fletcher/Plakoyiannaki 2008). The sampling is purposive, which includes the selection of information-rich cases for study in depth. Maximum variation sampling seeks to incorporate as much diversity as possible into the research design (Ibid: 6) The study is based on case studies of twenty firms from China, India, New Zealand, and Singapore, providing a high variation while still allowing for meaningful pattern-matching.

Data collection and analysis

Multiple histographic case studies serve to explore characteristics of the four sub-processes of globalisation and their facilitation by ICT. Interviews with the founders and/or top managers provide an overview on the globalisation processes from (pre-) foundation to the current activity profile. The data were coded and analysed by the use of a software package for qualitative data analysis (QSR NVivo 8).

Results

The results suggest a distinction of three structural dimensions that define the evolutionary interplay of firms and their environment in the globalisation process four processual dimensions that drive the change of the structural elements and their interplay. Structural dimensions include (1) the influence of ICT on the global environment, (2) the influence of ICT on the global social system structure, and (3) ICT in the global interplay of system and environment**.** In the environment, markets, industries, cooperations, and external

integration of the international supply chain turned out to be dominating dimension ISC. There is a mutual relationship of these variables with ICT. All sample firms observe increasing dynamics and complexity in their environment and some even with an accelerating speed. From a systemic perspective, more than half of the sample firms even integrate their whole supply chain through CRM or ERP - pointing to an increasing 'deepening' in the use of ICT and an increasing informational integration of operations both internally and externally. For most sample firms, ICT has an influence on organisation by flattening structures, increasing the quality of the information flow and decision-making as well as by enhancing the coordination capabilities. These results confirm suggestions from existing literature (e.g., Juliussen 2000). Particularly significant is the influence on business and management processes through ICT such as ERP. More specifically, those firms depending on project management all report a shortening of project life cycles and a better multiproject management. ICT has a very different influence on strategy depending on the industry, mind-set of top management, and business model of the sample firms. The contribution of ICT ranges from providing informational input to decision-making to being the backbone of the whole business. Contrary to early studies connecting ICT to strategy trying to establish a link between norm strategies are not deliberately following any norm strategy. The firms follow a rather customised approach using external sources and to a large extent 'off-the-shelf' ICT without significant internal resources. ICT-based communication is the backbone for all sample firms. While most firms still prefer to establish contacts and trust on a face-to-face basis, once established, communication is overwhelmingly based on ICT. To close the gap between sources of rich and more codifiable information, the trend goes to more collaborative forms of communication. In the interplay of ICT, strategy, and structure the strategic mind-set of the management crystallises as the strongest influence.

The results suggest that ICT offers the possibility to begin with the development of the international establishment chain earlier, or – in other words – to form 'born globals' (Knight/Cavusgil 2004). On the other hand, ICT allows exporters to remain in the first stage of the establishment chain, much longer as markets can be served through e-commerce, remote connectivity, ERP, or CRM. The study provided evidence that as predicted by most research (e.g., Berry/Brock 2004), the mind-set of the founder and/or management has a decisive influence on the adoption and use of ICT in the globalisation process. Most sample firms have very sophisticated information systems that perfectly match their needs, even integrating the whole international supply chain. The sample firms are well aware of risks caused by the use of ICT. Many firms are to a large extent dependent on ICT and thus the failure of ICT is seen as a major risk in itself. The reliability and trustfulness of information and actors in the Internet are seen as further obstacles. The problem of sending rich and contextual information is a source of limited exchange of such information. This may also cause misinterpretations and even changes in the communication behaviour. Due to these potential risks, the selection of reliable and substantive information is a key practice at the sample firms. All sample firms choose an active but cautious approach to the use of ICT rather than following every emerging new trend

The three structural dimensions are driven by four processual dimensions in globalisation: (1) Global foundation of firm (e.g., born globals), (2) internationalisation, (3) Global network development, and (4) Global evolutionary dynamics.The sample firm showed a wide range of motives and prior international experiences of the founders. The scope and speed of globalisation among the sample firms was influenced by four key factors. These are the nature of the products, the mind-set of the founders and/or top management, the global experience of key decision-makers, and incentives and assistance by government (particularly in three Asian sample firms). ICT has a strong impact on the internationalisation of most sample firms. The Internet allows for business with foreign markets through exports without the need to set up affiliates abroad in order to serve these markets and also for a prolongation of the export phase for most sample firms with foreign markets. All sample firms use ICT to get and to diffuse information. The most common source of information about foreign markets is the Internet. All sample firms also use ICT for international learning processes. Several sample firms are already running sophisticated intranets and knowledge management systems. The influence of ICT on network dimensions is the strongest as expressed by the respondents of the sample firms. The results provide evidence for the high importance of ICT in developing and coordinating inter- and intraorganisational networks. Larger sample firms already integrate their operations and supply chains completely through ERP and CRM. The sample firms thus display a much higher degree of advancement on the four stages continuum in Internet commerce development of: (1) presence, (2) use of portals, (3) transaction integration, and (4) enterprise integration as identified by Jaw/Chen (2006) only a few years earlier. This fact points to a rapid pace and increase in the diffusion and adoption of ICT among SMEs. The adoption, development, and diffusion of ICT pretty much reflect their influence on evolutionary dynamics. ICTs are increasingly important in providing transparency and in controlling the more complex and dynamic environment that they helped to create. Life cycles are getting shorter and ICTs help in keeping pace. Dialectical relations in business are getting more complex and dynamic. The results of this study clearly suggest that ICTs have a strong impact on the recursive interplay of firms and their environment in the process of globalisation, fuelling a positive feedback loop of accelerating change and increasing complexity.

Discussion

The results of this study support research indicating that ICT and particularly the Internet constitute a major engine of the process of globalisation). The literature review and the empirical analysis suggest a distinction of three structural dimensions (environment, system, and ICT) and four processual dimensions (global foundation, internationalisation, global networking, and global evolutionary dynamics). The structural dimensions are transformed through the process dimensions. This finding confirms literature, stating that technology is both driven by, and itself a key driver of globalisation (e.g., Bradley, Jerry/Richard 1993: 3) supporting the increasing influence of the network, knowledge, and ISC perspectives in internationalisation theory. The same applies to the development of ICT infrastructure. The results also confirm institutionalisation theory (e.g., Westney 1993) as the creation of an ICT infrastructure such as that in Singapore leads to a deeper and more rapid diffusion of practices, which may serve to develop competitive advantages. On firm level, the diffusion of ICT is fast and reached a considerable level in the sample firms as compared to SMEs in former studies (e.g., Nieto/Fernández 2006). Examples from the sample firms

suggest that mechanisms described in population ecology (e.g., McKelvey 1982) apply to the use of ICT as firms that did not adopt ICT in time went out of business. The same may apply to practices as the results of the life cycle perspective have shown. This finding supports that dynamics are gaining momentum and have moved to the centre stage (Earl/Khan 2001).

On firm level, the results support the early observation of Picot/Reichwald (1994) that ICT is penetrating all functional activities in the supply chain. In many key activities ICTs are simply a necessity by now. The results about the influence of ICT on strategy reflects prediction of strategic contingency theory (e.g., Doz/Prahalad 1991) about the influence of context on strategy and decision-making. However, industry or size do not have a major influence on the strategic position among the sample firms so that the strategic mind-set of the management crystallises as the strongest influence on relating ICT, strategy, and structure. This confirms results from previous studies (e.g., Moen 2002) The results of this study further suggest that contrary to former studies indicating a more operative rather than strategic use of ICT by SMEs in their globalisation process (e.g., Loane 2006) an already "deep integration" of ICT into the global activities of the sample firms and most could not even exist. This finding supports studies from institutionalisation theory (e.g., Westney 1993) and technology diffusion (e.g., Cantwell 2002).

While the structural dimensions define a firm and reflect its resource endowment, structure, and strategic position at a given point in time, the processual dimensions reflect processes of change along the life span of a firm, beginning with the foundation of a firm, processes of internationalisation, network development, and evolutionary drivers. The results on global foundation (global experience, contacts as founding capital) show the high importance of these resources for the globalisation of a firm. This finding contradicts the Uppsala model of incremental learning (e.g., Johanson/Vahlne 1977). The finding rather confirms results from research in international entrepreneurship (e.g., Knight/Cavusgil 1996). The findings of this study confirm the results of previous studies indicating that the mind-set of the entrepreneur and/or the management has a dominating influence on both the habit towards globalisation and towards the adoption of ICT (e.g., Fillis/Wagner 2005: 609). As link to the next section, the results of the study suggest that the sample firms become involved in international trade very early but do not follow the development of the establishment chain in the same pace as described by incremental stage models. This finding supports the findings of a study by Arenius et al. (2006) that although born globals internationalise may still expand incrementally but at a faster pace. This finding confirms literature, stating that technology is both driven by, and itself a key driver of globalisation supporting the increasing influence of the network, knowledge, and ISC perspectives in internationalisation theory. The same applies to the development of ICT infrastructure, which in turn also confirms institutionalisation theory (e.g., Westney 1993) as the creation of an ICT infrastructure such as that in Singapore leads to a deeper and more rapid diffusion of practices. On balance, the study confirms results from previous studies that indicate a positive correlation between the degree of internationalisation and the degree of e-business (e.g., Jaw/Chen 2006). The results of the study suggest that the internationalisation process of the sample firms is faster through ICT, confirming results of previous studies (e.g., Shneor 2009). The results of this study thus confirm studies covering born globals in terms of the speed of internationalisation due to ICT but not in terms of skipping stages in the establishment chain as argued in the incremental stage models. The study also provides evidence that the internationalisation process of Internet-related firms is fast and discontinuous.

The results support the observation of increasing dynamics in the globalisation of firms (e.g., Jaw /Chen 2006). Independent of their business, all sample firms report a strong impact on how they design and develop their business. This finding confirms former studies indicating that particularly the Internet has created completely new dynamics in the way SMEs can globalise (Mukherji 2002: 505) and expand their boundaries (Samii 2004: 15). The results that strategic decision-making is perceived by the sample firms to be least affected by ICT is contrary to the position of information processing in organisation theory (Galbraith 1977). The sample firms perceive different influences on the life cycles of products, practices, and organisational units confirming economic theories (e.g., Vernon 1966), organisational change literature (e.g., Greiner 1972), and population ecology (e.g., Aldrich 1999). The genuine contribution of this study in this regards is to observe not only life cycles of whole organisations or populations of organisations but also of organisational units and practices, which proved to be particularly valuable in observing and explaining change processes.

All sample firms perceive an increase in the amount of available information and innovative activity through the Internet and most of them even observe an acceleration of this process providing evidence for the 'evolutionary theory of globalisation' (Borghoff 2005). The evolutionary process of variation and retention of social systems seems to be accelerating emerging change. Particularly the Internet is perceived by the sample firms to be a very strong driver of evolutionary change. Regarding cognitive change, strategic thinking among the sample firms is increasingly more focused on network relations, dynamics, and change - with the rate of change itself as most critical dimension. These key dimensions are increasingly dominating decision-making and action.

On balance, the results of the study regarding the influence of ICT on the evolutionary dynamics in globalisation reflect the observation of researchers such as Eisenhardt (2002: 91) that the economics of disequilibrium and information have moved to the centre stage, or of Earl/Khan (2001: 66) that dynamics are gaining momentum and that the new currency is time, not cost. In addition to the increasing importance of dynamics, this study further clearly identified the increasing importance of relationships and network management. The results of this study indicate that all these dimensions are strongly enhanced and even driven by ICT. Given that former studies have identified the strength of ICT in 'levelling the playing field' in internationalisation or building up international activities, the results of this study suggest that this process has come to the next stage. As the importance of ICT in internationalisation is now perceived as a given or 'necessity' by the sample firms, the focus is now on strengthening relations, network and to build up capabilities inn the dynamics of the globalisation.

References

Aldrich, H.E. (1999): *Organisations evolving*; London: Sage Publications.

Arenius, P, Sasi, V & Gabrielsson, M. (2006): Rapid internationalisation enabled by the Internet: The case of a knowledge intensive company. *Journal International Enterprise*, Vol.3, pp. 279-290.

Ashby, W.R. (1956): *An Introduction to Cybernetics*, Chapman Hall, London, Vol. 2.

Berry, M./Brock, B. (2004): Marketspace and the internationalisation process of the small firm, *Journal of International Entrepreneurship*, Vol.2, Vol.3, pp.187-216.

Borghoff, T (2005): Evolutionary theory of the globalisation of firms, Wiesbaden: Gabler.

Cantwell, JA (2002): Innovation, profits and growth: Penrose and Schumpeter, in Pitelis, C. (Ed.): The theory of the growth of the firm: the legacy of Edith Penrose. Oxford University Press, pp. 215-248.

Chen, SH (2002): Global production networks and information technology: the case of Taiwan, *Industry and Innovation*, 9(3), pp. 249-265.

Doz, YL & Prahalad, CK (1991): Managing DMNCs: A search for a new paradigm, *Strategic Management Journal*, Vol. 12, Special Issue Summer, pp. 145-164

Earl, M & Khan, B (2001): E-commerce is changing the face of IT, *MIT Sloan Management Review,* 42(1), pp. 64-78.**Eisenhardt, KM** (2002): Has strategy changed? *MIT Sloan Management Review*, Vol.43, No.2.

Fillis, A & Wagner, B (2005): E-business development: An explanatory investigation of the small firm, *International Small Business Journal,* Vol.23, p. 604- 618.

Fletcher, M & Plakoyiannaki, E (2008): Case Study Selection: An overview of key issues for international business researchers", *Proceedings of the 34th EIBA conference*, Tallinn, Estonia, 11-13 December

Galbraith, J (1977): *Organization design*, Reading, MA: Addison-Wesley.

Greiner, LE (1972): Evolution and revolution in organizations' growth, *Harvard Business Review*, 50(4), pp. 35-46.

Jaw, YL & Chen, CL (2006): The influence of the Internet in the internationalization of SMEs in Taiwan. *Human Systems Management,* 25(3), pp. 167-183.

Johanson, J & Vahlne, JE (1977): The internationalization process of the firm: A model of knowledge development and increasing foreign market commitments, *Journal of International Business Studies*, 8(1), pp. 23-32.

Knight, G & Cavusgil, ST (2004): Innovation, organizational capabilities, and the born-global firm, *Journal of International Business Studies*, Vol. 35, No. 2, pp. 124-41.

Loane, S (2006): *The role of the Internet in the internationalisation of small and medium-sized companies*, University of Ulster.

Luhmann, N (1995): *Social Systems*, Stanford, CA: Stanford University Press.

McKelvey, B. (1982): *Organizational systematics: Taxonomy, evolution, classification*, Berkeley: University of California Press.

McMahon, P. (2002): *Global control,* Cheltenham, UK: Edward Elgar.

Moen, Ø. (2002): The born globals: a new generation small European exporters, *International Marketing Review*, Vol. 19 No.2, pp.156-75.

Mukherji, A. (2002): The evolution of information systems: Their impact on organizations and structures, *Management Decision*, Vol. 40, No.5, pp.497-507.

Nieto, M.J./Fernández, Z. (2006**):** The role of information technology in corporate strategy of small and medium enterprises, *Journal of International Entrepreneurship*, Vol. 3, pp. 251-262.

Picot, A./Reichwald, R. (1994): Auflösung der Unternehmung?, in: Zeitschrift für Betriebswirtschaft, Vol. 64. No. 5, p. 547-570.

Samiee, S. (1998): Exporting and the Internet: a conceptual perspective, *International Marketing Review*, Vol.15, No.5, pp.413-426.

Vernon, R. (1966): International investment and international trade in the product cycle, *Quarterly Journal of Economics*, Vol. 80, No. 2, pp. 190-207.

Westney, D.E. (1993): Institutionalization theory and the multinational corporation, In: Ghoshal, S./Westney, D.E. (eds.): *Organizational theory and the multinational corporation*, Houndsmill/London: Macmillan, pp. 53-76.

Title: An interdisciplinary perspective on the interplay of information and communication technologies (ICTs) and the globalisation of firms

Dr Thomas Borghoff

School of Marketing and International Management, PO Box 600, Victoria University of Wellington, New Zealand

Email: Thomas.Borghoff@vuw.ac.nz

Abstract

Accelerated globalisation since the 1980s and particularly the 1990s and the development of web-based ICT go hand in hand. Nonetheless, there has been little explicit research on the influence of ICT on the globalisation of firms. Despite a rich literature of the implementation ICT and the design of global information systems in firms, the influence of information and communication technologies (ICT) on the globalisation of firms has not been explicitly researched from a management perspective. This paper serves to provide an overview on existing literature in this field and to develop a basic framework for the study of the influence of ICT on the globalisation of firms. Specifically, the paper reflects on the influence of ICT on the three sub-processes of globalisation: internationalisation, global network building, and global evolutionary dynamics.

The role of ICT in the globalisation of firms

A major driver of globalisation is technological progress. The rise and commercialisation of the Internet and the maturing of ICT are making organisations' business environments increasingly more international, and as a consequence also their communication and business processes (Bicak 2005: 5). ICT encompass the full range of the production, distribution, and consumption of information, across all media from radio and television to satellites and the Internet. The information revolution facilitated the shift from analogue to digital technologies; convergence merges computers, telecommunication, television, and the Internet into a single multimedia environment (Wilson III 1998: 6). The radical development of ICT is an essential factor for the continuing globalisation of organisations' political, social, and economical environments. The most significant factor is the continuous development of the Internet and the WWW as the fundamental infrastructure for e-business (Ibid. 7).

Knowledge and innovation have taken a quantitative jump over the last decade in the wake of the "explosion" of ICT, the globalisation process, and dramatic advances in the life, materials, and energy sciences. These developments have led to new industries and new services, as well as to the renewal of established ones (Aubert/Reiffers 2003: 9). Industry boundaries are easily crossed as value chains are being redefined (Amit/Zott 2001: 495). The knowledge economy develops high-tech industries, particularly in ICT and services (Ibid. 10). The development and diffusion of ICT is a prerequisite and facilitator of globalisation and the transformation into a knowledge-based economy. The most significant advancement in recent times is the emergence of the Internet and the subsequent evolution of electronic commerce (Melewar/Stead 2002: 29).

ICT has supported, facilitated, and often provided the impetus for global business development (Nelson/Clark, Jr. 1994: 19). ICT is both a catalyst of globalisation and a solution base from which to address international main challenges. ICT can provide the strongest link in the business chain of partners, products, and suppliers, and is the basis for doing business around the clock and around the world (Deans/Kane 1992: 1). The network-centred phase we are in since the 1990s induces (1) an increase in the transparency of information on global markets and activities, (2) a decrease in the cost of information, facilitating global activities for an increasing number of firms, and (3) an increase in the speed and volume of communication, both internally and externally, making co-ordination of globally dispersed activities much easier (Samii 2004: 11).

ICTs reduce transaction and co-ordination costs in all forms of organisation, increase productivity, and accelerate the dynamics of innovation (ifo 1999). ICTs affect the cost and efficiency of the external marketplace (Blaine/Roche 2000: 4-6). ICT have the potential to dramatically reduce market imperfections and lowers transaction costs and co-ordination costs (Blaine/Bower 2000: 27). The combination of the evolution of cross-border networking and the increasing use of ICT also has far-reaching implications for the study of industry dynamics as the structures of value-added chains are changing and even boundaries between industries are blurring (Ernst/Kim 2002: 147). ICTs increase boundary spanning (Dewett/Jones 2001: 323). The tremendous advances in ICT are leading to an entirely different type of industrial structure with mutually beneficial co-operations and networking (Roche 2000: 82). In most industries, supply chains become more elastic and flexible but at the same time also more integrated. However, ICT will not eliminate the importance of distance and location, and in fact in some cases makes proximity and clustering even more important (De la Torre/Moxon 2001: 630). Due to the globalisation of local markets and the emergence of the global electronic markets, worldwide acquisitions and cooperation strategies gain importance (Bicak 2005: 14).

Within organisations, electronic technologies are stimulating changes in productivity, management practices, and corporate culture. By enabling instantaneous communication, ICTs allow firms to co-ordinate and control actions in distant locations, thus expanding the potential reach of the firm. They also lower transactions costs and facilitate networking. The Internet provides the possibility of distributed project teams, pooling of expertise worldwide and communicating electronically, rather than being bound to a single physical location (Gable 2006: iii). ICTs increase the information-processing capacity and thus the decision-making capacity. ICTs drive internal and external change and increase environmental complexity. In order to manage high levels of uncertainty, various subunits are driven toward greater differentiation and specialisation. This in turn forces firms to develop strong integrative mechanisms. ICT supports both the standardisation of products and the coordination of business processes across border (Schober 1993: 213). ICT thus can improve efficiency of business processes (Blaine/Bower 2000: 37).

Externally, by linking intranets to the Internet, organisations integrate their internal operations more closely with their vendors, partners, and customers (Bollier 1998: 2-3). ICT can support vertical quasi-integration, outsourcing, and quasi-diversification (all cooperative modes) (Clemons/Row 1992: 12). For example, "virtual consulting" and back office services can now often be provided from lower-cost countries or from the head office. ICTs allow much more cost-effective monitoring of cooperative arrangements. The value of the

network even increases with network size ("increasing returns") (Ibid: 19). Due to the described developments induced by ICT, De la Torre/Moxon (2001:8) state that ICT development leads to the "end of geography", which is marked by a redefinition of corporate boundaries and the development of flexible network structures (Borghoff 2005). There is also an increasing emergence of "born globals" (Melewar/Stead 2002). There is a progressive transformation of business into relations of information exchange, leading to globalisation and network building. Increasing globalisation and the growth and spread of ICT will continue to dominate the world economic scene for many years and their importance will grow as they are driving each other (Samii/Karush 2004).

From an international management perspective, there is a paucity of directly relevant empirical research, which illuminates how MNEs develop and manage their ICT capabilities in their different and complex circumstances (Roche 2000: 137). A fundamental gap in the research on global ICT is the static character of concepts and empirical studies. Del Águila et al. (2002: 32) hence remark that it is necessary to introduce a greater dynamic component in the analysis of ICT in a global environment both by using dynamic theories and by applying techniques of longitudinal empirical research. A significant gap thus exists in the research of the process dimension in the globalisation of firms. ICT have a large potential to facilitate the development of globalisation capabilities. The influence of ICT on the development of new international activities, their global coordination, and the adaptation to global competitive processes has been explored only rudimentary and will thus be the main focus of this paper.

The influence of ICT on the globalisation of firms

While internationalisation theories illuminate the development of firms from a national to an international level, they generally neglect the network building process, which is a central characteristic of globally operating firms and an evolutionary driver of this process. The internationalisation process is basically described as a life cycle (Vernon 1966), an incremental, stage-based process (e.g., Luostarinen 1980), a discontinuous process (Kutschker 1996, and the emergence of "born globals" (Knight/Cavusgil 1996) or "international new ventures" (Oviatt/McDougall 1994). ICT can help to gain competitive advantage and to re-engineer business processes but few researchers have attempted to move this research to a global context (Sakaguchi/Dibrell 1998). However, the very dynamic of ICT works against it being a source of unique, competitive advantage for any single company and provide competitive advantages only when combined with other organisational resources.(Manheim 1990: 147). Until the early 1990s, most research on the subject of ICT in management stopped short of looking at impact measures and was often limited to addressing the question of 'fit' (Jarvenpaa/Ives 1993). Then, a stream of research focused on the analysis of the correlation between economic performance and ICT investment (Brynjolfsson/Hitt 1996). Interdependencies between ICT and other organisational variables are difficult to prove due to problems in their identification, causal relations, and complexity. The missing positive influence of ICT on productivity is called the 'productivity paradox of IT'. Despite the difficulty of testing the direct effect of ICT on performance measures such as profitability, a study by Whitworth et al. (2005) provided evidence that ICT particularly facilitate the development of global activity structures. ICTs are essential ingredients for business expansion, providing strategic competitive advantage in worldwide markets (Ives/Jarvenpaa 1993) and facilitating globalisation (Palvia 1997). Therefore, these studies suggest a fit of strategy and ICT in order gain competitive advantages. Other studies extent this contingency perspective to the organisational structure (e.g., Del Águila et al. 2002). There are effects on an on-going basis and Rindova/Kotha (2001) even term this on-going strategic and organisational change "continuous morphing".

The influence of ICT on globalisation capabilities

Globalisation processes are constituted by three subprocesses: (1) internationalisation (changes in the level and dispersion of activities in different national markets) (2) global networking: (development of internal and external network structures in the global context), and (3) evolutionary dynamics as firms are in a co-evolutionary process with their environment. Firms develop respective characteristics and capabilities in their globalisation, which reflect these subprocesses (Borghoff 2005). ICT have a significant influence on the development and application of the three globalisation capabilities.

The Internationalisation of a social system can be conceived as a trajectory of changes in its geographical or cultural extension. Internationalisation literally induces changes in the system's boundaries and its relationship with its environment. In the case of expansion, environmental complexity increases, inducing an increase of the system's internal complexity and requisite variety as well. The use of ITC has enabled many companies to expand their international presence and international trading capabilities (Collins 2004: 67). The use of the Internet tends to expand the geographic market, bringing many more companies into competition with one another (Porter 2003: 381). The Internet enables firms to identify new market opportunities leading to business expansion. Specifically, it allows SMEs to gain deeper knowledge of target markets, to select suppliers and to establish direct contact with clients using a low cost medium. The Internet makes it easier for firms to expand internationally (Nieto/Fernández 2006: 254). Similarly, the Internet reduces the entry barriers to international markets, which in turn encourages the firm's international expansion and minimises the importance of the local market (Ibid: 252). ICT-intensive firms internationalise faster and more extensively than less ICT-intensive firms (Ibid. 96).

In terms of global networking, ICT brings extended connectivity with speed and will expand boundaries of firms and networks (Samii 2004: 15). ICT foster external alliances through interorganisational information systems for information partnerships (Earl/Feeny 1996: 79). The Internet can be used to enhance information flow and collection, as well as co-ordination among firms, a necessary tool for international expansions. Co-operations provide useful information and reduce the perceived risk of internationalisation significantly (Nieto/Fernández 2006: 254). ICT is more than a transaction facilitator and is promoted as an enabling technology for collaborative commerce amongst firms, involving not only interorganisational coordination of the supply chain but also cooperation in product definition, design, and R&D (Chen 2002: 253). ICT also increases the innovativeness of firms (Blaine/Roche 2000: 8). Therefore, Zaheer/Manrahakhan (2001) assume that the importance of coordination skills is a defining competitive competence for the Internet age.

ICT specifically has an impact on the on the global evolutionary capability of firms. ICT has the capacity to enable dramatic organisational transformation (Boudreau et al.

1998: 123). Typically, advanced ICTs play a central role in virtual and learning organisations because technology permits organisational designs to overcome the spatial and temporal dispersion that accompanies increased global reach (Ibid.: 120). ICT enhances the ability to combine distant learning processes in formerly separate activities. Subsidiary networks are increasingly used to source new technology. Global learning has become an important mechanism for corporate technological renewal within MNEs (Cantwell 2002: 238). MNEs have recently shifted to a closely integrated network of subsidiaries designed to facilitate complementary paths of innovation and new competence creation (Ibid. 244). On balance, ICT strengthens the evolutionary mechanisms in social systems and thus their capability to change and transform (Van de Ven/Poole 1995, Borghoff 2005).

Conclusions

There is a clear gap in the research of ICT in the globalisation of firms. A rich fund of literature exists on the technical side of ICT and information systems in firms. From a management perspective, there is almost no explicit research on the influence of ICT on the globalisation of firms. This influence is only implicitly included in terms of a better information processing capacity, time zone economies, or the contribution to innovation and knowledge management. There are assumptions that ICT facilitate a faster internationalisation and the emergence of "born globals". Future research could provide some transparency in this field.

References:

Amit, R./Zott, C. (2001): Value creation in e-business, Strategic Management Journal, Vol. 22, pp. 493-520

Aubert, J.-E./Reiffers, J.-L. (2003): *Knowledge Economies in the Middle East and North Africa: Toward new development strategies*. World Bank, Washington.

Bicak, K. (2005): *International knowledge transfer management: concepts and solutions for facilitating knowledge transfer processes in multilingual and multicultural business environment*, Herzogenrath: Shaker.pp. 5,7.

Blaine, M.J./Roche, E.M. (2000): Introduction, in: Roche, E.M./Blaine, M.J. (Eds.): *Information technology in multinational enterprises*, Cheltenham, UK: Edward Elgar, pp. 3-18.

Blaine, M.J./Bower, J. (2000): The role of IT in international business research, in: Roche, E.M./Blaine, M.J. (Eds.): *Information technology in multinational enterprises*, Cheltenham, UK: Edward Elgar, pp. 21-56.

Bollier, D. (1998*): The advance of electronic commerce*, Washington, DC: The Aspen Institute, pp.2-3

Borghoff, T. (2005): *Evolutionary theory of the globalisation of firms*, Wiesbaden: Gabler.

Boudreau, M.-C./Loch, K.D./Robey, D./Straud, D. (1998): Going global: Using information technology to advance the competitiveness of the virtual transnational organization, *Academy of Management Executive*, Vol. 12, No. 4, pp. 120-128.

Brynjolfsson, E./Hitt, L. (1996): Paradox lost? Firm-level evidence on the returns to information sstems spending, *Management Science*, Vol. 42, No. 4, pp. 541-558.

Cantwell, J. (2002): *Innovation, profits and growth*: *Penrose and Schumpeter*; in: Pitelis, C. (ed.): *The growth of the firm: The legacy of Edith Penrose*, Oxford: Oxford University Press, pp. 215-248

Clemons, E.K./Row, M.C. (1992): Information technology and industrial cooperation: the changing economics of co-ordination and ownership, Journal of Management Information Systems, Vol. 9, No. 2, pp. 9-28

Collins, J.S. (2004): IT infrastructure and global operations, in: Samii, M./Karush, G. (Eds.): *International business and information technology*, New York: Routledge, pp. 67-82.

Deans, P.C./Kane, M.J. (1992): *Information systems and technology*, Boston: PWS-Kent.

Del Águila, A.R./Bruque, S./Padilla, A. (2002): Global information technology management and organisational analysis: Research issues, *Global Information Technology Management*, Vol. 5, No. 4, pp. 18-37.

De la Torre, J./Moxon, R.W. (2001): Introduction to the symposium e-commerce and global business: The impact of the information and communication technology revolution on the conduct of international business, JIBS, Vol. 32, No. 4, pp. 617-639.

Earl, M.J./Feeny, D.F. (1996): Information systems in global business: Evidence from European multinationals. In: Thomas, N./O'Neal, D./Kelly, J. (eds.), *Strategic renaissance and business transformation*. John Wiley and Sons, Chichester, pp.183-210.

Ernst, D./Kim, L. (2002): Introduction: Global production networks, information technology and knowledge diffusion, Industry and Innovati0n, Vol. 9, No. 3, pp. 147-153.

Gable, G. (2006): The Internet, globalization , and IT professional services, Journal of Global Information Management, Vol. 14, No. 2, pp. i-vi

Ifo Institut für Wirtschaftsforschung (1999): *ifo Studien zur Strukturforschung : Tertiarisierung und neue Informations- und Kommunikationstechnologien*:, München: ifo Institut für Wirtschaftsforschung.

Ives, B./Jarvenpaa, S.L./Mason, R.O. (1993): Global business drivers: Aligning information technology to global business strategy, *IBM Systems Journal*, Vol. 32, No. 1, pp. 143-161.

Jarvenpaa, S.L./Ives, B. (1993): Organizing for global competition: The fit of information technology, *Decision Sciences*, Vol. 24, No. 3, pp. 547-580.

Knight, G.A./Cavusgil, S.T. (1996): The born global firm: A challenge to traditional internationalization theory, *Advances in International Marketing*, Vol. 8, pp. 11-26

Kutschker, M. (1996): Evolution, Episoden und Epochen: Die Führung von Internationalisierungsprozessen; in: Engelhard, J. (Ed.): *Strategische Führung internationaler Unternehmen*, Wiesbaden: Gabler, pp. 1-38.

Luostarinen, R. (1980): *Internationalization of the firm*, Helsinki: Acta Academicae Oeconomicae Helsingiensis.

Melewar, T.C./Stead, C. (2002): The impact of information technology on global marketing strategies, Journal of general management, Vol. 27, No. 4, pp. 29-40.

Nelson, K.G./Clark, J. (1994): Cross-cultural issues in information systems research: a research program, *Journal of Global Information Management*, Vol.2, No.4, pp. 19-29.

Nieto, M.J./Fernández, Z. (2006): The role of information technology in corporate strategy of small and medium enterprises, Journal of International Entrepreneurship, Vol. 3, pp. 251-262.

Oviatt, B.M./McDougall, P.P. (1994): Toward a theory of international new ventures, *Journal of International Business Studies*, Vol. 25, No. 1, pp. 45-64.

Palvia, P.C. (1995): Global management support systems: a new frontier. In*: Journal of Global Information*, Vol., No. 1, pp 3-4

Palvia, P.C. (1997): Developing a model of the global and strategic impact of information technology, *Information & Management*, Vol. 32, pp. 229-244.

Palvia, P.C. (1998): "Global Information Technology Research: Past, Present and Future." *Journal of Global Information Technology Management*. Vol 1, No 2, pp. 3-14.

Porter, M.E. (2003): The strategic potential of the Internet, in: Galliers, R.D./Leidner, D.E./Baker, B. (Ed.): *Strategic information management*, 3. Aufl., Oxford: Elsevier Science, pp. 376-403.

Roche, E.M. (1994): Finding application families and application fragments in global systems, in: Deans, P.C./Karwan, K.R. (Eds.): *Global information systems and technology: Focus on the organization and its functional areas*, Harrisburg: Idea Group, pp. 540-557.

Roche, E.M. (2000): Information technology and the multinational enterprise, in: Roche, E.M./Blaine, M.J. (Hrsg.): *Information technology in multinational enterprises*, Cheltenham, UK: Edward Elgar, S. 57-89.

Sakaguchi, T./Dibrell, C.C. (1998): Measurement of the intensity of global information technology usage: Quantitizing the value of a firm's information technology, *Industrial Management & Data Systems*, No. 8, pp. 380-394.

Samii, M. (2004): Globalization and IT, in: Samii, M./Karush, G. (Hrsg.): *International business and information technology*, New York: Routledge, pp. 9-20.

Samii, M./Karush, G. (2004: International business and information technology, in: Samii, M./Karush, G. (Hrsg.): *International business and information technology*, New York: Routledge, pp. 1-8.

Schober, F. (1993): The strategic role of information and communication technology for international business co-ordination, in: Matsugi, T./Oberheuser, A./Schober, F. (Eds.): *Integration and adjustment of global economies*, Berlin: Duncker & Humblot, pp. 213-227.

Vernon, R. (1966): International investment and international trade in the product cycle, *Quarterly Journal of Economics*, Vol. 80, No. 2, pp. 190-207..

Whitworth, J.E./Palvia, P.C./Williams, S.R./Aasheim, C. (2005): Measuring the impact of global information technology applications, *International Journal of Technology Management*, Vol. 29, No. 3/4, pp. 280-294.

Wilson, E.J. III (1998): Globalization, information technology, and conflict in the second and third worlds, Working Paper, Rockefeller Brothers Fund, New York

Zaheer, S./Manrakhan, S. (2001): Concentration and dispersion in global industries: Remote electronic access and the location of economic activities, *Journal of International Business Studies*, Vol. 32, No. 4, pp. 667-686

MEERKAT PROJECT
Institutionalization of Integrated Examination Policy

Sebastião Helvecio Ramos de CASTRO
Councilor of Minas Gerais Court of Auditor
Minas Gerais Court of Auditors.
Belo Horizonte, Minas Gerais- 30380-435, Brasil

Marília Goncalves de CARVALHO
Worker at Sebastião Helvecio Office
Minas Gerais Court of Auditors.
Belo Horizonte, Minas Gerais- 30380-435, Brasil

ABSTRACT

This paper intends to discuss the increase of knowledge management and technological resources in Brazil, showing the concern of Minas Gerais Court of Auditors (TCEMG), agency of external control that manages public resources of Minas Gerais States, with the subject, through the description of the Meerkat Project, a public policy that aims to improve the use of integration solution, helping TCEMG fulfill its constitutional mission. It is known that TCEMG has several computer systems, however, they don't communicate/integrate with each other. The idea of the Project is to build integration solutions to cross and check information from different systems and sources, being able to identify and to action at illegal action or misuse of public resource in real time. Always looking toward new technologies, systems and tools, TCEMG along with Information Technology Department, aims to become reference in the activity of external control as well as in the transformation of the current external control paradigm.
Keywords: TCEMG, Information Technology, Knowledge Management, Integration Tools, External Control.

INTRODUCTION

Minas Gerais Court of Auditors (TCEMG) has taken actions to expand the culture of strategic management within its operations. The approval of the Strategic Plan 2010-2014 (TCEMG, 2010) gives the institution the mission to perform external control of public resources effectively, efficiently and effectively in benefit of society, in order to become a reference in ensuring the right of the society to the regulate and effective management of public resources, based on the values of ethics, justice, effectiveness, transparency and social commitment.

International auditing standards, such as COSO and INTOSAI, emphasize the better management of risks in the development of control actions, which is consistent with the objective of this Project to give greater consistency to the selectivity procedures and to the planning monitoring activities.

Modern society is impacted by the development of new technologies and the improvement of knowledge management within the organizations. TCEMG is no stranger to this phenomenon and plans to develop the Integrated Examination Policy, a public policy that aims to improve the use of integration tools and the knowledge management, organizing and crossing data and information available internally and externally, looking toward to ensuring all citizens a fair and proper use of public resources.

THEORETICAL REFERENCE

Contemporary society has experienced the development and fast expansion of computers and communications processing capacity. These change impacts the economy, politics, academic field, culture and consequently, how the general population sees the government.

The new perspective has forced countries and international organizations to develop programs and initiatives to dominate and / or to democratize the process of information. In Brazil, there is a national project coordinated by the Brazilian Institute of Information Science and Technology (IBICT), giving rise to the Information Society Program, launched by the Federal Government in 2000.

According to Fayard (2000), information is power if properly used, however if stored, there is not value - an idea that originates a discussion of trends in knowledge management, which predispose the need for flexibility in the pursuit of collective growth.

Baran (1997) represents the relationship between data, information, knowledge and wisdom through the following scheme:

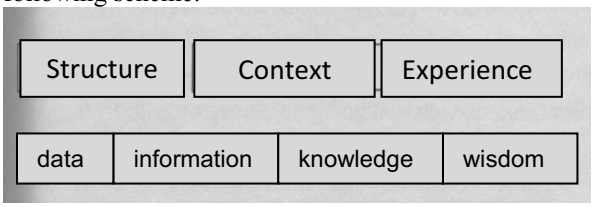


Figure 1: Baran (1997)

Though we live in the called information society, the real asset is not the information, but the knowledge which is information edited, putted into a context and analyzed in order to make sense and have value to the organization.

The Executive Committee of the Electronic Government uses as definition for knowledge management:

(...) a set of systematic processes, articulated and intentional, capable of enhancing the ability of public managers to create, collect, organize, transfer and share information and knowledge that can be used to strategic decision made by the public management and for inclusion of the citizen as a producer of collective knowledge.

The technology resources facilitate the networking and keep knowledge where it is generated and / or used (Davenport et al, 1998) and improve the interactivity of the knowledge with its users (Davenport & Prusak, 1998), and can actually be useful for knowledge management, if employed in a systematic interference / human interactivity (Davenport et al, 2001).

According to Batista (2004), some Brazilian public companies as SERPRO (Federal Service of Data Processing), Central Bank of Brazil, Bank of Brazil, Caixa Economica Federal, EMBRAPA (Brazilian Agricultural Research Corporation) and PETROBRAS (Brazilian Oil S/A) have already implemented these new model of management ,incorporating management knowledge and new technologies to its operation.

The challenge is to rethink public organization, directing it towards the knowledge and adopting new business models. According to Giacomini (2001), it is necessary, in the functional structure, a working group dedicated to: development and dissemination of new technology and its transformation into applied knowledge.

INTEGRATED EXAMINATION POLICY

The integrated examination policy was born in Sebastião Helvecio's office and accepted for the Court. The house voted and passed resolution regarding to the Project:

1) Resolution n.06 of 05/04/2011: legalization of the policy;

2) Resolution n.10 of 07/05/20011: legalization of the actions made by the policy;

3) Resolution n. 10 of 07/05/2001; legalization of the members of the new department.

Also, a Decree of the Presidency number 82 of 05/18/2011 establishes the Project as a priority for the year of 2011.

The central question of this Public Policy is to propose solutions for the low use of tools and integration technologies in the actions of external control in order to ensuring all citizens a fair and proper use of public resources, and to respond to demands and current offerings in relation to new information technologies.

To characterize the situation, we start from the observation that TCEMG has several computer systems that collect and store data and information and also has agreements with institutions whose goal is to obtain data of interest for the external control, however, it turns out that such knowledge potential has not been explored.

Among 193 nations included in the UN, only 24 are federations and Brazil is one of them. The Brazilian Constitution in its article 18 ensures: "The political and administrative organization of the Federative Republic of Brazil comprises the Union, States, Federal District and Municipalities, all autonomous under this Constitution". Thus, TCEMG has the mission to monitor the state of Minas Gerais and its municipalities without distinction, judging all public managers.

Minas Gerias state is one of the 27 federative units of Brazil and it is the fourth state with bigger territorial extension – 586.523 Km^2, equivalent to the France territorial extension. TCEMG is responsible for analyzing and controlling the revenue and expenditure of 853 municipalities. In total, TCEMG has the obligation to monitor and examine 2.292 public agencies, totalizing, in 2010, the amount of 76 billions of reais; approximately 47 billions of dollars. Below it shows the growth of the budget that TCEMG has to monitor:

Billon(R$)

Year	State	Municipalities
2008	35,60	27,64
2009	39,97	32,46
2010	41,11	35,54

Figure 2: Minas Gerais budget

The numbers show the importance of a proactive, constant and an efficient monitoring, what is the objective of this Project. To reach it, it's vital the use of technological resources and the development of tools that enable the analysis of the information in a way that it can serve as a basis for decision-making.

The main tool used in the Project is the use of integration solution to cross and check information from different sources. The idea is to compare, to analyze data and information to ensure its accuracy and, within the external

control activity, to ensure the proper implementation of public resources.

According to Carlos Nogueira (2009), crosschecking is currently used by various government institutions to monitor and curb the mismanagement of public resources, having always in mind that the idea is to confront information available internally with other information available.

Caiçara Junior (2006) says that there are numerous problems that arise in the scenarios of organizations because of the lack of integration of their systems and databases that do not communicate. Aspects such as rework, redundancy of data and lack of completeness of the information occurs as a consequence of this lack of integration.

First of all, the idea is to build integration tools that help TCEMG to monitor closely areas that involve a large amount of money such as public purchase and public construction. Nowadays all public agencies inform about their spending, but TCEMG does not check if that information is real, unless it goes in locus or check paper work; what demands time, work and money. With the integration solutions, information will be compared and checked and some illegal actions or misuse of public resource will be easily detected without waste of time and in real time.

As an example, in 2010, Minas Gerais state and its municipalities spent the total of R$ 2.794.379.585,48, approximately U$ 1.746.487.240,93 with public purchase.

The use of technological resources and the application of new systems and technological knowledge are essential for dealing with such important subject. In this way, TCEMG has invested in technological resources and training for the implementation and the success of the Meerkat Project.

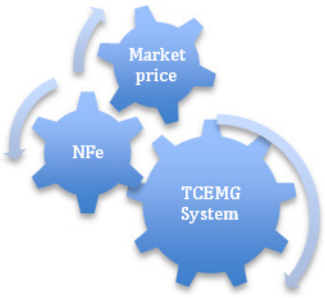
INTEGRATION SOLUTION

According to Alter (1996) "Information Techonology is the set of hardware and software that enable operation of information systems." According to Davenport (1996), since it entered at business environment, computers hooked up closely to the way that work is done. Information Technology has change radically the work - its location, speed, quality and other key features. To choose and to implement properly the best technologies within the organizational context, supporting their strategies, is a challenged activity for managers.

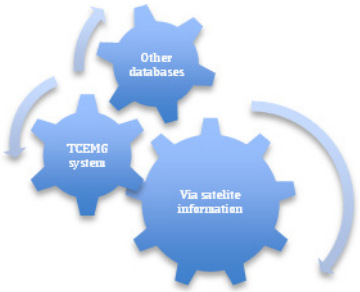
With the emergence of new technologies, the increase of legal requirements, the pressure from society and the transparency of public resources, Information Technology has become an important tool. Thus, the resources invested in that sector by public sector in Brazil have grown significantly (Cunha 2004).

INTEGRATION TOOLS

1) Public Purchase
A challenging issue and it will be the first target. The idea is build an integration solution crossing information from different databases as showed below:



2) Public Construction:
Another challenge faced by TCEMG is to monitore public construction due to Minas Gerais sizes. So, the idea is use technology resource to make it possible.



Important information as overpriced, failure mode of bidding, cartelization and others will be easily detected using integration solution. But the real innovation is the creation of a database of price per areas in Minas Gerais state. The idea is set up a mapping of regional costs, an important database when monitoring government actions. TCEMG will gain agility as to receive the information as well as to analyze and decided about it.

The pictures below illustrate the main idea of the Project:

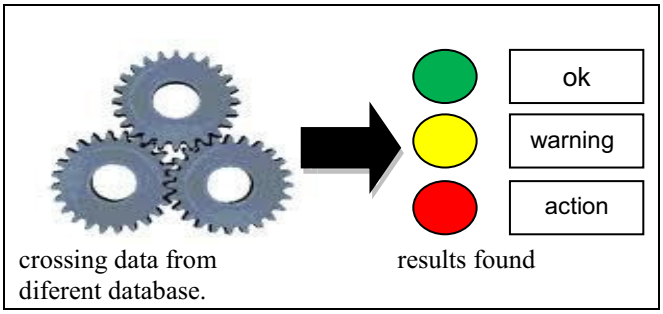


Figure 3: Integration toll

Using new technologies, systems and tools, and also building a trustful database, the Project aims to bring a new perspective on the role of controlling public resource.

Such important subject as public resource and the fair use of it in the benefit of society deserves a careful attention.

CONCLUSION

TCEMG is always looking forward news technology that could help it fulfilling its obligation with ethics, justice, effectiveness, transparency and social commitment. Brazilian society has the desire for better public services, good governance and honesty in dealing with public affairs. In this context, TCEMG has an important role in monitoring public resource, ensuring an effective and regular application of public money always in benefit of the society.

Because of the key role that the use of information technology currently takes in the institution, TCEMG has developed a strong use of computer technology to make its operations more agile. The Meerkat Project is based on technology resource being the main tool the use of integration solution to crosschecking information from different sources. The idea is to compare, to analyze data to ensure its accuracy and, within the external control exercised by the Court, to ensure the proper implementation of public resources.

The Audit Courts are gaining importance in the context of Public Administration, and the effectiveness of their actions are vital for the democracy and citizenship actions, and its performance cannot ignore the considerable changes at knowledge management and technology information fields.

We understand that the actions proposed by the Meerkat Project can contribute for the improvement of the activity of external control, as well as the transformation of the current control paradigm.

REFERENCES

Batista, Fabio and others. Knowledge Management in Public Administration. **The work resulting from the search: Government that learns - Knowledge Management in the public sector.** Brasilia: IPEA - Institute of Applied Economic Research, June 2005.

Baran, U. **Helping retailers generate customer relationships. I**CL System Journal, v. 11, n. 2, jan. 1997.

Caicara Junior, Cícero. **Integrated Management System - ERP: a management approach**. Curitiba: Ibpex, 2006.

Camatti, Tassiara Baldissera; Fachinelle, Ana Cristina. **Communication as a strategic differential in the knowledge management of organizations**. Connection, Communication and Culture. Caxias do Sul, UCS, v. 9, No 17, jan. / jun. 2010.

Costa, Marília, Damiani and others. **Information Management or Knowledge Management?** R. ACB. Santa Catarina: Santa Catarina in Librarianship, v. 5, No 5, 2000.

Cunha, M. A., Marques, E. V. & Meirelles, F. S. **Models of Information Technology Management in Brazilian public sector**. Salvador: I ENAPG 2004. Proceedings of the Event, September 2004.

Davenport, T. H. **Managing customer support knowledge.** California Management Review. v. 40, n. 3, p. 195-208, Spring 1998.

Davenport, T .H; PRUSAK, L. **Knowledge enterprise**. Rio de Janeiro: Campus, 1998.

Davenporte, T. H. et al. **Data to knowledge to results: building an analytic capability.** *California Management Review*, v. 43, n. 2, p. 117-138, Winter 2001.

Fayard, P. M. **The game of interaction: communication and information strategy**. Caxias do Sul: Educs, 2000.

Giacomini, C. H. and others. **The quality of life of municipals employee and the public productivity in Curitiba**. RAP. Rio de Janeiro, v. 35, No 6, Nov. / Dec. 2001.

Loureiro, Maria Rita and others. **Democratization and State Reform: the institutional development of the Audit Court in Brazil**. Journal of Public Administration. Rio de Janeiro, 43 (4): 739-72/ago.2009.

Prates, Cristiana de Lemos Souza; Rocha, Heloisa Helena Nascimento; Carvalho, Jaqueline Grossi Fernandes; Pinheiro, Valquíria de Sousa – **Research on the macro trends of external control. Minas Gerais Court of Auditors**. Belo Horizonte, in April 2009.

Schlesinger, Cristina C. B. and others. **Knowledge Management in Public Administration.** Curitiba: Municipal Institute of Public Administration - IMAP, 2008.

Silva, Sergio L. **Knowledge management: a critical review-driven approach to knowledge creation**. Information Science. Brasília, v.33, n. 2, May / August 2004.

Stewart, Thomas A. **The wealth of knowledge: Intellectual capital and the new organization**. Rio de Janeiro: Campus, 2002.

Minas Gerais Court os Auditors. **Strategic Plan 2010-2014**. Belo Horizonte, February 10, 2010.

Vidal, Patrícia G and others. **Knowledge management: two unique cases**. Electronic Journal of Administrative Science. Cenecista Faculty of Campo Largo, v. 5, No 1, May 2006.

E-Government as a Vehicle for Promoting and Improving Governmental Performances with Yardstick Competition Model

Yasuyuki Nishigaki
Faculty of Economics, Ryukoku University
67 Fukakusa-tsukamotocyo, Fushimiku,
Kyoto, 612-8577 Japan

Yuzo Higashi
Graduate School of Economics, University of Hyogo
1-3-3 Higashikawasaki, cyuo-ku,
Kobe, 650-0044 Japan

Wong Meng Seng
University of Nottingham Malaysia
Jalan Broga 43500 Semenyin
Selangor Darul Ehsan, Malaysia

and

Hideki Nishimoto
Faculty of Economics, Ryukoku University
67 Fukakusa-tsukamotocyo, Fushimiku,
Kyoto, 612-8577 Japan

ABSTRACT

In this paper, we investigate the empirical importance of providing policy information and understanding residents' needs to public goods by local government and introducing new elements of evaluating e-government and performance of local governments. As Besley and Case indicated empirically by using the U.S. data, local tax level and the residents' evaluation of their government through their voting behavior are closely connected by the nexus of yardstick competition. We examine, first, the yardstick relationship between local government activity and the residents' evaluations of their governments through voting behaviors using Japanese data. Then, we introduce "penetration rate of e-government" which we can use as a new index of e-government evaluation. By utilizing this index, we implement a new empirical study concerning the efficiency and the optimality of the local government performance under yardstick competition.

Keywords: E-government evaluation, Yardstick competition model, Public goods strategy, Policy evaluation, Social decision making

1. INTRODUCTION

In today's world, Information and Communication Technology (ICT) has come to be recognized as one of the drivers in promoting economic growth. ICT has been playing an important role in Japanese economic growth, and according to an e-government report by UNPAN (2010), Japan is considered as one of the South East Asia countries that has a high level of fixed line, mobile phones and Internet usage. Such world first class infrastructures have facilitated government providers to implement e-government services for the use of their citizen. Since 2000, Japan central and local governments have been actively promoting e-government projects by introducing new IT systems all over the country, and this is seen as part of the performance improving scheme of public sector. Nowadays, almost all levels of governments, except a few cases, are providing e-government services to the public in Japan. However, according to an e-government survey conducted by UNPAN (2010), it reveals that although the e-participantion index of Japan is 0.7571 which is ranked top 6 in the world, Japan's development index of e-government remains in relatively low score (0.7152 and ranked seventeenth). Therefore, in

order to achieve a higher e-government development index, Japanese governments needs to evaluate the performance of e-government and to identify factors that will contribute to such achievement. Unfortunately, there are not many e-government evaluation research has been conducted in Japan and hence there is a need to address this research gap (Wong et al., 2011).

However, at international level, we have identified quite a number of e-government evaluation (or benchmarking) research, for examples, studies conducted by Brown University (West, 2007), United Nation of Public Administration Network (UNPAN, 2010), Taylor Nelson Research (Dexter and Parr, 2003), and Accenture (Cole and Jupp, 2005). Bannister (2007) argues that some of the existing e-government benchmarking researches could be biased and unreliable for evaluating e-government progress because the result of such activity can favor the country that initiated the benchmarking research. In addition, some government might distort government policies to compete for a higher e-government rank at international level, neglecting the actual needs of their citizen at local and national level. Our research addresses this problem by looking at the importance of e-government evaluation at local level.

We indicate, first, that there is an economic and political significance of providing policy information and understanding of residents' needs concerning public services by governments and then we investigate, utilizing Japanese data, the empirical importance of the role of e-government in offering policy information and understanding residential needs of public services.

In many advanced countries, more than 60 % of public goods are provided by the local or regional governments. Tiebout (1956), in his pioneering work, indicated that **"**voting with feet" leads to optimal provision of local public goods, and this is clearly shown when residents emigrate from one local government to another in order to maximize utility. Since then, Tiebout hypothesis has been used as a theoretical base for decentralizing responsibilities in the provision of public goods, although it requires extreme or unrealistic prior conditions including "free mobility".

Seabright (1996), on the other hand, constructed a "yardstick competition" model of incomplete contracts under asymmetric information by introducing election at local government, and indicated that voting behaviors of residents ensure the ultimate efforts of local government to provide local public goods, if residents take a vote after evaluating their public goods in comparison with that of neighbor jurisdictions. Inter-governmental competition results in equilibrium in the same way as in yardstick competition among public utility enterprises (Konishi, 2009). Therefore, yardstick competition can be expected to depict a more realistic inter-governmental competition. However, because individual choice of private and public goods is excluded in their model, the efficiency of the yardstick equilibrium has not been discussed.

In our recent research, (Nishigaki, Higashi, and Nishimoto, 2011), we introduced residents' consumption choice and tax into the yardstick competition model, and examined the efficiency of local public goods provision under yardstick competition. In the paper, we obtained the following results. First, if we ignore residents' consumption choice, local governments tend to over-supply local public goods, since local government heads attach more importance to re-election. Second, in order to improve the efficiency of the yardstick equilibrium, local governments need to supply local public goods after considering regional disparities regarding residents' preferences, exogenous environmental conditions, and other factors. That is, policies for diminishing asymmetric information between the local government and residents are effective in improving efficiency. Our results suggest that views of enhancing the effectiveness and optimizing the public services give a new index to e-government evaluations which are also led to providing new elements to governmental performance evaluations.

2. YARDSTICK COMPETITION MODEL AND EFFICIENCY OF LOCAL PUBLIC GOODS PROVISION

Consider a simplified nation that comprises two symmetrical regions where a total of N identical immobile households reside and n_i ($i = 1,2$) out of N reside in region i. We assume that each region has identical land and production technology.

Households living in region i derive utility from the consumption of private goods x_i and the public goods g_i supplied in region i[1]. The residential utility is also affected by unobserved locality-specific shocks ε_i. Thus, the residential utility is represented as below:

$$U_i = u(x_i, g_i) + \varepsilon_i, \quad i = 1,2. \tag{1}$$

where ε_i is the noise that is independently drawn from a continuous density function $D(\varepsilon)$ with zero mean. We assume that $D(\varepsilon)$ has an identical distribution between regions.

Eq. (1) means that the residential utility is affected not only by the consumption of the private goods and public goods supplied in the region but also by locality-specific shocks ε_i. Here, Random noise ε_i is considered as a distinctive natural environment or economic environment.

By these assumptions regarding the constructions of the utility function, we assume the following asymmetric information structure: the values of g_i chosen by the local government are not directly observable by the

residents and remain the private information of the governments and the utility of the residents, while observable by both the residents and local government, is not verifiable. This means that the local governments do not know their residents' true public goods preferences, and the residents are unaware of the actual level of local public goods provided by their local government; that is, they cannot determine the exact efforts of their agent.

The residents supply 1 unit of labor per capita to a regional firm and gain a fixed wage w_i, which is the sole income earned by them. Suppose that their local government levies a lump-sum tax t_i on the residents within the region, and the residents spend all the income left over after deducting t_i on private goods. The residents' budget constraint, therefore, is indicated through the following equation:

$$x_i = w_i - t_i, \quad i = 1,2 . \tag{2}$$

The local government in region i supplies local public goods g_i that benefit only residents of region i. As g_i is subject to the lump-sum tax t_i, the local government's budget constraint is indicated as below:

$$g_i = t_i n_i \ , \ i = 1,2 \ . \tag{3}$$

In order to investigate the efficiency of yardstick competition, we introduce individual choice regarding private and public goods into the conventional yardstick competition model. By doing so, we attempt to clarify whether or not our new yardstick competition model can achieve the social optimal level of the local public goods.

The residents' utility is decided in the next procedures. To begin with, the local government i chooses the supply level of local public goods g_i. When g_i is decided, the lump-sum tax t_i is also decided. The residents decide their level of private goods consumption x_i in accordance with their budget constraints. Next, the noise terms ε_1 and ε_2 are independently drawn from the density function $D(\varepsilon)$ with zero mean, after which residents' utility is decided.

In this model, we assume that the residents know the utility level of the rival region's residents. They can compare their own utility with that of the rival region's residents and re-elect their incumbent government if own utility level surpasses at least C. That is, residents re-elect the local government in their own region when their own utility level surpasses at least C. The condition for the residents to re-elect the local government in their own region is

$$u(x_i, g_i) + \varepsilon_i \geq C , \tag{4}$$

where C, could be interpreted as the utility level of the residents in the rival region or the residents' past utility level gained from the local government in their own region.

Although the condition for re-election (Eq. (4)) is very similar to that in the conventional yardstick model, the utility function in this model is not a simple increasing function of local public goods. Since an increase in the public goods supply means a decrease in the private goods consumption in accordance with the residents' budget constraints, the yard stick equilibrium has more interesting and intricate elements.

Residents are not able to observe their local public goods supply. While the re-election of a local government depends upon its residents' utility, its re-election does not directly depend on the local public goods supply. As noted above, this fact indicates that local governments are unaware of the true preferences of their residents; additionally, the residents are unaware of the actual local public goods supply. This fact, in turn, causes asymmetricity of information between the local government and the residents in yardstick competition model.

The re-election rent for the incumbent government is R here as well, and the local government's utility associated with the local public goods supply g_i is $\upsilon(g_i)$, where $\upsilon(g_i)$ is again assume to be decreasing and convex function with respect to g_i, that is, we assume $\upsilon'(g_i) < 0$, $\upsilon''(g_i) > 0$.

In the abovementioned yardstick competition model into which the factor of individual choice of private and public goods has been incorporated, the local governmental problem is formulated as follows.

$$\max_{\{g_i, t_i\}} \ E[\upsilon(g_i) + R] = \upsilon(g_i) + R \cdot pr[C - u(x_i, g_i) \leq \varepsilon_i]$$
$$s.t. \quad U_i = u(x_i, g_i) + \varepsilon_i ,$$
$$x_i = w_i - t_i , \ i = 1,2 ,$$

where, by definition of the distribution function, $pr[C - u(x_i, g_i) \leq \varepsilon_i]$ is represented by

$$pr[C - u(x_i, g_i) \leq \varepsilon_i] = \int_{C-u\left(w_i - \frac{g_i}{n_i}, g_i\right)}^{\varepsilon_i} D(\varepsilon_i) d\varepsilon_i$$

Substituting the constrained conditions with the objective function in Problem) and setting up a first-order condition with respect to g_i, we have

$$\frac{\partial E[\upsilon(g_i) + R]}{\partial g_i} = \frac{d\upsilon(g_i)}{dg_i} + R \cdot -D[C - u(x_i(g_i), g_i)] \cdot \left(- \frac{\partial u}{\partial x_i} \frac{\partial x_i}{\partial g_i} - \frac{\partial u}{\partial g_i} \right) = 0$$

. (5)

Assembling Eq. (5), we have the following condition of local public goods supply:

$$-\frac{d\upsilon(g_i)}{dg_i} - R \cdot D[C - u(x_i(g_i), g_i)]\frac{\partial u}{\partial x_i}\frac{\partial x_i}{\partial g_i} = R \cdot D[C - u(x_i(g_i), g_i)]\frac{\partial u}{\partial g_i} \quad (6)$$

The intuitive interpretation of Eq. (6) is as follows. The first term of the left- hand side of Eq. (6) stands for the disutility of local government caused by the supply of the local public goods g_i. The second term of the left-hand side of Eq. (6) denotes the decrease in the re-election probability associated with the decrease in the residents' utility level that is caused by the decrease in the private goods consumption due to the rise in the lump-sum tax. The right-hand side of Eq. (6) denotes the rise in the local government's re-election probability associated with the resident's utility level increase caused by the increase in local public goods.

The left-hand side of Eq. (6), therefore, shows the marginal cost of supplying local public goods. The right-hand side of Eq. (6) on the other hand shows the marginal benefit of supplying local public goods. Eq. (6) denotes that the equilibrium level of local public goods is determined when marginal cost is equal to marginal benefit.

If we compare the local public goods level achieved in a yardstick competition model with that in the social optimal, we obtain Proposition 1 as follows.

Proposition 1 The level of local public goods achieved in a yardstick competition model incorporating residents' choice between private goods and local public goods indicates tendency to undersupply relative to the social optimal level of local public goods.

The intuitive interpretation of Proposition 2 can be stated as follows. Yardstick competition provides the local government with the incentive to increase its effort level since elections by residents give rise to competition between local governments. From the viewpoint of resource allocation, however, under yardstick competition, the local government exerts an effort to seek re-election but cannot achieve the optimal level of local public goods supply that the residents actually demand for. That is, under yardstick competition, the local government shows the tendency to undersupply public goods.

As stated above in Propositions 1, does yardstick competition inevitably lead local governments to oversupply or undersupply public goods? To consider the problem, we specify the density function $D(\varepsilon)$ as follows.

$$\int D(\varepsilon)d\varepsilon = \frac{1}{2\sigma\sqrt{\pi}}$$

With the specified density function, we clarify the situation as follows. As the standard deviation σ rises, the probability of ε_i being located much closer to mean zero rises; conversely, as the standard deviation σ comes down, the probability that ε_i being distant from mean zero rises. In summary, we have the following proposition:

Proposition 2 The level of local public goods supplied under yardstick competition depends upon the standard deviation σ. That is, when the standard deviation σ rises (comes down), the level of local public goods decreases (increases).

The intuitive interpretation of Proposition 2 can be stated as follows. When the standard deviation σ rises (comes down), the range within which the noise ε_i undergoes a change expands (diminishes). As the range expands (diminishes), the local government's probability of re-election decreases; thus, the local government decreases (increases) the local public goods supply since its efforts do not pay off (pay off).

On the basis of Proposition 3, we immediately get the following corollary:

Corollary 1 Yardstick competition usually leads the local government to undersupply local public goods; however, when standard deviation σ diminishes adequately, yardstick competition leads the local government to supply more local public goods

This result may have unfavorable effect to the efficiency of equilibrium derived from the conventional yardstick competition model. In the conventional yardstick model, when the standard deviation σ is small enough, the local government can expect the probability of its own re-election to rise due to the rise in the probability that ε_i is located much closer to mean zero. This leads the local government to increase local public goods supply. Thus, conventional yardstick competition raises the local government' efforts to the excessive level when the standard deviation σ is small enough.

Here, attention should be paid to the implication of the small value of the random noise ε_i. This suggests that all the regions become identical. That is, as all regions become identical, yardstick competition stimulates local competition between local governments with regards to the public goods supply

On the contrary, by Corollary 1, there is a possibility that yardstick competition involving residents' choice between private goods and local public goods achieves the social optimal level of local public goods, thus alleviating the problem of the undersupply of local public goods, when

the standard deviation σ is small. This result is opposite to the result observed under conventional yardstick competition. Thus, in yardstick competition involving residents' choice between private goods and local public goods, excessive efforts to on the part of the local government toward being re-elected may be alleviated when regions are identical to each other.

Next, rearranging Equation (9), we have

$$n_i \frac{u_g^i(x_i(g_i), g_i)}{u_x^i(x_i(g_i), g_i)} = 1 - \frac{d\upsilon(g_i)}{dg_i} \frac{1}{R \cdot D[C - u(x_i(g_i), g_i)]} \frac{n_i}{u_x^i(x_i(g_i), g_i)} \quad (7)$$

From Eq. (7), we obtain Proposition 3 as follows.

Proposition 3 When the local government' utility $\upsilon(g_i)$ is constant with respect to the variation of g_i, the yardstick competition involving the residents' choice between private goods and local public goods can achieve the social optimal local public goods supply.

The intuitive interpretation of Proposition 3 can be stated as follows. The second term of the left-hand side in Eq. (7) represents the marginal disutility of local public goods measured by the marginal utility of private goods. Proposition 3 suggests that yardstick competition incorporating residents' choice between private goods and local public goods can achieve the social optimal local public goods supply if the local government's disutility is not increased. That is, even if no local government is a complete benevolent government, the social optimal level can be attained if the government considers only the resident's welfare.

3. YARDSTICK COMPETITION AND ROLE OF INFORMATION POLICY

As there are two regions in the model, a strategic interdependence between them should be taken into consideration. If we assume that the residents and government of each region first consider their counterparts' utility level and then decide their behavior, the deterministic reservation level of voters C will be their counterparts' utility level under the assumptions of yardstick competition.

$$u(x_j, g_j) + \varepsilon_j = C \quad (8)$$

Eq. (8), therefore, can be considered as the systems of two equations that simultaneously determine the Nash equilibrium level of public goods in two regions.

The utility level of both regions is expected to be improved since yardstick competition leads to greater efforts on the part of both governments, i.e.

$$\frac{dpr_i}{dg_j} = -\frac{du_j}{dg_j} \cdot D(g_i) < 0 . \quad (9)$$

For the discussion of empirical significance of yardstick competition among local government, we implemented preliminary statistical analysis by using Japanese provincial data. According to the assumptions of yardstick competition, incumbent government execute policies such as tax cut or increasing their expenditure which attract their voters in order to re-elect in the next election.

Besley and Case (1995) showed, by using U.S. local data, that tax cut could attract voters and incumbent government had tendency to cut residential or income tax to seek their vote for re-election. However, the operating margin of local tax rate is relatively narrow in Japan, and the local tax rates of most local governments are about the same.

By considering these situations, we framed a hypothesis of yardstick competition by local expenditure, and constructed a statistical model which explains the number of prefectural governors' sequential re-election by the growth rate of their local expenditures. The estimation results are as follows.

Table 1 Estimation Results

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(constant)	0.549	0.525	1.047	0.311
EXPCH(Total Expenditure)	0.619	2.754	0.225	0.825
PLEXPCH(Social Welfare & Security	1.344	0.608	2.212	0.042
HYEXPCH(Health & Sanitary Expen	1.068	0.459	2.329	0.033
LAEXPCH(Labor Expenditure)	−0.011	0.064	−0.169	0.868
AFFEXPCH(Agriculture & Fishery)	−3.089	1.427	−2.165	0.046
CIEXPCH(Commercial & Industry)	0.251	0.334	0.753	0.462
CEEXPCH(Public Work Expenditure	−0.011	1.305	−0.009	0.993
HOEXPCH(Public Housing)	0.559	0.314	1.783	0.094
POEXPCH(Police Depertment Expe	−0.031	2.850	−0.011	0.991
EDEXPCH(Education Expenditure)	−2.070	3.764	−0.550	0.590
NDREXPCH(Disaster Restration)	0.077	0.049	1.589	0.132
PBEXPCH(Public Bond and Loans)	0.862	1.183	0.728	0.477

R-squared	0.661	Mean dependent var	2.172
Adjusted R-squared	0.407	S.D. dependent var	0.658
S.E. of regression	0.507	Akaike info criterion	1.782
Sum squared resid	4.114	Schwarz criterion	2.394
Log likelihood	−12.832	Hannan-Quinn criter	1.973
F-statistic	2.601	Durbin-Watson stat	2.132
Prob(F-statistic)	0.038		

The estimation results shows that the correlation coefficients of the growth rate of Social Welfare and Security Expenditure, Public Health Expenditure are significantly positive and the correlation coefficients of the growth rate of the Agriculture and Fishery Expenditure is significantly negative. These results suggest that the yardstick competitions among the prefectural governor seeking for re-election are observable in Japanese data and the competition leads to the expenditure growth in the field of social welfare and public health, and expenditure cut, on the contrary, in the field of agriculture and fishery.

As is shown in Proposition 2 and its corollary, the decrease of the standard deviation (σ) of the regional environmental shock parameter (ε_i) improve the efficiency of yardstick equilibrium. Since this is the proxy

for the asymmetricity of information between residents and the local government, the information policy for reducing the asymmetricity proves effective in improving efficiency. It means that local governments need to supply local public goods after considering regional disparities regarding residents' preference, exogenous environmental conditions, and other factors. Information technology such as an e-government could be one of the factors facilitating the policy.

7. CONCLUSIONS

In this paper, we obtained the following results. First, if we ignore residents' consumption choice, local governments tend to over-supply local public goods, since local government heads attach more importance to re-election. Second, in order to improve the efficiency of the yardstick equilibrium, local governments need to supply local public goods after considering regional disparities regarding residents' preference, exogenous environmental conditions, and other factors. That is, policies for diminishing asymmetric information between the local government and residents are effective in improving efficiency. Our results suggest that views of enhancing the effectiveness and optimizing the public services give a new index to e-government evaluations which are also led to providing new elements to governmental performance evaluations.

8. REFERENCES

[1]Besley, T. and Case, A. (1995),"Incumbent behavior: vote seeking, tax setting and yardstick competition."American Economic Review, vol.85, pp.25-45.

[2]Besley, T., Coate, S., (2003),"Centralized versus decentralized provision of local public goods: A political economy approach."*Journal of Public Economics*, vol.87, pp.2611-2637.

[3]Gibbons, R. (1992), Game Theory for Applied Economists, Princeton University Press.

[4]Konishi, H., (2009), *Economic Analysis of Public Choice*, University of Tokyo Press, Tokyo (in Japanese).

[5]Lazear, E. P. and Rosen, S (1981),"Rank-Order Tournaments as Optimum Labor Contracts."*Journal of political economy*, vol.89, pp.841-864.

[6]Nalebuff, B. and Stiglitz, J.E. (1983),"Prizes and incentives: towards a general theory of compensation and competition."*The Bell Journal of Economics*, vol.14, pp.21-43.

[7]Oates, W. E. (1972), *Fiscal Federalism*, Harcourt Brace Javanovich Inc. New York.

[8]---------- (2005), "Toward a Second-Generation Theory of Fiscal Federalism," *International Tax and Public Finance*, Vol.12, pp.349-373.

[9]Seabright, P. (1996),"Accountability and decentralisation in government: An incomplete contracts model."*European Economic Review*, vol.40, pp.61-89.

[10]Shleifer (1985), "A theory of yardstick competition."*The RAND Journal of Economics*, vol.16, pp.319-327.

[11]Tiebout, C. M. (1956), "A pure theory of local expenditures."*Journal of Political Economy*, vol.64, pp.416-424

[12]Wong, M. S., Nishimoto, H., Philip, G (2011) "The Use of Importance-Performance Analysis in Evaluating Japan's E-Government services". *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 6, no. 2, pp.17-30.

[13]UNPAN (2010) Leveraging E-government at a Time of Financial and Economic Crisis, New York: United Nations Public Administration Network.

[14]Dexter, A. and Parr, V (2003) "Government Online: An Online Perspective 2003 – Global Summary, Taylor Nelson Research".

[15]West, D. M. (2007) "Global E-Government 2007", Center for Public Policy, Brown University, Providence, RI.

[16]Cole, M., and Jupp, V. (2005) "Leadership in Customer Service: New Expectations, New Experiences", The Government Executive Series, Accenture

[17]Bannister, F. (2007) "The curse of the benchmark: an assessment of the validity and value of e-government comparisons", *International Review of Administrative Sciences*, vol. 73, no. 2, pp. 171-188.

.

Development of an information system to evaluate vaccine loss (wastage)

Samia Abdul Samad
Ministry of Health of Brazil
Brasília, Distrito Federal, Brazil

and

Antonia Maria Teixeira, Brendan Flannery, Ricardo Gonçalves, Consuelo Freiria

ABSTRACT

Vaccine wastage may occur due to physical damage including breakage or because fewer vaccine doses are administered during an immunization session than included in multi-dose vials. Methods: We analyzed reported data from 2,553 vaccination centers distributed in 600 municipalities for 4 vaccines: measles-mumps-rubella (MMR), diphtheria-tetanus-pertussis-Hib (DTP-Hib), Bacille Camille Guerin (BCG) and oral rotavirus vaccine. Vaccine costs were provided by the immunization program. Results: Mean vaccine wastage was 65.7% for MMR (range, 46.1% to 72.4%), 23.9% for DTP-Hib (range, 10.3% to 32.6%), 74% for BCG (range, 64.4% to 79.9%), and 3.2% for rotavirus (range, 1.3% to 4.8%). The ratio of doses in opened vials to doses administered was almost 3:1 for MMR, 1.3:1 for DTP-Hib, 3.8:1 for BCG and nearly 1:1 for rotavirus. Conclusion: Multi-dose vials are often preferred by immunization programs for requiring less space for transport and vaccine storage but may result in higher cost per dose administered. Calculation of vaccine wastage is an important component of planning immunization program requirements; incorrect estimates can result in purchasing too few or too many doses of vaccines, resulting in stock-outs or large quantities of expired vaccines.

Keywords: vaccines, wastage, utilization

1. INTRODUCTION

With the expansion and increasingly high profile of Brazil's National Immunization Program (abbreviated PNI), information was needed to improve management of vaccine supplies and control the movement of vaccines throughout the country. PNI prioritized the development of an information system that would identify the percentage of vaccine doses distributed that were acutally administered, with the objective of estimating financial costs involved in vaccine supply. The information system needed to provide information on vaccine wastage at all three levels (national, state and municipal) in Brazil's public health system, referred to as the Unified Health System (or SUS).

In 2006, an information system for monitoring of vaccine utilization (referred to as AIU) was created in Delphi, with an Access database. In 2010, the system was upgraded to Java language, using a PostgreSQL database with an interactive site in PHP (Oracle database). The AIU information system is based on data provided by individual vaccination posts (point-of-use), and provides information for routine evaluation of doses received at the vaccination post, doses contained in opened vials (utilization of vaccine vials), doses applied and doses discarded (wastage) for 44 immunobiologicals (vaccines and antisera), as well as reasons for vaccine wastage. Data from the system were used to calculate costs of vaccination.

As part of system implementation, data were analyzed from the first four states to begin using the AIU system (Amazonas [AM], Mato Grosso do Sul [MS], Rio Grande do Norte [RN], Santa Catarina [SC]) to estimate the prevalence and causes of losses at the three levels of the immunization program. The objective of this analysis was to compare information on the movement of vaccines obtained from four states in Brazil that have different characteristics, and to evaluate the influence of these characteristics on vaccine utilization. The evaluation was supported by Brazil's National Immunization Program, which is responsible for purchase of recommended vaccines and delivery to state immunization programs. More accurate data on vaccine utilization is important for planning, procurement and distribution.

2. OBJECTIVES

To evaluate and use data from the AIU system in the first four Brazilian states to adopt the system, for calculation of the prevalence and reasons for wastage of four recommended vaccines provided by Brazil's national immunization program: MMR, DTP-Hib, BCG and oral rotavirus vaccine.

3. METHOD

We abstracted AIU fields corresponding to the type of vaccine and presentation (number of doses per vial), vaccine stocks, opened vials (utilization), doses administered, physical losses (discarded, unopened vials) and technical loss (discarded doses in opened vials). Data were considered valid if the number of doses administered was equal to or less than the number of doses in opened vials, if the total number of vaccine doses utilized was equal to doses administered plus losses, and if doses registered in the AIU system were equal to the number of doses administered at the vaccination post (according to the monitoring system for vaccination coverage, or API). We calculated the percentage of technical loss according to the following equation:

$$100 \times \frac{(n \text{ doses in opened vials}) - (n \text{ doses administered})}{n \text{ doses in opened vials}}$$

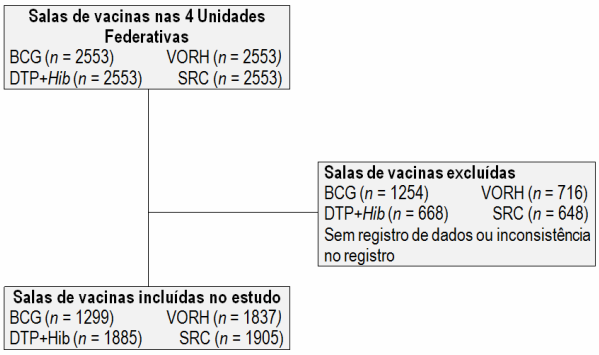
Physical loss is defined as the total number of discarded doses in unopened vials due to expiration, breakage and conservation outside recommended temperatures. Brazil does not use Vaccine Vial Monitors on vaccine vials to determine whether vaccines have been damaged by exposure to high or low temperatures.

Costs of wastage and costs of vaccine per dose administered were obtained by multiplying vaccine losses by the per-dose cost of vaccine vials, excluding storage and transportation costs. The per-dose cost of each vaccine in 2008 was US$3.50 for MMR in 10-dose vials, US$3.25 for DTP-Hib in 5-dose vials, US$0.31 for BCG in 10-dose vials and US$9.22 for oral rotavirus vaccine in single-dose vials.

4. RESULTS

We received data from 2553 registered vaccination posts in 600 municipalities. Non-valid data were excluded for 1254 vaccination posts for BCG vaccine, 668 posts for DTP/Hib, 716 for oral rotavirus vaccine and 648 for MMR. For the four vaccines, and average of 1731 vaccination posts had valid data for evaluation. Of these, 502 (30%) vaccination posts submitted data for twelve months (January to December) in 2008 and 1229 (70%) submitted valid data for one or more months (range, 1 to 11).

Figure 1: Flow rooms and selection of study vaccines



Mean vaccine wastage was 65.7% for 10-dose vials of MMR vaccine (range, 46.1% to 72.4%), 23.9% for five-dose vials of DTP-Hib (range 10.3% to 32.6%), 74% for 10-dose vials of BCG (range 64.4% to 79.9%), and 3.2% for single dose vials of oral rotavirus vaccine (range 1.3% to 4.8%).

Table 1. Prevalence of losses by type of vaccine and UF in 2008

UF	VACINA	N	TAXA	IC (95%)
	SRC			
AM		344.350	0, 465	(0, 463, 0, 466)
MS		297.370	0, 703	(0, 701, 0, 705)
RN		86.760	0, 648	(0, 645, 0, 651)
SC		884.970	0, 728	(0, 727, 0, 729)
Total		1.613.450	0, 641	(0, 500, 0, 760)
	DTP+Hib			
AM		212.660	0, 110	(0, 109, 0, 111)
MS		116.370	0, 285	(0, 283, 0, 288)
RN		68.560	0, 358	(0, 355, 0, 362)
SC		324.835	0, 314	(0, 312, 0, 316)
Total		722.425	0, 251	(0, 156, 0, 379)
	BCG			
AM		253.370	0, 702	(0, 701, 0, 704)
MS		179.060	0, 818	(0, 817, 0, 820)
RN		52.770	0, 692	(0, 688, 0, 696)
SC		344.280	0, 776	(0, 775, 0, 778)
Total		829.480	0, 751	(0, 691, 0, 803)
	VORH			
AM		88.867	0, 020	(0, 019, 0, 021)
MS		54.711	0, 040	(0, 038, 0, 041)
RN		25.719	0, 071	(0, 068, 0, 074)
SC		150.553	0, 029	(0, 028, 0, 030)
Total		319.850	0, 036	(0, 022, 0, 057)

Source: SI_AIU, MS

Legend: SRC (measles mumps and rubella); DTP+*Hib* (diphtheria, tetanus, pertussis and haemophilus influenzae tipo b); BCG (Bacilo Calmett Guérin); VORH (rotavírus).

The ratio of doses in opened vials to doses administered was approximately 3:1 for MMR, 1.3:1 for DTP-Hib, 3.8:1 for BCG and 1:1 for oral rotavirus vaccine. For MMR, the ratio indicates that for opened, 10-dose vials, two doses were discarded for each dose administered (or child vaccinated).

Table 2: Loss technique, due to the doses used doses technical losses and expenses, by vaccine, state, year 2008

	Doses utilizadas	Doses aplicadas	Taxa de perda técnica	Razão de doses aplicadas pelas doses utilizadas	Valor ideal da dose aplicada	Valor real da dose aplicada	Gasto ideal da aplicação	Gasto real da aplicação	Diferença
SRC									
AM	341.830	184.387	46,1	1,9	R$ 6,210	R$ 11,51	R$ 1.145.043,27	R$ 2.122.764,30	R$ 977.721,03
MS	289.650	88.331	69,5	3,3	R$ 6,210	R$ 20,36	R$ 548.535,51	R$ 1.798.726,50	R$ 1.250.190,99
RN	82.310	30.551	62,9	2,7	R$ 6,210	R$ 16,73	R$ 189.721,71	R$ 511.145,10	R$ 321.423,39
SC	872.340	241.006	72,4	3,6	R$ 6,210	R$ 22,48	R$ 1.496.647,26	R$ 5.417.231,40	R$ 3.920.584,14
TOTAL	1.586.130	544.275	**65,7**	2,9	R$ 6,210	R$ 18,10	R$ 3.379.947,75	R$ 9.849.867,30	R$ 6.469.919,55
DTP+*Hib*									
AM	210.890	189.255	10,3	1,1	R$ 6,034	R$ 6,72	R$ 1.141.964,67	R$ 1.272.510,26	R$ 130.545,59
MS	113.435	83.179	26,7	1,4	R$ 6,034	R$ 8,23	R$ 501.902,09	R$ 684.466,79	R$ 182.564,70
RN	65.240	43.999	32,6	1,5	R$ 6,034	R$ 8,95	R$ 265.489,97	R$ 393.658,16	R$ 128.168,19
SC	319.280	222.848	30,2	1,4	R$ 6,034	R$ 8,65	R$ 1.344.664,83	R$ 1.926.535,52	R$ 581.870,69
TOTAL	708.845	539.281	**23,9**	1,3	R$ 6,034	R$ 7,93	R$ 3.254.021,55	R$ 4.277.170,73	R$ 1.023.149,18
BCG									
AM	241.730	75.387	68,8	3,2	R$ 0,5703	R$ 1,83	R$ 42.993,21	R$ 137.858,62	R$ 94.865,41
MS	162.100	32.526	79,9	5,0	R$ 0,5703	R$ 2,84	R$ 18.549,58	R$ 92.445,63	R$ 73.896,05
RN	45.650	16.256	64,4	2,8	R$ 0,5703	R$ 1,60	R$ 9.270,80	R$ 26.034,20	R$ 16.763,40
SC	323.940	76.989	76,2	4,2	R$ 0,5703	R$ 2,40	R$ 43.906,83	R$ 184.742,98	R$ 140.836,16
TOTAL	773.420	201.158	**74,0**	3,8	R$ 0,5703	R$ 2,19	R$ 114.720,41	R$ 441.081,43	R$ 326.361,02
VORH									
AM	88.622	87.071	1,8	1,0	R$ 17,8587	R$ 18,18	R$ 1.554.974,87	R$ 1.582.673,71	R$ 27.698,84
MS	53.815	52.549	2,4	1,0	R$ 17,8587	R$ 18,29	R$ 938.456,83	R$ 961.065,94	R$ 22.609,11
RN	25.112	23.899	4,8	1,1	R$ 17,8587	R$ 18,77	R$ 426.805,07	R$ 448.467,67	R$ 21.662,60
SC	148.134	146.191	1,3	1,0	R$ 17,8587	R$ 18,10	R$ 2.610.781,21	R$ 2.645.480,67	R$ 34.699,45
TOTAL	319.850	309.710	**3,2**	1,0	R$ 17,8587	R$ 18,44	R$ 5.531.017,98	R$ 5.712.105,20	R$ 181.087,22
TOTAL GERAL	3.388.245	1.594.424					R$ 12.279.707,69	R$ 20.280.224,65	R$ 8.000.516,96

Source: SI_AIU, MS

Legend: SRC (measles mumps and rubella); DTP+*Hib* (diphtheria, tetanus, pertussis and haemophilus influenzae tipo b); BCG (Bacilo Calmett Guérin); VORH (rotavírus).

During the period of evaluation, a total of 150,000 vials of MMR were opened (for a total of 1.5 million doses of MMR vaccine) in the four state immunization programs, while 550,000 MMR doses were administered. The actual vaccine cost per dose administered was US$10.06. The total expenditure for MMR vaccine in the four states was US$5.4 million versus costs in an ideal scenario without wastage of US$1.8 million (excluding costs of vaccine storage and transportation), a difference of US$3.6 million.

For 5-dose vials of DTP-Hib, cost per dose administered was US$4.40, which was US$1.15 more than the per-dose purchase price. Actual expenditures based on the number of DTP-Hib doses administered was US$1.0 million in excess of the ideal per-dose cost assuming no wastage (excluding storage and transportation costs). For BCG, the cost per dose administered was US$1.22. Of a total number of 770,000 doses in opened vials, only 200,000 doses were applied, resulting in a difference between ideal and actual spending of US$181,000. For oral rotavirus vaccine, actual per-dose costs were similar in the four states, with approximately US$0.50 difference between ideal and actual costs, resulting in a total excess expenditure of US$100,000. In the four states evaluated, expenditure for the four vaccines totaled 20 million Brazilian reais (~US$10 million), while expenditure whithout wastage would have been US$4.4 million.

In relation to physical losses (discarded, unopened vials), the main reasons varied among the four states. For the states of Santa Catarina and Amazonas, physical losses for 10-dose vials of MMR resulted from lack of electricity (35.7% and 41.2% of physical losses, respectively). In Mato Grosso do Sul and Rio Grande do Norte, the most common reason provided for physical loss of 10-dose MMR vials was "other reasons" (38.6% and 40.4%, respectively), suggesting confusion about the concept of physical loss. Physical loss due to problems with maintaining recommended temperatures in the cold chain had the lowest frequency. For DTP-Hib, physical losses due to interruptions in electricity and temperature changes were most common reasons in SC (50.7%) and MS (26.4%), expiration was most common in AM (36.2%) and "other reasons" were most common in RN (40.7%). For BCG, expiration of vaccine was the most common reason for physical losses in RN (65.2%) and AM (88%). In addition to low frequency of loss due to problems with transportation of vaccines within recommended temperatures (0% - 0.4%), inadequate procedure was an infrequent cause of wastage for BCG (0.3% - 3.5%). Physical losses of oral rotavirus vaccine were mainly due to lack of electricity (18% - 47.8%), expiration (38%) and refrigeration equipment failure (10.8% - 20.6%). In addition to vaccine wastage in opened multi-dose vials, principal reasons for vaccine loss were refrigeration problems including interruptions in the supply of electricity and past expiration dates.

Table 3: Proportional distribution of physical losses due to vaccines and state in the year 2008

	QF (%)	EE (%)	FE (%)	VV (%)	PI (%)	FT (%)	OM (%)
SRC							
AM	7,9	35,7	9,5	14,3	2,0	5,2	25,4
MS	3,9	15,3	14,2	21,2	4,4	2,3	38,6
RN	4,5	16,6	4,5	32,8	1,1	-	40,4
SC	5,9	41,2	7,8	28,5	4,6	2,6	9,4
DTP+*Hib*							
AM	15,3	12,7	17,5	36,2	8,5	0,8	9,0
MS	6,6	26,4	25,9	18,6	10,2	2,2	10,1
RN	3,5	20,8	4,1	26,7	3,6	0,8	40,7
SC	7,2	50,7	12,6	9,5	10,0	2,2	7,9
BCG							
AM	4,3	2,1	1,5	88,0	0,3	-	3,9
MS	6,5	5,0	7,3	69,0	2,4	0,4	10,6
RN	6,0	10,1	0,1	65,2	1,0	-	17,6
SC	4,2	13,2	2,9	73,2	3,5	0,2	2,9
VORH							
AM	6,1	25,3	18,4	38,0	4,9	-	7,3
MS	4,9	18,0	20,6	25,4	15,1	5,6	10,4
RN	3,0	47,8	10,8	3,3	9,4	-	25,7
SC	3,3	34,6	16,9	14,8	11,9	4,8	13,7

Source: SI_AIU, MS

Legend: SRC (measles mumps and rubella); DTP+*Hib* (diphtheria, tetanus, pertussis and haemophilus influenzae tipo b); BCG (Bacilo Calmett Guérin); VORH (rotavírus).

QF (loss of bottle breakage);EE (loss due to power outages); FE (loss due to equipment failure refrigeration); VV (loss lost validity); PI (loss due to inadequate procedures); FT (loss for failure to transport) e OM (loss for other reasons than the above described).

5. CRITICAL FACTORS

This analysis is subject to several limitations. Valid data were not provided from all vaccination posts, suggesting problems with the use of the information system or confusion regarding the concepts. Many vaccination

posts did not send complete data throughout the entire period evaluated. The evaluation was only performed with data from four Brazilian states and is not representative of all 27 state immunization programs. Frequency of wastage and reasons for physical losses may vary between reporting and non-reportings sites within the states, as well as between state immunization programs.

6. CONCLUSION

This was the first evaluation of vaccine utilization and losses to be conducted by Brazil's National Immunization Program. The creation of an information system for this purpose made it possible to estimate the frequency of vaccine wastage and its causes at the point-of-use.

Results of this evaluation showed that vaccine wastage due to technical losses (discarded doses in opened, multi-dose vials) were greated than expected. Wastage rates in the system are related to vial volume and technical specifications for use of multi-dose vials—i.e. how long a multi-dose vial may be used after opening. Prior to implementation of the AIU system, vaccine wastage was estimated by subtracting the number of doses administered from the number distributed. However, frequency of vaccine wastage in opened, multi-dose vials were almost two-times higher than estimates of wastage used by PNI for procurement of BCG and MMR. For DTP-Hib, estimated wastage from the AIU evaluation was similar to previous estimates used by PNI for vaccine procurement. Despite having the lowest wastage of the vaccines evaluated, the wastage of oral rotavirus vaccine was concerning due to the use of single-dose vials and high vaccine cost, indicating the need for better training and monitoring. Results of the AIU evaluation provided specific knowledge about vaccine utilization at public health care centers, which may inform decisions regarding vaccine supplies in the future.

Results show that for multidose vaccines with limited shelf lives after opening (BCG and MMR), technical losses due to discarded doses in opened vials were much greater than physical losses due to breakage, expiration of vaccine and temperature fluctuations (loss of cold chain). Two doses from opened vials are discarded for each one dose administered. On the other hand, multi-dose vials are preferred for use in the public network for facilitating distribution, reducing storage space requirements and assocciated costs. Losses due to transportation problems were minimal. Data from the AIU system will contribute to analyses that consider all the cost implications of vaccine formulations on distribution, cold chain capacity and wastage.

Previously, it had been assumed that not all vials distributed were opened, and that physical losses (discarded, unopened vials) predominated. The quantity of doses discarded for technical reasons and reasons for vaccine wastage were not monitored to identify and correct problems. The AIU represents a breakthrough in controlling the movement of vaccines from central stores to the point-of-use at vaccination posts, enabling better management and site assessment in the central cold chain.

Brazil's National Immunization Program seeks an appropriate use of vaccines and minimal losses, to reduce costs while expanding access to vaccinations. Data from the AIU system in four states demonstrate the utility of the system for estimating the frequency of vaccine wastage and its causes. Ongoing evaluation of data from the AIU system will provide input for planning vaccine requirements for production and procurement, as well as for distribution of these products.

An SMS Server Prototype For Supporting Medical Prescription Adherence

Abdel Ejnioui
Division of Information Technology, University of South Florida Polytechnic
Lakeland, Florida 33803, U.S.A

and

Mathieu Morjaret
Department of Computer Science and Networks, ESISAR
Valence, France

ABSTRACT

Barriers to prescription adherence among patients have been shown to have significant impact on service quality and cost in the healthcare system. To minimize this impact, many stakeholders in the healthcare industry are highly interested in supporting prescription adherence among patients. Most of these stakeholders believe that information technology in general, and mobile technology in particular, can help in developing medical practices that can be highly conducive to prescription adherence by enhancing communication between patients and healthcare providers. To this end, a number of pharmacy management benefit companies plan to adopt SMS communication to reach their customers given the wider acceptance of SMS messaging among cell phone users. However, most of these pharmacies are reluctant to purchase service agreements from SMS aggregators without a complete understanding of user, service and business requirements related to SMS messaging. Hence, many are in dire needs for prototypes of SMS servers that can help them define and refine these requirements before committing to costly agreements with SMS aggregators. This paper describes such a prototype for a pharmacy benefit management company located in central Florida.

Keywords: Mobile technology, Prescription adherence, Short message service, and SMS aggregator.

1. INTRODUCTION

Adherence is defined as the extent to which patients take medications as prescribed by their healthcare providers. A 2001 survey showed that although 62% of physician office visits generate a prescription, these prescriptions are not always adhered to [1, 2]. Poor adherence tends to be serious among patients who suffer from chronic diseases since these diseases require long-term treatments (e.g., HIV infections, hypertension, asthma, diabetes, heart disease and psychiatric illness). This is even more critical considering that 75% of all health expenditures in 2000 went to care for individuals with chronic illnesses although these individuals represent only 45% of Americans [3, 4]. In fact, non-adherence to prescribed medication is responsible for 10% of hospital admissions and 25% of nursing home admissions. It is estimated that the healthcare system in the U.S. incurs a cost of $300 billion annually due to non-adherence to essential medications. Patient surveys about non-adherence reveal an array of barriers to adherence such as costs of drugs, forgetfulness (e.g., it is practically difficult for a patient to remember to take medication several times a day), lack of clarity in the purpose of treatment, perceived lack of medication effect, debilitating side effects (e.g., for some professionals such as doctors, lawyers, professors and writers, the side effect of taking anticonvulsant drugs can interfere severely with abstract thinking), complicated regiment, lack of clarity in administration instructions, physical difficulty in handling medication (e.g., opening containers, handling small tablets), and unattractive formulation (e.g., unpleasant taste). According to the World Health Organization, increasing the effectiveness of adherence interventions may have a far greater impact on the health of world populations that any improvement in specific medical treatments [5]. For healthcare providers such as hospitals and insurance companies, strong adherence can lead to improved performance, which in turn can generate financial incentives for these providers. Providers can use improved performance as a metric to determine whether their services meet the expectation of their customers or not. Today, most people own a cell phone. It is conceivable to design mobile applications with user-friendly interfaces to help interested users in restoring their good health. Considering the current advantages of mobile technology and its communication facilities, it is clear that this technology can be exploited to help people learn to live healthy. In addition, it can be used effectively to personalize the therapy offered to a given individual considering his/her needs. To do so, a mobile application can be readily conceived as a medication adherence management tool on the go for the patient.

2. PHARMACIES AND PRESCRIPTION ADHERENCE

For pharmacies, strong adherence can lead to a volume increase in prescriptions refills as well as access to patients who were otherwise invisible to drug manufacturers for marketing promotions. Of special note is the importance of employers and pharmacies in augmenting adherence if both stakeholders collaborate in designing smart pharmacy benefits. These benefits can increase prescription use without impacting overall drug expenditures in the healthcare system. Most pharmacy benefit management companies prefer to use Short Message Service (SMS) messaging to communicate with their customers

considering its simplicity and wider acceptance. However, these companies lack a suitable infrastructure of information technology to do so. They can solicit the services provided by SMS aggregators by negotiating a cost-effective service level agreement with these aggregators that meets the requirements of SMS communication between the pharmacy and its customers. Worse yet, most pharmacies do not know what requirements must be taken in consideration to insure a successful message service with their customers. These requirements can be related to user interaction with the service, characteristics of the message service, and requirements related to business criteria as shown in Table 1. In the absence of well-defined requirements, a pharmacy benefit management company might make its best effort to purchase a service package with an aggregator only to realize later that the purchased package does not satisfy the requirements of its SMS communication with its customers. There is always a risk of over- or under-shopping for these service packages. Hence, it becomes reasonable for such a company to develop a prototype of an SMS service in order to define and refine these requirements. Such a prototype can be used as an exploratory tool for developing requirements that can be used as guidelines to purchase the most suitable service package from an aggregator. In this context, an emerging pharmacy benefit management company in central Florida decided to build an SMS server prototype to generate such requirements. This paper describes the architecture and design of this SMS server prototype.

3. SMS SERVER ARCHITECTURE

Although the pharmacy benefit management company mentioned above did not have a complete understanding of requirements in each category, it opted to base the design of the message server on the following requirements:

Table 1. SMS service requirements.

Category	Requirement
User	• Number of messages per hour or day • Appropriate delivery time of messages (before midnight) • User responsiveness to messages • Sequence of messages in prescription adherence scripts • Suitability of interaction with prescription script messages
Service	• Message sending (batch, number of retries, queuing, etc…) • Retrieving messages • Checking delivery of message status • Error and exception handling • Logging and tracking • Service configurability • Data storage (messages, customers, message traffic, etc…)
Business	• Number of short codes • Provisioning of short codes • Type of network connections to the SMS gateway • Transactions per day • Transactions per second • Message content

- *Self-Containment*: The message server must contain all the computing resources it needs to separate its responsibilities from those of the software applications of the company.
- *Logging*: This capability is needed to keep track of various events taking place between the software applications and the message server. The purpose of this tracking is to help the company determine the most important user, service and business requirements.
- *Error Handling*: This capability is need to record all errors and exceptions between the SMS gateway server of the network service provider and the software applications of the company. Recording these errors can provide a rich perspective on the reliability of the service offered by the network service provider.
- *Configurability*: This capability allows the benefit management company to manage the message service in different ways in order to explore requirements that are not clearly understood in normal operating conditions of the message server.

Based on these initial requirements, the architecture of the message server has been developed and refined over several iterations to include the following components as shown in Fig. 1:

- *Front Interface*: This interface provides methods that can be called by the software applications of the pharmacy benefit management company to perform tasks related to communication with its customers via SMS messages.
- *Message Database*: This database is a persistent store sued to record sent messages, received messages, errors and exceptions generated during SMS message exchange between the software applications and customer cell phone.
- *Message Server*: This server is a process that runs continuously to record all the events related to SMS communication between the software applications the customers such as sending message, retrieving reply messages and checking the delivery status of sent messages.
- *Data Access Layer*: This layer consists of dynamic libraries responsible for passing data from and to the message database on behalf of the front interface, the message server and the back interface.
- *Back Interface*: This interface provides methods to the message server for sending messages, retrieving messages, and checking the delivery status of sent message via the application programming interface (API) of the SMS gateway of the network service provider.

4. SMS SERVER IMPLEMENTATION

The architecture shown in Figure 1 was used to derive a class hierarchy for implementing the components seen in the architecture. This hierarchy was implemented using C# on .NET [6]. Fig. 2 shows the classes of the message server.

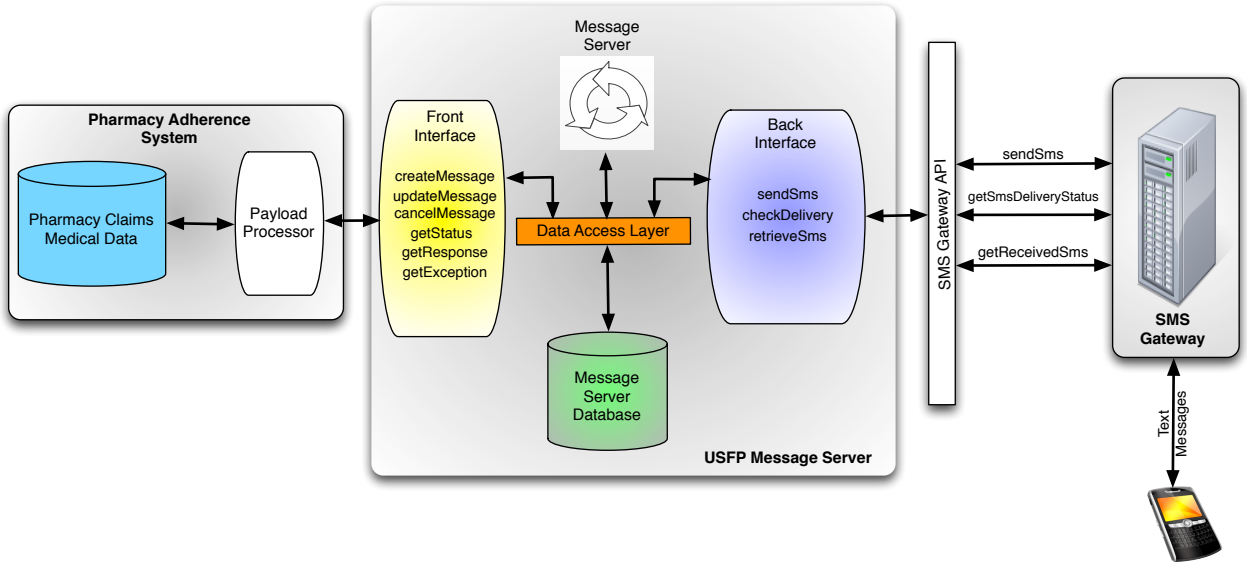


Fig. 1. Architecture of the message server.

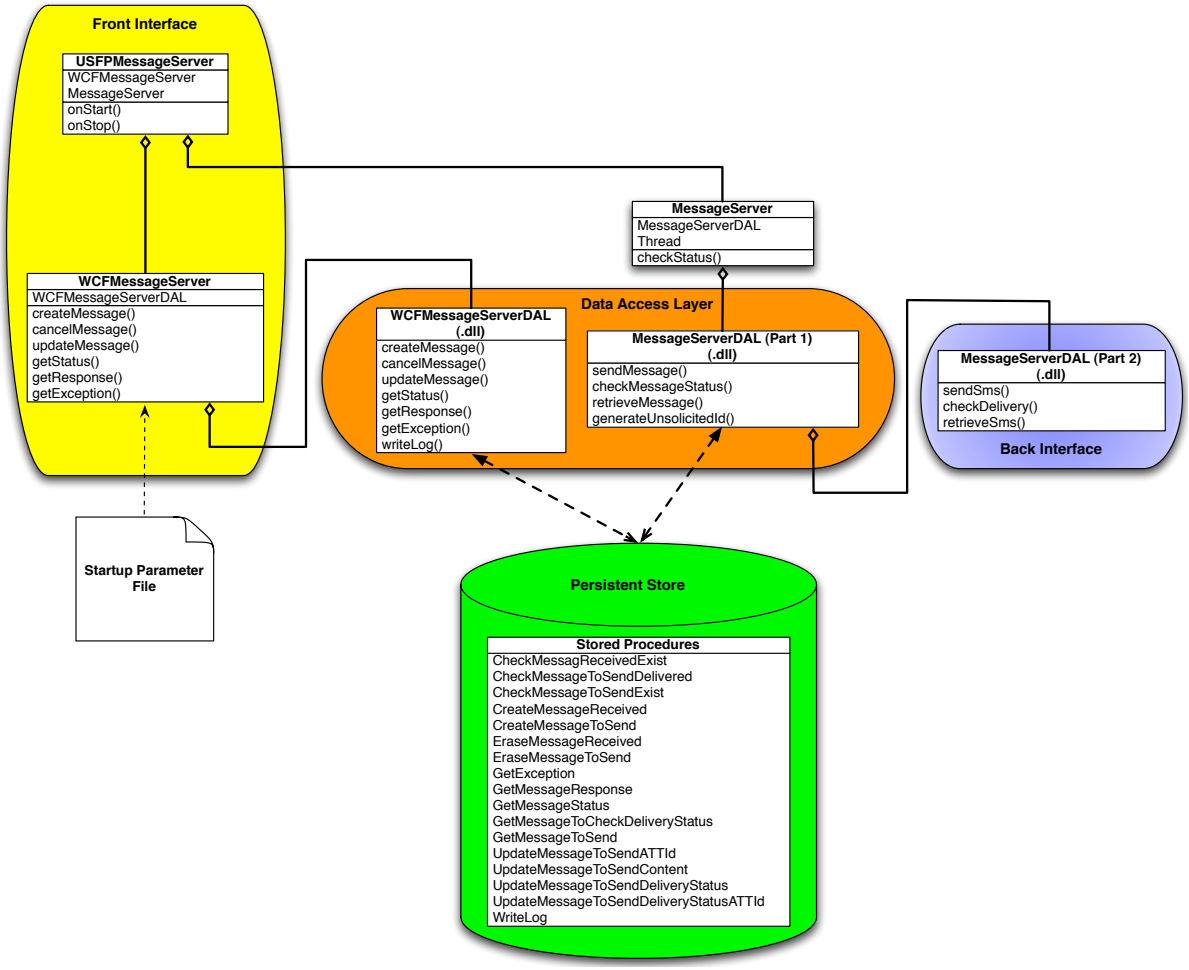


Fig. 2. Class mapping on the architecture of the message server.

The Front Interface

The front interface consists of the `WCFMessageServer` and `MessageServer` classes. Table 2 summarizes the methods of these two classes.

The Message Server

The message server consists of the `MessageServer` class. This class spawns a thread responsible for calling repetitively the `sendMessage`, `checkMessageDelivery` and `retrieveMessage` methods in the `MessageServerDAL`

class. These repetitive calls are performed as long as the number of transactions does not exceeded the maximum number of transactions allowed per day by the network service provider. Most network service providers impose limits on the number of transactions completed between a company server and their own SMS gateways based on the service level agreement purchased by the company in need of SMS services. It is meant by a transaction any call to the SMS gateway server of the network service provider.

The Data Access Layer

The data access layer consists of the `WCFMessageServerDAL` and the first part of the `MessageServerDAL` classes. Table 3 summarizes the methods of these two classes.

The Message Database

The message database consists of the following tables:

- Table of messages to send: This table contains records of the messages that need to be sent to the SMS gateway.
- Table of received messages: This table contains records of received messages. These messages are reply messages sent by customers are replies to messages sent by the pharmacy benefit management company.
- Message table: This table contains records of messages generated for events that took place while the message server is in operation. These events can be errors, exception or log entries recording specific tasks completed by one of the classes in the message server.

In addition to these tables, this database stores a number of procedures shown in Fig. 2.

The Startup Parameter File

When the message server starts, it needs to upload several parameters for proper functionality. These parameters are:

- Database connection settings: These are the settings necessary for the server to establish the connection with the database. They consist of the location path of the database and its security settings.
- Short code: This is the code assigned by the network service provider to the customer. This code is used to address SMS messages coming to or leaving from the servers of the pharmacy benefit management company.
- Endpoint send address: This is the URL address to which messages must be sent as required in the SMS gateway API.
- Endpoint receive address: This is the URL address to which message must be received as required in the SMS gateway API.
- Number of transactions per day: This number is fixed by the network service provider based on the service level agreement purchased by the pharmacy benefit management company.
- Number of transactions per second: This number is fixed by the network service provider based on the service level agreement purchased by the pharmacy benefit management company
- Transaction counter: This is a software counter generally initialized to 0 unless specified otherwise at startup time.

These parameters are stored in a file that is used by the message server during startup to load these parameters.

4. SMS COMMUNICATION VIA THE MESSAGE SERVER

Communication between the pharmacy benefit management company and its customers intended to enforce prescription adherence mostly of scripted dialogs between the company and its customers. The dialog below illustrates a simple adherence communication session between the pharmacy and a customer named DuPont.

Sending A Reminder Message To The Customer

In the first step, the pharmacy sends a message to Mr. DuPont to remind him to take his medication as shown in Fig. 3. Before the message is forwarded to the SMS gateway, it is inserted in the tables of messages to send in the database. Fig. 4 shows that the first entry is the entry of this message in the table. This entry shows that this message has reached the cell phone of Mr. DuPont since its delivery status has been automatically updated to 'DeliveredToTerminal'.

Receiving A Reply From The Customer

After customer DuPont receives the reminder message, he replies affirmatively by sending a "Yes, I did." reply message as shown in Fig. 5. As soon as the message server receives a this reply, the reply is immediately inserted in the table of received messages as shown in Fig. 6. The reply is passed back to the software applications of the pharmacy.

Sending An Acknowledgement To The Customer

When the pharmacy receives the reply message, its script dictates that it sends an acknowledgment to the customer as shown in Fig. 7. Before the message is forwarded to the SMS gateway, it is inserted in the tables of messages to send. Fig. 8 shows that the second entry is the entry of this acknowledgement message. This entry shows that this message has reached the cell phone of Mr. DuPont since its delivery status has been automatically updated to 'DeliveredToTerminal'.

6. CONCLUSION

This paper presented the prototype architecture and implementation of an SMS server intended to help a pharmacy benefit management company to define its SMS service requirements. These requirements can be used to shop for a service level agreement from an SMS aggregator that is highly suitable to the needs of the benefit management company.

Table 2. Front interface classes and methods.

Class	Method	Description
USFPMessageServer	onStart	It starts the Windows service.
	onStop	It stops the Windows service.
WCFMessageServer	createMessage	It calls the createMessage method in the WCFMessageServerDAL class by passing a message object.
	cancelMessage	It calls the cancelMessage method in the WCFMesssageServerDAL class by passing a message id.
	updateMessage	It calls the udpateMessage method in the WCFMessageServerDAL class by passing a message object.
	getStatus	It calls the getStatus method in the WCFMessageServerDAL class by passing a message id.
	getResponse	It calls the getReponse method in the WCFMessageServerDAL class by passing a message id.
	getException	It calls the getException method in the WCFMessageServerDAL class by passing a start and end dates.

Table 3. Classes and methods of the data access layer and back interface.

Class	Method	Description
WCFMessageServerDAL	createMessage	It calls the CreateMessageToSend stored procedure to insert each message in a batch of messages if the message does not already exist in the table of messages to send in the database.
	cancelMessage	It calls the EreaseMessageToSend stored procedure to remove the message from the table of messages to send and EraseMessageReceived stored procedures to remove the message from the table of received messages.
	updateMessage	It calls the UpdateMessageToSendContent stored procedure to update the message contents in the table of messages to send in the database.
	getStatus	It calls the getStatus stored procedure to obtain the status of a sent message from the table of message to send in the database.
	getResponse	It calls the getMessageResponse stored procedure to extract the reply messages from the table of received messages.
	getException	It calls the GetException stored procedure to extract exceptions between two timestamps from the table of exceptions.
	writeLog	It class the WriteLog stored procedure to write relevant information to the messages table about an event taking place while the message server is in operation.
MessageServerDAL (Part 1)	sendMessage	It extracts the messages that need to be sent from the table of messages to send and calls the sendSms method in the MessageServerDAL class.
	checkMessageStatus	It calls the checkDelivery method in the MessageServerDAL class for each message whose status needs to be checked from the table of message to send.
	retrieveMessage	It calls the retrieveSms method in the MessageServerDAL class to retrieve reply messages from the SMS gateway.
MessageServerDAL (Part 2)	sendSMs	It creates a connection to the SMS gateway and calls the sendSms method in the SMS gateway API in order to send a batch of messages.
	checkDelivery	It creates a connection to the SMS gateway and calls the getSmsDeliveryStatus method in the SMS gateway API in order to check the delivery status of a batch of sent messages.
	retrieveSms	It creates a connection to the SMS gateway and calls the getReceivedSms method in the SMS gateway API in order to retrieve a batch of reply messages.

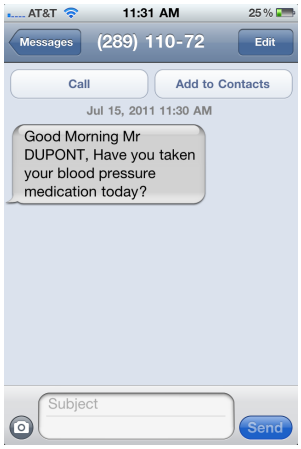
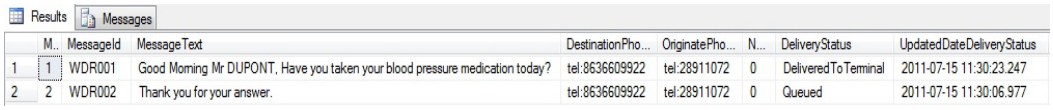


Fig. 3. Reminder message to Mr. DuPont.

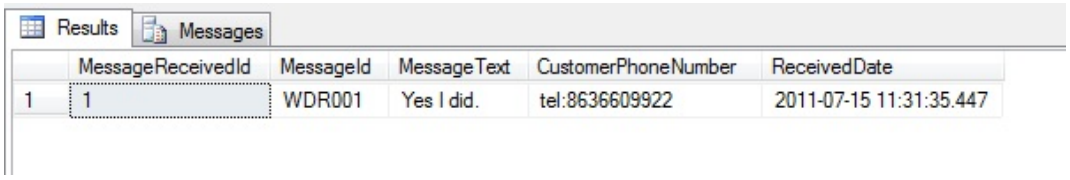
Fig. 5. Rely message from Mr. DuPont.

Fig. 7. Acknowledgement message from the pharmacy.



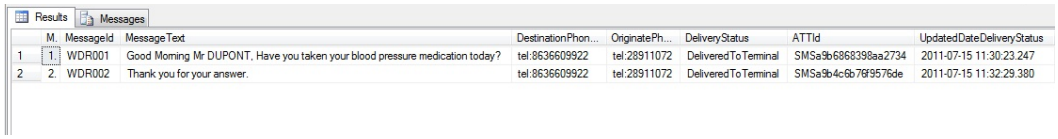
	M..	MessageId	MessageText	DestinationPho...	OriginatePho...	N...	DeliveryStatus	UpdatedDateDeliveryStatus
1	1	WDR001	Good Morning Mr DUPONT, Have you taken your blood pressure medication today?	tel:8636609922	tel:28911072	0	DeliveredToTerminal	2011-07-15 11:30:23.247
2	2	WDR002	Thank you for your answer.	tel:8636609922	tel:28911072	0	Queued	2011-07-15 11:30:06.977

Fig. 4. Contents of the table of messages to send.



	MessageReceivedId	MessageId	MessageText	CustomerPhoneNumber	ReceivedDate
1	1	WDR001	Yes I did.	tel:8636609922	2011-07-15 11:31:35.447

Fig. 6. Contents of the table of received messages.



	M.	MessageId	MessageText	DestinationPhon...	OriginatePh...	DeliveryStatus	ATTId	UpdatedDateDeliveryStatus
1	1.	WDR001	Good Morning Mr DUPONT, Have you taken your blood pressure medication today?	tel:8636609922	tel:28911072	DeliveredToTerminal	SMSa9b6868398aa2734	2011-07-15 11:30:23.247
2	2.	WDR002	Thank you for your answer.	tel:8636609922	tel:28911072	DeliveredToTerminal	SMSa9b4c6b76f9576de	2011-07-15 11:32:29.380

Fig. 8. Contents of the table of message to send.

7. REFERENCES

[1] D. K. Cherry, C. W. Burt, D. A. Woodwell, "National Ambulatory Medical Care Survey," *Advanced Data from Vital Statistics*, 2003.

[2] Boston Consulting Group and Harris Interactive, "The Hidden Epidemic: Finding a Cure for Unfilled Prescriptions and Missed Doses," December 2003, available at http://www.bcg.com/publications/files/TheHiddenEpidemic_Rpt_HCDec03.pdf.

[3] G. Anderson, J. Krickman, "Changing The Chronic Care System to Meet People's Needs," *Health Affairs*, vol. 20, no. 6, pp. 146-160, 2001.

[4] C. Hoffman, D. Rice, H.-Y. Sung, "Persons with Chronic Conditions: Their Prevalence and Costs," *Journal of American Medical Association*, vol. 276, no. 18, pp. 1473-1479, 1996.

[5] World Health Organization, "Adherence to Long-Term Therapies: Evidence for Action," 2003, available at http://www.who.int/chronic_conditions/en/adherence_report.pdf.

[6] Microsoft, "Introduction to Windows Service Applications," Microsoft Developer Network, 2005, available at http://msdn.microsoft.com/en-us/library/d56de412(VS.80).asp.

Assembling an IT Infrastructure in Data Intensive Collaborative Projects in the Life Sciences

Stathis KANTERAKIS*
EMBL-EBI, European Molecular Biology Laboratory, Wellcome Trust Genome Campus
Hinxton, Cambridge, CB10 1SD, UK.

and

Maria KRESTYANINOVA
Institute for Molecular Medicine Finland FIMM, University of Helsinki
Helsinki, FI-00290, Finland.
*Corresponding author (kanterae@ebi.ac.uk)

ABSTRACT

Wide adoption of novel analytical technologies in genetics and advancements in data analysis that allowed to pool data from several data collections, have created a need for web-based data integration and harmonisation services [1-4]. Research consortia are seeking for IT platforms that can enhance the communication between biologists, statisticians, geneticists and clinicians and through this reduce the burden of data management and administration tasks. Design and implementation of a communication and data exchange platform is crucial for the efficient resource allocation and fruitful data analysis in large research consortia [5,6,7].

So far, most information systems for molecular, genetic, clinical and life-style data have been designed as long-term repositories with strict formats and requirements [8, 9]. The primary mission of such repositories is to preserve data for posterity in the most uniform fashion and make it available to researchers worldwide. Demands for short-to-medium term data deposition and assistance in data handling during the creative, discovery phase of a research project have not yet been addressed. Project-specific data management platforms scalable to population-size datasets and flexible enough to deal with very diverse biological, medical and life-style data are vital for researchers, since they speed up data exchange, annotation and integration for the context of a specific study [10, 11].

We present a novel framework for data intensive communications in large collaborative projects. First, we analyse current trends in collaborative knowledge generation, online information exchange and peer design, and then apply these principles in the life sciences, in the context of large collaborative studies. We come up with common use cases and corresponding software modules to enable such usage, and propose a vision for the design principles that should permeate collaborative biomedical software in the future.

Keywords: collaborative research, open design, omics, information management, biomedical studies, knowledge creation

1. INTRODUCTION

"On the one hand", claims writer Steward Brand in the late 1980s, "information wants to be expensive, because it's so valuable. The right information in the right place just changes your life. On the other hand, information wants to be free, because the cost of getting it out is getting lower and lower all the time"[12]. The latter is more relevant than ever, a quarter of a century later. We live in an era of abundant information. Since the sequencing of the first human genome, a map of our genetic "blueprint", and the advent of high-throughput technologies in biomedical research, information has exploded in the life sciences. This information is expensive; a stand-alone research laboratory can no longer afford the machinery and expertise to generate it. The cost of sequencing a single sample, while steadily declining, adds up as we try to detect increasingly rare genetic variants using larger cohorts. Information is also free; a push for open-data and open-science in the recent years, and the adoption of such standards from scientific journals as a requirement for publication means there is a lot of "free" information out there. Data volume and complexity are soaring faster than any analysis can hope to tackle. How do we distribute the financial burden of data that is costly to generate? How do we handle the increasing volume and complexity of available data that is sprinting further and further from our capacity to examine it? Furthermore, what does it take to convert information to knowledge, or to translate scientific findings to clinically relevant results? [13]

Open innovation

The revolution of open innovation in business, which arrived a decade ago, is only recently appearing in the life sciences. Open innovation refers to the model of allowing ideas to flow freely outside a firm's boundaries, becoming licensing deals or spin-offs, as well as from the outside into a firms boundary, becoming part of its IP portfolio and core development efforts [14]. What made this possible? The availability of private Venture Capital funding allowed ideas to materialise independently of a firm's primary focus. If the R&D department was not willing to fund a project, visionaries could look outside for their own funding. Good ideas got a chance of survival, while bad ideas were quickly replaced by better ones from the

market. The dawn of such funding opportunities in science, notably with SciFlies.org [15], a website which lists and promotes peer-reviewed projects for funding directly from the online community, means science is becoming in part more independent. Think tanks such as the Open Science Working Group advocate open publication (free of charge to the reader) as well as open deposition of scientific data to public repositories in a useable form [16]. The European Bioinformatics Institute (EMBL-EBI) is in the forefront of the open data effort, creating standards such as minimal information requirements, to make published data interpretable and usable, without the need to contact the primary source [17,18,19]. It also offers access to state-of-the-art data archives, free of charge [20,21].

Biological science is becoming more distributed. Because of the cost and complexity of modern "omics" research [22], it is unlikely that all resources and expertise to tackle a biological problem will exist under a single roof. For that reason, pooled research, such as in forming consortia, has become popular in modern biological research [1-4]. What technically makes data pooling and integration possible is consistent annotation. For that reason, terminologies, nomenclatures, ontologies and other types of controlled descriptors are being developed and maintained centrally, through curation and community involvement. When several datasets are merged for the sake of meta-analysis, this is often done through the application of a standard format (e.g. MAGETAB [23], ISATAB [24]) or simply through re-annotation and transformation of data to a common lexical denominator (DataShaper [11], SAIL [25], or tools for semantic tagging). Many such tools empower researchers with the means to carry out annotation tasks in their local laboratories, thus distributing the burden of curation from central repositories to the data source and local expertise. The differences, benefits and possible shortcomings of the open innovation model in biological research are outlined in Table 1.

Table 1. Comparison between open, distributed research and the traditional closed discovery model from the perspective of a single research institution.

	Closed model	**Open model**	**Benefits of the open model**	**Drawbacks of the open model**
Funding	Individual public funding	As part of a consortium / directly from the community	Increased accountability, peer pressure	Less competition, motivation
Scientific community involvement	Research behind "closed doors"	Intermediate results immediately available to the community	More discussion, feedback, cross-disciplinary communication	Proper attribution of credit for discovery, security, ethical issues
Curation	At central repositories, if at all	By local experts, at central repositories and community	Higher quality research	Increased scrutiny, especially against challenging the status quo
Visibility	Not a high priority	Standards enforced, tools for better contextualization, cross-linking of information	Objectivity, re-use of data and resources	More effort upfront
Value	Realized by a single laboratory	Realized by the whole community	Better return for funders' investment, better knowledge generation potential	Increased stress, push for publication rather than knowledge generation

Finally, promising community authorship attempts, such as wikigenes.org, which is built around the creed of proper author attribution, hint at the way scientific knowledge might be organised in the future. It is not a stretch of imagination to think of an era where the journal publication will be a from of rhetoric art, serving its own purpose, while scientific facts will be organised in a highly inter-connected, immediately and openly accessible medium and in a useful manner. This would not only save time and energy, but also propel scientific discovery. It would be analogous to a family collaboratively solving a gigantic puzzle, placing each piece straight into its proper place during each move.

A central point in the discussion about open data has always been the fear of its "incompetent use", who should be concerned about it, and which data is to be released: raw data, processed data or results? Interestingly, these three "levels" correspond to the three stages of knowledge creation, from raw material to information to knowledge. Scientists are often happy to disclose the latter two levels but reluctant to do so with the first, be it in fear of being swept by competition, uneasiness in releasing "garbage" data, infringement of legal or ethical regulations, such as to research subjects, or concern regarding proper acknowledgement for generating the data. This, by definition, cripples the ability of further information and knowledge creation since it forces data to be adopted in the view that the generator intended. This is understandable, in part, due to disincentives in place to promote such behaviour: scientists being judged by volume of publications and not necessarily quality, the inability to externalize costs related to production and enforcement of intellectual property rights, consent by research subjects to only particular kinds of research (e.g. diabetes) and the lack of apparent benefit (to the producing individual or group) in releasing well annotated data into the public domain. This problem is very complex to tackle in an atomistic way; there needs to be a coordination of policy making, IT support and culture change to achieve openness. However, we would argue that Information Technology is currently lagging furthest of the three and has the most potential to support such a change.

A shift from data generation to knowledge creation

In the model of expansive learning, the community is in the centre-point of knowledge-creation. It is not sufficient for individual players to interact; but to co-create shared artefacts of knowledge. When we speak of such "trialogical" [26] mode of learning, we refer to combining the following three elements: (a) individual competencies, (b) communal participation and (c) co-creation. The danger of focusing on the first two without the latter is that knowledge may become reductionist, not expansive. When researchers tackle scientific questions in isolation or rely on a community to help answer them, inquiry is reduced to an individual mental process or a social process of participation. In "trialogical" learning, individual initiative serves the communal effort to create something new, and the social environment feeds individual initiative and cognitive growth. Thus, communal knowledge becomes materialized as common objects of activity are developed. An application of the above principles has powered the creation of the Knowledge Practices Environment (KPE) platform [27], used successfully in numerous collaborative learning projects [28,29,30].

Given time and resource restrictions few software providers can afford to develop from scratch. Thankfully, the open-source movement has now reached the mainstream and with immense success. If we are to expand on current knowledge using funds

in an efficient way, our utilisation of software cannot be any different. Recently available software toolkits such as Jango, jQuery, Google tools such as Google Docs or Google Refine and collections of widgets such as Bootstrap or BioPortal [31], to name a few, make it easy to develop functional software with little overhead. Even better yet, existing biomedical information management suites such as SIMBioMS [32], ISA tools [23], OBiBa (obiba.org) or BioMart [33] implement much of the functionality that will be discussed later. The main focus of software in the biomedical segment should therefore be integration, adherence to standards and interoperability, rather than lengthy generalised solutions, which risk being rendered obsolete as high-throughput technologies progress. If facilitating knowledge creation in highly complex environments is the goal, software groups should think more about designing for integration and communication rather than for solutions to problems.

DESIGN IS FUNCTION

In this section we wish to introduce recent advances in the design field, relevant to collaborative biomedical communities, such as open peer-to-peer design [34]. This design methodology is a promising community-based organizational form, building short and long collaborative networks with high probabilities of achieving success in society. Popular examples of similar efforts include Linux, Wikipedia, YouTube and other Web 2.0 communities. They represent, maybe, the only participation-based organizational forms with high scalability: the more the participants, the faster they achieve success. High participation however, means high complexity and to understand such complexity means to design in and for complexity itself. The Linux community arguably succeeded in this, because it faced the challenge of designing an operating system without reducing it, but by leveraging its own intrinsic complexity. The complexity of the project thus reflects the complexity of the community, and both strengthen each other.

The open peer-to-peer design process is a co-design process, where designer and participants collaborate in a wider design community (a collective intelligence). Designing software in such a context is an *enabling* act; not an act of delivering a solution. The chances of coming up with a better design are higher in such a scenario; and the better the design the more use the community extracts out of it. Thus design and function become synonymous terms.

COLLABORATIVE SOFTWARE IN BIOMEDICAL SCIENCES

We envision the next generation of information management software in the life sciences to be inspired by the following ideas:

Enabling complexity. Scientific communities should be given tools to self-organise and harness collective intelligence. These tools should not only support the exchange of data and information, but also enable the co-creation of knowledge, for example by means of group document editing, group discussion or distributed annotation.

A highly ethical exchange. While it is beneficial for data, information and knowledge to flow freely within a scientific community, there need to be means for acknowledging authorship of these scientific artefacts in a trackable way.

Standards, integration and networking. Each knowledge artefact needs to be organised into an appropriate context. Adherence to standards, such as minimum annotation requirements or exchange formats and integration with existing resources, such as through the use of URIs should be the bare minimum. Developing a network of partners can also enhance this process, as expertise and software components can flow in and out of the network for the benefit of integrating the most relevant software modules.

Commitment to openness. Open-source software and online collaborative communities are an unprecedented example of how Information Technology can harness collective intelligence. The next generation of information management software in biomedical studies should promote this idea in this field as well, trying to alleviate the shortcomings previously outlined.

We now shift our focus on how these requirements can be put into practice through the design of an IT infrastructure to support collaborate biomedical projects.

Software design in biomedical studies

Since the demand for distributed data management software grows constantly -both in terms of number of project and number of knowledge domains- it has become possible to present a consensus workflow. It includes researchers, data managers, information administrators (e.g. handling confidentiality information) and system administrators. If a discussion of such a generic workflow can be taken beyond the specifics of a particular knowledge domain, it can pave the way to creating a more sustainable communication infrastructure for the research community.

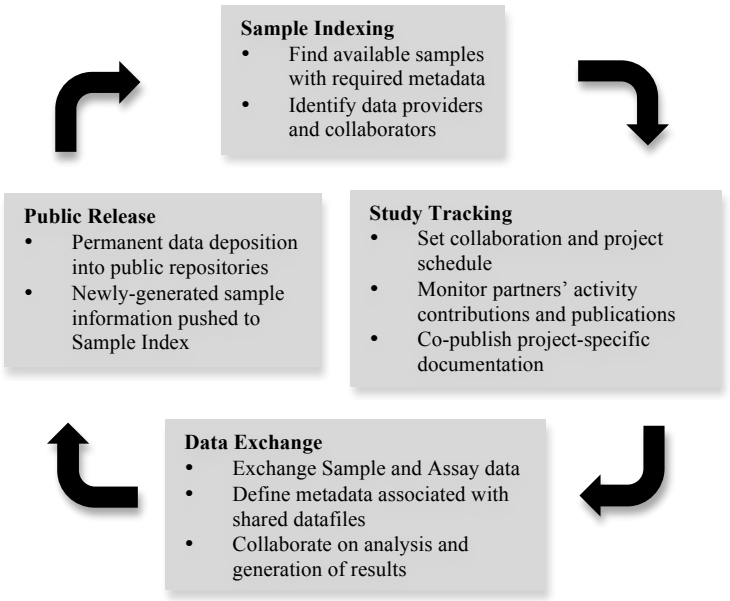


Figure 1. A set of common data management use-cases in biomedical studies, modelled into four distinct modules: Sample Indexing, Study Tracking, Data Exchange and Public Release

3.1.1. Sample Indexing. Research coordinators, looking for partners that can provide samples with high quality annotation that fulfils the needs of a study, start by querying a publicly accessible sample database. After selecting the list of required parameters (specific phenotypes, availability of

genotype data, etc) from the query interface, researchers are presented with a report that will show the number of available samples that satisfy the list of requirements from all registered data providers. Using the same system, the research coordinator is able to contact the data providers in question to initiate a collaboration and gain access to the sample data.

3.1.2. Study Tracking. On definition of study collaboration, the research coordinator makes use of a study-tracking module to add the list of participants to the project and to establish the schedule for the project. The study coordinator and the researchers are able to follow the progress of the project and monitor the activity of each of the partners. Each collaborator also uses this tool to publish and retrieve procedural information related to the study (protocols, publications, description of study, data access application, etc). The system will also allow study coordinators to add or remove partners and to set the access rights for each partner. Investigators may interact with various modules of the software depending on their role; *e.g.* biomaterial management, lab analysis, data analysis, *etc*. to access multiple services (databases, sftp, mailing lists, *etc.*).

3.1.3. Data Sharing. This module helps research coordinators to define their required data structure and system configuration through personal consultations, the study of data files and the capture of critical metadata descriptors. The product of this initial analysis is a customized installation of sample and/or assay databases with web forms and standard templates for data capture. Changes to web forms and vocabularies can be made by a researcher or an administrator using the graphical user interface. Configuration changes can be made quickly, making it possible to optimize the system in a timely manner. Once the system is configured, data upload/download and browsing is straightforward. Users can upload data manually using input web forms or they can use a customised template to batch upload sets of files of raw or processed data.

Additionally, analysis and visualisation pipelines can be connected to the data module as plug-ins. These plugins cater to needs for specific pre-processing, quality control and production pipelines, but also to individual research labs, as a means to publish in-house tools in a consistent and shareable manner. Results of these tools are fed back into the database and shared within the research community.

3.1.4. Public Release. Due to the requirement of many scientific journals of public access to research data, data sharing modules provide a solution to export the selected information directly to public repositories, such as the European Genome-phenome Archive, eliminating the need for the user to resubmit all the data to the final public repository manually. Ideally the use cycle of a biomedical study will close with the addition, by the research coordinator, of any newly generated sample data back into the Sample Index. This will increase the chances of reusing the sample data in new projects, improving the return of investment to public research funds by maximizing the number of publications that come out of a set of data.

CONCLUSIONS

Availability of Information Technology to enable it, pressure from public and industrial domains that have already adopted it, as well as increasing complexity, costs and demand for sustainability, are calling for open innovation in the life sciences. While there are tremendous benefits from harnessing the collective intelligence of communities to create knowledge from data collaboratively, there are disincentives in place to prevent open innovation from hitting the mainstream in life science research. Most notably, increased competition, fear of infringement of legal or ethical regulations, such as subject consent, or improper acknowledgement to the group that generated the research data. We believe IT can be in the forefront of the push for open biomedical science, by designing software pervaded by the following principles: enabling complexity, supporting ethical exchange of information, adhering to standards, integrating and networking whenever possible and being committed to openness by actively advocating it. We have provided a use-case information flow based on our experiences with data-intensive collaborative projects in the life sciences and hope to generate awareness and support in the biomedical software design community for enabling and integrating such beneficial collaborative research workflows.

ACKNOWLEDGEMENTS

We would like to thank Yulia Tammisto and Ugis Sarkans for fruitful discussions, advice and guidance. We also thank entire simbioms.org network for hard work and systematic gathering of usage data from 9 EU projects.

Conflict of interest statement: None declared.

Funding

This work has been funded by IP EC projects SIROCCO (LSHG-CT-2006-037900) and ENGAGE (grant agreement No: 201413).

REFERENCES

[1] T. Thorgeirsson, et al. **Sequence variant at CHRNB3-CHRNA6 and CYP2A6 affect smoking behaviour**. Nature Genetics (2010), 42(5):448 - 453

[2] M. Kolz, et al. **Meta-Analysis of 28,141 Individual Identifies Common Variant within Five New Loci That Influence Uric Acid Concentrations**. PLOS Genetics (2009), 5(6):e1000504.

[3] I. Prokopenko, et al. **Variants in MTNR1B influence fasting glucose levels**. Nature Genetics (2008), 41:77-81.

[4] A. Ripatti, et al.: **Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts**. Nature Genetics (2008), 41:47-55.

[5] A. Burgun, O. Bodenreider. **Accessing and Integrating Data and Knowledge for Biomedical Research in IMIA Yearbook 2008: Access to Health Information**, (2008) pp. 91-99.

[6] S. Oster. **CaGrid 1.0: an enterprise grid infrastructure for biomedical research**, JAMIA 15 (2008), pp. 138–149.

[7] P. McConnell, et al. **The cancer translational research informatics platform**. BMC Med Inform Decis Mak (2008) 24;8:60.

[8] C.F. Taylor, et al. **Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project**. Nature Biotechnology (2008), 26(8):889-896.

[9] B. Smith, et al. **The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration**. Nat. Biotechnol. (2007), 25:1251–1255.

[10] D.B. Keator, et al. **Derived Data Storage and Exchange Workflow for Large-Scale Neuroimaging Analyses on the BIRN Grid**. Front Neuroinformatics. (2009) 3:30. Epub 2009 Sep 7.

[11] I. Fortier, et al. **Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies**. Int. J. Epidemiol. (2010), 39:1383-1393.

[12] S. Brand. **The Media Lab: Inventing the Future at MIT**. New York: Viking/Penguin, 1987.

[13] N.R. Anderson, et al. **Issues in Biomedical Research Data Management and Analysis: Needs and Barriers**. J Am Med Inform Assoc. (2007), 14(4):478–488.

[14] H.W. Chesbrough (2003). **The era of open innovation**. MIT Sloan Management Review, 44 (3), 35–41

[15] **Gift System for Science**. Nature, Vol. 480, No. 7376. (7 December 2011), pp. 281-281.

[16] Molloy JC. **The open knowledge foundation: open data means better science**. PLoS Biol. 2011 Dec;9(12):e1001195. Epub 2011 Dec 6.

[17] A. Brazma, et al. **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data**. Nat Genet. 2001 Dec;29(4):365-71.

[18] N. Le Novère, et al. **Minimum information requested in the annotation of biochemical models (MIRIAM)**. Nat Biotechnol. 2005 Dec;23(12):1509-15.

[19] C.F. Taylor, et al. **The minimum information about a proteomics experiment (MIAPE)**. Nat Biotechnol. 2007 Aug;25(8):887-93. Review.

[20] A. Brazma, et al. **ArrayExpress--a public repository for microarray gene expression data at the EBI**. Nucleic Acids Res. 2003 Jan 1;31(1):68-71.

[21] M. Kapushesky, et al. **Gene expression atlas at the European bioinformatics institute**. Nucleic Acids Res. 2010 Jan;38(Database issue):D690-8. Epub 2009 Nov 11.

[22] D. Field, et al. **'Omics Data Sharing**. Science (2009), 326(5950):234-236.

[23] T.F. Rayner, et al. **A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB**. BMC Bioinformatics (2006), 7:489.

[24] P. Rocca-Serra, et al. **ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level**. Bioinformatics (2010), 26(18):2354-2356.

[25] M. Gostev, et al. **SAIL: A software system for sample and phenotype availability across biobanks and cohorts**. Bioinformatics (2011), 27(4):589-591.

[26] S. Paavola and K. Hakkarainen. **The Knowledge Creation Metaphor - An Emergent Epistemological Approach to Learning**. Science & Education (2005), 14(6):535-557.

[27] H. Markkanen, at al. **The Knowledge Practices Environment: a Virtual Environment for Collaborative Knowledge Creation and Work around Shared Artefacts**. In J. Luca & E. Weippl (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (2008) pp.5035-5040. Chesapeake, VA: AACE.

[28] K. Hakkarainen. **Three generations of research on technology-enhanced learning**. British Journal of Educational Technology (2008). Volume 40, Issue 5, pp.879–888, September 2009

[29] A. Lund. **Assessment made visible: individual and collective practices**. Mind, Culture, and Activity (2008). 15:32-51.

[30] L. Lundvoll Nilsen and A. Moen. (2008). **Teleconsultation – collaborative work and opportunities for learning across organizational boundaries**. Journal of Telemedicine and Telecare, 14, 377-380.

[31] N.F. Noy, et al. **BioPortal: ontologies and integrated data resources at the click of a mouse**. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W170-3. Epub 2009 May 29.

[32] M. Krestyaninova, et al. **A system for Information Management in BioMedical Studies-SIMBioMS**. Bioinformatics (2009), 25(20):2768-2769.

[33] D. Smedley, et al. **BioMart – biological Queries made easy**. BMC genomics (2009), 10:22.

[34] M. Menichinelli. **openp2pdesign.org_1.1. Design for Complexity**. (2008), openp2pdesign.org.

Biobank Metaportal to Enhance Collaborative Research: sail.simbioms.org

Maria KRESTYANINOVA,
FIMM Institute for Molecular Medicine Finland, Helsinki University
Helsinki, FI-00014, Finland

And

Ola SPJUTH
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet,
Box 281, SE-171 77, Stockholm, Sweden

And

Janna HASTINGS, Jörn DIETRICH, Dietrich REBHOLZ-SCHUHMANN
EMBL-EBI, European Bioinformatics Institute, Wellcome Trust Genome Campus,
Hinxton, CB10 1SD, United Kingdom

ABSTRACT

In order to identify new ways to prevent, diagnose and treat diseases, biobanks systematically collect samples of human tissues and population-wide data on health and lifestyle. Efficient access to population biobank data and to biomaterial is crucial for development and marketing of new pharmaceutical products, especially in the area of personalised medicine. However, such access is hindered by legal and ethical constraints, and by the huge semantic diversity across different biobanks. To address these challenges, we have developed SAIL, a sophisticated metaportal for biobank data annotation across different collections and repositories, harmonised to allow cross-biobank searchability, while preserving the anonymity and privacy of the underlying data such that legal and ethical requirements are met. We describe the technological architecture and design of SAIL that allows us to meet these pressing challenges, and give an overview of the current functionality of the application. SAIL is available online at sail.simbioms.org, and it currently contains around 200 000 samples from 14 collections.

Keywords: Biobanks, Resource Discovery, Biomedicine, Metadata, Ontologies, Semantics

1. INTRODUCTION

Biobanks

Research at the frontier in the fight against pressing human conditions such as cancer relies heavily on the availability of sample biomaterial for broad populations in order to adequately evaluate research hypotheses and develop novel treatments [1]. Biobanks are large-scale sample repositories addressing this need with the objective of identifying new ways to prevent, diagnose and treat diseases, as well as that of gaining a better understanding of the lifestyle and nutrition factors that optimize human health.

Biobanks systematically collect population-wide samples of human tissues together with data on health and lifestyle, and make these materials available to the scientific research community, while guarding the privacy of the sample donors by navigating the challenging ethical and legal considerations involved in dealing with human samples. Such collections contain millions of tubes with primary biomaterial in a storage container (freezer), and associated information records about millions of people and thousands of measurements, often carried out in a longitudinal fashion.

The outcomes of biobank-based studies are of great value for healthcare, academia and biomedical industry [2, 3].

Ethical and Legal Considerations Affecting Access

Efficient access to population biobank data and to biomaterial is crucial for realization of the research potential of the valuable samples, in particular in the development and marketing of new pharmaceutical products, with population-wide samples delivering breakthroughs especially in the area of personalised medicine [4].

However, due to ethical and legal constraints, biobanks are not at liberty to release their data or share biomaterial without the approval of a local access committee, tasked with ensuring that ethical considerations are met and that legal and privacy requirements will be addressed, on evaluation of an intended research proposal. This leads to a "Catch22" situation, since a biobank is not in a position to release any data until the purpose and design of the study is presented and approval is granted,

while parties interested in performing studies need to know what data is available at the time of study design in order to inform their research proposal and determine which biobanks contain data which are suitable for the scope of a proposed study [5]. This processual challenge directly impacts the translational value of the sample collection, but has the potential to be addressed by a sophisticated technological solution, one example of which we will present here.

Semantic Diversity across Biobanks

The challenge of obtaining access to the data and biomaterials from a single biobank is not the only challenge which researchers need to overcome in the pursuit of research involving human samples. To obtain statistical effectiveness for a particular research question, it is often necessary to utilize samples and data from more than one biobank [1], exposing a difficult challenge in semantic heterogeneity across different biobanks. Biobanks have to meet diverse research targets, collecting different sorts of samples and data points from populations in order to address varying issues, and furthermore are situated in differing countries with differing regulatory contexts and languages.

Different types of biobank include population banks, prioritizing biomarkers of susceptibility and population identity for a concrete country, region or ethnic cohort; disease-oriented epidemiological banks, focused on biomarkers of exposure, with specifically designed often longitudinal samples and data; and disease-oriented general biobanks such as tumour banks, focused on biomarkers of disease through tumour and non-tumour samples associated to clinical data and sometimes associated to clinical trials [1]. The diversity of types of biobanks, and the diversity of populations and diseases for which samples and data are being collected, easily result in excessive diversity across the sample annotation leading to low interoperability.

Furthermore, original sample annotations, captured at the time of collection, come in a variety of formats and languages, with there being no universal standard in common use [6,7]. This issue is further complicated by the fact that various types of specialists (medical doctors, statisticians, geneticists and others) are accustomed to different technical vocabularies and the use of differing language conventions to communicate about their work [8]. The inevitable result is that sample annotations can diverge even when those annotations are intended to capture the same semantic semantics (meaning). Thus, in order to determine whether data exists for a particular research question across different biobanks, there is a costly and repetitive data management process involved at every stage: selective tagging, mapping and interlinking of various types of sample descriptions, commonly referred to as *harmonisation* [1]. Technically, these descriptions are implemented via ontologies, controlled vocabularies, free text, database identifiers and other reference utilities, and may come in a multitude of underlying formats (RDF, XML, OWL). Such vocabularies may be internal (biobank-specific) or external (such as when using community standards).

This semantic diversity of biobank annotations is a fundamental problem for the exposure of biobank content to meet research needs and harness the potential of the biobanking for translational research.

2. SAIL – A TECHNOLOGICAL SOLUTION

Sail is the biomedical informatics solution to the abovementioned problems of access to biobank information and semantic diversity across different biobanks, which we believe will assist in building efficient research communities and ultimately lead to a more efficient translation of biobank resources into improved healthcare and treatment options for patients, which takes the form of a central and controlled metaportal for data release by biobanks to potential and existing partners.

SAIL (sail.simbioms.org), the Sample avAILability System, is an web-based resource, which allows researchers to locate and estimate the amount of relevant biomaterial available from a sample collection. SAIL provides information for each sample on whether a value for a given phenotypic variable exists or not, without storing or disclosing the value per se. Phenotypic variables are organised in controlled vocabularies, taxonomic structures and studies.

The resource has been successfully used for retrospective harmonisation of phenotypic information from hospitals and biobanks, and it currently contains references to 200 000 samples from 14 collections [9]. The current version of SAIL allows creating, editing and relating new terms and vocabularies with subsequent loading of sample availability data annotated with these descriptors. Due to the links between synonymous variables, e.g. equivalent measurements with different labels, and to the annotation structure (timepoint, type of measurements etc), samples can be searched for by a variable per se, e.g. ‘glucose’, as well as by a more specific statement, e.g. ‘fasting glucose’. Furthermore, the visibility of samples from a certain collection can be increased by additional classification of variables that are used to characterize the samples: by assigning a variable to a vocabulary, a study or a canonical phenotype. Such visibility reveals new opportunities to highlight the scientific value of biobank content, e.g. identifying samples that have been used in many studies or those which have rare phenotypes or data associated with them.

The SAIL mission as an online resource is to increase the visibility of the biobank content and to ease the set-up of population-wide genetic and molecular studies and to enhance collaborative research. In the remainder of this communication, we describe the features of the SAIL system and show how technological solutions are found for the underlying challenges of access and diversity.

3. HARMONISATION AND SEARCHABILITY

SAIL provides 1) an interface for harmonisation and submission of sample and phenotype information that is available in various biobank collections; and 2) a search engine for surveying which data from which cohorts could be combined for specific tasks such as study construction and sample selection. SAIL is a database that is populated with information about metadata and availability of biomaterial at within various collections. To enable early access or gradually adjusted access to the data and avoiding the “Catch-22” limitation, SAIL makes the data *discoverable* – that is, it is possible to search for samples which contain annotations of a specified type – without making the data *publicly available* (which would, of course, violate the legal and ethical constraints governing the use of such sensitive data).

To our knowledge, SAIL is the first platform that facilitates resource discovery across biobanks at the level of a single individual samples, rather than presenting summary content of for an entire collection, as well as being the first comprehensive solution for semantic indexing and harmonisation of sample and phenotypic variables between different repositories. It assists in the set-up of large scale genetic studies and raises awareness about the scientific value of biobank data by making the data easy to locate, interpret and incorporate into a study.

The database consists of two parts: vocabularies and samples. 'Vocabularies' are collections of terms which are specific to a study (medical topic) or to a collection of samples. The syntax used for description of terms is universal throughout the database, thus allowing linking terms across vocabularies or studies. In this fashion, external shared vocabularies and ontologies can be integrated with internal biobank-specific vocabularies. The use of external vocabularies and ontologies for semantic annotation conveys several benefits: firstly, the external vocabularies are often already shared across a community and may be used in annotation of knowledge base resources such as pathway, gene and protein databases, easing the path from hypothesis generation to sample selection; secondly, the external vocabularies are maintained outside of the biobank project thus easing the burden of internal maintenance; and finally, being community-wide, the resource is neutral between the different biobanks, easing the burden on integrated searching. Examples of relevant external ontologies are the Gene Ontology (GO; [10]), the Phenotype and Trait Ontology (PATO; [11]) and the Human Phenotype Ontology (HPO; [12]). However, gaps in external resources can still be filled by internal biobank-specific and SAIL-wide vocabularies, as the system is flexible enough to accommodate both, thus preventing any delays to annotation that might have been caused by dependence on external resources.

The other component of the database is the 'samples', which are references to sample IDs through vocabulary terms, allowing semantic searchability across the wide range of different samples from different biobanks.

4. SYSTEM DESCRIPTION

The SAIL software is implemented as a client-server application. The client part is developed with Google Web Toolkit (GWT) and the Ext-JS widget library, and runs in a regular web browser. The server part is written using Java servlet specifications and runs within a Tomcat web application container.

The first prototype of the system was released after the initial dataset was collected. All subsequent developments and implementations have been done as a continuous iterative process of consultations with users, uploading data, testing and releasing upgraded versions of the interface. SAIL has been designed particularly for availability data, and to answer questions such as 'How many samples across all available cohorts have measurements available for plasma levels of fasting glucose and HDL cholesterol, and records of clinical diagnosis of type 2 diabetes, as well as a body mass index (BMI)?' Each such variable describing a sample, a cohort, an experiment or a measurement type is stored in the SAIL system as a parameter. Sets of parameters can be grouped together, such as parameters annotated using the same vocabulary. Parameters can contain information beyond simple descriptive annotations by using qualifiers and variables. These can store assay and sample preparation information, or specify different measurement types associated with each parameter. To facilitate the harmonisation of sample parameters contributed from different sources, it is possible to define relations between parameters, specifying the level of synonymy or overlap in parameter definition.
The main view of the SAIL system is the Report Constructor (Figure 1).

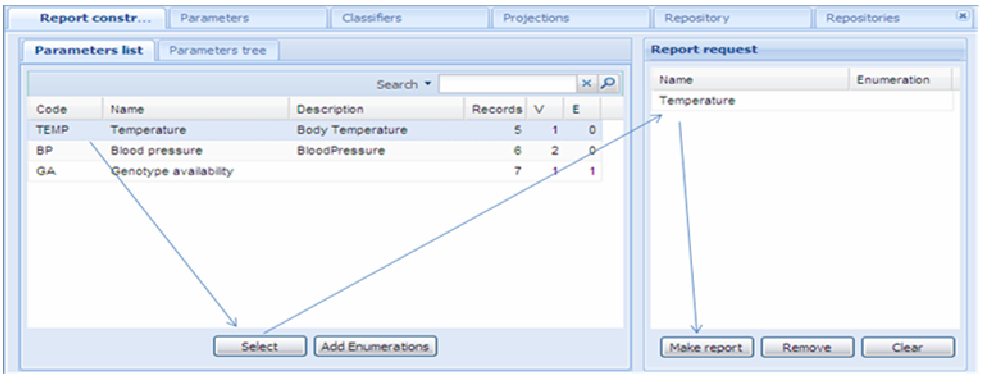
This view consists of a parameter list and a report request. Queries are constructed by selecting parameters in the list, and adding them to the query structure which will appear in a graphical manner within the report request window. Complex queries can be formulated by addition of many parameters, selected variants of parameters (such as only samples with fasting glucose concentration), and by combining AND and OR logic. Very complex queries can also be pre-defined to facilitate later analysis. Quick single-parameter queries across all cohorts are available. The query result is reported as a table (Figure 2), detailing the number of samples for each cohort fulfilling the query criteria and the final result of the combined parameters.

The list of parameters can be additionally filtered by free text filter, as well as filters for specific tags of classifiers, such as for a specific vocabulary. Filters for samples only included in specific studies or specific cohorts can also be added. Overviews are also available, providing full information about all available phenotypes for samples included in a study or a cohort. An important part of the functionality is the parameter view, where new parameters can be added and edited, creating the annotation structure. The flexibility of the data structure allows for complex parameters with layers of annotations and relations to other parameters. This allows for import of any hierarchy or directed acyclic graph (DAG) structured ontology.

In addition to availability data, the SAIL system can also handle actual data values, and contains tools for using, extending and harmonising vocabularies that describe the samples, experiments and phenotypes. Ontologies such as the Experimental Factor Ontology (EFO) [http://www.ebi.ac.uk/efo] or the ontologies developed under the Open Biomedical Ontologies (OBO) [http://www.obofoundry.org] umbrella can be uploaded, as well as user defined vocabularies. It benefits from other data harmonisation efforts, such as the DataSHaPER project at the Public Population Project in Genomics (P3G) [http://www.datashaper.org] and Promoting Harmonisation of Epidemiological Biobanks in Europe (PHOEBE) [http://www.phoebe-eu.org].
For a more detailed description of the functionality and specific features of the system, see User Guide at http://www.simbioms.org/software/SAIL .

The SAIL system is developed as open source and distributed by SIMBioMS with the AGPL license. Code, tutorials and documentation are available at http://www.simbioms.org/software/SAIL/ which also hosts an installation containing availability data contributed for the European Network for Genetic and Genomic Epidemiology (ENGAGE) project [http://sail.simbioms.org/]. We encourage cohort owners and study co-ordinators to contact us at support@simbioms.org for submissions.

a)



b)

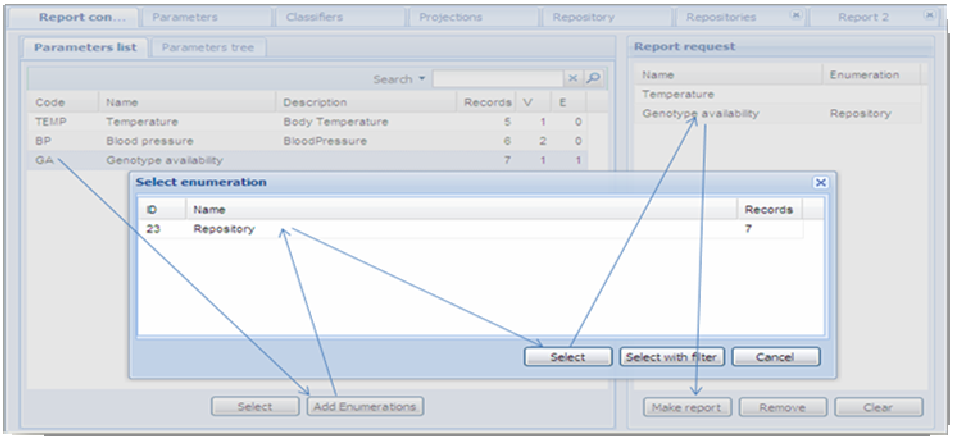


Figure1. *Constructing a report. a) parameter as a filter: all samples which have value recorded for this variable are counted in b) enumerated values as a filter: for each of the values number of samples is calculated;*

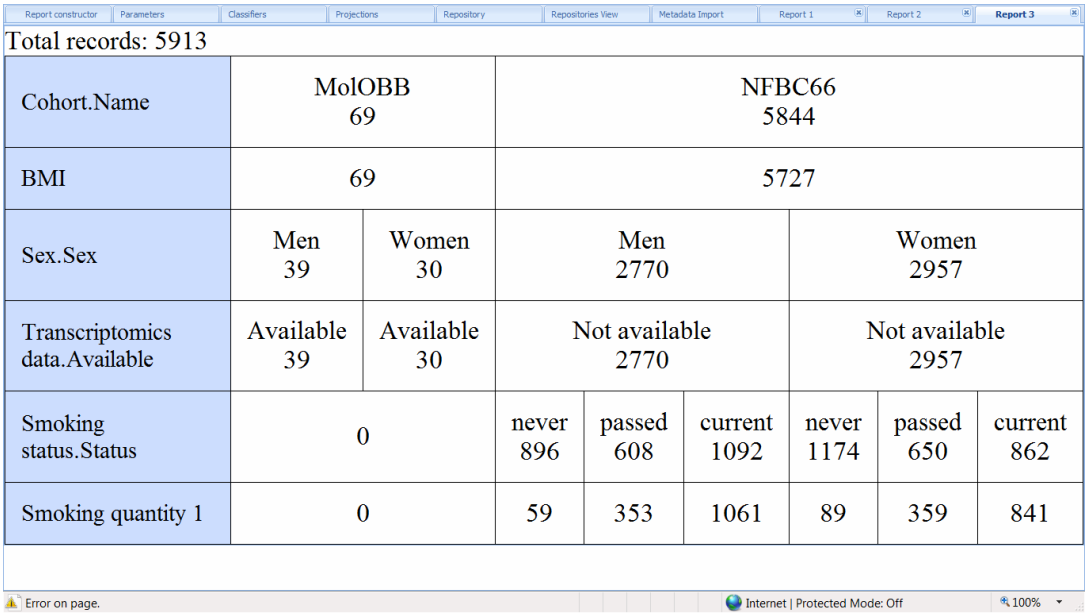


Figure 2. *Viewing report*

5. SAMPLE INCORPORATION PROCESS

Incorporation of biobank sample metadata into the SAIL system allows exposure of that data to the broader research community, increasing the impact of the biobank resources. However, to fully maximise the benefit of the searchability and harmonisation of the metadata across the SAIL database, it is often necessary to *re-annotate* the data as it is being incorporated, in order to enhance searchability and maximise exposure of samples. This is particularly the case where, for example, original sample annotation is in a national language and not enhanced with internationally accessible synonyms. Re-annotation also allows maximum application of shared controlled vocabularies and ontologies, pre-harmonising and thereby reducing the subsequent time taken for harmonisation in early phase study preparation.

The first prototype of SAIL was test-run on a cumulative index of samples from 10 collections. The index was based on 87 variables, which were suggested by data analysts from Oxford University and FIMM working on identification of genetic markers for such diseases as type 2 diabetes and cardio vascular disease. Selected variables of interest were grouped in a Metabolic Syndrome (MetS) vocabulary. The initial format for the description of terms (name, definition, unit, time point, etc.) was suggested by epidemiologists and subsequently cross-checked against the standard format proposed by DataSHaPER [7], the major international provider of standardised dataschemas for harmonisation in population genetics and epidemiology.

Upon finalisation of the harmonised MetS vocabulary, the local data managers at each collection mapped local sample descriptions (variables) to MetS, extracted sample data from the biobank database for those samples which were relevant to at least some of the variables in MetS, in the extracted matrix replaced the values with 1 and missing values with 0, and sent the availability matrix to the SAIL development team.

The second batch of data was submitted by cohorts which were not part of the ENGAGE consortium. Data was either provided in the MetS vocabulary or in case of a different clinical scope in other vocabularies. In the latter case, related variables from different vocabularies were linked in SAIL.

A pressing concern for the usability of the informatics solution provided by SAIL is the ease with which data providers (submitters) are able to re-annotate their data in the submission process, in particular considering that biobanks are frequently not resourced for on-going metadata management. We are presently in the process of developing a sophisticated intelligence-based annotation suggestion facility, based on the NCBO BioPortal collection of biomedical ontologies and controlled vocabularies [13]. The facility will combine a search across term names and synonyms throughout the BioPortal collection of ontologies with a sophisticated ranking system which places the most relevant terms highest.

6. DISCUSSION

As more effort and resources are brought together to increase the scientific value of biomedical samples, it is important to address the new information management needs created by the size and complexity of the collected data, and by the increasingly distributed character of research projects. With great disparity between different cohorts and biobanks, there is a risk that existing data or biomaterial are not used to the extent that they could be, or that the results from studies based on these collections are not comparable or combinable. The efforts to collect and record highly complex data must be complemented with systems that can make this content accessible and understandable, maximising its value and usability.

While structures of biobank databases are usually optimised for keeping information consistent and complete in the long-term, architecture of a system for cross-biobank harmonisation has to facilitate the mapping process in a variety of contexts, and therefore has to offer a semantically normalised structure, e.g. controlled vocabularies or taxonomic structure, suitable for phenotypic variables of wide variety. In order to keep track of harmonised variables and interlink vocabularies, classification of variables and their relationships has to be multidimensional, in a sense of multi-label classification, and has to allow for rich biomedical contextualisation. In SAIL we have attempted to provide in a single software application a solution for creating a semantic space, tagging samples with various standardised terms including those sourced from external ontologies and vocabularies, and enabling sophisticated querying and searching, thus facilitating resource discovery.

It would be of great benefit to integrate data from different quality registries, as this not only enables merging and comparison of data from different diseases but also allows linking clinical observations to biobank data. Such solutions open up opportunities for new types of studies, such as including genotype data when studying treatment success. As registries and biobanks traditionally are both geographically as well as operationally separated, SAIL has the possibility to enhance biobank research by bringing these data into a single platform, and we envision that this will be widely adopted in the future.

Facilitation of resource discovery in a cross-disciplinary fashion for the data that requires controlled access is a task that is currently being solved across many knowledge domains. The holy grail of communicating across borders brings a difficult choice between the tedious work of describing in great detail, and often in several languages, ‘what is stored where’, or making everything available to everyone. In the case of biobanks the data access is restricted for ethical and legal reasons, so full open access is not possible. At the same time the potential brought by the data and biomaterial for health and pharmaceutical research cannot be overestimated. Thus, the SAIL system enhances the communication between biobanks and the research community, enables collaborative research, and facilitates the maximal impact of the valuable resources stored in the biobanks for translation into primary research results and ultimate patient benefits.

7. CONCLUSIONS

By operating on the metadata level, SAIL enables harmonisation of biobank data and assists in the construction of population-wide meta-studies. This places SAIL in a new informatics niche, not focusing on recording all data at the finest level of detail, but instead providing a way to browse, summarise and manage results from such databases, even if these are individually complex and highly diverse.

Much of the success of SAIL depends on harnessing the ongoing community efforts to build biomedical ontologies and vocabularies. Annotation with community-wide ontologies allows integrated searches to be performed across disparate data sources, and maximizes visibility for both primary data and research results. SAIL itself is not an ontology-building tool, but a semantic annotation and indexing platform that can be used to extend, and interlink the semantic information from associated with biobank data in such a fashion as to enable the sort of wide-ranging and interdisciplinary studies to be performed using biobank data that will drive the next generation of medical science.

Acknowledgements
This work was supported through funds from the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE consortium, grant agreement HEALTH-F4-2007-201413. JH is partially supported by the European Union under EU-OPENSCREEN, Work Package 2 (Standardization). OS is supported by the Swedish e-Science Research Center (SeRC).

We thank the projects and centres that have so far provided data: Diabetes Genetics Initiative (DGI, Broad Institute of Harvard and MIT, Lund University, Novartis Institute of Biomedical Research); Erasmus Rucphen Family (ERF, Erasmus Medical Centre, Rotterdam); Cooperative Health Research in the Region of Augsburg (KORA-gen, Helmholtz Zentrum München); UK Twin database, King's College London; Estonian Genome Project (EGP, University of Tartu); Swedish Twin Registry (STR, Karolinska Institutet); Metabolic Syndrome subcohort of the Health 2000 Survey (GenMetS, National Institute for Health and Welfare, Finland, and Queen's University Belfast) (please check affiliations); Northern Finland Birth Cohort 1966 (NFBC 1996, Imperial College London, University of Oulu); Oxford Biobank (MolOBB, Oxford University).

8. REFERENCES

[1] P. H. J. Riegman et al., "Biobanking for better healthcare", **Molecular Oncology,** Vol. 2 No. 3, 2008, pp. 213–222.

[2] M. I. McCarthy, et al., "Genome-wide association studies for complex traits: consensus, uncertainty and challenges", **Nature Reviews Genetics**, Vol. 9 No. 5, 2008, pp. 356-369.

[3] M. Yuille, et al., "Biobanking for Europe", **Briefings in Bioinformatics**, Vol. 9 No. 1, 2007, pp. 14-24.

[4] F. Kauffman and A. Cambon-Thomsen, "Tracing Biological Collections: Between Books and Clinical Trials". **Journal of the American Medical Association**, Vol. 299, No. 19, 2008, pp. 2316-2318.

[5] G. Helgesson et al., "Ethical framework for previously collected biobank samples", **Nature Biotechnology**, Vol. 25 No. 9, 2007, pp. 973-976.

[6] P. Founti, et al., "Biobanks and the importance of detailed phenotyping: a case study-the European Glaucome Society GlaucoGENE project". **British Journal of Ophthalmology**, Vol. 93 No. 5, 2009, pp. 577-581.

[7] I. Fortier, et al. "Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies", **International Journal of Epidemiology**, Vol. 39 No. 5, 2010, pp. 1383-1393.

[8] I. Hirtzlin et al., "An empirical survey on biobanking of human genetic material and data in six EU countries", **European Journal of Human Genetics**, Vol. 11 No. 6, 2003, pp. 475-488.

[9] M. Gostev et al., "SAIL—a software system for sample and phenotype availability across biobanks and cohorts". **Bioinformatics**, Vol. 27 No. 4, 2011, pp. 589-591.

[10] M. Ashburner et al., "Gene Ontology: tool for the unification of biology", **Nature Genetics**, Vol. 25, 2000, 25-29.

[11] G. V. Gkoutos et al., "Using ontologies to describe mouse phenotypes", **Genome Biology**, Vol. 6 No. R8, 2004.

[12] P. N. Robinson, S. Mundlos, "The Human Phenotype Ontology", **Clinical Genetics**, Vol. 77, 2010, pp. 525–534.

[13] N. F. Noy et al., "BioPortal: ontologies and integrated data resources at the click of a mouse", **Nucleic Acids Research**, 2009, doi:10.1093/nar/gkp440.

An Acoustic Monitoring System for Aircraft using Multiple Microphones

James Gil de Lamadrid
Bowie State University
14000 Jericho Pk Rd
Bowie, MD 20715

ABSTRACT

A pilot can glean significant amounts of information concerning the status of an aircraft from simple audible sound data. In this paper we discuss a system being developed for detecting acoustic anomalies in an aircraft, using several microphones.

The system we are building is capable of separating background noise from acoustic anomalies. With several microphones strategically placed throughout the aircraft, the geographic source of the anomaly can be determined. Our system will also attempt to identify the anomaly cause, using a small dictionary of known anomaly templates.

This paper provides a software overview of the system, and a description of the hardware we have developed to implement the system.

Keywords: Acoustic Anomaly, Detection, localization, Aircraft.

1. INTRODUCTION

Audible sound often conveys a great deal of information about events. In aircraft sound can indicate the shifting of cargo, a leak in a hydraulic line, or a simple shift of the co-pilot in his seat. The types of sounds heard can be classified using two different dichotomies.

- Anomaly versus background. The acoustic signal could be interpreted as background noise that would be considered normal in the operation of the aircraft. On the other hand, it could be considered anomalous, meaning a noise which is out of the ordinary.
- Expected versus unexpected. The signal could represent an event that, although it is not part of the background, is identifiable, and considered part of the normal operation of the aircraft. On the other hand, the noise could represent an unexpected event that needs to be addressed by the crew.

Sounds that the pilot can hear in the cockpit represent only a portion of useful acoustic information. A sound in the cargo hold may be significant, but inaudible from the cockpit.

Our system attempts to gather acoustic data from all parts of the aircraft, process it, and present useful information to the pilot, to enable him to make informed decisions concerning the status of the aircraft. We use multiple microphones placed throughout the aircraft, providing the pilot with information on areas which he normally would not be able to monitor acoustically. The system can provide location information on acoustic events, using the several microphones to pin-point the source. The system can also categorize the events as either known or unknown, providing the pilot with information on how urgently the information must be analyzed.

The system that we describe is currently under development. When finished it must be able to perform the following tasks.

- Discern anomalies from background signals.
- Locate the source of the anomaly.
- Categorize an anomaly as either unknown, or one of several commonly occurring events.
- Present event information to the pilot for analysis.

2. RELATED WORK

The first task required in anomaly detection is, in fact, detecting an anomaly from background noise. Chandola et al. [1] have written a survey on anomaly detection, mostly geared toward intrusion detection, but general enough to be applicable to the acoustic domain. They classify anomalies by characteristics and discuss their detection. The anomalies in which we are interested are classified as contextual anomalies, or patterns that are anomalous from their context. The context in this case is the time of the datum in the time series.

A natural way to analyze acoustic data is using spectral analysis. Rabiner and Shafer [2] present a survey of digital processing techniques, with an emphasis on human speech. The first phase of analysis is often short term Fourier transformation, in which small windows of the time series are transformed from the time domain to the spectral domain. Spectral coefficients can then be used as features representing a window.

Once features are extracted, anomalous windows must be identified. To do this background noise must be modeled. Once background noise is modeled, anomalies may be defined as differences from the background model. A common technique for modeling background noise is described by Harbin and Hauk [3]. This article describes the analysis of underwater noise, using hydrophones. They model the background as a set of frequency bins, from which they develop a probability distribution. A variation to this, used in our work, would be to aggregate background readings into a frequency prototype, which is used as the model.

Once the background is modeled, an acoustic reading can be compared with the background, and classified as either background, or anomaly. Phyu [4] gives a survey of common classification techniques. The techniques presented include decision trees, Bayesian networks, and nearest neighbor. For classifying in two class situations, such as determining if a noise is background or not, The problem of classification is much simpler than the general case, and a simplified form of the nearest neighbor technique suffices.

Location of an event can proceed after the sound has been identified as an anomaly A simple procedure for localization uses time delay of arrival (TDOA) of the signal. Systems using this technique have been built by researchers like Dostalek et al. [5]. Their paper describes the hardware and software used in a typical TDOA system, in which time delays, which are proportional to distance, are used in triangulation.

Our system, in addition to localization, also attempts to classify an anomaly The techniques described in Phyu can easily be used not only to distinguish a signal from the background, but also classify it in terms of known common events.

The last job of our system is presentation of the results to the pilot in a useful form. Some research has been done on visualization aimed specifically at aircraft pilots. A thesis by Aragon [6] has done a fairly extensive study of visualization for helicopter pilots. Many of the same results are useful in general. The thesis stresses the need for a user centric design. It examined the importance of several variables, including color, transparency, depth cues, animation, texture, and shape.

3. SYSTEM DESIGN

Our system design is illustrated in Fig. 1. Only a portion of the system has been completed. This paper covers spectral analysis, anomaly detection, and localization. These stages are described in more detail below. The classification stage and reporting stage are still being designed.

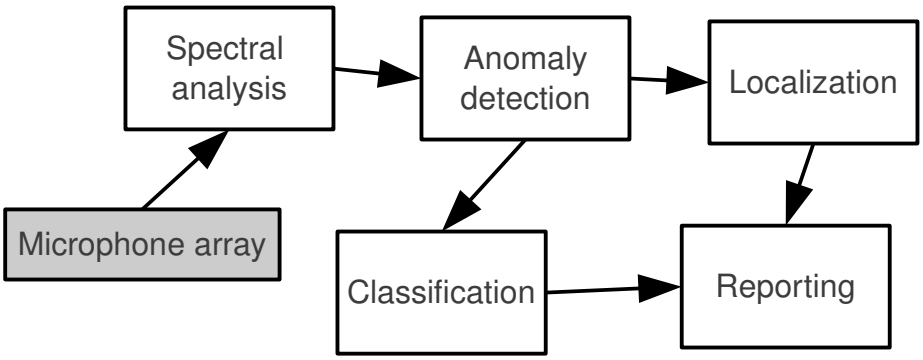


Fig. 1. System data flow

3.1. Sampling and Spectral Analysis

Sampling is done by polling each microphone in an array, at a given frequency. Amplitude samples are collected into *sample windows*, which are collections of samples from a small time interval.

The samples in the window are run through a fast Fourier transform (FFT), to convert them into spectral samples. After conversion, a window is transformed into a *gross frequency window*, which is the collection of harmonic coefficients for the original sample window. The gross frequency window is then pruned by eliminating all coefficients for frequencies above the *kth* harmonic, where *k* is a fixed constant. The resulting lower dimension vector is called the *net frequency window*.

3.2 Background Modeling

When the system first starts up, a set of *m* windows are collected for each microphone, where *m* is a fixed constant. We operate under the assumption that initially there is a quiet time in which no anomalies are present, and all acoustic information represents background noise.

To model the background noise the collected windows are combined into a single window template. This template is represented by a mean value matrix and a variance matrix. Let W_{ijf} be the complex Fourier coefficient for the *fth* harmonic, where f k, in the *jth* net frequency window for microphone *i*, where j m. Then the background model used for microphone *i* consists of

$$\overline{W_{if}} = \frac{1}{m}\sum_{j=1}^{m} W_{ijf} \tag{1}$$

$$\hat{W}_{if} = \frac{1}{m}\sum_{j=1}^{m} \left(W_{ijf} - \overline{W_{if}}\right)^2 \tag{2}$$

3.3. Anomaly Detection

To determine if an anomaly has occurred, a frequency window sampled at time *t*, U_{if}, is compared against the background model. The scheme used involves calculating the distance from the sample to the background, and then comparing this to the variance. More precisely, let $i(a)$ and $r(a)$ denote the imaginary and real components of the complex number a, respectively. Then the distance

$$d_{if}(t) = \left(U_{if}(t) - \overline{W_{if}}\right)^2 \tag{3}$$

is calculated, and if $r\left(d_{if}(t)\right) > r\left(g\hat{W}_{if}\right) \vee i\left(d_{if}(t)\right) > i\left(g\hat{W}_{if}\right)$, for some f k, where g is a real gain constant, then the window is considered as an anomaly. This rule implements a policy that considers a sample anomalous if the sine or cosine coefficient at any harmonic exceeds a threshold variance.

3.4. LOCALIZATION

The work on localization is still in progress, and we are currently working with a simplified system. In our simplified system we use chirps, or single frequency sounds, as our anomalies. Detecting these chirp events is much easier than the process previously described for dealing with a noisy background. In fact, given a chirp frequency of f^*, detection is simply the process of determining if the coefficient for the chosen frequency, U_{if^*}, exceeds a threshold value.

In the general case in which anomalies can be noises other than pure chirps, sounds must be matched between different microphones in the array. Matching two windows can be done using several procedures. In an environment in which anomalies are sparse occurrences, it may be possible simply to assume that if two windows from two different microphones are anomalous, it is probably the same sound. In situations in which anomalies are dense occurrences, a better strategy would threshold the distance between the two frequency windows. Of course, to achieve meaningful results the two sample windows would first need to be normalized with respect to sound volume.

Localization is performed via triangulation. The requirement is that at least two microphones in the array must detect the same anomalous windows, and accurate pin-pointing the anomaly requires three microphones. Triangulation uses the fact that the TDOA is proportional to distance. Fig. 2 shows the triangulation situation for three microphones.

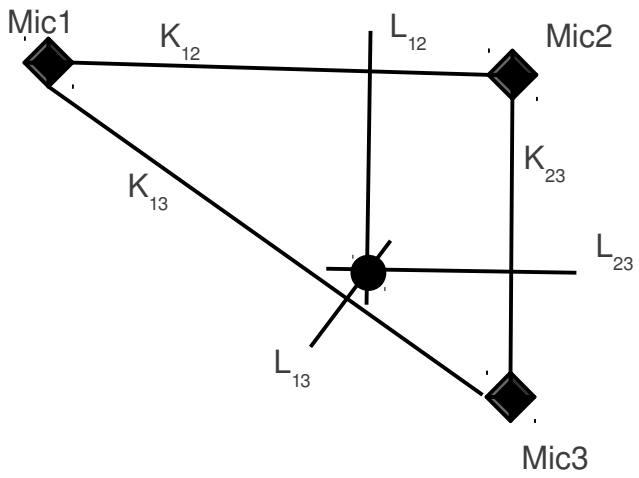


Fig. 2. Triangulation with three microphones.

The diagram indicates the placement of three microphones, Mic1, Mic2, and Mic3, with the lines K_{12}, K_{13}, and K_{23} connecting the microphones and forming a triangle. If the times for arrival of a chirp for Mic1, Mic2, and Mic3 are t_1, t_2, and t_3, respectively, this information can be used to calculate the lines L_{12}, L_{13}, and L_{23}, which are perpendicular to the lines K_{12}, K_{13}, and K_{23}, respectively. The intersection of the three L lines is the source of the acoustic event.

The L lines are defined by their intersections with the K lines. These intersections can be discovered using the proportionality between distance and TDOA. This proportionality can be stated as

$$\frac{d_{i,Lij}}{d_{j,Lij}} = \frac{t_i}{t_j}\ , \qquad (4)$$

where $d_{i,Lij}$ is the distance from Microphone i to any particular point on the line L_{ij}, including the intersection of L_{ij} with K_{ij}.

4. CUSTOM HARDWARE

In the design of our system, particularly the work on localization, it became apparent that our system needed complete control over the microphone array. This control consists of determining exactly when a microphone is polled, and how many readings are taken from the microphone at a time. With many sound packages, this type of control is difficult to achieve. Often the sound package controls sampling frequency, and is set up to do batch sampling. Particularly in TDOA work this is not appropriate, since the sample windows for each microphone must be taken simultaneously, to allow for accurate timing.

Most computers are not configured to allow them to be connected to large sets of microphones. They are usually equipped with no more than 3 acoustic input ports. As a consequence, our system implementation was required to use the USB capability of the host computer, to run a hub with each microphone connected through a USB sound card. This configuration further complicates precise control of the polling process by introducing a layer of processing by the sound card.

It was decided to build a controller for the microphone array, called the *acoustic multiplexer* (ACMUX). This device connects to a USB port on the host computer, controls an array of microphones, and allows the host to read a single signal value at a time from a selected microphone, using high-speed USB communication. The structure of the ACMUX is shown in Fig. 3. The demonstration model that was constructed controls four microphones. Each microphone signal is filtered

through a capacitor, and fed into an op-amp. The four amplified signals are funneled into an analogue to digital converter (ADC), with a selectable input channel. This ADC is part of the PIC-18f2455 micro-controller. The micro-controller contains a firmware stack which handles communication with the host computer using USB.

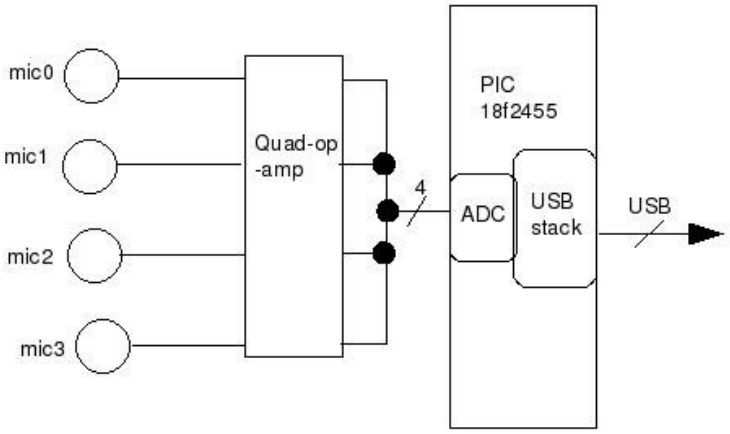


Fig. 3. Structure of the ACMUX

The host computer communicates with the ACMUX by sending a channel number. In response, the ACMUX polls the requested microphone, and returns the signal amplitude as a ten bit unsigned integer. The micro-controller's USB communication is clocked by a 48 MHz clock, easily capable of our system's sampling frequency of less than 10KHz.

5. CONCLUSION AND FUTURE WORK

We have presented the structure of a system to detect sound events, and report them to an aircraft crew. The system can identify common events, and not only report to the crew that an event has occurred, but also what type of event has occurred.

Our system is under construction, We have developed software to detect acoustic anomalies. This was done by moving from the temporal domain into the spectral domain, and then comparing frequency windows with a background model. The background model is constructed by sampling a quiet period, collecting several windows of samples, and aggregating the windows into a single window.

We have developed software to do localization. The localization software polls the microphones until an anomaly is detected, records the time of arrival at each microphone, and then uses the time delays as an indication of the distances to the anomaly source, and triangulates to find the location of the source.

To complete the system we will be required to construct software that classifies an acoustic event as a particular common event type, or as an unknown event. This can be done in a similar fashion as that which we currently use to distinguish an event from the background noise. This would involve comparing windows against frequency window models of the different known events.

Another piece of the system that needs to be built is the reporting software. This software would present the results of the system to the pilot. The presentation must be non-intrusive, but at the same time effective, in the sense that it calls the pilots attention to truly significant and problematic events.

The total system will be an important tool for the aircraft crew. It will increase aircraft safety by notifying the pilot of events he would not normally be aware of, giving the pilot information on event location, and providing the pilot with information helpful in determining the significance of an event.

ACKNOWLEDGMENTS

This research is supported by a NASA Grant.

REFERENCES

[1] V. Chandola, A. Banerjee, V. Kumar, “Anomaly Detection: A Survey”, **ACM Computing Surveys**, Vol. 41(3), July 2009.

[2] L. R. Rabiner, R. W. Shafer, “Introduction to Digital Speech Processing”, **Foundations and Trends in Signal Processing**, Vol. 1, Nos. 1–2, 2007).

[3] P. E. Harbin, T. F. Hauk, **Background Acoustic Noise Models for the IMS Hydroacoustic Stations**, Lawrence Livermore National Laboratory, 2010.

[4] T. N. Phyu, “Survey of Classification Techniques in Data Mining”, **Proceedings of the International Multiconference of Engineers and Computer Scientists**, Hong Kong,March, 2009.

[5] P. Dostalek, V. Vasek, J. Dolinay, “Acoustic Source Localization Based on Time-delay Estimation Method”, **Proceedings of the 13th WSEAS International Conference on CIRCUITS**, Rodos Island., Greece, July, 2009.

The 5th Umpire: Automating Cricket's Edge Detection System

R. Rock, A. Als, P. Gibbs, C. Hunte
Department of Computer Science, Mathematics and Physics
University of the West Indies (Cave Hill Campus)
Barbados

ABSTRACT

The game of cricket and the use of technology in the sport have grown rapidly over the past decade. However, technology-based systems introduced to adjudicate decisions such as run outs, stumpings, boundary infringements and close catches are still prone to human error, and thus their acceptance has not been fully embraced by cricketing administrators. In particular, technology is not employed for bat-pad decisions. Although the snickometer may assist in adjudicating such decisions it depends heavily on human interpretation. The aim of this study is to investigate the use of Wavelets in developing an edge-detection adjudication system for the game of cricket. The role of Artificial Intelligence (AI) tools, namely Neural Networks, in automating the detection process will also be implemented. Live audio samples of ball-on-bat and ball-on-pad events from a cricket match will be recorded. DSP analysis, feature extraction and neural network classification will then be employed on these samples. Results will show the ability of the neural network to differentiate between these key events. This is crucial to developing a fully automated edge detection system.

Keywords*:* Cricket, Wavelets, Neural Networks, Edge-detection, feature classification

I. INTRODUCTION

The revenue generated from sport globally is estimated to reach over $130bn US dollars by the year 2013 [1, 2]. In 2010, soccer, the world's most popular sport according to [3], was reported by its international governing body, FIFA, to have generated US$1bn on the strength of the successful world cup in South Africa [4]. Cricket is the second most popular sport, and the Indian Premiere League's (IPL) 20/20 format boasts of being the second highest paid sport ahead of the football's English Premiere League (EPL) [5]. In 2009, the Indian Premier League (IPL) offered pay checks as high as US$1.55 million to top class cricketers for a five week contract [6]. This figure was eclipsed in 2011 when Gautam Gambhir of the Kolkata Knight Riders was awarded a contract for US$2.4 million [7].

It is common knowledge that bookmakers have also capitalised on cricket's wide fan-base. The plethora of online betting sites dedicated to cricket, such as bet.com, cricketworld.com, cricket.bettor.com and cricketbetlive.com, to list a few, are evidence of this practice. Unfortunately, the sport has gained notoriety with several of its elite players being charged with bringing the game into disrepute. In the 1999-2000 India-South Africa match fixing scandal, Hansie Cronje, the South African captain admitted to accepting money to throw matches and was subsequently banned from playing all forms of cricket [8, 9]. In August 2010 during the match between England and Pakistan at Lord's Cricket Ground two Pakistani players were accused of match fixing by deliberately bowling three illegal deliveries (i.e. no-balls) at pre-determined times during their bowling spells. It was alleged that Mr. Mazhar Majeed, a property developer and sports agent, orchestrated the events and tipped off betting syndicates so they could place "spot" bets and make profits of millions of pounds [10, 11].

To deter match fixing, and ensure legitimate results, it is not surprising that the use of technology in cricket has steadily increased over the years and now has a major role in adjudicating the outcome of events. However, although the use of technology serves to protect both players' careers by avoiding incorrect decisions and the reputation of the game, its adoption has not been fully embraced by the cricket's administration body. There are a number of devices being used to assist umpires in the adjudication process and for the entertainment of television audiences. One such device is the Snickometer (also known as 'snicko'). English Computer Scientist, Allan Plaskett, invented this in the mid-1990s. The Snickometer is composed of a very sensitive microphone, located behind the stumps, and an oscilloscope (wirelessly connected), which displays

traces of the detected sound waves. These traces are recorded and synchronized with the cameras located around the ground. For edge-decisions, the oscilloscope trace is shown alongside the slow motion video of the ball passing the bat. By the transient shape of the sound wave, the viewer(s) first determines whether the noise detected by the microphone coincides with the ball passing the bat, and second, if the sound appears to come from the bat hitting the ball or from some other source. Unfortunately, this technology is currently only used as a novelty tool to give the television audience more information regarding if the ball actually hit the bat. Umpires do not enjoy the benefit of using 'snicko' but must rely instead on their senses of sight and hearing, as well as personal judgment and experience. In many instances, there are coinciding events that may be confused with the sound of ball-on-bat. These include the bat hitting the pad during the batman's swing or the bat scuffing the ground at the same time the ball passes the bat. The shape of the recorded sound wave is the key differentiator as a short, sharp sound is associated with bat on ball. The bat hitting the pads, or the ground, produces a 'flatter' sound wave. The signal is purportedly different for bat-pad and bat-ball however, this is not always clear to the natural eye [12]. It is our submission that as the final decision requires human interpretation of the signal traces, it may be subject to error.

The aim of this paper is to employ wavelet analysis, feature extraction and artificial neural networks to implement a fully automated decision making system for bat on pad and bat on ball (i.e. edge) decisions, thereby extending the work done by Rock et al in [13]. It is expected that this will give teams a fairer chance on the outcome of a match (game) by minimising the number of these decision errors currently observed in the game.

II. BACKGROUND

It is well known that the continuous wavelet transform (CWT) may be used to analyse audio signals [14, 15]. The CWT provides another view of temporal signals as it transforms the regular time vs. amplitude signal to time vs. scale, where scale can be converted to a pseudo-frequency. This method allows one to examine the temporal nature of audio events and the corresponding frequencies involved. In essence, the correlation values, produced during the transformation process, provide critical information on the characteristics of the signal. By exploiting these characteristics a distinction can be made between different audio events. In particular, the five (5) features extracted from the wavelet transform are the **maximum correlation coefficient** and its associated **pseudo-frequency** for several CWT scale ranges, along with the **standard deviation**, **kurtosis** and **skewness** of the said correlation coefficients. These features were fed into a Neural Network to produce the final result.

Neural networks have been used over the years in a wide range of areas. These areas range from forecasting to extracting patterns from imprecise or complicated data, which human and other pattern recognition techniques may have missed. For the purpose of this paper we will be examining the pattern recognition and classification capabilities of an Artificial Neural Network (ANN). An ANN is an information-processing system that has certain performance characteristics in common with biological neural networks. It consists of a large number of interconnected neurons, each with an associated weight. These neurons work together to help solve various problems [16]. One of the main features of the ANN is its ability to take a set of features it has not encountered before and accurately output the desired output after it has been well trained.

There are many instances where Neural Networks are being applied. There has been extensive research of their use in the medical arena, specifically in the classification of heart and lung sounds. There has also been extensive research in the Computer Science field. Hadi [17] used a Multilayer Perceptron Neural Network to classify features obtained from heart sounds by the Wavelet transform, and a high correct classification rate of 92% was achieved. Borching [18] developed a chord classification system using features extracted from the wavelet transform and classified by a neural network. A high recognition rate was achieved even under a noisy situation.

Kandaswamy [19] using feature extraction from the wavelet transform and classification found that results using the Neural Network out performed conventional methods of classification of lung sounds. Though all classes of lung sounds were not used in the experiment, results showed this method was worth exploring.

No known instances where neural network classification has been applied in the area of cricket were uncovered in the literature. The approach adopted in this work is to utilise neural networks to classify features that have been extracted from bat-on-ball and bat-on-pad sound files using the CWT. These features can then be used to accurately determine the source of the noise in a

cricket match. The automated sound detection technique can greatly decrease the number of incorrect decisions being made in the game, which may ultimately protect a player's career. This approach can then be expanded throughout the sporting world improving the quality and minimising the errors observed at various events.

III. METHODOLOGY

The equipment setup, shown in Fig. 1, is identical to that used for international matches and was configured at various hardball cricket grounds throughout Barbados. The microphone transmitter is covered in a small hole directly behind the stumps. The receiver and the laptop are assembled inside the players' pavilion. The recordings, made using the laptop's sound recorder program, are stored as a 16-bit pulse coded modulation (PCM) .WAV file, sampled at 44,100 kHz (stereo) for later processing.

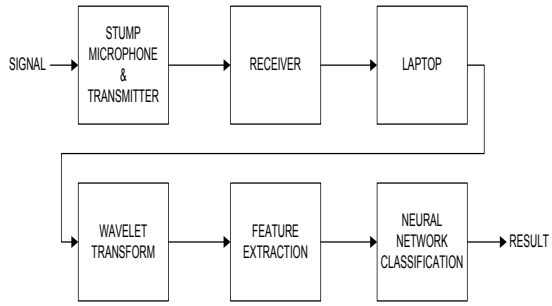


Figure 1: Schematic of experimental setup.

The key specifications for equipment used in recording the audio data are listed in Table 1.

TABLE I. EQUIPMENT PARAMETERS

EQUIPMENT	KEY PARAMETERS
	WL184 Supercardiod Lavalier Condenser Mic:
	Supercardioid pickup pattern for high noise rejection and narrow pickup angle
Shure SLX14/84 Wireless Lavalier Microphone System	**SLX1 Body pack Transmitter:**
	518 - 782 MHz operating range
	SLX4 Wireless Receiver:
	960 Selectable frequencies across 24MHz bandwidth
Mobile Precision M6400 Notebook Computer	Precision M6400, Intel Core 2 Quad Extreme Edition QX9300 2.53GHz, 1067MHZ

MATLAB programs were written to perform the CWT analysis and extract the following features: maximum correlation coefficient (x_{corr}) and the associated pseudo-frequency (P_{freq}) from selected CWT scale ranges along with the standard deviation (σ), kurtosis (k) and skewness (sk_n) of the said correlation coefficients. These features were used as input to the fully connected 5-input Multi-Layer Perceptron neural network depicted in Fig. 2. The network consists of a single hidden layer with three neurons each of which employed the tanh transfer function.

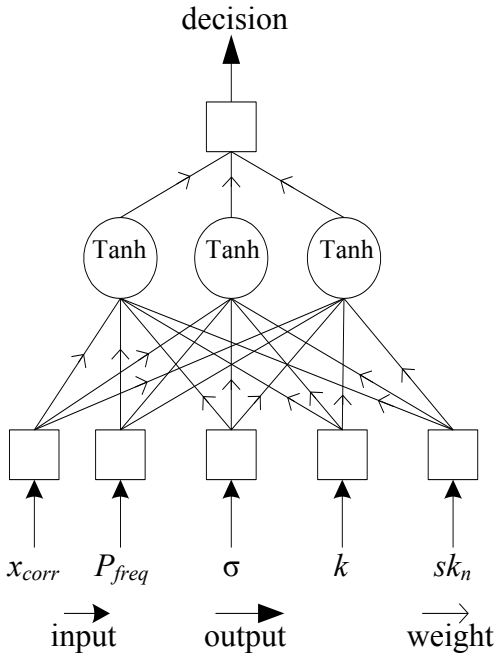


Figure 2: 5-Input Multi-Layered Perceptron

The network was trained with 260 data samples using a backpropagation algorithm. The data set was divided into 130 incidences of bat-on-ball signals and 130 of bat-on-pad. Moreover, testing was done on 40 previously unknown signals. The output from the network was a decision on whether the ball hit the bat (1) or the pad (0).

IV. RESULTS

Recordings of the impact of ball hitting bat and ball hitting pad were successfully compiled and analysed. Figure 3 shows the plot of the mean squared error (MSE) of the network after each complete presentation of the training data to the network (i.e. epoch). The epoch number is shown on the X-axis and the MSE is shown on the Y-axis. The MSE of the training (T) set is shown in white diamonds and that of the cross validation (CV) set is shown in black squares. Ideally, a neural network is deemed to be well trained when both lines gradually decrease to zero. Results from the graph showed that the neural network has trained reasonably well.

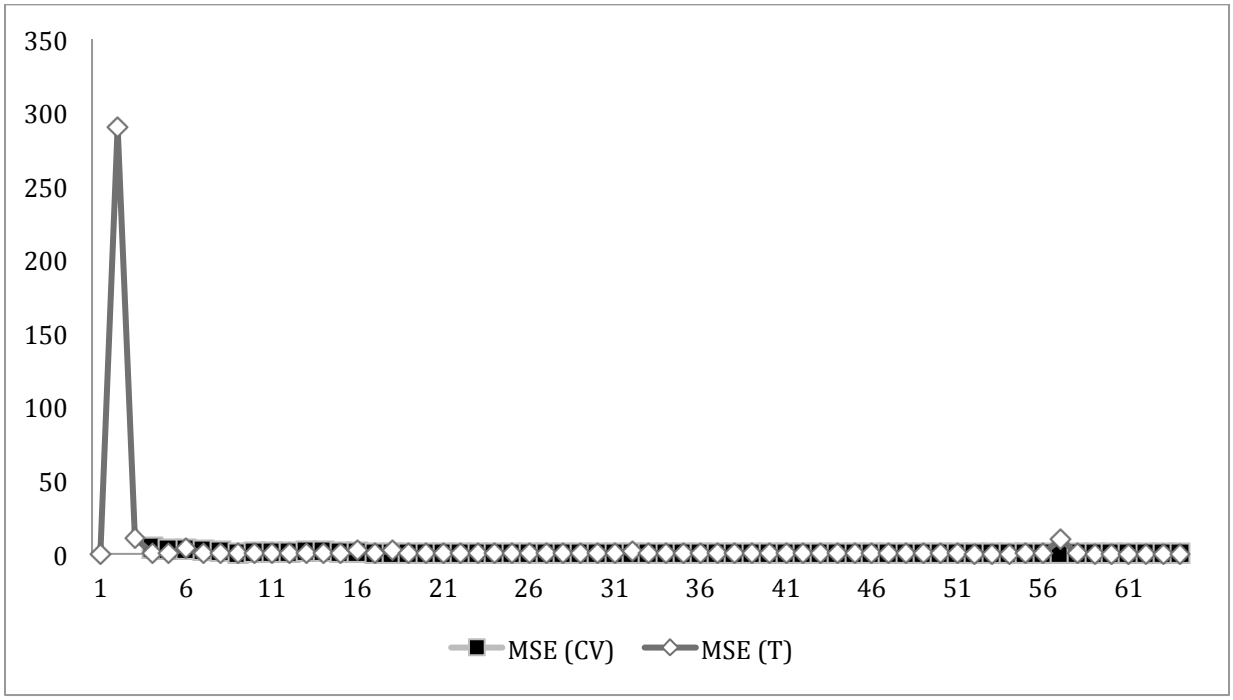


Figure 3: Graph of mean error versus number of epoch

Figure 4 shows the plot of desired output and actual output versus number of samples used for testing. The samples (40) used for testing originated from data the neural network was not exposed to previously. Note there is only one error (26th sample, white diamond) thus indicating 97.5% accuracy.

V. CONCLUSION

Results show the neural network performed exceptionally well rendering a correct classification of 97.5% for data not previously encountered. It is believed that better results may be obtained by optimising the choice of features that are extracted using the wavelet transform and employed to train the neural network. This will be the subject of future work. Technology must be used judiciously if it is to gain support of the players and administration. For example consider the case in the recently concluded 2nd Test match between India (I) and the West Indies (WI) in Barbados were the on field umpire consulted the third umpire regarding the legality of a delivery from Fidel Edwards (WI), which ultimately resulted in Raul Dravid (I) being given out to a no-ball. Ironically, the wrong television replay of the bowling delivery was used. This supports the efforts pursued in this paper to completely remove the human factor from the data gathering and information-processing portion of the adjudication process. The 5th umpire system will provide a decision, which will greatly assist the on-field umpire, with whom the final decision resides.

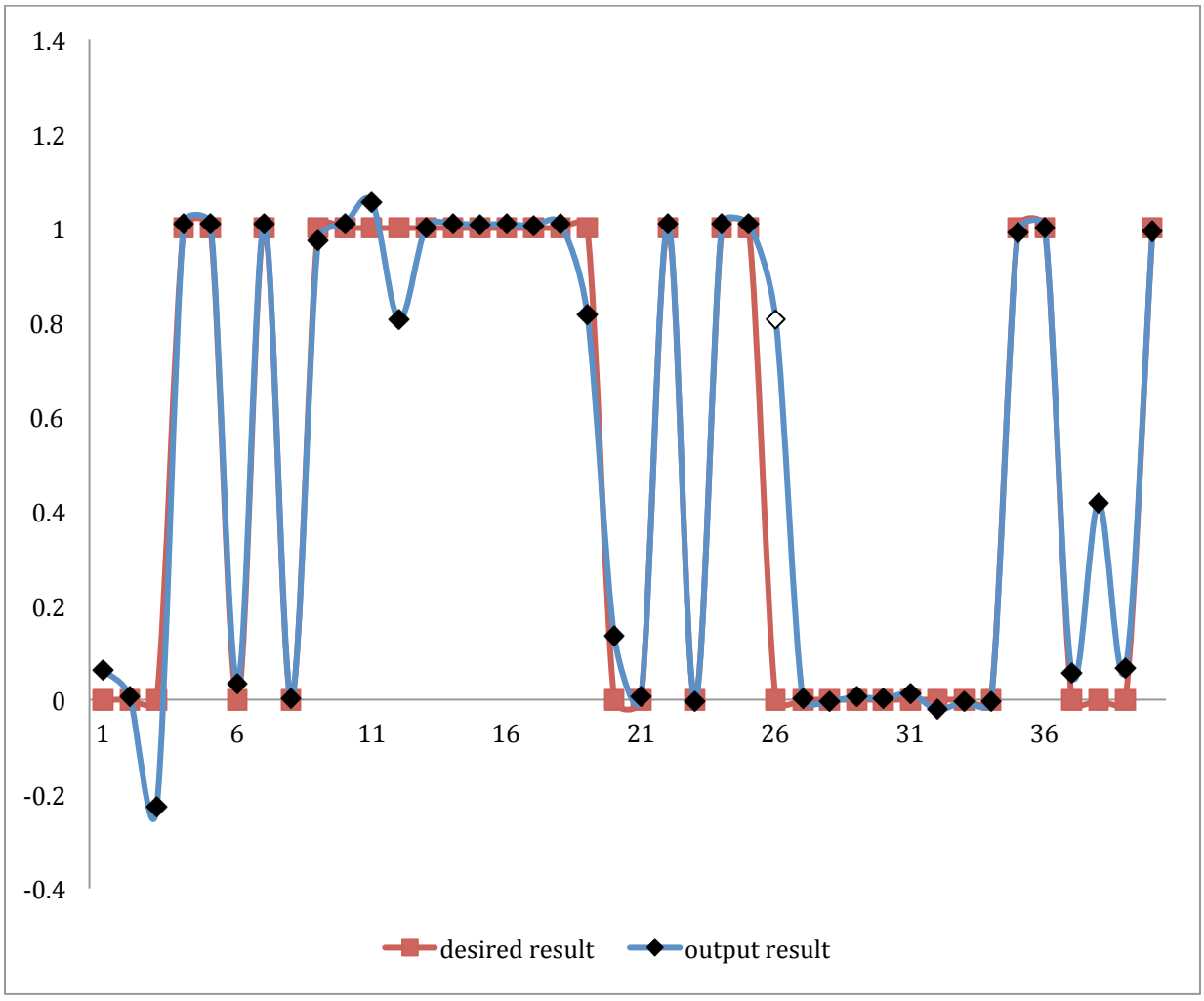


Figure 4: Graph and results of actual and desired results for the forty test data samples

VI. ACKNOWLEDGMENT

The authors would like to acknowledge Mr Simon Wheeler (Executive Producer/Director of TWI and IMG media company), Mr Mike Mavroleon (Senior Engineer IMG media) and Mr Collin Olive (Asst. Sound Engineer IMG media) for their direction in acquiring the equipment needed to record the data samples used in this paper. These are the persons responsible for technological aspects of the broadcast of international cricket (i.e. snickometer, hawk-eye, TV replays).

VII. REFERENCES

1. Ben Klayman, L.G. *Global sports market to hit $141 billion in 2012*. 2008 [cited 2011 17/08]; Available from: http://www.reuters.com/article/2008/06/18/us-pwcstudy-idUSN1738075220080618.
2. Clark, J. *Back on track? The outlook for the global sports market to 2013*. [cited 2011 17/08]; Available from: http://www.scribd.com/doc/46262710/Global-Sports-Outlook#outer_page_2.
3. *World's Most Popular Sports* [cited 2011 17/08]; Available from: http://www.mostpopularsports.net/.
4. Zurich, *61st FIFA Congress FIFA Financial Report 2010* 2011.
5. *IPL 2nd highest-paid league, edges out EPL*. 2010 [cited 2011 17/08]; Available from: http://timesofindia.indiatimes.com/iplarticleshow/5736736.cms.
6. Peter J. Schwartz, C.S. *The World's Top-Earning Cricketers*. [cited 2011 January, 1st]; Available from: http://www.forbes.com/2009/08/27/cricket-ganguly-flintoffl-business-sports-cricket-players.html.
7. *Income of IPL 2011 Players* [cited 2011 17/08]; Available from: http://www.paycheck.in/main/salarycheckers/copy_of_income-of-ipl-2011-players.
8. *Hansie Cronje*. 2003 [cited 2011 17/08]; Available from: http://www.espncricinfo.com/southafrica/content/player/44485.html.
9. Varma, A. *Match-fixing - At the turn of the century, cricket felt a tremor stronger than any it had known till then*. 2011 June 11, 2011 [cited 2011 17/08].
10. Quinn, B. *Match-fixing allegations hit England v Pakistan Test at Lord's, guardian.co.uk*. 2010 [cited 2011 15/08]; Available from: http://www.guardian.co.uk/sport/2010/aug/29/match-fixing-allegations-england-pakistan.
11. Richard Edwards, M.B., Murray Wardrop. *Cricket match-fixing: suspicion falls on 80 games as 'fixer' bailed*. 2010 [cited 2011 15/08]; Available from: http://www.telegraph.co.uk/sport/cricket/7971107/Cricket-match-fixing-suspicion-falls-on-80-games-as-fixer-bailed.html.
12. Wood, R.J. *Cricket Snicko-Meter* 2005 [cited 2010 January, 21st]; Available from: http://www.topendsports.com/sport/cricket/equipment-snicko-meter.htm.
13. R. Rock, A.A., P. Gibbs, C. Hunte, *The 5th Umpire: Cricket's Edge Detection System*, in *CSC'11 - 8th Int'l Conference on Scientific Computing* 2011: Las Vegas, Nevada. p. 6.
14. Lambrou, T., et al., *Classification of audio signals using statistical features on time and wavelet transform domains*, in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. 1998.
15. Schclar, A., et al., *A diffusion framework for detection of moving vehicles*. Digital Signal Processing. **20**(1): p. 111-122.
16. Fausett, L., *Fundamentals of Neual Networks Architectures, Algorithms and Applications*. 1994: Prentice Hall
17. Hadi, H.M., et al., *Classification of heart sounds using wavelets and neural networks*, in *Electrical Engineering, Computing Science and Automatic Control, 2008. CCE 2008. 5th International Conference on*. 2008.
18. Borching, S. and J. Shyh-Kang, *Multi-timbre chord classification using wavelet transform and self-organized map neural networks*, in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*. 2001.
19. Kandaswamy, A., et al., *Neural classification of lung sounds using wavelet coefficients*. Computers in Biology and Medicine, 2004. **34**(6): p. 523-537.

PREDICTING HOME SERVICE DEMANDS FROM APPLIANCE USAGE DATA

Kaustav Basu*, Mathieu Guillame-Bert†, Hussein Joumaa*, Stephane Ploix* and James Crowley†
*G-SCOP lab
46, avenue Flix Viallet F-38031 Grenoble cedex 01
Email:Kaustav.Basu@ensimag.imag.fr, Hussein.Joumaa@imag.fr, Stephane.Ploix@inpg.fr
†INRIA Grenoble - Rhone-Alpes
655 avenue de l'Europe 38 334 Saint Ismier Cedex France
Email:Mathieu.Guillame-Bert@inrialpes.fr ,James.Crowley@inrialpes.fr

***Abstract*—Power management in homes and offices requires appliance usage prediction when the future user requests are not available. The randomness and uncertainties associated with an appliance usage make the prediction of appliance usage from energy consumption data a non-trivial task. A general model for prediction at the appliance level is still lacking. In this work, we propose to enrich learning algorithms with expert knowledge and propose a general model using a knowledge driven approach to forecast if a particular appliance will start at a given hour or not. The approach is both a knowledge driven and data driven one. The overall energy management for a house requires that the prediction is done for the next 24 hours in the future. The proposed model is tested over the Irise data and the results are compared with some trivial knowledge driven predictors.**

***Keywords*-Appliance Usage Prediction, Enriched Learning Algorithm, Energy Management in Homes, Data Mining.**

I. INTRODUCTION

Reducing housing energy costs is a major challenge of the 21^{st} century. In the near future, the main issue for civil engineering is the thermal insulation of buildings, but in the longer term, the issues are those of "renewable energy" (solar, wind, etc) and "smart buildings". Home automation system basically consists of household appliances linked via a communication network allowing interactions for control purposes [1]. Thanks to this network, a load management mechanism can be carried out: it is called *distributed control* in [2]. Load management allows inhabitants to adjust power consumption according to expected comfort, energy price variation and CO_2 equivalent emissions. A home energy management system able to determine the best energy assignment plan and a good compromise between energy production and energy consumption [3]. In this study, energy is restricted to the electricity consumption and production. [4], [3] present a three-layers (anticipative layer, reactive layer and device layer) household energy control system. This system is both able to satisfy the maximum available electrical power constraint and to maximize a ratio between user satisfaction and cost. The objective of the anticipative layer explained in [5] is to compute plans for production and consumption of services.

Uniqueness of housing systems involves a set of new issues in control system science: it is necessary to develop new tools [6], [7], [8] and algorithms [9], [10] for globally optimized power management of the home appliances, able to anticipate difficult situations but also able to take into account the actual housing system state and the occupant expectations.

Anticipating problematic situations require also prediction capabilities. Even if it is easier to predict overall consumption, it is important to be able to predict the consumption of each appliance because, regarding dynamic demand side management, it is also important to evaluate how much energy can be saved thanks to request to customers like unbalancing requests or energy price variations. The energy savings depend on appliances: some can be unbalanced, some can be postponed and some cannot be changed. The overall goal of the prediction is described in figure 1. It also includes an user interface where the user may provide his plans for the future. The proposed approach is restricted to the prediction of appliance usage using only appliance consumption data and time of the event.

The problem of appliance usage prediction through consumption data is new. [11] deals with the problem of the user behavior prediction in a home automation system using a Bayesian network for a single appliance but a general model for appliance prediction is still lacking. Short term load forecasting (STLF) at the grid level has been there for some time but at the appliance level, these techniques are yet to be tested. Though STLF uses regressive approaches whereas the proposed approach is based on classification but the strategies used in the domain of energy load prediction led to the choice of input to the predictor.
[12] does a study on the approaches used in load prediction. The approaches range from using methodologies such as similar day, expert knowledge and linear and non linear learning algorithms. [13], [14], [15], [16] gives details of implementation of neural networks in the domain of energy load forecasting and [17] proposed a SVM model to predict daily load demand for a month.

The objective of this work is to build an enriched learning

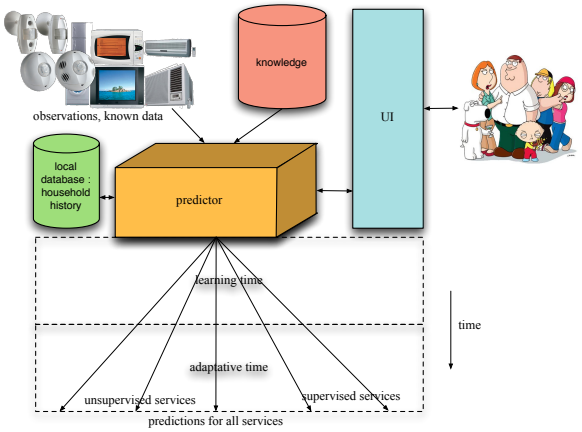


Figure 1. Goal of the prediction system

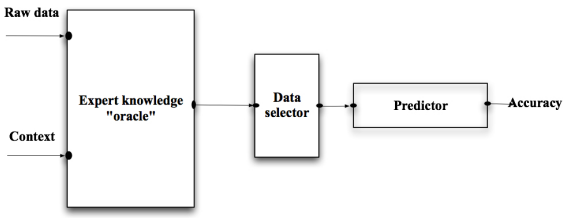


Figure 2. Schematic representation of the prediction system

algorithm which takes knowledge into account and formalize it to statistically predict the user energetic service request for the next 24 hours. For the prediction of appliances from consumption data we first reduce the problem to a two class classification problem, i.e if an appliance is consuming at a particular hour or not. The model for the prediction is for each appliance in each house. The time space is sampled into 24 hours which aim's to predict the user appliance usage requirement for a particular hour. At each point of time the Prediction system will predict for the following 24 hours and then shift to the next hour and predict the following 24 hours.

In the approach an expert proposes certain knowledge based on his domain expertise and then to formalize and represent this knowledge. The knowledge representation is considered in an incremental manner and at every stage validate the knowledge in terms of accuracy of prediction. The organization of the paper is as follows, firstly, the Proposed model (section II) is discussed in details followed by choice of classifier and parameters (section III). The Oracle results as well as the overall results are presented in (section IV and V). Finally discussion about the results and the conclusion is drawn in (section VI and VII).

II. Proposed Model

The proposed model consists of enriched learning algorithm which proposes a general way to take expert knowledge into account. The proposed model divides the task into modules each of which has its own purpose. The general model is shown in figure 2 and in the following sub-sections each of the processing modules is discussed in details.

- Raw data contain
  - Energy consumption for an appliance .
  - Contextual information (Time, Date, Weather).
- Oracle is composed of statements leading to entities (factors) that may be taken into account.
- Data Selector is a processor which stores, selects and structures the data in order to present them to the classifiers.
- Predictor.

A. Database

A database is obtained from Residential Monitoring to Decrease Energy Use and Carbon Emissions in Europe (REMODECE) which is a European database on residential consumption, including Central and Eastern European Countries, as well as new European Countries (Bulgaria and Romania). This database stores the characterization of residential electricity consumption by end-user and by country. The IRISE project has been chosen from REMODECE which deals only with houses in France. Each database concerns one house. In such a database, information is recorded every 10 minutes for each appliance in house and over one year. This information represents the consumed energy by each service, its data and its time. Moreover, it is possible to know the number of people who live in each house. However, this data is not directly available. Let us notice that appliances are just involved in services: they are not central from the inhabitant point of view. Consequently, they are not explicitly modeled. The presence of the user is important but it is not predictable at the moment.

B. An expert system that generates knowledge : the Oracle

We define Oracle knowledge as statements leading to generated data (or factors) that may be taken into account. The Oracle receives the raw data from the database giving the consumption at an particular hour and the date, time and weather information at that hour. The Oracle proposes knowledge and then gives the necessary function which represents the data in a form interpretable by the Prediction system. The knowledge which are relevant for a particular appliance in a house might not be relevant for another house using the same appliance. So all knowledge proposed by the Oracle have to be validated and knowledge which doesn't increase or reduces the accuracy of prediction for a particular appliance have to be rejected. The part of validating and structuring of the Oracle output is done in the subsequent processing module as seen in figure II.

Statements and the functional representation inside the Oracle:

- Immediate past history of consumption is meaningful to appliance usage prediction.
- Hour of the day is meaningful to appliance usage prediction.
- Day of the week is meaningful to appliance usage prediction.
- Season of the year is meaningful to appliance usage prediction.
- Previous days same hour is meaningful to appliance usage prediction.

In the following sub-sections each of the proposed knowledge by the Oracle and their representation is looked in details.

1) Past Consumption History: The Oracle proposes that the past sequence of energy consumption prior to an event is meaningful in appliance usage prediction. It is represented mathematically followed by an illustration.

Mathematically, it is formalized by the following predicate function

Inputs: Consumption(*H*-1); Consumption(*H*-2);...; Consumption(*H*-*n*) ;

where,

n is the size of the past time history

H is the Current hour

Output : $\{0,1\}^n$

Here the output is a thresholded binary vector of size *n* which signifies if there is consumption at the hours prior to the event on not.

2) Hour of the Day: The Oracle proposes that the "time of the day" is meaningful to appliance usage prediction. In practice, by this knowledge the time space is discretized into 24 hour slots and the "actual time" the event occurs assumes priority. Firstly it is represented mathematically and then provided some illustrations to better understand the representation.

Predicate function HOD

Inputs : Current Hour in the day, X

$X \in \{(0-1),(1-2),(3-4),(5-6),...,(23-24)\}$ Hours

Output : $\{0,1\}^{24}$

This is an orthogonal representation of an hour rather than a numeric value.

Illustration :

If the Time of the day is 6.00 am,

instead of using the numeric value 6 we use "0, 0, 0, 0, 0, 1, 0,...,0" to represent the same.

So for a day example, the representation will be :

Hour 0 — (1,0,0,0,0,...,0)
Hour 1 — (0,1,0,0,0,...,0)
. — (0,0,1,0,0,...,0)
. — (0,0,0,1,0,...,0)
Hour 23 — (0,0,0,0,0,...,1)

3) Day of the Week: The Oracle proposes that "Day of the week" is meaningful in appliance usage prediction. Similar to the way in II-B2 by taking this knowledge into account the whole week is discretized into 7 days. Instead of representing this with a numeric value, we use the orthogonal representation as in II-B2.

Predicate function DOW

Inputs : Current day of the week, X

$X \in \{Sunday, Monday, ..., Saturday\}$

Output : $\{0,1\}^7$

4) Season of the Year: Similar to the prior sub-sections the Oracle proposes that season of the year is meaningful in appliance usage prediction. There are appliances in houses which show distinctive different behavior depending on the season of the year. As like the prior representations an orthogonal representation is chosen over a numeric one.

Predicate Function SOY

Inputs : Current season of the year, X

where $X \in \{Spring, Summer, Autumn, Winter\}$

Output : $\{0,1\}^4$

So the season Oracle output will be:

5) Previous Days Same Hour: Here the Oracle proposes that what happens on previous days on the same hour is important in appliance prediction. So we look if there is consumption or not in the previous days for the same hour. The output is a vector of thresholded binary values of previous days at the same hour.

Predicate function

Inputs: consumption(*H*-24); Consumption(*H*-48);... Consumption(*H*-*n*);

where *n* is typically taken as 168 (one week)

H is the Current Day

Output : $\{0,1\}^7$

6) Oracle Output: The overall Oracle output after the representation of all the knowledges proposed by the Oracle is shown in table II-B6 where each row represent the proposed knowledge at a particular time in a incremental manner. The table is obtained by the incremental addition of knowledge proposed by the Oracle, so the knowledge proposed in section II-B1 to II-B5 are added incrementally in order. The Oracle has an available memory, thereby every hour in the history is represented by table II-B6.

C. Data Selector

The data selector is defined as an non-temporal matrix processor which stores, selects and structures the data for the predictor. This matrix is the input to the predictor. The

Knowledge	*H*
Consumption(*H*-1)	0/1
Consumption(*H*-2)	0/1
...	0/1
Consumption(*H*-n)	0/1
Hour of the day(0-1)	0/1
Hour of the day(1-2)	0/1
...	0/1
Hour in the day(23-24)	0/1
Day of the week(Sunday)	0/1
Day of the week(Monday)	0/1
...	0/1
Day of the week(Saturday)	0/1
Season of the year(Summer)	0/1
...	0/1
Season of the year(Winter)	0/1

Table I
OVERALL ORACLE OUTPUT

data selector may choose the whole or the subset of the output from the Oracle. It should be noted that all the knowledge proposed by the Oracle might not be useful for a particular appliance in a house and there are different possible structuring of the knowledges proposed by the Oracle. All the outputs of the Oracle is stored in the data selector, but only those which are validated by the predictor are selected for the overall prediction. In this work, only one of the possible structuring is implemented, which is done by taking the knowledges proposed by the Oracle as a single unit after selection. There are other possible structuring which will be looked in the future.

D. Predictor

This module consists of the classifiers commonly used in Machine Learning such as the Neural networks. The classifier gets its input from the data selector. It first validates the knowledges proposed by the Oracle and then the prediction for the next 24 hours.

III. CHOICE OF CLASSIFIER AND PARAMETERS

In this section the justification of using a neural network classifier for such an application is discussed. Choice of neural networks are more on the basis of past literature than on the initial results seen in table II, where different non-linear classifiers are compared. The comparison is based on past consumption history and then prediction for the next hour. The comparison with other classifiers such as Support vector machines, Naive Bayes and K-nearest Neighbors are given. It must be mentioned, that at no point the fact that other classifiers may perform better is disregarded, these are initial results with suitable parameters. The results are the accuracy for the next hour.

In table IV the parameters of the neural network classifier is given. The number of hidden layers are chosen to be one and the number of hidden neurons to be half of the number of input nodes. This choice is to avoid the over fitting or under-fitting of the network. The choice of training algorithm is also shown in III. The results of the choice of architecture is shown in IV. The final choice of all the parameters are shown in table V.

Appliance	SVM	Naive Bayes	KNN	Neural Network
900 lamp	82.40	60.1	79.72	82.94
932 oven	84.42	84.25	83.51	84.95

Table II
CLASSIFIER COMPARISON

Training Algorithm	Accuracy
Gradient descent	57.20
BGFS	82.94
Conjugate entropy	83.08

Table III
TRAINING ALGORITHM

Architecture	Accuracy
RBF	57.20
MLP	83.22

Table IV
NEURAL NETWORK ARCHITECTURE

The scoring is done in terms of accuracy, where accuracy is the number of correct classification to the total number of classifications.

Parameter	Selection
Sampling Method	Random
Train sample size	75
Test sample size	25
Network Type	MLP
Activation function(hidden unit)	Tanh
Activation function(output unit)	Softmax
No of hidden neurons	no of input/2
Error Function	Cross entropy
Training Algorithm	BGFS
Learning Rate	0.1

Table V
NEURAL NETWORK PARAMETERS

IV. ORACLE KNOWLEDGE RESULTS FOR DATA SELECTION

In this section each of the proposed knowledge is validated in an incremental manner. The results indicate that all the knowledges proposed by the Oracle might not result in increase in performance of the prediction system. So for each appliance in a house a subset of the knowledge proposed by the Oracle is selected. Therefore,

only knowledge which increase in prediction performance is selected for a particular appliance. All the predictions are done using a Neural Network Predictor whose parameters are discussed in III. It must be mentioned, that the knowledge proposed by the Oracle are prioritized on the basis of domain knowledge. It can be seen from table VI that due to our incremental approach the knowledge which appears first has a higher chance of getting selected than the next one.

Knowledge	Neural Network Prediction	Selected
Past consumption	82.94	✓
+ Time of the day	83.45	✓
+ Day of the week	83.73	✓
+ Season in the year	84.14	✓
+ Same hour previous 7 day	83.50	

Table VI
ORACLE RESULT : HOUSE- 900; APPLIANCE-LAMP

V. OVERALL RESULT

After the selection of the data, prediction is done for the following 24 hours at each hour and the results in terms of accuracy are shown in table VII. Here the prediction system is scored by two methods, one is by simple averaging all the accuracy for the 24 hours and the second one is a weighted average. The proposed weighing scheme is expressed by the following equation :

$$\Rightarrow \sum \frac{2(24-i)}{25*24} * Accuracy[i] \text{ for i=0,1,...,23}$$

It is a linear weighting scheme giving more importance to the first hour and least to the 24th hour. This is done due to the fact that all the hours other than the next hour will be predicted again in the next step. Results contain both the scoring methods and also the results of some trivial knowledge based predictors. The trivial knowledge based predictors are

- The predictor that always predicts the appliance wont start : Never starts.
- The predictor that always predicts the appliance will start : Always start.
- The predictor that predicts "what happens the previous day at the same hour happens the next day", i.e 24 hour similarity.
- The predictor that predicts "what happens the previous week at the same hour happens the next day", i.e 168 hour similarity.
- The predictor that predicts "what happen a random hour back happens the next hour" , i.e random hour similarity.

These estimates are important because they give an overall idea of the performance of the proposed model.

VI. DISCUSSION

The results indicates that the proposed model works better than other trivial knowledge based predictors. Previous works on appliance usage prediction from consumption data relied heavily on the assumptions expressed in the trivial knowledge based predictor. The assumption of 24 hour or 168 hour similarity is intuitive but other knowledges also need to be incorporated to make the system dynamic. The incorporation and representation of the expert knowledge helps the system to perform better as seen in the table VII. Now, from the results an overall idea about the predictability of the appliance is seen. Though it must be mentioned here that the high prediction for some appliances at homes are due to the fact that some appliances are ON or OFF at most of the time. Results show that both the categories (ON and OFF) for the appliances is predicted. Appliances which are started very few times seem to require less knowledge for prediction.

Among appliances, from the results, it indicates that the lamp requires most of the knowledges proposed by the expert among other appliances. The applicability of expert knowledge varies not only from appliance to appliance but also from House to House as the user behavior is different.

VII. CONCLUSION

To anticipate the energy needed for a service in a home automation system, the system must take into account the actions which will be done by the inhabitants. In this context, a proper prediction of energy demand in housing sector is very important. This work focuses on the prediction of the appliance usage in housing because it is a very important problem in a home automation system. The objective is to construct a model able to predict the appliance usage in housing which help the system to organize energy production and consumption and to decide which appliance will be used at each hour (energy planing). In this work we tried to predict if a particular appliance will be used at a particular hour looking 24 hours in the future. The proposed approach tried to formalize expert knowledge using predicate functions and also find a suitable data structuring for the classifier. The model is validated using an IRISE database which contains the consumption record of 100 houses for a period of 1 year. Our initial results indicate that the approach is useful in appliance usage prediction and its comparison with other trivial knowledge based predictor validates our approach. This model is applied to a wide range of appliances and houses and the initial results are encouraging.

Going into the future our aim is to build a general, fully automated and user interactive prediction system for home automation and simulate how the prediction is actually helping energy management in homes. By user interface we mean prediction also controlled by inhabitants where the users calender can be incorporated.

Appliance	Never Start	Always Start	24 hour similarity	168 hour similarity	Random hour similarity	Neural Networks (Average accuracy)	Neural Network (weighted accuracy)
900 Lamp	43.73	56.26	71.25	66.99	50.36	78.25	78.85
983 Electric Heater	71.76	28.23	94.38	89.39	92.35	96.31	96.68
925 Lamp	32.02	67.97	88.014	83.35	80.33	90.55	90.82
932 Oven	84.46	15.53	80.32	80.76	72.59	86.00	86.03
986 TV	72.13	27.86	71.36	69.43	57.30	76.64	77.18
951 Cooker	93.94	6.05	89.72	89.68	88.80	94.18	94.21
939 Washing machine	88.88	11.11	86.25	86.79	78.45	89.55	89.89

Table VII
OVERALL RESULT

REFERENCES

[1] P. Palensky and R. Posta, "Demand side management in private home using lonworks," in *Proceedings of the IEEE International Workshop on Factory Communication Systems*, 1997.

[2] K. Wacks, "The impact of home automation on power electronics," in *Applied Power Electronics Conference and Exposition*, 1993, pp. 3–9.

[3] S. Ha, H. Jung, and Y. Oh, "Method to analyze user behavior in home environment," *Personal Ubiquitous Comput.*, vol. 10, pp. 110–121, January 2006. [Online]. Available: http://dx.doi.org/10.1007/s00779-005-0016-9

[4] S. Abras, S. Ploix, S. Pesty, and M. Jacomino, "A multi-agent design for a home automation system dedicated to power management," in *Artificial Intelligence and Innovations 2007: from Theory to Applications*, ser. IFIP International Federation for Information Processing, C. Boukis, A. Pnevmatikakis, and L. Polymenakos, Eds. Springer Boston, 2007, vol. 247, pp. 233–241.

[5] S. Abras, S. Pesty, S. Ploix, and M. Jacomino, "An anticipation mechanism for power management in a smart home using multi-agent systems," in *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, april 2008, pp. 1 – 6.

[6] S. Abras, S. Ploix, S. Pesty, and M. Jacomino, "Advantages of mas for the resolution of a power management problem in smart homes," in *8th International Conference on Practical Applications of Agents and Multi-Agent Systems, PAAMS'2010*. Salamanca, Spain: Springer Verlag, 26-28 April 2010.

[7] S. Abras, S. Ploix, and S. Pesty, *Housing, Housing Costs and Mortgages: Trends, Impact and Prediction*, ser. Housing Issues, Laws and Programs. Nova Publishers, 2010, no. ISBN 978-1-60741-813-9, ch. Managing Power in a Smart Home Using Multi-Agent Systems.

[8] A. M. Elmahaiawy, N. Elfishawy, and M. N. El-Dien, "Anticipation the consumed electrical power in smart home using evolutionary algorithms," in *MCIT 2010 conference*, 2010.

[9] L. D. Ha, S. Ploix, F. Wurtz, P. Perichon, and J. Merten, "Energy management system for a photovoltaic grid-connected building," in *24th EU PVSEC and 4th World Conference on Photovoltaic Energy Conversion*, Hamburg, Germany, September, 21-26 2009.

[10] L. D. Ha, S. Ploix, M. Jacomino, and H. Le Minh, *Energy Management*, ser. ISBN 978-953-307-065-0. INTECH, 2010, ch. A mixed integer programming formulation of the home energy management problem.

[11] L. Hawarah, S. Ploix, and M. Jacomino, "User behavior prediction in energy consumption in housing using bayesian networks," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, Eds. Springer Berlin / Heidelberg, 2010, vol. 6113, pp. 372–379.

[12] E. A. Feinberg and D. Genethliou, "Load forecasting," in *Applied Mathematics for Restructured Electric Power Systems*, ser. Power Electronics and Power Systems, J. H. Chow, F. F. Wu, and J. Momoh, Eds. Springer US, 2005, pp. 269–285.

[13] H. Hippert, C. Pedreira, and R. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 44 –55, feb 2001.

[14] A. Bakirtzis, V. Petridis, S. Kiartzis, M. Alexiadis, and A. Maissis, "A neural network short term load forecasting model for the greek power system," *Power Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 858 – 863, may 1996.

[15] D. Park, M. El-Sharkawi, I. Marks, R.J., L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural network," *Power Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 442 –449, may 1991.

[16] A. Khotanzad, R. Afkhami-Rohani, and D. Maratukulam, "Annstlf-artificial neural network short-term load forecaster generation three," *Power Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 1413 –1422, nov 1998.

[17] B.-J. Chen, M. wei Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on eunite competition 2001," Tech. Rep., 2001.

Risk-based VV&A Assessment and Mitigation: A Naval Network Security System Test Facility Case Study

James Elele, Ph.D.; Naval Air Systems Command; Patuxent River, MD, USA

David Hall; SURVICE Engineering Company; Ridgecrest, CA, USA

Keywords: Model, Simulation, Risk Assessment, Verification, Validation, Accreditation, Risk Mitigation

Introduction

Models and Simulations (M&S) are used throughout the DOD systems engineering (SE) process, for all aspects of system development and deployment, from requirements definition through operational testing and mission rehearsal. DOD policies require that these M&S be accredited before the results are used for decision-making. The term Accreditation is used to define the process of making a decision that adequate information is available to demonstrate that the M&S and its results are credible enough for the intended purpose. But how can we really know that the M&S we use in this process will give us good enough answers? What do verification and validation (V&V) activities do for us in this process? And if I only have a minimal amount of resources (money and time), how can I do something meaningful to help me decide if the M&S and its results are credible enough? These same questions can also be asked about facilities that are designed to test and evaluate networked systems. Such facilities address critical security and suitability issues as part of the design and testing of systems of networks.

One needs to recognize that these networks are only representations of the actual systems in a laboratory environment; hence laboratories are not too different from M&S by definition. We are therefore inclined to ask the following questions: Is the lab environment adequately representative of the installed network system? If one conducts the test in the limited lab environment, how can we be sure that the system will work in the larger "real-world" environment? How much information does one need to answer these questions, and with what level of fidelity? How much credibility is enough for one to be able to use the facility results with confidence? What then is the risk associated with believing and using the results from these facilities, if their outputs/results are wrong? The following sections will describe the M&S VV&A processes, how they are applied, and one example case study of their application to networked security system testing. In particular, we will describe how we have applied the M&S Risk-based VV&A processes that we have developed over the last 20+ years to addressing these questions about laboratory test facilities.

Verification, Validation and Accreditation, and Credibility

But what are verification, validation and accreditation? And how do they relate to M&S (or test facility) credibility? The official definitions of M&S verification, validation and accreditation can be found in the Department of Defense VV&A Instruction (ref. 1):

- Verification is the process of determining that a model implementation and its associated data accurately represent the developer's conceptual description and specifications.
- Validation is the process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.
- Accreditation is the official certification (determination) that a model, simulation, or federation of models and simulations and its associated data are acceptable for use for a specific purpose.

In other words, Verification addresses whether the model does what the originator intended: "Did you build the model right?" Validation addresses how well the model matches the real world: "Did you build the right model?" And Accreditation addresses whether there is sufficient evidence to use it: "Did the decision-maker accept it?"

The following three basic criteria are those we use to describe the credibility of a model and hence to determine its viability for accreditation:

1. **Capability** – the functions it models and the level of detail with which they are modeled should support its anticipated uses.
2. **Accuracy** – how accurate it must be should depend on the risks involved if the answers are incorrect. M&S Accuracy is composed of three elements: *Software Accuracy* (including verification results and software testing), *Data Accuracy* (including input and embedded data V&V) and *Output Accuracy* (including comparisons with representations of the real world).
3. **Usability** – the extent of available user support to ensure it isn't misused should also derive from the importance of its application.

The essence of M&S accreditation is an objective comparison between the application requirements for the M&S with the information that is known about its credibility (capability, accuracy and usability) in the context of the problem at hand. The ultimate goal of M&S VV&A efforts is to demonstrate that M&S results have "good enough" credibility to provide confidence in program decisions. A cost-effective VV&A process cannot aim at securing the perfect model answer. It must recognize that the VV&A process can be resource intensive if not carefully tailored, and thus should only be geared towards demonstrating that the M&S and its outputs are "good enough" for the particular intended use at hand. "Good enough" is therefore determined for the most part by considering the risk involved if M&S results are in error. Thus accreditation is basically a decision that the user is willing to accept whatever residual risk remains after all V&V and related activities are completed, if decisions are based on M&S outputs that are wrong. Consequently, basing the VV&A process on risk allows us to focus activities on the areas of greatest potential impact to the program, and to those areas that reflect the most uncertainty in M&S outputs. By focusing VV&A efforts in this way we can allocate VV&A resources in the most cost-effective manner.

M&S VV&A as Risk Assessment and Mitigation

This risk-based VV&A process has been developed from the principles of risk assessment and mitigation in use by the system safety community. We have applied the risk assessment process described in MIL-STD-882 (ref. 2) to M&S and have developed a guide to conducting a cost-effective VV&A program based on the level of risk identified. Figure 1 illustrates the steps followed to arrive at an accreditation decision. The first two steps, "Analyze Intended Use" and "Develop M&S Requirements and Accreditation Information Requirements" are designed to produce an assessment of the risk of using the M&S to support the proposed decision (i.e. the "Intended Use"), and the application of that risk assessment in the development the M&S Accreditation Plan.

The U.S Defense Department's Model and Simulation Coordination Office (MSCO) VV&A Recommended Practices Guide (RPG) has this to say about risks associated with M&S: "Risks associated with simulation development and use can be categorized as either **Development Risk** or **Operational Risk**. **Development risks** are related to the simulation development itself and typically relate to potential problems in meeting technical, schedule, or cost aspects of the simulation development or modification program. **Operational risks** are those arising from using the incorrect outputs of a simulation that are believed to be correct." (ref. 3) Since V&V activities conducted during M&S development are intended to discover defects, they do therefore help mitigate developmental risk. But V&V efforts also compile the information needed to assess how well the model mimics the underlying concepts of operations, as well as how well the outputs from the M&S reflect operational data from the actual /"real" system. V&V support the assessment of operational risk, and hence **accreditation activities mitigate operational risk**.

The rest of the steps identified in Figure 1 involve the execution of the Accreditation Plan, including verification and validation activities and development of the case for accrediting the M&S for the intended use. As such, these activities comprise the risk mitigation actions for use of the M&S for that intended use, and the Accreditation Report will document the residual risks remaining after all VV&A activities have been completed.

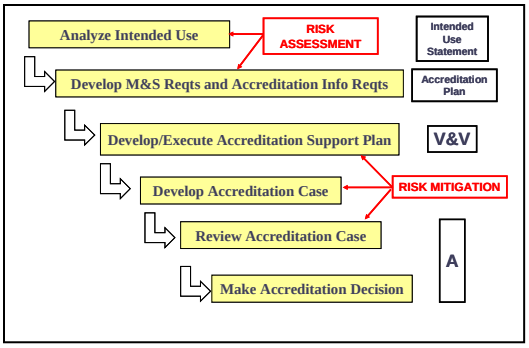


Figure 1 - Steps to an Accreditation Decision

How much V&V is enough to adequately support an accreditation?

The correct answer to this question is that it depends on the risk associated with using the M&S to support program decisions (the intended use). In order to evaluate that risk, we have developed a risk assessment technique for M&S. As defined in MIL-STD-882 for system safety, risk is composed of two elements: the impact (or consequences) of an event, and the likelihood that the event occurs. If you could assign a number to each of those components, you could express risk with the following equation:

Risk = (Consequence) x (Likelihood) (1)

For M&S, likelihood is the probability that the M&S and/or its input data are incorrect or inappropriate to the intended use; V&V activities address the likelihood of M&S errors. Consequence is the impact if the M&S output is wrong but you believe it and act on it; the consequence of using wrong M&S outputs depends on the role of the M&S outputs in making decisions and the importance of those decisions. It is usually not possible to derive a numerical value for consequence and likelihood, but it is possible to derive subjective estimates using standardized techniques that have been developed over the years and described in MIL-STD-882.

The approach for quantifying risk *likelihood* is based on the extent of known credibility information about the M&S when considering the requirements of the intended use. A model that has very little documented V&V information, especially if there is poor or no model documentation, would pose a likely source of errors (Level 5 in Table 1). As more information is known about the credibility of the model through documented V&V results, especially if model documentation is complete, the risk of error is reduced. M&S with extensive V&V documentation, adequate model documentation, and a history of effective configuration management of model software and data would provide the least likelihood of error, especially if the information is specifically evaluated in light of the intended use.

The approach we recommend for estimating M&S error *consequences* considers two elements of the impact of errors: (1) the *level of reliance* on M&S outputs in making a decision, and (2) the *importance of the decision*. This approach is in concept similar to the concept of "exposure", where one option for reducing risk is to reduce the potential exposure to the risk. One way to reduce the risk associated with using M&S is to limit the role that M&S outputs play in the decision-making process. More detail on this approach is provided in Reference 4, which leveraged the work documented informally in Reference 5.

Table 1 – Quantifying Likelihood of Error for M&S

Likelihood of Error	Level	Sample Criteria
Likely	5	No known V&V history; for new M&S developmental and V&V plans may have been written; limited or no M&S documentation
Probable	4	Documentation of Capabilities and Limitations and known Errors of candidate simulation; written V&V plans; some M&S documentation
Occasional	3	Level 4 + Adequate M&S documentation + Usage History + V&V history + Configuration Management Plan
Remote	2	Level 3 + SME face validation relevant to current intended use + Evidence of effective configuration management
Improbable	1	Level 2 + Extensive body of documented verification, validation & accreditation; extensive, disciplined M&S development history including technical and managerial review over time

The *level of reliance* on M&S outputs can be described in terms of the variety of information and methods that are available to support the decision in question. Table 2 illustrates four different levels of reliance and their definitions. At the highest level of reliance M&S are the only method of developing information to support a decision. At the lowest level M&S are only a supplement to other types of information, and the outputs from M&S can be verified against other sources.

Table 2 – Quantifying Level of Reliance on M&S

Level	Reliance on M&S to Make Decision
4	M&S will be the only method employed
3	M&S will be the primary method, employed with other non-M&S methods
2	M&S will be a secondary method, employed with other non-M&S methods, and will provide significant data unavailable through other means
1	M&S will be a supplemental method, employed with other non-M&S methods, and will provide supplemental data available through other means

The *level of importance* of the decision that is being supported by M&S outputs often cannot be reduced by the decision authority, since he or she is consigned to making the decision and has no other course of action. Reducing the importance of the decision would likely require a major change to the program approach, and would likely be a high-cost and schedule-busting alternative. Table 3 illustrates one scheme for categorizing the importance of the decision, based on the risk to the program associated with an unfortunate decision, and how many aspects of the program the decision may impact.

Table 3 – Quantifying Level of Decision Importance

Level	Importance of Decision
4	Intended use addresses multiple areas of significant program risk, key program reviews and test events, key system performance analysis, primary test objectives and test article design, system requirements definition, and/or high software criticality, used to make a technical or managerial decision
3	Intended use addresses an area of significant program risk
2	Intended use addresses medium or low program risk, other program reviews and test events, secondary test objectives and test article design, other system requirements and system performance analysis, and medium or low S/W criticality used to make technical or managerial decisions
1	Intended use addresses program objectives or analysis that is not a significant factor in the technical or managerial decision making process.

Once the level of reliance on M&S and the importance of the decision based on M&S results have been determined, we can combine those two elements into an overall "consequence value" for errors in the M&S results. Table 4 illustrates a recommended combination matrix and the resulting "scores" for M&S error consequence on the decision. Here the consequence ratings are broken into five categories, from 1 = Negligible to 5 = Catastrophic. As can be seen in the table, if M&S are the only method for supporting a decision, and the decision affects areas of significant risk to the program, then the resulting consequence of an error is a "5", or catastrophic. Once the likelihood and consequence ratings have been developed, MIL-STD-882 suggests a matrix whereby they can be combined into an overall risk assessment as illustrated in Table 5.

The MIL-STD-882 example of this table in fact divides risk into four categories, rather than three; however, the discussion below will describe the development of a VV&A process for M&S that is based on only three levels of risk. The matrix illustrated in Table 5 is tailorable to each M&S application.

Table 4 – Quantifying Consequence for M&S

Consequence		Level of Reliance on M&S			
		Supplemental Method	Secondary Method	Primary Method	Only Method
Importance of Decision		1	2	3	4
Multiple Areas of significant risk	4	3	4	5	5
An area of significant risk	3	2	3	4	5
Med-low risk areas	2	2	2	3	4
Not a significant factor	1	1	2	2	3

Table 5 – Quantifying M&S Risk Level

Likelihood of Error	Consequence (Impact) of Error				
	1 Negligible	2 Marginal	3 Moderate	4 Critical	5 Catastrophic
5 Likely	Low	Moderate	High	High	High
4 Probable	Low	Moderate	Moderate	High	High
3 Occasional	Low	Low	Moderate	Moderate	High
2 Remote	Low	Low	Low	Moderate	Moderate
1 Improbable	Low	Low	Low	Low	Moderate

Figure 2 illustrates the two basic risk reduction approaches that can be taken for M&S use. As shown in the figure, the risk can be reduced either by reducing the reliance placed on M&S results, or by increasing the credibility of the M&S results (or both). If we reduce reliance on M&S results, we move from right to left in the figure and lower the consequence of incorrect M&S results to the program (by making use of other information in addition to M&S results for decision-making). If we improve the credibility of the M&S we move from top to bottom on the figure by reducing the likelihood of a wrong M&S result (by conducting V&V activities, Subject Matter Expert (SME) reviews, improving or enhancing the M&S algorithms, improving M&S documentation – anything that improves the credibility of the M&S). Most programs consider both approaches in their risk reduction plans.

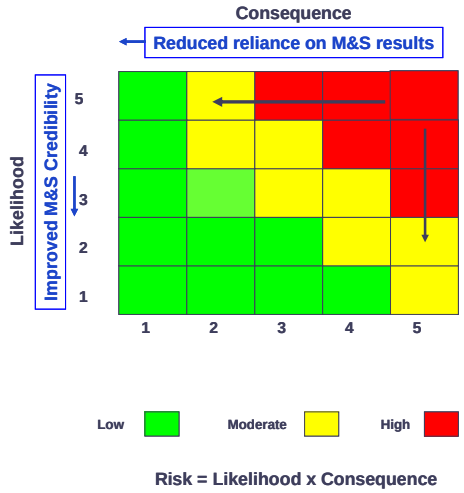


Figure 2 - M&S Risk Reduction Strategies

A Network Security System Case Study

We have applied the M&S risk-based VV&A process to an example network security system test facility in order to demonstrate the broader applicability of the approach beyond M&S. The example presented here is for a land-based test facility (LBTF) of a specific real-world Naval Network Security System (NNSS). The NNSS is intended to enforce network security policy for IT systems: detect malicious code, detect unauthorized intrusions, and provide virus protection and recovery for the IT systems in question. The NNSS is used as the LBTF for part of the operational test and evaluation (OT&E) of the facility. The LBTF simulates the naval operational environment for the network security system as part of the operational assessment (OA) of NNSS. Guidelines for OT&E of information and business systems is provided by a Director, Operational Test and Evaluation (DOT&E) Memorandum (ref. 6), which says that, "The degree of independent operational testing appropriate for each software increment or capability can be tailored by using risk analysis…". That guidance goes on to say that the level of operational testing that is required for an IT (or business) system is a function of its overall risk:

- Level I (low risk): the assessment may be based on integrated (that is, non-OT&E dedicated) testing and information
- Level II (moderate risk): the assessment must include an independent operational test event but can also include integrated non-OT specific tests and information
- Level III (high risk): an evaluation of the operational effectiveness, operational suitability, and survivability & security of the operational capability using the critical operational issues (COI) must be conducted, including an independent dedicated operational test

Risk assessment details

The following tables and discussion describe in more detail the assessment of risk for each of the facility characteristics shown in Table 6. In each case a table is provided that describes the rating scale for each of the seven criteria, the overall LBTF rating for each criterion, the justification for that rating, additional information that if provided could improve the rating, and suggestions for specific risk mitigation actions that should be taken to reduce the rating to "green". The rating scales have been developed and improved over the last several years via their use in support of a number of programs, M&S and test/laboratory facilities.

LBTF Risk Assessment Summary

Table 6 shows a summary risk assessment of using that facility for operational assessment of the NNSS system. Seven characteristics of the laboratory facility were examined for compliance with criteria we have developed as a result of working with a wide variety of defense systems and associated M&S and test facilities. As will be discussed below, the overall result of this risk assessment was "Moderate (Yellow)," meaning that there are identified areas that require improvements or additional information to justify an unqualified accreditation recommendation to the Accreditation Authority. The Table provides a list of the facility characteristics, evaluation criteria, and ratings.

Table 6 - Risk Assessment Summary

Facility Characteristic	Criteria	Rating
Intended Use	The specific intended use(s) is/are clearly stated.	YELLOW
Input Data	For each facility or laboratory, input data are credible and subject to review and revision.	GREEN
Facility Design	The facility design produces expected results.	YELLOW
System Verification	The laboratory or facility has been formally tested or reviewed and has been demonstrated to accurately represent the specific intended use(s) and requirements.	YELLOW
Results Validation	The facility's or laboratory's responses have been compared with known or expected behavior from the subject it represents and has been demonstrated to be sufficiently accurate for the specific intended use(s).	YELLOW
Configuration Management	For each facility or laboratory, components and their integration are supported by a sound Configuration Management (CM) Process.	YELLOW
User Community	Each facility or laboratory is designed and developed for the level of competency of the users for its intended purpose. The facility is supported by documents such as user's manual, technical manual, and/or reference guide.	GREEN

Intended Use:
*RATING: **Yellow** (program goals not yet established*

1. Justification for Rating: The intended use of the facility seems clear for this application: the Test and Evaluation Master Plan (TEMP) shows the relationships between critical operational issues (COI), Measures of Effectiveness (MOE), Measures of Suitability (MOS), planned test methodologies, and the decision or milestone supported. The TEMP also lists some of the limitations on the test due to the use of the laboratory equipment, and it lists the test objectives in detail along with detailed requirements and evaluation procedures. However, the intended use is not specifically stated as such, and must be inferred from several documents, and details are missing.

2. Information not previously provided that may affect the accreditation rating assigned for the facility's intended use. (What extra information can the Developer or the M&S SME provide?) Specify in more detail the intended use statement in the request for accreditation. Currently in the V&V Plan (and Report) the statement of the laboratory application is: *"The LBTF is used to conduct post-development integration of engineering improvements and configurations and to establish operational baselines. Lab engineers also use the LBTF to provide Tier 2 and Tier 3 Help Desk support to the fleet."* More details would be useful to better describe the intended use.

3. Recommend any mitigation (VV&A) steps necessary to bring the rating up to green. Provide a detailed intended use statement including how the laboratory facility will be used to provide specific OA and OT&E COI information.

Input Data:
*RATING: **Green** (All data are valid or certified)*

Justification for rating: The V&V Plan (and Report) lists a table of all data verification activities and results as well as a table of all data validation activities. These were done by both the program V&V team and by IV&V agents, primarily via visual inspection of the facility equipment and the results. Some discrepancies were noted and recommendations made to ensure a successful OA. While most of these tests were designed to verify operation of the system, the IV&V agent was able to compare results to a real-world data set capture report. The V&V and IV&V agents conducted a "side by side" comparison of the NNSS as-built and the LBTF as-built and verified that the LBTF input dataset is a representative build-out of the Operational Site.

Facility Design:
*RATING: **Yellow** (The design requires some improvement to improve results credibility)*
1. Justification for rating: Design Verification: the V&V Plan (and Report) contains a table with planned activities and results comparing the diagrams of the facilities with the actual hardware layout (of the facilities). The V&V Plan also lists limitations and it lists risks and impacts, but from a programmatic standpoint rather than from the technical limitations of the lab facility. The TEMP has a good list of the system's components and their application to the systems capabilities.

2. Information not previously provided that may affect the accreditation rating assigned for the facility design. (What extra information can the Developer or the M&S SME provide?) Provide a detailed comparison between the facility design and the actual on-shore and off-shore equipment and linkages.

3. Recommend any mitigation (VV&A) steps necessary to bring the rating up to green. Better document the detailed rationale behind the facility design.

System Verification:
*RATING: **Yellow** (the system has been tested informally and represents the intended use and requirements)*

1. Justification for rating: Implementation Verification: the V&V Plan (and Report) contains a table with planned activities and results, basically reviews of documentation by the IV&V team and of tests to ensure backup and restoration capability.

2. Information not previously provided that may affect the accreditation rating assigned for the system's verification. (What extra information can the Developer or the M&S SME provide?) The system has been formally tested and appears to be operable and represent the intended use (as far as is has been stated). However, most of the tests appear to have been primarily of operation and backup of the system and not specifically geared toward verification that the system meets the intended use. The documentation review likely was geared toward that verification, but if so the result was not clearly spelled out.

3. Recommend any mitigation (VV&A) steps necessary to bring the rating up to green. Provide documentation of actual system verification results (how the laboratory setup compares with the intended use and requirements). This will depend on obtaining a more specific and detailed definition of the intended use.

Results Validation:
*RATING: **Yellow** (The results have been examined and are not sufficiently accurate for the intended use)*

1. Justification for rating: Implementation Validation: the V&V Plan (and Report) contains a table summarizing the V&V testing performed; they also executed backup and restore tests and validated the results by visual inspection. Operational Concept (Conceptual Model) Validation – the V&V Report compares the lab setup with the desired "Tier 3" capability – and identified some limitations and discrepancies. Appendix B of the V&V Plan identifies a number of limitations to the test matrix, along with a risk assessment for each (they all were identified as low risk with only one exception regarding a backup system; two high risk faulty hardware issues were resolved with additional equipment procurements), and actions to mitigate the risk for those items that can be mitigated (basic lab limitations were not assigned a mitigation action, presumably because they would have been time and cost prohibitive).

2. Information not previously provided that may affect the accreditation rating assigned for results validation. (What extra information can the Developer or the M&S SME provide?) Most of the testing appears to be related to ensuring proper operation of the system, rather than tests of how well the system replicates the real-world systems in operation. It would be useful for the IV&V agent to provide an independent assessment of how the laboratory system compares with the actual system in operation.

3. Recommend any mitigation (VV&A) steps necessary to bring the rating up to green. Provide a documented independent assessment comparing real-world results to the results of the laboratory setup.

Configuration Management:
*RATING: **Yellow** (Some CM processes exist for all major upgrades)*
1. Justification for rating: There was no CM Plan provided; however the V&V Plan (and Report) says that the LBTF personnel use CMPro ™ as the primary configuration management (CM) tool. The Configuration Control Board (CCB) meets regularly to ensure issues are tracked accordingly.

2. Information not previously provided that may affect the accreditation rating assigned for the facility's configuration management processes. (What extra information can the Developer or the M&S SME provide?) Need to document the CM process and demonstrate implementation.

3. Recommend any mitigation (VV&A) steps necessary to bring the rating up to green. LBTF personnel should Develop (and demonstrate that they are following) a Configuration Management Plan.

User Community:
*RATING: **Green** (User community has the ability and tools to fully utilize the facility)*

1. Justification for rating: Documents appear to be available describing the operation and design of the laboratory facility (from the list of documents reviewed by the IV&V agent). In addition, only qualified personnel from the laboratory facility will be using the facility to support testing and OA of the NNSS system.

Overall LBTF Risk Assessment

The overall risk assessment for the LBTF as applied to operational assessment of the NNSS is illustrated in Figures 3 and 4. Figure 3 is the consequence determination chart: the level of reliance on the facility was judged to be "3" based on it being the primary method to support the operational assessment; the importance of the decision was judged to be a "4", since the decisions based on LBTF results will affect multiple areas of significant program risk. The overall result as shown on the consequence matrix is that the resulting consequences of erroneous results from LBTF are severe (5).

The likelihood of erroneous results was judged to be minimal (likelihood value of 2), since there has been extensive facility system testing, IV&V reviews of available results, expert users, and known (albeit undocumented) configuration management processes, along with a history of prior successful use by other programs.

Consequence		Level of Reliance on M&S			
		Supplemental Method	Secondary Method	Primary Method	Only Method
Importance of Decision		1	2	3	4
Multiple Areas of significant risk	4	3	4	5 X	5
An area of significant risk	3	2	3	4	5
Med-low risk areas	2	2	2	3	4
Not a significant factor	1	1	2	2	3

Figure 3 - Consequence Determination Chart

As the risk matrix in Figure 4 shows, a likelihood of 2, and consequence of 5 yields a moderate overall risk of using LBTF to support the operational assessment of NNSS. The result of this risk assessment is that the program may make use of existing non-dedicated OT&E test results along with other information, in addition to a dedicated OT&E test to develop the operational assessment of NNSS.

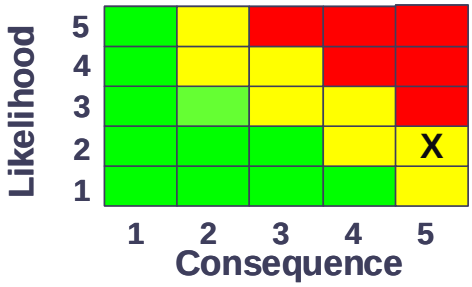


Figure 4 Overall Risk Assessment

Summary

Cost-effective M&S VV&A programs are based on an assessment of the risk of using M&S results to support program decisions, and risk-mitigation actions should drive development of M&S Accreditation Plans. This risk-assessment, risk-mitigation approach to VV&A has been borrowed from system safety risk principles described in MIL-STD-882 and adapted to the unique requirements of M&S credibility. Application of risk principles to M&S VV&A is particularly appropriate in those cases where errors in M&S results could potentially affect personnel or equipment safety.

We have applied the M&S VV&A Risk process to assessing the risk of using the land-based test facility in support of an operational assessment of NNSS. The risk assessment was carried out against specific facility credibility characteristics, which were then combined into an overall risk assessment based on the M&S approach. The resulting "moderate" risk requires that the program conduct at least one dedicated OT&E test event, but it does allow for extensive use of existing non-dedicated test and other information to support the OA. This approach was successful even though there were very limited funds and time available to carry out the risk assessment.

The likelihood of error is reduced by conducting V&V as part of the risk-mitigation process. The consequence of error can only be reduced by lessening the dependence of the decision-making process on test facility or M&S results. Ultimately the Accreditation Authority must decide whether the residual risk achieved after the completed VV&A tasking is acceptable for the particular intended use. The residual risk may also determine at what level in the approving organization is required of the accrediting authority. That is, a high residual risk may require a higher authority for approval as is the case with system safety risk. Ultimately, the level of risk that the accreditation authority is willing to assume drives the overall V&V effort required to support an accreditation decision.

References

1. Department of Defense, *DOD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)*, DOD Instruction 5000.61, December 9, 2009.
2. Department of Defense, *Standard Practice for System Safety*, MIL-STD-882D, 10 February 2000
3. Model and Simulation Coordination Office, *Verification, Validation and Accreditation Recommended Practices Guide (RPG) Special Topic Paper on Risk Assessment and Mitigation*, 11/30/2000.
4. International Test and Evaluation Association, *Assessing Risk Levels of Verification, Validation, and Accreditation of Models and Simulations*, James N. Elele, PhD, ITEA Journal Volume 29-2, June 2008; pp 190–200.
5. Naval Air Warfare Center, Weapons Division, China Lake CA, *Risk Based VV&A: A Step-by-Step Guide to Developing and Implementing a Cost Effective M&S VV&A Strategy for Your Acquisition Program*, May 2008, Viewgraph Presentation by Michelle Kilikauskas
6. Department of Defense, Director, Operational Test and Evaluation (DOT&E) Memorandum, *Guidelines for Operational Test and Evaluation of Information and Business Systems*, September 14, 2010

Social Media as Legal Evidence: The Quest for Online Social Capital at the Expense of Privacy Causing Offline Legal Consequences in Family Court

Dana C. Hackley, M.S.
Department of Communications Media, Indiana University of Pennsylvania
Indiana, PA 15705, USA

and

Carrie West, M.A.
Department of English, Communications, and Fine Arts, Wheeling Jesuit University
Wheeling, WV 26003

ABSTRACT

Online social media has allowed users to share more information than ever before in order to obtain various degrees of social capital. However, this new level of exposure brings with it possible legal ramifications as attorneys take advantage of e-discovery now available from social networking sites particularly in family court cases. The following is a content analysis of U.S. published divorce and child custody court cases involving the utilization of either personal Facebook or MySpace pages as legal evidence within proceedings. The findings indicate thus far, individual desire for online social capital has outweighed any possible court room consequence.

Keywords: Social media, Facebook, MySpace, Divorce, Child custody, Law

INTRODUCTION

In divorce and child custody cases, the gloves come off with attorneys from both sides attempting to dig up dirt on the other's client in order to discredit them in court. As we enter a new era of multimedia and computer mediated communication, the modern day "War of the Roses" has taken on new meaning. The U.S. judicial system has been setting new precedents in regards to the admissibility of e-discovery and electronically stored information (ESI). In addition, social media mining has become common practice for attorneys and firms.

As social media is becoming more utilized within civil court proceedings, particularly divorce and child custody cases, the motivation behind social media usage and its relationship to online social capital as well as the resulting impact on privacy are in need of evaluation.

REVIEW OF LITERATURE

Through legal journals and continuing legal education seminars attorneys are learning how to utilize and defend against electronic data discovery. Akin (2011) as well as several other legal journals (Dysart 2011) provide tips for locating damaging information including cached or hidden content as well as how to get that information into admissible form for trial or pre-trial motions. Some firms are even using specialized software to monitor social networking sites (Akin 2011).

The American Bar Association and the American Academy of Matrimonial Lawyers (AAML) have surveyed attorneys regarding their use of social media within legal proceedings. The 2010 AAML study found Facebook is the, "unrivaled leader for online divorce evidence" with 66 percent of attorneys citing it as a primary source. MySpace followed with 15% of attorneys utilizing the site as evidence. Also, 81% of AAML members cited an increase in the use of evidence from social networking websites during the past five years. However, the legal utilization of e-discovery is not a new phenomenon (Leroux 2004).

E-Discovery & Technology as Evidence

Society's information structure changed following World War II, but it wasn't until the 1970's and 1980's with the initiation of personal computers and faster processing that electronically stored and

created information became reality, having an impact on the legal and justice system (Paul & Nearon 2006). Federal and state laws as well as international rules have been developed requiring companies to preserve specific data for specific periods of time (Paul & Copple 2005). But it is user generated content that has had the greatest impact of late. According to Demay (2011), social media is changing the way individuals, companies, and the government organizes, navigates and shares information as well as the very nature of privacy in society. While electronic discovery tends to benefit defendants rather than plaintiffs, social media evidence may prove to be beneficial to plaintiffs (Demay 2011). Gartner Research predicted in a 2010 study that half of all companies will be required to produce social media records for e-discovery requests by 2013 saying, "'if it exists, it is discoverable" (Gartner 2011).

Social media posts present similar admissibility issues as e-mails or text messages, because of the difficulty in authenticating the source, yet the practice of using such postings as evidence in proceedings has become frequent (Zemlicka 2010). Social media is most widely utilized as evidence in defamation (Azriel 2011), healthcare privacy (Hader & Brown 2010), worker's compensation, sexual harassment, divorce and child custody cases.

Social Capital

Theorists have developed the concept of social capital to describe the value placed on or attributed to networks (Coleman 1988). Putnam (2000) even went as far as to differentiate between bonding and bridging forms of social capital in regards to the strength of ties or connections created between people. Leiner, Hohlfeld, & Quiring (2009) determined that the relevance of social media on social capital is really decided by whether the capital can be consumed and thus converted into cultural knowledge, social power or economic advantage. Young (2011) similarly concluded that among adult Facebook users, the SNS is an efficient and convenient way to maintain relationships with a larger and more diverse group of acquaintances thus expanding one's social capital. But with the integration of SNS, social science researchers have just really begun investigating the impact of offline interaction with online social networking and vice versa. Matzat (2010) found that people are held more accountable for their online activity within the same group of people. Additionally, Vergeer & Pelzer (2009) found that online networking augments and increases offline socializing and Valenzuela, Park & Kee (2008) concluded that a positive relationship exists between Facebook use and a student's life satisfaction, social trust, civic and political participation. Although, SNS does provide a double-edged sword in regards to social capital.

Researchers have determined that social networking has an influence on an individual's self-worth (Muise, Christofides & Desmarais 2009; Schouten, Valkenburg & Peter 2009) while others raise concern that it heightens emotions such as jealousy (Foulger, Ewbank, Kay, Popp & Carter 2009) and a feeling of less engagement in work and life (Junghyun, LaRose & Wei 2009). In addition, the relationship between online social capital gained or lost from social networking sites and offline social capital is being investigated by communication researchers (Vergeer & Pelzer 2009).

Privacy and Motivation

Communication uses and gratifications theory seeks to understand why people become involved in one particular type of mediated communication or another and what gratifications they receive from it. Guosong (2009) found three reasons why people embrace user-generated content:

1. Fulfilling their information, entertainment, and mood management needs.
2. Taking advantage of user-generated sites to interact with the content and other human beings.
3. Allowing self-expression and self-actualization, both of which may ultimately be aimed at constructing their own identity.

Social media users often willingly give up privacy rights without realization that online content is admissible in a court of law. The most notable case regarding social media privacy is *Crispin v. Christian Audigier, Inc*. In the case, the magistrate judge ruled the Stored Communications Act applied to online social media, meaning sites like Facebook and MySpace are not compelled to comply with a subpoena for information. However, attorneys have been locating loopholes around the act, such document discovery requests made under Federal Rule of Civil Procedure 34.

While the Federal Rules of Civil Procedure have been in use for decades, rules specifically pertaining to the discovery of electronically stored information (ESI), made by amending Rules 16, 26, 33, 34, 37 and 45 and Form 35, became effective on December 1, 2006. ESI includes e-mail, instant messaging chats, documents, accounting databases, CAD/CAM files, Web sites, and any other electronically stored information that could be relevant evidence in a law suit. Following the federal court, according to K & L Gates, an international law firm, forty-one state courts have also developed civil procedure rules pertaining to e-discovery (K&L).

But the ethics involved for attorneys within social media mining is still being debated. Ethics committees have generally agreed that attorneys must avoid engaging in deception when attempting to obtain information via social media sites from parties to litigation, but it is not illegal.

The New York City Bar Association Ethics Committee even determined that an attorney may, directly or through an agent, "friend" an unrepresented party to litigation without disclosing the reason for the request, but "[r]ather than engage in 'trickery,' lawyers can — and should — seek information maintained on social networking sites, such as Facebook, by availing themselves of informal discovery, such as the truthful 'friending' of unrepresented parties, or by using formal discovery devices such as subpoenas directed to non-parties in possession of information maintained on an individual's social networking page" (Black 2011) On the other hand, other ethics committees such as the Philadelphia Bar Association Professional Guidance Committee and San Diego County Bar Association Legal Ethics Committee ruled that an attorney may "friend" an unrepresented party only if the reason for the communication is disclosed (Black 2011).

When Dwyer, Hiltz, Passerini (2007) compared perception of trust and privacy between Facebook and MySpace, Facebook users were more willing to share information while MySpace users considered themselves more experienced with social media in general. Additionally, users had no problem or resistance to making new online relationships in a world of perceived weak privacy controls (Dwyer, Hiltz, Passerini 2007).

METHODS

As demonstrated in the review of literature SNS users seek social capital online, but the legal profession is utilizing the content offline in a court of law. Thus, focusing on published court cases, the following research questions were addressed:

RQ1: With what frequency are U.S. divorce and child custody cases utilizing either Facebook or MySpace as evidence?

RQ2: How are Facebook and MySpace being utilized as evidence in divorce and child custody cases?

RQ3: Which social networking site, Facebook or MySpace is more frequently utilized as evidence in divorce or child custody cases to discredit one's reputation or credibility?

In order to answer the above research questions, the study investigated published U.S. court cases in all jurisdictions. Opinions are published at the direction of the court if they add something to the current body of law.

The Fastcase database was used through the West Virginia Bar Association's website to search published court cases from 1925 to September 12, 2011. However, because the social network MySpace was launched toward the end of 2001 and Facebook in 2004, the earliest published case was in 2001. Cases were analyzed using the following keywords: Facebook, MySpace, divorce, and child custody. Duplicate cases were excluded from analysis.

In total 211 cases were examined by one of the authors, as sole coder. Themes were created to descriptively analyze the findings as means, frequencies and proportions.

RESULTS

The keyword searches of published court cases spanning all jurisdictions resulted in a total of 264 cases containing either divorce and Facebook (23), child custody and Facebook (47), divorce and MySpace (53) or child custody and MySpace (141).

As seen in Table 1, 2010 showed the most (36%) of divorce or child custody cases involving the social

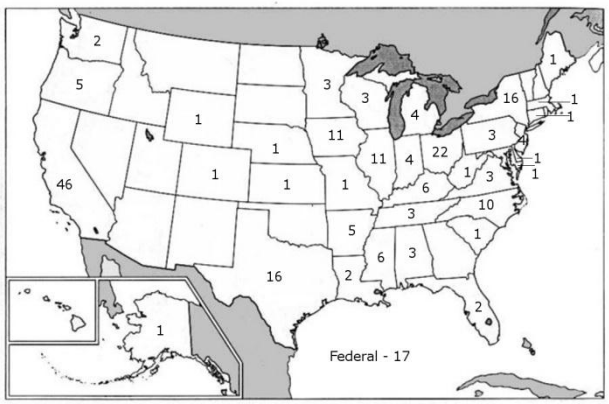
media Facebook or MySpace. According to Facebook, the site reached over 500 million active users in 2010 increasing by 150 million active users from 2009's number while MySpace according to media reports dropped from around 76 million users in 2009 to around 64 million users in 2010.

Table 1: Frequency distribution of published court cases by year

Year	Published Cases
2011	61 (29%)
2010	77 (36%)
2009	48 (23%)
2008	15 (7%)
2007	8 (4%)
2003	1 (.5%)
2001	1 (.5%)

As evident in Figure 1, California and New York saw the highest number of published divorce and child custody cases utilizing social media.

Figure 1: Frequency distribution of published court cases by state



The utilization of either Facebook or MySpace within divorce or child custody proceedings as evidence was categorized into the following descriptive themes: content sexual in nature, content to prove a claim made in court false, content to refute one's credibility or tarnish their reputation, as a means of communication or contact, attorney use of social media, juror use of social media and reference to a legal precedent in which Facebook or MySpace are named. Table 2 indicates the frequency of these themes within the analyzed cases.

Table 2: Frequency distribution of published court cases by social media and content themes

	Sexual Content, Affair	Proving False Claims	Credibility / Reputation	Contact
facebook	2	4	28	18
myspace.com	29	10	91	30

	Attorney on Social Media	Juror on Social Media	Legal Precedent
facebook	1	2	2
myspace.com	0	2	5

***12 cases cited both SNS**

MySpace (74%) was disproportionately referenced more than Facebook (26%). Additionally, cases involving sexual content cited MySpace (94%) more often than Facebook (6%). However, both sites were utilized as evidence of immoral or prejudicial conduct making up 56% of the total content.

DISCUSSION

The content evaluated in regards to social media utilization as legal evidence demonstrated online users are frequently not taking into consideration the possible offline repercussions of their online actions. The gratification that comes from sharing information has outweighed the reality that one's online footprint is everlasting and admissible in a court of law. Social gratification has become more important than abstract potential implications for SNS users. The published court cases analyzed within the course of this study mentioned the use of social media to threaten, expose, and demonstrate lewd activity. The very use of social media by children was provided by parents as leverage hoping to discredit the other's parenting skills and children posted messages painting their parents as irresponsible, alcoholic, abusers. The published court proceedings provided greater insight into these families than any of the parties involved most likely realized when they initially posted the content online. Future research may consider whether social media uses by litigants were to change post litigation.

References

A.L. Hader & E.D. Brown, "Legal briefs: Patient Privacy and Social Media." **AANA Journal**, Vol. 78 No. 4, 2010, 270-274.

A. Muise, E. Christofides & S. Desmarais, "More Information than You Ever Wanted: Does Facebook Bring Out the Green-Eyed Monster of Jealousy?", **CyberPsychology & Behavior**, Vol. 12 No. 4, 2009, 441-444. doi:10.1089/cpb.2008.0263.

A.P. Schouten, P. Valkenburg & J. Peter, "An experimental test of processes underlying self-disclosure in computer-mediated communication", **Cyberpsychology**, Vol. 3 No.2, 2009.

C. Dwyer, S.R. Hiltz & K. Passerini, "Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace", Proceedings of the Thirteenth Americas Conference on Information Systems, Keystone, Colorado August 09 – 12, 2007.

C.J. Akin, "How to Discover and Use Social Media-Related Evidence". **Litigation**, Vol. 37 No. 2, 2011, 32-34.

Crispin v. Christian Audigier, Inc., 717 F. Supp. 2d 965 - Dist. Court, CD California 2010.

D. Leiner, R. Hohlfeld & O. Quiring, "What People Make of Social Capital Online: Empirical Study on Capital Conversion via Networking Sites", **Conference Papers -- International Communication Association**, 2009, 1-27.

Erase Your Own Online Evidence, **Annals of the American Psychotherapy Association**, Vol. 12 No. 3, 2009, 9.

"Gartner: Firms Must Manage Social Media Better", **Information Management Journal**, Vol. 45 No.3, 2011, 7.

G.L. Paul and R.F. Copple, "Dealing with Data: No, you can't call them documents anymore", **Business Law Today**, 2005.

G.L. Paul and B.H. Nearon, "The Discovery Revolution", Chicago, IL: American Bar Association Publishing, 2006.

G. Rosenberg, "Electronic Discovery Proves an Effective Legal Weapon", **New York Times**, 1997, p.5.

J.N. Azriel, "Using social media as a weapon to harm victims: Recent court cases show a need to amend section 230 of the communications decency act", **Journal of Internet Law**, Vol. 15 No. 1, 2011, pp. 3-10.

J.Dysart, "The Trouble with Terabytes"**. ABA Journal**, Vol. 97 No. 4, 2011, 32-62.

J.E. Demay, "The Implications of the Social Media Revolution on Discovery in U.S. Litigation". **Brief**, Vol. 40 No.4, 2011, 55-64.

J.S. Coleman, "Social capital in the creation of human capital", **American Journal of Sociology**, Vol. 94, 1988, S95-S120.

J. Zemlicka, "High-tech hearsay? Credibility and admissibility of Facebook posts are debatable", **Wisconsin Law Journal**, 2010.

K. Junghyun, R. LaRose & P. Wei, "Loneliness as the Cause and the Effect of Problematic Internet Use: The Relationship between Internet Use and Psychological Well-Being", **CyberPsychology & Behavior**, Vol. 12 No. 4, 2009, 451-455. doi:10.1089/cpb.2008.0327.

K & L Gates. "Updated List: Local Rules, Forms and Guidelines of United States District Courts Addressing E-Discovery Issues", **Electronic Discovery Law Blog**. http://www.ediscoverylaw.com/articles/resources/

K. Young, "Social Ties, Social Networks and the Facebook Experience", **International Journal of Emerging Technologies & Society**, Vol. 9 No. 1, 2011, 20-34.

M. Vergeer & B. Pelzer, "Consequences of media and Internet use for offline and online network capital and well-being. A causal model approach, **Journal of Computer-Mediated Communication**", Vol. 15 No. 1, 2009,189-210. doi:10.1111/j.1083-6101.2009.01499.x.

N. Black, "Commentary: Legal Currents: California Ethics Committee on social media mining". **The Daily Record**, Rochester, NY. 2011.

N.B. Ellison, C. Steinfeld & C. Lampe, "The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites**, Journal of Computer-Mediated Communication**, Vol. 12 No. 4, 2007, 1143-1168.

O. Leroux, "Legal admissibility of electronic evidence", **International Review of Law, Computers & Technology**, Vol. 18 No. 2, 2004, 193-220.

R. C. Losey, "Electronic Discovery: New Ideas, Case Law, Trends and Practices", Eagan, MN: West Publishing, 2010.

R.D. Putnam, "Bowling Alone: The collapse and revival of American community", New York: Simon & Schuster, 2000.

S. Guosong, "Understanding the appeal of user-generated media: a uses and gratification perspective", **Internet Research**, Vol. 19 No. 1, 2009, 7-25, doi:10.1108/10662240910927795.

S. Valenzuela, N. Park, & K.F. Kee, "Is There Social Capital in a Social Network Site?: Facebook Use and College Students' Life Satisfaction, Trust, and Participation, **Journal of Computer-Mediated Communication**, Vol. 14 No.4, 2009, 875-901. doi:10.1111/j.1083-6101.2009.01474.x.

T.S. Foulger, A. D. Ewbank, A. Kay, S.O. Popp, H.L. Carter, "Moral Spaces in MySpace: Preservice Teachers' Perspectives about Ethical Issues in Social Networking", **Journal of Research on Technology in Education**, Vol. 42 No. 1, 2009, 1–28.

U. Matzat, "Reducing Problems of Sociability in Online Communities: Integrating Online Communication With Offline Interaction", **American Behavioral Scientist**, Vol. 53 No. 8, 2010, 1170-1193.

META: a new hybrid methodology to software development created to suit the current needs in Mexico

for ICTA 2011

E.Miriam JIMENEZ-HERNANDEZ
Technology Software and Databases, Center for Research in Computing (CIC), IPN.
Mexico City, 07738 / Mexico.

and

Sandra D. ORANTES-JIMENEZ
Technology Software and Databases, Center for Research in Computing (CIC), IPN.
Mexico City, 07738 / Mexico.

ABSTRACT

This article presents a new hybrid methodology to develop software projects named META (*MEtodología Tradicional y Ágil*, Agile and Traditional MEtodology); which was created taking into account the current needs of software development companies in Mexico.

Keywords: agile methodologies, hybrid methodologies, META, object-oriented methodologies, Software Engineering, traditional methodologies.

1. INTRODUCTION

According to the Dictionary of the Royal Spanish Academy, the word *methodology* means a set of methods that are followed in a scientific research or in a doctrinal exposition [1].
But in Software Engineering the term refers to a framework used to structure, plan and control the process of developing computer systems [2].
So, it is expected that by using a software development methodology, this can provide a set of practices and tools that facilitate the development process, getting a product with high quality, safe and meets customer expectations [3].
There are many different methodologies which are included in two major types, traditional and agile; however the hybrid methodologies are setting a new trend in the area of Software Engineering by merging the best those types of methodologies.
So taking into account the new trend, META was designed as a hybrid methodology, which could be a good option for solving the current problems present in the software development companies in Mexico. Also could be applied to companies in other countries, just is important to know the main characteristics for which are ideal the use of META.

2. TRADITIONAL METHODOLOGIES

Traditional methodologies have two types: structured programming methodologies and object-oriented methodologies.
With the emergence of structured programming, were born some methodologies for this paradigm, as the created by Yourdon [4], in whom the use of Data Flow Diagrams (DFD), Context Diagrams (CD), State Transition Diagrams (SDT) and Entity-Relationship diagrams [5] appeared as a constant in the system modeling.
In the same way, when the paradigm of object-oriented programming appeared, specific methodologies emerged too for this new paradigm, as OMT (Object Modeling Technique) [6], RUP (Rational Unified Process) [7], Metric 3 [8], among others.
This kind of methodologies positively contributed to the existing traditional methodologies being interactive and incremental. They also promoted the allocation of roles, facilitated the division in subsystems and promoted the reuse of components [9].
So generally, traditional methodologies considered the importance of system documentation, allowing understand, extend and maintain the software; because these methodologies provide well-defined structure and order [10].
However they also have some disadvantages, for example, require a high degree of discipline, there not exist a rapid respond to change, they generates unnecessary documentation and they require to invert a lot of time in the system modeling. Besides these methodologies have a strict project plan; therefore these do not consider the unpredictable that can be the analysis, design and construction.

3. AGILE METHODOLOGIES

The scheme proposed in previous methodologies has proven to be effective and necessary in large projects (with respect to time and resources). However, this approach is not the most suitable for many existing projects, in which the system environment is changing and requires, reduce development time.

In this scenario the agile methodologies were born, as XP (eXtreme Programming) [11], Crystal Clear [12] and Scrum [13], which have certain advantages and disadvantages.
Some of the advantages are that they have quick and effective responses to change. Also, they have a flexible project plan and present some simplicity in the development process.
However, agile methodologies generate little documentation and do not make use of formal methods.
The main characteristics of agile with traditional methodologies are summarized in **Table 1**.

Table 1: Agile vs. Traditional Methodologies

Agile methodologies	Traditional methodologies
Heuristic-based from practices of production code	Rules-based from followed standards by the development environment
Less process controlled, whit few principles	Much more controlled process, with many policies/standards.
Few artifacts	More artifacts
Few roles	More roles
Less emphasis on software architecture	The software architecture is essential
Produce little documentation	Produce a lot documentation
Easy to learn and implement	Difficult to learn and implement
They do not require a lot of discipline	Requires a great discipline
Specially prepared for changes during the project	Some resistance to change
Do not have a traditional contract or at least is quite flexible	There is a fixed contract

4. HYBRID METHODOLOGIES

As shown, the methodologies have evolved, as the programming paradigms did. An evidence of this is the new tendency: the hybrid methodologies. With the hybrid methodologies have sought take the advantages of each of the previous methodologies, to create a combination of the best practices described in each.
It could be mentioned EssUP (Essential Unified Process) [14] as the pioneer of hybrid methodologies. EssUP was created by Ivar Jacobson, and it is based on the Unified Process (UP) [15], agile methods and maturity of processes.
EssUP tries to be agile, because EssUP does not intend to impose a specific process, also it considers the necessity to be flexible and have a rapid response to changes. But it makes clear the necessity to documenting, as the traditional methodologies do, and modeling with UML [16]. However, it is conceptually a hybrid methodology, because in practice the software development team whose intends to use this methodology must select the lifecycle model of software development that best suits to the necessities. Thus presents a major problem if it does not have the experience and knowledge to choose the best practices existing in Software Engineering and appropriately to apply them to each software project.

5. META

META (*MEtodología Tradicional y Ágil*, Traditional and Agile Methodology) is a hybrid methodology for developing software projects. META is a methodology that combines some existing practices within RUP (Rational Unified Process), XP (eXtreme Programming) and Scrum, making it a hybrid between traditional and agile.
META was created to suit the current needs of Mexican companies engaged in software development. It was decided to take Mexico as a study case, but META can be used in other countries, as long as it takes into account META´s characteristics.
So, part of META development was to investigate the real and current situation regarding the usage of methodologies and some practices of Software Engineering in such companies.
According to INEGI (*Instituto Nacional de Estadística y Geografía,* National Institute of Statistics and Geography) [17], in 2010 there were 9540 companies in Mexico dedicated to software development, so to calculate the size of the sample, it was applied a quiz to a pilot group of 20 companies, with the information of such evidence it was obtained that the sample size should be 86 companies.
So, 86 companies were surveyed randomly selected from the list provided by INEGI. With the obtained data, it was performed a hypothesis testing for a proportion, obtaining the next result:

"50% or more of the software development companies have an inclination towards the use a hybrid methodologies. "

To corroborate the hypothesis test a confidence interval was performed for a confidence level of 95%.
Once it was proved the hypothesis, some important additional information was obtained in the survey, it is summarized as follows:

1) Most software projects developed in Mexico are Web Applications.
2) Most companies have a seniority in the market between 1 to 5 years. So, they are still very young companies.
3) The time to develop the most of the projects is between 2 to 6 months.
4) 10 members form development teams at most.
5) The companies do not apply management models and quality assurance in most cases.
6) The programming paradigm most widely used is object-oriented.
7) The requirements acquisition technique most used is the brainstorm.
8) In most companies do not exist certified people in some methodology.
9) The most commonly used tests are the black box.
10) The most widely used diagrams are the UML modeling and Entity-Relationship diagrams.
11) The companies usually do not follow a project plan during development.

With this information, it was designed the methodology. Thus, the characteristics of software projects for which is ideal META:

1. Web Application development projects.
2. Projects that can be developed between 2 to 6 months.
3. Development teams must to be comprised of up to 10 members (excluding the users and the client).

Also, and to use META is enough to know the next three elements:

1. The roles that should be exist.
2. The methodological principles of META.
3. The development process of META.

Next, these three elements will be described in some detail.
The development team must be conformed by seven roles whose features are described in **Table 2.**

Table 2: META's Roles

Name	Description
Customer	It specifies the requirements and pays the software project.
Project Leader	It is responsible for finding new projects, gather the necessary information to establish requirements and negotiate projects, and it is the member of the development team that has more to do with the customer/users, so it must have facility to communicate with them. It is also the intermediary between the client and the project manager, so it should know how to build the software. It must have the ability to connect ideas, people and resources. And have the facility for making decisions.
Project Manager	It coordinates the programmers, the tester and documenter. It also organizes the meetings needed to analyze the requirements, the design, the project plan and testing. It must accompany the project leader in the meetings with the client.
Programmers	They are responsible for coding the design.
Tester	It is responsible for testing at any time. This consists on verify that activities are properly carried out at every stage of software development process. It must also comply with the task of ensuring quality by relying on the 14 Deming Quality Principles [10].
Documenter	Its main function is to generate documents that can support and substantiate what is generated during the software process.
End users	They are the people who interact with the software once it is released for productive use.

In general, the communication flow among development team members can be resuming as shown in **Figure 1.**

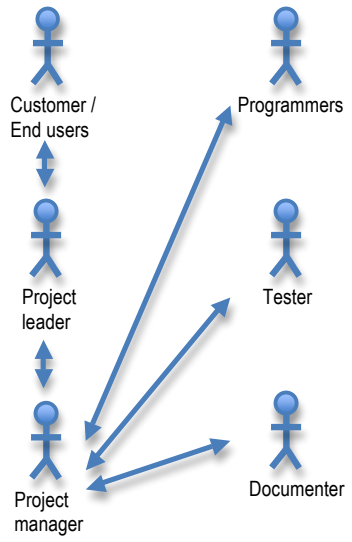


Figure 1: communication flow between development team members.

To properly use META it should be known the following methodological principles:

- Make use of the newsprint. It is a bond paper of measures 63.5cm x 77.47cm approx. It is used as a communication tool for the job applied for "face to face" because it facilitates the interaction and discussion. It is convenient to use this type of paper because it has a low cost, easy to transport, encourages participation in a group and the contents can be resumed due of the permanence of the message.
- Brainstorming [18] during meetings.
- Make use of the Deming Cycle [19] to Risk Management throughout the development process of META.
- Consider the 14 Deming Quality Principles [19] throughout the development process META, to ensure product quality in the software.

These principles shall guide the activities within the process of development of META.

META has 4 phases or stages in its process:

1. Approach.

At this stage the first activity consists in establishing the requirements that the software must meet, through a series of visits to the client, in which must always be present the client, the project leader and project manager.

Despite the presence of these three members of the team, the leader and the client are who really engage the communication, because the administrator should not intervene unless that it was necessary.

During the meetings, the participants need to do a brainstorming, and the agreement should be captured in newsprint. The use of newsprint is convenient because it allows everyone involved seeing the ideas embodied in the paper.

At this stage it should also be applied the Use Cases [16] as a tool to verify that the leader and the administrator understand what the customer expected to do the software. In this stage is necessary to estimate the costs in time, resources and money, as well as make a "mini cost-benefit analysis" that evaluates the convenience of buying an existing software (if any), or building the software from scratch; in this phase is also developing the contract and the negotiations required to sign the contract.

The stage ends when the contract is signed and develops a comprehensive project plan, which includes all information obtained during this time.

2. Preparation

At this stage, it is necessary that the project manager, the project leader, the tester, documenters and programmers, meet to analyze requirements, identify the subsystems and modules, modeling through Entity-Relationship and UML Diagrams, as well as to assign tasks and responsibilities within the development team.

So, after all this was prepared, it is necessary to put this information in a visible place for all members of the development team, for faster communication among them. Whereby, is very important that all the mentioned members can be present in the meetings, because all of them could be contribute with ideas and experiences to bring a better solutions. However, if there were the case that in the meetings, the members could not attained a consensus and they were wasting a lot of time, the project leader must make a firm decision to not falling into this problem.

3. Construction

In this stage, is it performs the coding of the design that was raised during the preparation phase. For this, it could be used

pair programming [11]. In this phase is important doing 15 minutes sessions, as Scrum proposes [13].

4. Implantation

At this stage the software is installed, and it carried out final tests with users. So there exists a feedback from the end users and costumer to rest of the development team.

The META´s process is summarized in **Figure 2**.

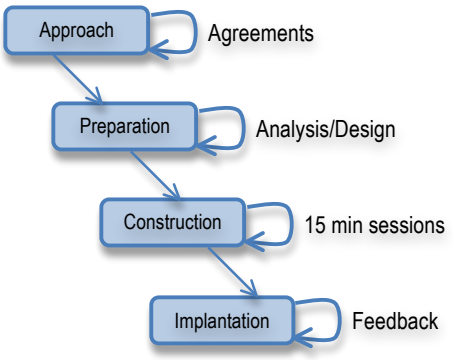


Figure 2: The META´s process.

One important aspect included in META is that it not imposes a rigorous quality standard; so it is not necessary invert a lot of time managing quality.

6. CONCLUSIONS

The methodologies to software development have provided some tools and practices that help in the process of development. They have evolved as also the programming paradigms did.

Despite the evolution that they have been suffered, there is not exist a methodology that can be applied to all types of software project, so it is necessary to investigate what are the real necessities of the scope where it wants to implement a specific methodology.

The hybrid methodologies have resolved a part of this problem, because they can combine some of the best practices including in the existing methodologies.

This is the reason that it decided to design META as a hybrid methodology. META was designed for try to solve the current problems in the software development companies in Mexico. But it is not represents that it cannot be used in other countries.

In fact, META could represent a good option for the software development companies existing in other countries, because its development proccess just have four phases, its principles could be apply without too much effort, also META does not impose a quality standard, instead, it proposes the usage of Deming System, which is the base model in the most of existing quality standards.

7. REFERENCES

[1] Dictionary of the Royal Spanish Academy (2011). ***Metodología.*** Retrieved September 3, 2011, from http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=cultura

[2] Pressman, Roger (2005). ***Software Engineering: a practitioner's approach (6th ed.).*** EUA: Mc Graw Hill.

[3] Piattini et al (2007). ***Análisis y Diseño de Aplicaciones Informáticas de Gestión. Una perspectiva de Ingeniería del Software***. Spain: Ra-Ma.

[4] Yourdon, Edward (1976). ***Techniques of Program Structure and Design***. EUA: Prentice Hall.

[5] Software Development Group of Liverpool John Moore University (1999). ***Software Systems: Planning & Design***. Retrieved September 15, 2011, from http://computing.unn.ac.uk

[6] Rumbaugh et al (1990). ***Object-Oriented Modeling and Design***. EUA: Prentice-Hall.

[7] IBM (2011). ***IBM Rational Unified Process (RUP)***. Retrieved September 18, 2011, from http://www-01.ibm.com/software/awdtools/rup/

[8] *Ministerio de Presidencia de España* (2011). ***Métrica v.3.*** Retrieved October 8, 2011, from http://administracionelectronica.gob.es/?_nfpb=true&_pageLabel=P60085901274201580632&langPae=es

[9] Sommerville, Ian (2009). ***Software Engineering***. (7th ed). EUA: Prentice Hall.

[10] Braude, Eric (2000). ***Ingeniería de Software. Una perspectiva orientada a objetos***. Mexico: Alfaomega.

[11] Beck, Kent (2011). ***Extreme programming (XP): a gentle introduction***. Retrieved Octubre 3, 2011, from http://www.extremeprogramming.org/

[12] Cockburn, Alistar (2004). ***Crystal Clear: a human-powered methodology for small teams***. EUA: Addison-Wesley Professional.

[13] Scrum group (2011). ***Scrum***. Retrieved October 2, 2011, from http://www.scrum.org/

[14] Jacobson, Ivar (2011). ***The Essential Unified Process (EssUP)***. Retrieved September 5, 2011, from http://www.ivarjacobson.com/process_improvement_technology/essential_unified_process_software/

[15] Jacobson, Ivar (2011). ***The Unified Process Lifecycle Practice***. Retrieved September 2, 2011, from http://www.ivarjacobson.com/Unified_Process_Lifecycle.aspx

[16] Object Management Group (2011). ***UML***. Retrieved September 14, 2011, from http://www.uml.org

[17] INEGI (2011). ***Estadística***. Retrieved February 14, 2011, from http://www.inegi.org.mx

[18] Piattini et al (2007). ***Análisis y Diseño de Aplicaciones Informáticas de Gestión***. *Una perspectiva de Ingeniería del Software*. Spain: Ra-Ma.

[19] Deming, Edwards (2011). ***The Deming System.*** Retrieved September 14, 2011, from http://deming.org/

Peer-review and ICT Practices of Mexican Mainstream Journals

Judith **LICEA DE ARENAS**

Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México, Ciudad Universitaria, México, D.F. 04510, MEXICO
jlicea@unam.mx

Sergio **RANGEL**

Dirección General de Bibliotecas, Universidad Nacional Autónoma de México, Ciudad Universitaria, México, D.F. 04510, MEXICO
smarquez@dgb.unam..mx

ABSTRACT

An analysis of Mexican mainstream journals is carried out in order to identify to what extent peer-review is an accepted practice in the paper selection process as well as the use of ICT in such journals.

Keywords: Peer-review; Mexico; Mainstream journals; ICT

1. INTRODUCTION

The peer-review system in science involves the systematic use of reviewers, referees, peers or judges to validate the acceptance of protocols or manuscripts submitted for publication [1]. The reviewer is therefore an example of the judge in charge of evaluating the quality in a social system. The presence of experts validating research results, therefore, seems essential.

The review of manuscripts before publication is not recent. It has its roots in the 17th century when the Philosophical Transactions were authorized by the Royal Society on the following terms: "It is ordered that Philosophical Transactions made by Mr. Oldenburg, be printed on the first Monday of each month, if there is enough material for it, and that the publication is authorized by the Board of the Society, the first being reviewed by any member of it. . . ". In Mexico, a member of the editorial board of the Anales de la Asociación Larrey, Manuel S. Soriano, noted in 1875 that "the papers would be chosen …" [2-6].

However, it is until recently that several Mexican journals are recognized as quality journals because their editors use a critical refereeing system [7], i.e. peer-review becomes a requirement to receive official funding.

According to the above, we attempted to determine whether peer-review is an accepted practice in the more visible Mexican journals for the social sciences and sciences, as well as identifying the use of ICT in the process of reviewing articles.

2. METHODS

The Journal Citation Reports (JCR) of ISI Web of Knowledge from Thomson Reuters' Social Science (SSE) and Sciences (SE) editions will be used to identify those Mexican journals that are part of the mainstream and that are in the Index of Mexican Journals of Scientific Research and Technology of the National Council of Science and Technology of Mexico. Also, the journals online portals will be used to determine the use of ICT in the editorial process.

3. RESULTS AND DISCUSION

JCR Social Sciences Edition includes 13 Mexican journal titles and Sciences Edition JCR lists 26 titles. The inclusion of Mexican journals in the Thomson Reuters core lists of journals indexed is determined by several factors, peer-review is among them: "Application of peer-review process is another indication of journal standards and signifies overall quality of the research presented and the completeness of cited references" [8].

Of the 13 titles in JCR Social Sciences Edition, one is not part of the Index of Mexican Journals of Scientific Research and Technology, of the 26 titles in JCR Science Edition, four are not included in the Index of Mexican Journals.

Journals in the Index, financed by the National Council of Science and Technology of Mexico have supposedly immediate social impact [9], and must be subjected to a strict peer-review process.

The mainstream journals are published by academic institutions (Table 1) located in the country´s capital (Table 2). Accepted papers are published mainly in Spanish language although seven journals are published in English cover to cover (Table 3).

Table 1. Typology of publishers of Mexican mainstream journals in JCR Sciences and Social Sciences Editions

Typology	Social Sciences Edition	Sciences Edition
Academic	5	12
Learned societies	1	7
Private health	-	1
Public health	2	1
Public sector	1	1
Research	4	4

Table 2. Geographic distribution of Mexican mainstream journals

	Social Sciences Edition	Sciences Edition
Country´s capital	11	21
Country´s states	2	5

Table 3. Language of publication of mainstream journals

Language	SSE	SE
Only Spanish	5	2
Only English	-	7
Spanish/English	5	15
Spanish/English/French/ Portuguese	3	2

Almost half of the journals in JCR Social Sciences Edition indicate explicitly in the Instructions to Authors, that the manuscripts will be reviewed by two experts and, in some cases that number increases to three. The remaining six titles only state that the articles submitted for publication will be submitted to review (Table 4). Three titles specify that the review will be double-blind.

Table 4 . Number of reviewers of Mexican mainstream journals in the JCR Social Sciences and Sciences Editions

Edition	1	2	3	No indication
SSE	0	7	0	6
SE	0	4	2	20

For titles in JCR Science Edition, except three, it is mentioned that the items will be reviewed prior to acceptance: one journal indicates that the authors may suggest reviewers; one has an open system of review. Three of the journals publish the reviewers' guide or instructions for referees.

It is agreed that peer review is a good way to assess research papers [10]. However, is it necessary to mention how many experts will review a paper or the type of review the manuscripts should by subjected? If most mainstream journals specify that submitted papers will be evaluated without indication of the number of reviewers why Mexican journals are very punctilious about the number?

With regards to the use of ICT, most of the journals in JCR Science Edition and Social Sciences Edition use e-mail for submitting manuscripts, but e-mail still coexists with traditional printouts sent by mail. Of the thirty nine JCR journals, only seven accept manuscripts online (Table 5). Twenty seven titles indicate that the files must be in Word, TXT or PDF (Table 6). Eleven titles indicate that images must be submitted in JPG, TIFF or EPS (Table 7). The availability of an editorial manager is not mentioned in any of the 39 mainstream journals as well as the use of LaTeX which is helpful for the publication of papers (LaTeX. Available: http:www.latex-project.org/) or recommendations about the use of software such as iThenticate (iThenticate. Available: http://printfu.org/ithenticate) or Turnitin (Turnitin. Available: https://turnitin.com/static/index.php) to ensure that authors have cited their sources properly and not plagiarized.

Table 5. Via of submission of papers to mainstream journals

Via	Social Sciences Edition	Sciences Edition
e-mail	5	9
e-mail/printout	4	8
Online	1	2
Printout	-	1
e-mail/online	-	1
e-mail/CD	-	2
online/e-mail/CD	-	2
printout/CD	-	1
fax/e-mail	1	-
e-mail/printout/CD	1	-
online/printout	1	-

Table 6. Document format of papers submitted for publication in mainstream journals

Format	Social Sciences Edition	Sciences Edition
Word	8	10
Word/RTF	1	1
Word/Open Office	1	-
Word/PDF	1	2
PDF	-	1
Word/CD	-	1
Postscript/PDF	-	1
Word/TXT	-	1
No indication	2	9

Table 7. Format of images of papers submitted to mainstream journals

Formal	SSE	SE
JPG/TIFF/EPS	2	-
JPG	1	1
EPS/WMF/CDR/AI		1
TIFF/printout		1
TIFF/JPG		2
CD/printout		1
TIFF		2
PDF		1
BMP/JPG		1

CONCLUSION

The results indicate rarely used practices in mainstream journals published abroad such as the indication of the type of review and the number of reviewers. Also, there is scarce mention of how ICT is applied.

REFERENCES

[1] H. Zuckerman, R.K. Merton, "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System", **Minerva**, Vol. 9, 1971, pp. 66-100.

[2] A.C. Miller, S.L. Serzan, "Criteria for Identifying a Refereed Journal", **Journal of Higher Education**, Vol. 55, 1984, pp. 673-699.

[3] M.S. Soriano, "[Presentación]", **Anales de la Asociación Larrey**, Vol. 1, 1875, p. 1.

[4] C. R. Weld, **History of the Royal Society**. London, 1848. v. 1, p. 177.

[5] J. M. Ziman, **Public Knowledge : the Social Dimension of Science**. Cambridge : Cambridge University Press, 1966.

[6] S. Zsindley, A. Schubert, T. Braun, "Editorial Gatekeeping Patterns in International Science Journals", **Scientometrics** Vol. 4, 1982, pp. 57-68.

[7] R. Rousseau, "Journal Evaluation: Technical and Practical Issues", **Library Trends**, Vol. 50(3), 2002, pp. 418-439.

[8] J. Testa, "The Journal Selection Process". Available: http:thomsonreuters.com/products_services/science/free/essays/journal_selection_process/.

[9] E. Loría "Un Debate sobre el Sistema de Evaluación de las Revistas Académicas Mexicanas", **Interciencia**, Vol. 25, 2000, pp. 165-169.

[10] D.Butler, "Expert Question Rankings of Journals: F1000 Scoring System Could Throw off Results, Say Critics", **Nature**, Vol. 478, 2001, p. 1.

Approximate Nearest Neighbor Search Small World Approach

Alexander Ponomarenko, Yury Mal'kov, Andrey Logvinov, Vladimir Krylov
MERA Labs LLC, Nizhny Novgorod, Russia
aponom@meralabs.com, ymalkov@meralabs.com, alogvinov@meralabs.com, vkrylov@meralabs.com

ABSTRACT

In this paper we propose a novel approach to solving the nearest neighbor search problem. We propose to build a data structure where the greedy search algorithm can be applied which is known to have logarithmic complexity in structures with navigable small world properties. The distinctive feature of our approach is that we build a non-hierarchical structure with possibility of local minimums which are circumvented by performing a series of searches starting from arbitrary elements of the structure. The performed simulation shows that the structure built using the proposed algorithms has navigable small world properties with logarithmic search complexity which is retained even for high-dimensional data.

Keywords: Similarity Search, Small World, Distributed Data Structure

1. INTRODUCTION

We present a new approach for solving nearest neighbor search problem in general metric space. This problem appears when we need to find a closest object $p \in X$ from finite set of objects $X \subseteq \mathcal{D}$ to given query $q \in \mathcal{D}$, where $\mathcal{D}$ is the set of all possible objects (data domain). Closeness or proximity of two objects $o', o'' \in \mathcal{D}$ is defined as distance function $d(o', o'')$.

In general, the search problem can be described as follows: Let $\mathcal{D}$ be a domain, d a distance measure on $\mathcal{D}$, and $(\mathcal{D}, d)$a metric space. Given a set $X \subseteq \mathcal{D}$ of n elements, preprocess or structure the data, so that proximity queries are answered efficiently.

The nearest neighbor search problem is relevant to many applications such as pattern recognition and classification [**1**], content-based image retrieval [**2**], machine learning [**3**], Recommendation systems [**4**], searching similar DNA sequence [**5**], semantic document retrieval [**6**].

In a trivial case data structure S is a simple linear list. The complexity of addition operation is $O(1)$, but searching for closest object for q requires evaluation of the metric function for every element from the set of objects X. This amounts to complexity $\theta(n)$, where n is the number of objects in X.

General way to reduce amount of distance measure calculations consists of building a set of equivalence classes, discarding some classes, and exhaustively searching the rest [**7**]. Authors also showed that two main techniques based on equivalence relations, namely, pivoting and compact partitions encompass all the existing methods. Pivot technique relies on taking k pivots and mapping the metric space onto $\mathcal{R}^k$ using the L_∞ distance and they can outperform a compact partitioning index if it has enough memory. Methods based on compact partition are more efficient for spaces with high dimensionality.

However, methods from both classes generally use either data structures with tree topology (GNAT, GHT, SAT, BST, VT, MT) or, in some cases, distance matrix (ALAESA, LAESA)

We suggest using for solving nearest neighbor problem a data structure with small word network topology presented by graph $G(V, E)$, where every object o_i from X is uniquely associated with vertex v_i from V. Thereby searching for the closest element to query q from the data set X will take the form of searching for a vertex in the graph $G(V, E)$.

Application of that approach is based on follows:

- There exist algorithms for building small world networks that have the ability to perform nearest neighbor and addition of a new object to the structure with complexity of $\log n$ [**8**].
- Small world networks have no root element.
- All operations (addition and search) use only local information and can be initiated from any element that has been added to the structure.

This gives opportunity for building decentralized similarity search oriented storage systems where physical data location doesn't depend on the content because every data object can be placed on the arbitrary physical machine and can be connected with other by links like in p2p systems. Such storage systems can provide simultaneous access to large numbers of users for performing data search and addition), have good fault tolerance and have unlimited scalability in terms of performance and capacity.

One of the basic vertex search algorithms in graphs is greedy search. This algorithm has simple implementation on the structure that has small world network topology and can be initiated from every vertex.

In order for the result of the algorithm to be the exact nearest element to the query, the network should contain the Delaunay graph as its subgraph, which is dual to the Voronoi tessellation [**9**].

However, the requirement of search in for the exact nearest neighbor can be excessive (optional) for the applications

described above. So the problem for finding the exact nearest neighbor can be substituted for the approximate nearest neighbor search, since we don't need to support whole/exact Delaunay graph.

For the search algorithm to be logarithmically scalable, the small world network should have navigation property that was already discussed in [**8**]

In this paper we present the algorithm for data structure construction based on small world network topology with graph $G(V, E)$ which uses greedy search algorithm for finding approximate nearest neighbor. Graph $G(V, E)$ contains approximate Delaunay graph and has navigation property. Search algorithm has the ability to change accuracy of search without modification of the structure. Presented algorithms do not use the coordinate representation and do not presume the properties of linear spaces, because they are based only on the metric computation between objects, and therefore is applicable to data from general metric spaces.

2. RELATED WORKS

Kd-tree [**10**] and quadra trees [**11**] were among the first structures for solving exact nearest neighbor search problem. They perform well in 2-3 dimensions (search complexity is close to $O(\log n)$), but analysis of the worst case for that structures [**12**] indicates $O(d * N^{1-1/d})$ search complexity, where d is dimensionality.

Other structures which have tree topology such as variants of kd-trees, R-trees and structures based on space-filling curves are surveyed in [**13**]. They also have good performance when searching in a low-dimension ($d < 4$) metric space, but they quickly lose their effectiveness with increasing number of dimensions [**14**]. A more effective data structure for exact nearest neighbor search in $\Re^d$ with search complexity $O(2^d \log n)$ has been described in [**15**]. But as can be seen, search complexity has exponential dependence from the number of dimensions.

Structures such as mvp-tree [**16**], vp^s-tree and vp^{sb}-tree [**17**] use "vantage point" technique, but no analysis has been provided for search complexity in spaces with high number of dimensions.

In general, presently there are no methods for effective exact nearest neighbor search in high-dimensionality metric space. The reason behind it lies in the "curse" of dimensionality [**7**].

To avoid the curse of dimensionality while retaining the logarithmic scaling of the number of elements, it was proposed to reduce the requirements for finding the nearest neighbor, making it approximate.

Thus a large number of papers appeared which proposed to search for nearest neighbor with ε accuracy (ε-NNS). For example, Arya and Mount proposed methods with search complexity $O(\log^3 n)$, but preprocessing requires $O(n^2)$ and algorithm was applicable only to data from E^d [**18**].

Kleinberg proposed two methods [**19**] for solving ε-NNS. First method requires $O(\mathrm{n} \log d)^{2d}$ preprocessing time and query time polynomial in d, ε and $\log n$. Another method with preprocessing polynomial in d, ε and n, but with query time $O(n + d\log^3 n)$. Also both methods are applicable only to data from E^d

The first algorithms with search complexity polynomial in d, $\log n$, ε^{-1} and polynomial preprocessing time for fixed ε were proposed by Indyk and Motwani in [**20**] and Kushilevitz, Ostrovsky and Rabani in [**21**]. Indyk and Motwani were the first ones to relax ε-ANN problem to approximate point location in equal balls (ε-PLEB). For the formulation of the problem in ε-PLEB points in metric space expand to the balls with center at this point and radius $(1+\varepsilon)r$, it is necessary to determine which ball belongs to the query q. Also in [**20**] proposed a second method, which uses the concept of locality-sensitive hashing regarding formulation of the problem ε-PLEB, with search time $O(n^{1/(1+\varepsilon)})$, however requires near quadratic memory (for small ε). In addition, the first method is applicable only for E^d, and the second for the Hamming space.

In general, the concept of locality-sensitive hashing has become popular in the last decade to solve the ANN problem. Other works using the concept of locality-sensitive hashing are [**22**], [**23**]. But they all have the same major drawback: each algorithm is focused on a narrow class of metrics such as Hamming distance, Jakarta or l_s norms for Euclidean space. Thus it is necessary to create digests in order to decide which method to choose.

The first structure for solving ANN in E^d with topology of small world networks is Raynet [**24**]. It is an extension of earlier work by the same authors Voronet [**25**], which solved the problem of the exact NN in E^2. Originally Voronet was envisioned as a p2p network, where every node has coordinates in E^2. In Raynet every node has the coordinates in E^d. The system supports two levels of links - short for correct work of the greedy search algorithm and long - for logarithmic search. Short links correspond to edges of Delaunay graph, i.e. each object has references to objects that are neighbors of its Voronoi region. The main difference of Raynet from Voronet is that in Raynet every object doesn't know all of its Voronoi neighbors, i.e. Raynet obtains neighborhood with approximately using the Monte Carlo method.

Raynet is the closest work to ours in terms of general concept. But unlike Raynet, we propose a structure that works with objects from arbitrary metric spaces.

3. STRUCTURE OVERVIEW

We solve the problem of approximate nearest neighbor search formulated as follows: given objects from domain $\mathcal{D}$ with

distance function $d: \mathcal{D} \times \mathcal{D} \longrightarrow \mathbb{R}$. For finite set $X = \{x_1, \ldots, x_n\}$, $X \subset \mathcal{D}$ an effective probability search method is required to find $x_i \in X$ which is closest to $q \in \mathcal{D}$. Effective method means that search complexity must scale logarithmically with the number of elements in X. The exact search is not guaranteed, i.e. the result of the algorithm may be an element that is not true nearest neighbor, nevertheless structure and algorithms are designed to minimize the probability of this and there is a possibility to adjust it by varying the parameter of the search algorithm without changing the structure**.**

The structure of S is constructed as a small world network described by a graph $G(V, E)$, where the objects from the set X are uniquely mapped to the vertices from the set V. The set of edges E is determined by the structure construction algorithm, so as to ensure correct operation of the greedy search algorithm.

Since in the proposed structure each vertex is uniquely mapped to an element from the set X, we will use the terms "vertex", "element" or "object" interchangeably. We will use the term “friends” for vertices that share an edge. List of vertices that share a common edge with the vertex v_i is called the friend list of vertex v_i.

4. SEARCH ALGORITHM

Greedy Search

The basic search algorithm traverses the edges of the graph $G(V, E)$ from one vertex to another. The algorithm takes two parameters: `query` and the vertex $V_{enter\_point} \in V[G]$ which is the starting point of search (the entry point). Starting from the entry point at each vertex the algorithm computes the metric value from query q to each vertex from the friend list of the current vertex and then selects the vertex with minimal value of the metric. If the metric value between the query and the selected vertex is smaller than between the query and the current element, then the algorithm moves to that vertex. After that the algorithm repeats. The algorithm stops at the vertex whose friend list doesn’t contain a vertex that is closer to the query than the vertex itself. That vertex is a local minimum.

```
Greedy_Search(q: object, v_enter_point: object)
1  v_curr ← v_enter_point;
2  d_min ← d(q, v_curr); v_next ← NIL;
3  foreach v_friend ∈ v_curr.getFriends() do
4     if d(query, v_friend) < d_min then
5        d_min ← d(q, v_friend);
6        v_next ← v_friend;
7  if v_next = Nil then return v_curr;
8  else return Greedy_Search(q, v_next);
```

The element which is a local minimum with respect to query q, can be either the true closest element to the query `q` from the entire set of elements of X, or a false closest..

If every element in the structure had in their friend list all of its Voronoi neighbors, then this would exclude the existence of false local minimums. Maintaining this condition is equivalent to constructing Delaunay graph, which is dual to the Voronoi diagram.

Because it is impossible to determine exact Delaunay graph [**26**] (excluding the variant of the complete graph) we cannot avoid the existence of local minimums.

But for the problem of approximate searching as defined above it is not an obstacle since approximate search does not require the entire Delaunay graph [**24**]. As shown below, the probability of finding the true nearest element tends exponentially towards 1 with increase of the average number of edges in the approximated Delaunay graph.

Multi Search

In order to be able to find the true closest element in a network with local minimums, we propose the following modification of the search algorithm. We propose to use a series of m searches initiated from random vertices and choose the result element that is closest to the query from the set of found elements. Since the greedy search `Greedy_Search(q, v_enterPoint ∈ V)` is deterministic for each entry point $v_{enterPoint} \in V$ it either results in a success - finding the true nearest neigbor, or with a failure - finding the element that is not the nearest neighbor of q.

Thus search of the closest element to the same query q may result in finding of the true nearest neighbor or a false nearest neigbor depending on the entry point from which the search algorithm started.

Since we can choose the entry point at random, there is a probability p of finding the true closest to the particular element q (but not to all elements). Moreover, this probability is always nonzero, because it is always possible to choose the exact nearest neighbor as the entry point, which subsequently will be returned by the greedy search algorithm.

If probability to find true closest in one search attempt is p then probability to find the same element in m search attempts is $1 - (1 - p)^m$, so failure probability decreases exponentially with the number of search attempts. Thus, we can improve search precision, increasing the parameter m - number independent searches.

```
Multi_Search(object q, integer: m)
1  results: SET[objects];
2  for (i ⟵ 0; i < m; i++) do
3     enter_point ⟵ getRandomEnterPoint();
4     local_min ⟵ Greedy_Search(query,
enter_point)
5     if local_min ∉ results then
6        results.add(result);
7  return results;
```

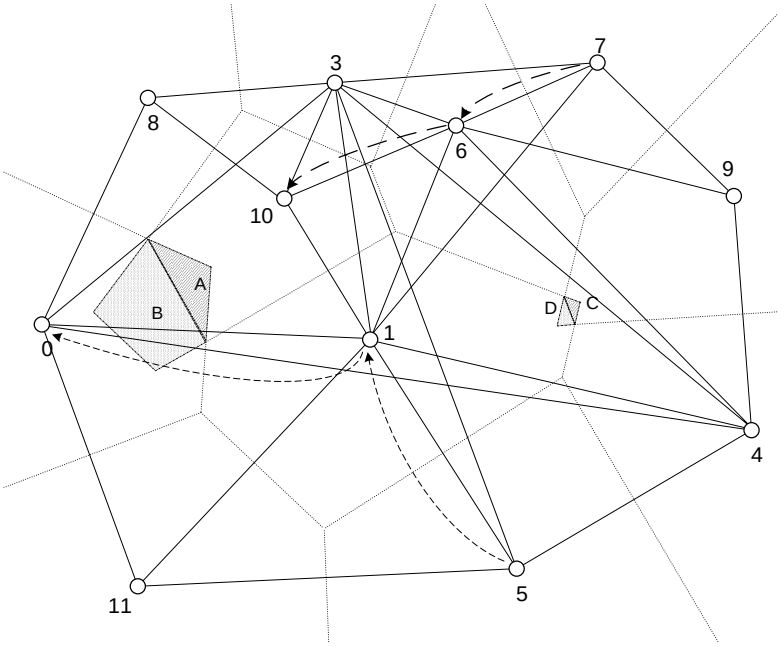


Fig 1

If $m = n$, where n is the number of elements in the structure, the algorithm becomes exhaustive search.

If the graph of the network has small-world properties, then it is possible tp choose a random vertex in a number of random steps proportional to $\log n$, which doesn't affect overall logarithmic search complexity.

Therefore the overall complexity of the search will increase no more than m times.

5. DATA ADDITION ALGORITHM

Since we build an approximation of the Delaunay graph, there is much freedom in the choice of construction algorithm. For example in [**24**] it is proposed to build approximate Delaunay graph which minimizes the volume of Voronoi region for a fixed number of edges for each vertex in the graph. In [**27**] it is proposed to connect new element with k closest objects which are already in the structure. It is based on the idea that intersection of the set of elements which are Voronoi neighbors and the k closest elements is large. In [**27**], [**28**] authors also have shown theoretically and confirmed by experimental results that graph which constructed by proposed algorithm has properties of small world network if elements arrive in random order.

We propose a modified variant of this algorithm which is distinguished by the fact that that the search for k nearest elements uses a series of searches.

The algorithm takes three parameters: the object to be added to the structure and two positive integers `k` and `init_attempts`. First, the algorithm determines a set of local minima, using the procedure `Multi_Search`, which produces a series of independent searches on init_attempts of randomly selected elements from the set of objects that already have been added to the structure. After that algorithm determines neighborhood `u`, which contains all neighbors of each found local minimums. Set `u` is sorted in ascending order by distance from the object `new_object` to be added. After that `new_object` is connected with the first `K` nearest elements from the set of `u`.

```
Nearest_Neighbor_Add(object: new_object,
integer: k, integer: init_attempts)
1  SET[object]: localMins ⟵
MultiAttempts_Search(new_object,
init_attempts);
2  SET[object]: u ⟵ ∅; //neighborhood;
3  foreach object: local_min ∈ localMins
do
4    u ⟵ u ∪ local_min.getFriends();
6  sort the set u so to satisfy the
condition d(u[i], new_object) < d(u[i+1],
new_object)
7  for (i ⟵ 0; i < k; i++) do
8    u[i].connect(new_object);
9    new_object.connect(u[i]);
```

Fig 1 shows the structure which is constructed by Nearest_Neighbor_Add algorithm for points from E2. Circles denote the elements. The numbers near thems correspond to the addition order. Solid lines show the links (edges) between elements. Dotted lines correspond to the borders of Voronoi tessellation. Delaunay graph edges between elements 0 and 10, 1 and 9 are missing. The structure obtained by the algorithm with parameters m = 3 and attemptsNumber = 5. Element with number 0 is a local minimum, which is not the closest to queries that fall into the

shaded area "A", respectively 10 for the "B", 9 for D and 1 for the region"C". Hatched lines show the paths of the two search algorithm runs for query q in the region "B". The algorithm run which starts from vertex 7 stops on the element 10 which is local minimum, but not the closest to the query q. However, the algorithm run which starts from vertex 5 finds the true closest vertex to the query q.

6. EXPERIMENT RESULTS

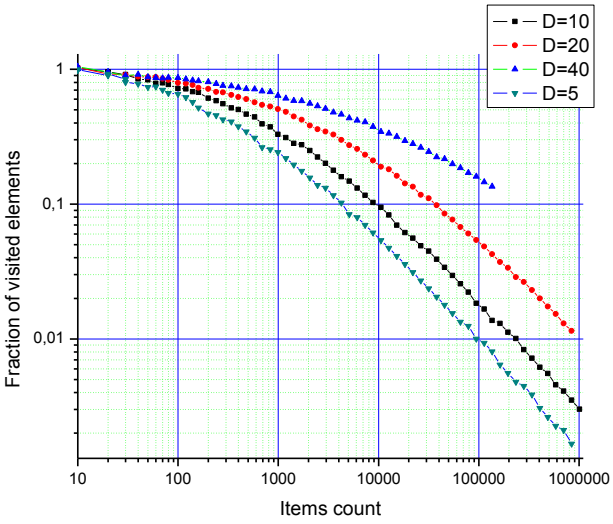


Fig 2

We have implemented the algorithms presented above in order to validate our assumptions about the logarithmic search complexity dependence of the total number of elements.

We used randomly selected points from the E^D as test dataset. L_2 (Euclidean distance) was selected as proximity function

n elements were added to the structure. We chose the number of search attempts `m` so that the probability of finding the true closest element to the query was not less than 95%. The number of metric calculations was measured. The graph shows the percent of scanned elements (vertical) with an increase in the number of added elements in the structure (horizontal).

The graph (**Fig 2**) shows that with the increase of number of elements in the structure, the percentage of visited elements decreases, and the curve becomes a straight line with angle 45 degrees. This gives us grounds to speak of logarithmic complexity of the search on the number of scanned elements.

The graph shows that the curve for higher dimensions behaves similarly. From this we can make the assumption that there is no exponential dependence from the dimension of space. But it requires more careful study.

7. CONCLUSION

We have proposed a method of organizing data into a small world topology data structure suited for approximate nearest neighbor search in metric space.

We have created a modified *k* nearest neighbor connection algorithm which is one of the possible algorithms for construction of small world data structures with navigation properties.

Simulation results confirm logarithmic dependency of search complexity from the number of elements in the structure.

All proposed algorithms use only local information on each step and can be initiated from any vertex.

All elements in the structure are of the same type, there is no central or root element.

Thus, all mentioned structure properties are a basis for using the structure for building totally decentralized data storage systems.

8. REFERENCES

[1] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21-27, Jan. 1967.

[2] M. Flickner, et al., "Query by image and video content: the QBIC system ," *Computer*, vol. 28, no. 9, pp. 23-32, Sep. 1995 .

[3] Salzberg, S. Cost, and Steven, "A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features," *Machine Learning*, vol. 10, no. 1, pp. 57-78, 1993.

[4] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, New York, USA, 2001, pp. 285-295.

[5] Rhoads, W. Rychlik, and R. Unknown, "A computer program for choosing optimal oligonudeotides for filter hybridization, sequencing and in vitro amplification of DNA," *Nucletic Acids Research*, vol. 17, no. 21, pp. 8543-8551, Oct. 1989.

[6] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," *J.*

Amer. Soc. Inform. Sci., vol. 41, pp. 391-407, 1990.

[7] E. Chávez, G. Navarro, R. Baeza-Yates, and J. L. Marroquín, "Searching in metric space," *Journal ACM Computing Surveys (CSUR)*, vol. 33, no. 3, pp. 273-321, Sep. 2001.

[8] J. Kleinberg, "The Small-World Phenomenon: An Algorithmic Perspective," *ANNUAL ACM SYMPOSIUM ON THEORY OF COMPUTING*, vol. 32, pp. 163-170, 2000.

[9] F. Aurenhammer, "Voronoi diagrams — a survey of a fundamental geometric data structure," *ACM Computing Surveys (CSUR)*, vol. 23, no. 3, pp. 345-405, Sep. 1991.

[10] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18 , no. 9, pp. 509-517, Sep. 1975.

[11] Bentley and R. Finkel, "Quad Trees: A Data Structure for Retrieval on Composite Keys," *Acta Informatica*, vol. 4, no. 1, pp. 1-9, 1974.

[12] Wong and Lee, "Worst-case analysis for region and partial region searches in multidimensional binary search trees and balanced quad trees," *Acta Informatica*, vol. 9, no. 1, pp. 23-29, 1977.

[13] H. Samet, *The design and analysis of spatial data structures*. Addison-Wesley, Reading, MA, 1989.

[14] Mount, Arya, and Narayan, "Accounting for boundary effects in nearest-neighbor searching," *Discrete & Computational Geometry*, vol. 16, no. 2, pp. 155-176, 1996.

[15] D. Dobkin and R. Lipton, "Multidimensional Searching Problems," *SIAM J. Comput*, vol. 5, no. 2, pp. 181-186, 1976.

[16] T. Bozkaya and M. Ozsoyoglu, "Distance-based indexing for high-dimensional metric spaces," in *Proceedings of the 1997 ACM SIGMOD international conference on Management of data*, New York, USA, 1997, pp. 357-368.

[17] P. Yianilos, "Data structures and algorithms for nearest neighbor search in general metric spaces," in *SODA '93 Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms*, Philadelphia, USA, 1993, pp. 311-321.

[18] S. Arya and D. Mount, "Approximate nearest neighbor queries in fixed dimensions," in *SODA '93 Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms*, Philadelphia, PA, USA, 1993, pp. 271-280.

[19] J. Kleinberg, "Two algorithms for nearest-neighbor search in high dimensions," in *STOC '97 Proceedings of the twenty-ninth annual ACM symposium on Theory of computing* , New York, USA, 1997, pp. 599-608.

[20] Motwani, P. Indyk, and Rajeev, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *STOC '98 Proceedings of the thirtieth annual ACM symposium on Theory of computing*, New York, USA, 1998, pp. 604-613.

[21] E. Kushilevitz, R. Ostrovsky, and Y. Rabani, "Efficient search for approximate nearest neighbor in high dimensional spaces," in *STOC '98 Proceedings of the thirtieth annual ACM symposium on Theory of computing*, New York, NY, USA, 1998, pp. 614-623.

[22] A. Gionis, P. Indyk, and R. Motwani, "Similarity Search in High Dimensions via Hashing," in *VLDB '99 Proceedings of the 25th International Conference on Very Large Data Bases*, San Francisco, USA, 1999, pp. 518-529.

[23] A. Andoni and P. Indyk, "Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions," in *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, Berkeley, California , 2006, pp. 459-468.

[24] O. Beaumont, A.-M. Kermarrec, and É. Rivière, "Peer to peer multidimensional overlays: approximating complex structures," in *Proceedings of the 11th international conference on Principles of distributed systems*, Berlin, Heidelberg, 2007, pp. 315-328.

[25] O. Beaumont, A.-M. Kermarrec, L. Marchal, and E. Riviere, "VoroNet: A scalable object network based on Voronoi tessellations," in *International Parallel and Distributed Processing Symposium*, Long Beach, USA , 2007, p. 20.

[26] G. Navarro, "Searching in metric spaces by spatial approximation," in *String Processing and Information Retrieval Symposium*, Cancun, Mexico, 1999, pp. 141-148.

[27] Krylov, Logvinov, Ponomarenko, and Ponomarev, "Metrized Small World Properties Data Structure," in *SEDE*, Los Angeles, California USA, 2008.

[28] V. Krylov, A. Ponomarenko, A. Logvinov, and D. Ponomarev, "Single-attribute Distributed Metrized Small World Data Structure‖," in *IEEE International Conference on Intelligent Computing and Intelligent Systems (CAS)*, 2009.

A Social Network Based CBR System for Quality Group Decisions

Wei-Lun Chang,
Department of Business Administration, Tamkang University,
Taipei, Taiwan, R.O.C.,
wlchang@mail.tku.edu.tw

ABSTRACT

Decision-making is a cognitive mental process which selects certain actions among several alternatives for a specific problem. Human beings sometimes make intuitive decisions without advanced analysis and judgment. However, some problems need sufficient information to generate superior solutions. Group decision-making eases personal biases, enabling participants to discuss, argue and coordinate a coalition of ideas. This research provides a social network based group decision model which accounts for collective intelligence. This work considers peers from a user's social network as experts in the group decision-making process and each peer has its own Internet agent. The purpose is to conduct a virtual group decision-making process over the Internet at anytime. Each agent can reason by enabling the CBR (case-based reasoning) approach and discuss by the proposed decision model. The proposed agent-based CRR approach for group decisions attempts to improve efficiency and effectiveness for a quality decision, which is extremely critical and crucial. The proposed approach (1) empowers collective intelligence from a social network, (2) enhances trustworthy group decisions of the social network (i.e., each level of social network has selected experts), (3) ensures heterogeneity of selected experts, and (4) diminishes the domination of certain experts.

Keywords: Social network, Case-Base Reasoning, Delphi method, Collective intelligence, Group decision

1. INTRODUCTION

Decision-making is a cognitive mental process which selects certain actions among several alternatives for a specific problem. The output of a decision making process should include actions or options. Human beings sometimes make intuitive decisions without advanced analysis and judgment. However, some problems need sufficient information to generate superior solutions. Theoretically, a person's decision-making style may cause cognitive or personal biases (e.g., information overload or selective perception). Group decision-making eases personal biases, enabling participants to discuss, argue and coordinate a coalition of ideas (Rohrbaugh, 1979). Laughlin et al. (2002) indicate that groups may be effectively superior to individuals in processing information.

Certain group decision-making methods are currently utilized (Brockhoff, 1983); for example, the Delphi method, the brainstorming method, and the nominal group technique. Jelassiand and Foroughi (1989) indicate the significance of negotiation support systems (NSS) and discuss negotiation-structuring issues that group decision should account for when designing NSS. Particularly, the Delphi method is a systematic interactive forecasting method for obtaining forecasts from a panel of independent experts. The Delphi method was developed, over a period of years, at the Rand Corporation at the beginning of the cold war to forecast technology impact on warfare (Helmer and Rescher, 1959). The technique is a method of obtaining an intuitive consensus of group expert opinions (Wedley et al., 1979; Morgan et al., 1979).

This research provides a social network based group decision model which accounts for collective intelligence (e.g., Wikipedia). Social network researches have been investigated for many years, but the concept of combining social network and group decision making is still lacking. This work considers peers from a user's social network as experts in the group decision-making process and each peer has its own Internet agent. The purpose is to conduct a virtual group decision-making process over the Internet at anytime. Each agent can reason by enabling the CBR (case-based reasoning) approach and discuss by the proposed decision model. Thus, this research devises a system (iGD), based on the proposed approach to verify feasibility and validity.

The proposed agent-based CRR approach for group decisions attempts to improve efficiency and effectiveness for a quality decision, which is extremely critical and crucial (Boje and Mumighan, 1982). In short, the proposed approach (1) empowers collective intelligence from a social network, (2) enhances trustworthy group decisions of the social network (i.e., each level of social network has selected experts), (3) ensures heterogeneity of selected experts, and (4) diminishes the domination of certain experts.

2. iGD SYSTEM FRAMWORK

2.1 Group Decision Making Concept

This study accounts for collective intelligence power from a social network. This work considers that people from a user's social network are valuable providing sufficient suggestions to the user. Consequently, this paper proposes a CBR-based decision concept for group decision-making based on the Delphi method. **Equation (1)** demonstrates the concept as follows. In **Eq. (1)**, *SN(Es)* are the selected experts through the user's social network. The *CBR* function enables each agent to reason through the solutions based on past experience. The *TR* function estimates the required number of solutions (ideas) as the threshold to ensure the width of solutions. The *DP* function furnishes the traditional Delphi method concept in terms of anonymity, group decision making, and systematic forecasting. Finally, certain decisions are the final outputs from this model.

$$GDs = DP(CBR(SN(Es)), TR(Es)) \quad \textbf{(1)}$$

2.1.1 Expert Selection Module

The expert selection module (ES Module) is the foundation among other modules, selecting appropriate experts from the user's social network. The current study assumes that the user has a pre-defined social network with a specific level as shown in Figure 2(a). Supposedly, the user has a social network including friends, relatives, and families. Persons in the first level are the most familiar peers for the user and considered as strong ties. Persons in the second level are familiar with the

ones from the first level and considered as weak ties. The same concept is expanded to the following levels. The given number of experts from our model generates the number of levels. The selected process should avoid a strong tie situation (Fig. 2(b)); for example, selecting user A in the first level but un-selecting user A's friends in the second level. This attains heterogeneity of selected experts and good quality decisions.

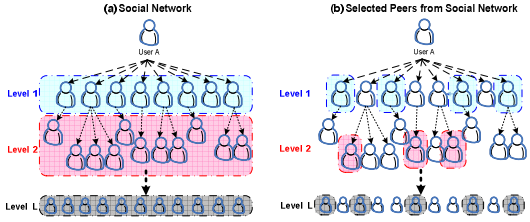


Figure 1 (a) pre-defined social network
(b) concept of the expert selection

To propose a heuristic model for expert selection, this work assumes N is the selected experts, L is the required levels of the social network, K is the actual numbers of friends for each level, n is the number of peers for each social network level, n' is the modified number of experts for each social network level, and i and p represent the current level. N is greater than three, ensuring the minimum number of experts in a group decision. That is, $N \geq 3, i \geq 2, p, i \geq 1$.

$$L = \left\lceil \frac{N}{3} \right\rceil \quad \textbf{(2)}$$

In **Eq. (2)**, we set the total number of levels as the ceiling of N divided by three. This ensures the minimum number of levels (L=1) if the number of experts is given as three (N=3). For example, if the person prefers to five experts (N=5) then the number of levels will be two (L=2). This also increases expert heterogeneity if the number of levels rises.

This research also separates two situations for selecting experts from each level: (1) the required number of experts is always equal to or lower than the actual number of friends for all levels and (2) the required number of experts is greater than the actual number of friends for an unknown level. However, the given number of experts as three (N=3) might be an exception for the following two situations. That is, ES module selects all three experts from the first level.

(a) if $n_i \leq K_i$ (assume all n_i is less than K_i)

In **Eq. (3)**, the number of selected experts for the first level is the floor of N divided by three. This ensures the minimum number of experts for the first level (n_1=1) if the number of experts is given as four (N=4). For example, if the person prefers to seven experts (N=7), then the number of experts for the first level will be one (n_1=2).

$$n_1 = \left\lfloor \frac{N}{3} \right\rfloor \quad \textbf{(3)}$$

Equation (4) represents the number of experts for each level except the first level. n_i is equal to the ceiling difference between the given number of experts and the number of experts for previous levels ($N - \sum_{z=1}^{i-1} n_z$) divided by the remaining number of levels ($L - \sum_{z=1}^{i-1} L_z$). For example, if the person prefers to seven experts (N=7), then the number of experts for the first level will be two (n_1=2) and the total number of levels is three (L=3). Conversely, the remaining number for the other levels (except the first level) is three and two (i.e., n_2=3, n_3=2). Meanwhile, **Eq. (5)** represents that the actual number of peers for each level (K_i) should be greater than the required number of experts (n_i).

$$n_i = \left\lceil \frac{N - \sum_{z=1}^{i-1} n_z}{L - \sum_{z=1}^{i-1} L_z} \right\rceil \quad \textbf{(4)}$$

$$n_i \leq K_i \quad \textbf{(5)}$$

(b) if $n_p > K_p$ (assume one level of n_p is greater than K_p)

The second condition assumes a level with an insufficient number of peers for selection. In **Eq. (6)**, K_p is the actual number of peers and n_p is the required number of experts for a specific level p. Once the required number is greater than the actual number ($n_p > K_p$), ES module needs to select a new level (p+1). The required number of experts for a new level is the difference between the required number of experts and the actual number of peers for the previous level (i.e., $n_p - K_p$).

$$n'_{p+1} = (n_p - K_p) \quad \textbf{(6)}$$

We utilize the same example in the aforementioned section. The required number of experts for the third level is two (n_3=2) but we assume the actual number is one (K_p=1). That is, one expert will be selected from level four (n_4=1) owing to the minus of n_3 and K_p. Moreover, **Eq. (6)** is the iterative process while the actual number of peers is lower than the required number of experts for any level. The iterative process does not terminate until all the required experts are selected from required levels.

2.1.2 Case-Based Reasoning Module

Researchers originally used Case-Based Reasoning (CBR) for specific or independent fields such as dynamic memory theory. Recent researches have used CBR in computer-based cognition processes. CBR applies artificial intelligence (Perner, 2008), which reasons through possible solutions from results of past cases or experience. CBR includes four steps: retrieve, reuse, revise, and retain (Aamodt and Plaza, 1994). This work assumes each expert as represented by an agent. Theoretically, each agent selects similar problems of past case(s) from the case base (retrieve) and reuses the solutions of case(s) to solve existing problems. The provided solutions are revised accordingly to form appropriate solution(s). If the process generates new problem(s) or solution(s), the agent retains it and stores it back to the case base for future utilization. CBR is constructed from past cases and has well-defined problems within attributes and solution(s).

The similarity of cases is significant as it may affect outcome quality. Changchien and Lin (2005) combine AHP and CAPM (Core Process Analysis Matrix) to apply CBR from a marketing perspective. They use the concept of minimum distance (Euclidean Distance) to estimate the similarity of two cases. This research utilizes the same concept to discover similar

cases and the equation is shown in **Eq. (7)**.

$$dist\ (Xg, Yg) = \sqrt{\sum_{g=1}^{f} \left(x_g - y_g\right)^2} \quad \textbf{(7)}$$

In **Eq. (7)**, we assume X is the new case, x_g is the set of attributes of case X, Y is the existing case, and y_g is the set of attributes of case Y. We estimate the distance by calculating the distance of every attribute represented by $dist\ (Xg, Yg)$. For example, we assume a new problem X1 with four attributes (i.e., $X1 = (x_1, x_2, x_3, x_4) = (2,2,1,1)$) and two existing cases Y1 and Y2 that both have four attributes (i.e., $Y1 = (y_1, y_2, y_3, y_4) = (2,2,2,1)$ and $Y2 = (y_1, y_2, y_3, y_4) = (2,2,3,1)$). The distances for X1/Y1 and X1 /Y2 are estimated by the concept of Euclidean distance (i.e., $dist(X1,Y1) = \sqrt{0+0+1+0} = 1$ and $dist(X1,Y2) = \sqrt{0+0+4+0} = 2$). Hence, the distance for X1/Y1 is shorter than X1/Y2, which means X1 is similar to Y1 and the Y1 solutions will be extracted to help solve the X1 problem.

2.1.3 Solution Threshold Determination Module

The solution threshold determination module (STD Module) is the component to estimate the minimum number of possible solutions. This ensures that selected experts provide sufficient numbers of decisions. The threshold signifies that the number of solutions should be large enough when the number of experts increases. This research proposes a solution threshold in the group decision model to attain better decision quality.

We assume α is the number of required solutions, and N is the given number of selected experts ($N \geq 3$). In **Eq. (8)**, if the required number of experts is greater than the average number of experts, α should be at least two-thirds greater than M. Otherwise, α should be at least greater than half of M. The rationale is that ideas should increase if the number of required experts rises.

If $N > \left\lfloor \frac{3+N}{2} \right\rfloor$ then $\alpha \geq \left\lfloor \frac{2N}{3} \right\rfloor$; otherwise $\alpha \geq \left\lfloor \frac{N}{2} \right\rfloor$ **(8)**

Using the same example, if we set the given number of experts as seven (N=7), the number of solutions should be at least greater than four ($\alpha \geq 4$). Similarly, if we replace the given number of experts by four (N=4), the number of solutions should be at least greater than two ($\alpha \geq 2$). In other words, the number of solutions (α) is subject to the given number of experts (N).

2.1.4 Decision Ranking Module

The decision-ranking module (DR Module) is based on a heuristic scoring concept. The estimated number of levels is the DR module input. The DR module also generates scores and weights for different expert levels. Meanwhile, the DR module calculates ultimate scores for each solution automatically by multiplying score by weight. Conversely, the top three ranking recommendations are the DR module outputs.

Theoretically, the Delphi method enables experts to propose and revise ideas iteratively for two rounds. In the proposed model, this work un-limits the number of rounds to gain feasible ideas for group-decision making. This work also proposes a decision weighting and ranking mechanism to ensure consensus of the selected decisions. This concept is similar to the Pretty Good Privacy (PGP) trust model, created primarily for encrypting e-mail messages using public or conventional key cryptography (Zimmermann, 1995). The traditional group decision-making approach utilizes questionnaires to attain a consensus. However, the approach needs a facilitator to gather the questionnaires and summarize the information manually (as shown in Fig. 3). The proposed model not only enhances efficiency (i.e., selects the solutions quickly) but attains effectiveness (i.e., selects the right solutions).

This research assumes W is the weight, I is the weight interval for each level, X is the total number of proposed solutions (ideas), E_y represents a specific expert y, q stands for the level of expert y, S is the solution (idea) set, $\bar{s}$ mean a specific solution, and r is solution ranking from the expert. In **Eq. (8)**, the weight interval of each level is estimated by the reversed total number of levels, plus one. This ensures that the top three solution weights are not equal to zero. The estimated weight interval sets the different decision weights of varied levels of experts in **Eq. (9)**. The decision weight supposedly decreases if the expert level increases. This distinguishes the differences among different levels of experts. The first level of experts should have higher decision weight than experts in the second level.

For example, if the total number of levels is three (L=3), the weight interval will be a quarter ($I = \frac{1}{4}$). The decision weight for first level experts is three-quarters ($W_1 = \frac{3}{4}$) and the second level is one-second ($W_2 = \frac{1}{2}$).

$$I = \left\lfloor \frac{1}{L+1} \right\rfloor \quad \textbf{(8)}$$

$$W_q = 1 - (q * I) \quad \textbf{(9)}$$

In **Eq. (10)**, the current study considers the top three solutions as the priority for from each expert's ranking list. If the solution is prioritized as first (r=1), the solution score will be 2X. This distinguishes the different scores of different solutions at various priorities. Hence, the final score of a solution ($\bar{S}r$) is estimated by multiplying the score ($\left\lfloor \frac{2X}{r} \right\rfloor$) and weight ($W_q^{E_y}$) of the solution. Finally, we sum and rank the solution scores appearing in each expert's list in **Eq. (11)**.

$$\bar{s}_r = \left\lfloor \frac{2X}{r} \right\rfloor W_q^{E_y} \quad \textbf{(10)}$$

$$Rank_{\bar{s}_r} = \sum_{r \in S} \bar{s}_r \quad \textbf{(11)}$$

3. iGD SYSTEM EVALUATION

This section provides two evaluating metrics to demonstrate iGD system performance: *Accuracy* and *Precision*. *Accuracy* is the percentage of the number of solutions the user selects to the number of recommended solutions. **Equation (12)** demonstrates the concept of *Accuracy*, where *i* denotes the ranking among recommended solutions, q_i specifies whether the user selected the recommended solutions for *i* (e.g., q_i =1 for selected and q_i =0 for unselect), and *system_R* indicates the number of top-ranked solutions (e.g., *system_R* =3 means there are three top-ranked solutions). For example, if the user selects

the recommended solutions from the top 1 and top 2 but do not select from the top 3, *Accuracy* value will be 66.7%.

$$Accuracy = \frac{\sum_{i=1}^{3} q_i}{system\_R} \quad \textbf{(12)}$$

$$\Pr ecision = \frac{\rho}{\theta} \quad \textbf{(13)}$$

Precision is the percentage of matches between the system's ranking and the user's ranking based on the number of solutions selected among all the solutions. **Equation (13)** reveals the concept of *Precision*, where θ denotes the number of selected solutions among all the solutions and ρ indicates the number of ranked solutions that both the system and the user selects. For example, if the user selects two solutions (ranks jn top-1)/1 solution (ranks in top-2) and recognizes only one solution from the top-1 ranking as the preferred top-1 ranking, the value of *Precision* will be 33.33%.

3.1 Experiment Setting

This section collects different social networks from three users and uses the experiment to verify *Accuracy* and *Precision* of the iGD system. The experiment collects users' preferences and their social networks and uses the information to conduct the experiment for each user. The system generates a list of recommendations (top 3) for each user and requests users to select and rank the recommendations in order to validate their performance. The number of experts is determined by the collected social network; that is, the range is 3 to 12 for user A, 3 to 9 for user B, and 3~12 for user C. The experiment also randomly assigns a number of missing attributes in the case matching process (CBR). Each case includes a total of sixteen attributes and this research allows a range of missing attributes between 0 to 8 to verify accuracy of the CBR approach. Hence, this work simulates ninety experiments for user A (i.e., ten different experts and nine different missing attributes), sixty-three experiments for user B (i.e., seven different number of experts and nine different number of missing attributes), and ninety experiments for user C. The simulated experiments also automatically generate the number of rounds. The results reveal that the discussion is averagely determined between three and five rounds, confirming the research of Rowe *et al.* (1999) who specified at least two rounds to qualify for the traditional Delphi method. The following sub-sections provide a cross analysis for three experiments and the summary of the two indicator results.

3.1.1 Background of three Social Networks

The experiment collects three users' preferences and social networks. The total numbers of peers for three users are different; for example, forty-three peers for user A, fifty-seven peers for user B, and thirty-five peers for user C. Structures of three social networks are particularly diverse. User A has the most peers in the 2nd level (61%), user B has the most peers in the 1st (39%) and 2nd levels (39%), and user C has the most peers in the 3rd level (40%) and 2nd level (31%). All users have few peers in the 4th and 5th levels and only user C has two peers in the 5th level.

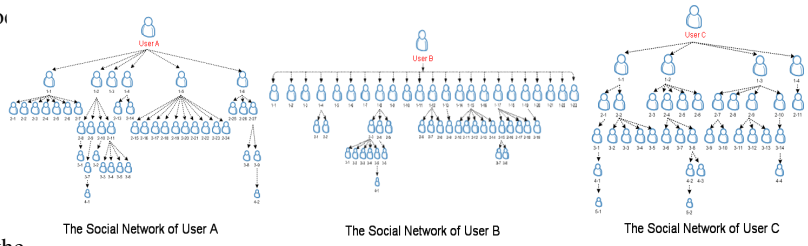


Figure 2 Collected social networks for three users

3.1.2 Comparison of Accuracy

This section compares the results of three users in terms of average accuracy for different number of experts and average accuracy for different levels of selected experts. The results (see Fig. 11) reveal that user A and C have superior average accuracy to user B. User A has nearly 70% accuracy and the accuracy enhances when the number of experts increases (e.g., after N=7). The variation of average accuracy for user A fluctuates (around 50% to 100%). Similarly, user C has at least 60% average accuracy. The variation of average accuracy for user C is stable (around 60% to 100%).

Conversely, user B also has at leat 60% average accuracy; however, the number of experts is limited to nine due to the insufficient number of experts. The variation of average accuracy for user B is stable (around 60% to 100%). In particular, all users have 100% average accuracy when the number of experts is three. This is because the experts are selected only from the 1st level and people always trust their friends in a strong-tie condition. That is, the average accuracy is extremely high in this situation. The trends for all users also reveal that average accuracy will enhance if the number of experts increases.

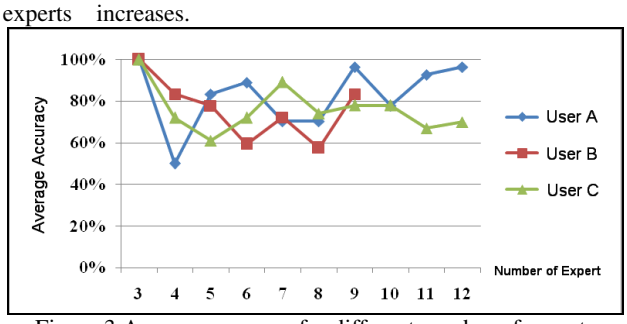
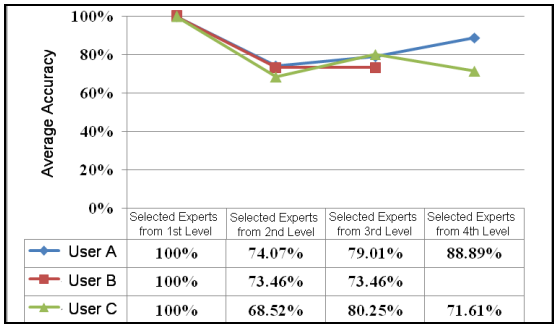


Figure 3 Average accuracy for different number of experts

The experimental results also reveal that the average accuracy decreases when the level of experts increases (Fig. 12). Experts for user A provide the most appropriate suggestions, especially for the 1st level of experts (100% of accuracy). Average accuracy decreases to 74.07% for the 2nd level but increases to 79.01% for the 3rd level and 88.89% for the 4th level. Conversely, the average accuracy of user B is also higher than 70% and even the 2nd and 3rd level both have 73.46% average accuracy. The 2nd level of user C has the lowest average accuracy (68.52%) in the experiment but the average accuracy is fair for other levels (80.25% for 3rd level and 71.61% for 4th level). In summary, our approach attempts to leverage trustworthy experts from different levels. The experimental results demonstrate that the number of levels affects average accuracy; in other words, average accuracy decreases when the level increases. However, average accuracy is superior for most experiments (i.e., higher than 70%).



	Selected Experts from 1st Level	Selected Experts from 2nd Level	Selected Experts from 3rd Level	Selected Experts from 4th Level
User A	100%	74.07%	79.01%	88.89%
User B	100%	73.46%	73.46%	
User C	100%	68.52%	80.25%	71.61%

Figure 4 Average accuracy for different levels of selected experts

In short, the structures of different users may result in the maximum number of experts. For example, the experiment results reveal that the maximum number of experts for user A is twelve and nine experts for both user B and C. The average accuracy for user A is between 60% and 100% and user B and C are both between 60% and 80%. Average accuracy for user A is incremental and declines for user C. Average accuracy for user B fluctuates. This is because the number of experts for all users and existing cases for selected experts are insufficient. However, all average accuracies are higher than 60%, which means the users accepted at least one or two suggestions.

Conversely, the number of levels also affects average accuracy for all users. The iGD system determines four levels for user A and average accuracy increases when the number of levels increases. The number of levels for user B is three and average accuracy diversity is not obvious. The average accuracy of the 2nd level and 4th level for user C is lower than other levels. Finally, the results provide evidence showing that our approach accurately and adequately recommends solutions (at least 60% of accuracy) within different social network structures.

3.1.3 Comparison of Precision

This section compares the results of three users in terms of average precision for different numbers of experts and average precision for different levels of selected experts. Precision is the indicator that measures if the rsolution ranks match the user's ideal list. Results reveal that average precision decreases for user A when the number of experts is between four and six (see Fig. 13). The lowest average precision value for user A is 40%. However, average precision increases when the number of experts increases after eight. The highest average precision value is 80%, which means our weighting and ranking approach attains 80% similarity compared to the user's ideal list.

The average trend for user B declines when the number of experts increases. This is because the experts for user B may be students and have insufficient knowledge about buying computers. User B may also prefer fancy shells rather than computer functions. The highest average precision value is 80% when the number of experts is three, which indicates that 1st level experts recommend solutions adequately to match user's ideal list. Average precision for user C also decreases when the number of experts is between four and seven. However, average precision increases when the number of experts raises after seven. The highest average precision value is 100% when the number of experts is three, which means 1st level experts recommend solutions adequately to match user's ideal list.

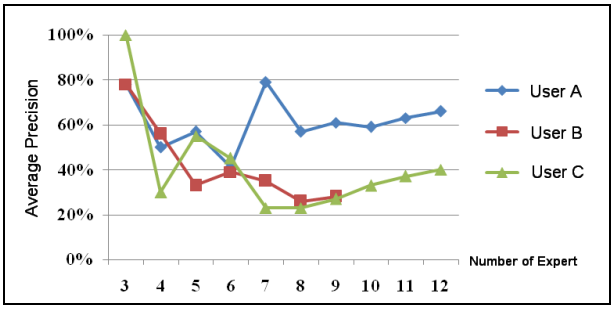
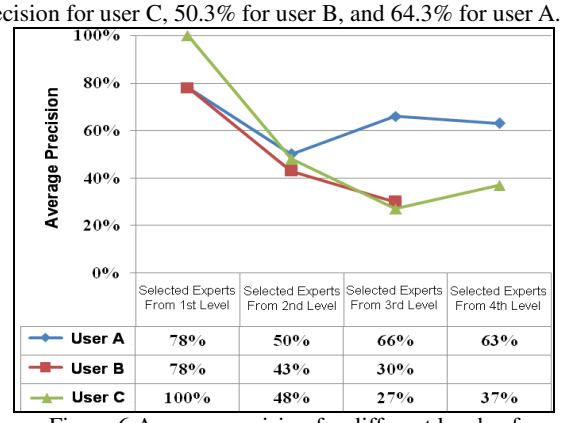


Figure 5 Average precision for different numbers of experts

The experimental results also reveal that average precision decreases when the level of experts increases (Fig. 14). In the experiments, 1st level experts recommend solutions adequately to match users' ideal ranking list (e.g., 100% for user C, 80% for both user A and B). The average precision dramatically decreases (e.g., 50% for user A, 43% for user B, and 48% to user C), which means that 2nd level experts only predict 50% of ranking. The average precisions of 3rd and 4th levels for user A are still higher than 60% (66% for 3rd level and 63% for 4th level). However, the average precisions of 3rd and 4th level for user B and C are both lower than 30%. This indicates that experts from the 3rd and 4th levels may provide totally different ranking from the users. The reasons for the user B result are that the background of experts is homogeneous and user B prefers a specific brand. In summary, different structures or numbers of experts may cause various results. In our experiments, the performance for user A is superior to user B and C. Yet, our approach predicts overall 53% of average precision for user C, 50.3% for user B, and 64.3% for user A.



	Selected Experts From 1st Level	Selected Experts From 2nd Level	Selected Experts From 3rd Level	Selected Experts From 4th Level
User A	78%	50%	66%	63%
User B	78%	43%	30%	
User C	100%	48%	27%	37%

Figure 6 Average precision for different levels of selected experts

4. CONCLUSION

This study proposes a system (iGD) for supporting a group decision-making process which is a type of modified Delphi method. Anonymity is the major strength of this model; meanwhile, the proposed system considers the power of collective intelligence from a social network. The approach also provides a way to prioritize and summarize ideas in a timely and efficient manner. Compared with the traditional group decision making methods (e.g., Delphi, brainstorming, and nominal group technique), this work combines the characteristics of these methods. For example, brainstorming enables a large number of ideas, the nominal group technique asserts ideas to be prioritized, and the Delphi method states the anonymity of participants.

The experimental results show that our approach has

superior average accuracy (overall higher than 60%), which means the iGD system generates adequate solutions from an automatic group decision-making process. Results also reveal that the iGD system produces at least 50% average precision, which indicates our weighting and ranking approach is valid and feasible in practice. This paper empowers collective intelligence from social networks (i.e., agent-based collective decision making), leverages the decision effort of experts (i.e., each level has selected experts to contribute together), ensures heterogeneity of experts (i.e., assure different levels of experts), and diminishes domination (i.e., different weights and ranks for various levels of experts).

REFERENCE

[1] A. Aamodt and E. Piaza (1994), "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," AI Communication, IOS press, 7(1), pp. 39-59.

[2] A. Kemgpol and M. Youminen (2006), "A framework for group decision support systems: An application in the evaluation of information technology for logistics firms," **International Journal of Production Economics**, 101, 59-171.

[3] A. M. Walker and J. Selfe (1996), "The Delphi method: a useful tool for the allied health researcher," **British Journal of Therapy and Rehabilitation**, 3(12), 677–80.

[4] B. F. Beech (1991). Changes: the Delphi technique adapted for classroom evaluation of clinical placements. Nurse Education Today. 11. 207–12.

[5] D. L. Passmore, E. D. Cebeci, and R. M. Baker (2005), "Market-Based Information for Decision Support in Human Resource Development," **Human Resource Development Review,** 4(1), 33-48.

[6] D. M. Boje and J. K. Mumighan (1982), "Group Confidence Pressures in Iterative Decisions," **Management Science**, 28(10). 1187-1196.

[7] D. R. Morgan, J. P. Pelssero, and R. E. England (1979), "Urban Planning: Using a Delphi as a Decision-Making Aid," **Public Administration Review**, 39(4), 380.

[8] E. Cornish (1977). The study of the future. Washington, D.C.: World Future Society.

[9] G. H. Baldwin (1982). The Delphi Technique and the Forecasting of Specific Fringe Benefits. Futures. 14. 319–325.

[10] G. Rowe and G. Wright (1999), "The Delphi Technique As a Forecasting Tool: Issues and Analysis," **International Journal of Forecasting**, 15, pp.353-375.

[11] H. A. Linstone (1978), "The Delphi technique," **Handbook of Futures Research**, Greenwood, Westport, CT, 271–300.

[12] H. A. Linstone and M. Turoff (1975), "Introduction to the Delphi method: techniques and applications," Addison-Wesley Publishing Company, MA, 3–12.

[13] H. Sackman (1975), Summary evaluation of Delphi, Policy Analysis, 1(4), 693–718.

[14] J. Pill (1971), "The Delphi method: substance, context, a critique and an annotated bibliography," **Socio-economic Planning Science,** 5(1), 57–71.

[15] J.Rohrbaugh (1979), "Improving the Quality of Group Judgment: Social Judgment Analysis and the Delphi Technique," **Organizational Behavior and Human Performance**, 24(1), 73.

[16] K. Brockhoff (1983), "Group Processes for Forecasting," **European Journal of Operational Research**, 13(2), 115-127.

[17] M. Adler and E. Ziglio (1996), Gazing into the Oracle: The Delphi Method and its Application to Social Policy and Public Health, London: Jessica Kingsley Publishers.

[18] M. T. Jelassi and A. Foroughi (1989), "Negotiation support systems: an overview of design issues and existing software," **Decision Support Systems**, 5(2), 167-181.

[19] Helmer (1977). Problems in futures research: Delphi and causal cross-impact analysis. Futures. 9. 17–31.

[20] Helmer and N. Rescher (1959). On the Epistemology of the Inexact Sciences. Management Science. 6(1). 25-52. M. T.

[21] P. Perner (2008), "Case-based Reasoning and the Statistical Challenges," Quality and Reliability Engineering International, 24(6), pp.705-720.

[22] P. R. Laughlin, B. L. Bonner, and A. G. Miner (2002), "Groups Perform Better Than the Best Individuals on Letters-to-Numbers Problems," **Organizational Behavior and Human Decision Processes**, 88(2), pp. 605-620.

[23] P. Zimmermann (1995), PGP Source Code and Internals. MIT Press.

[24] R. Phillips (2000), New applications for the Delphi technique, Annual San Diego Pfeiffer and Company, 12. 191-196.

[25] S. M. Campbell, M. Hann, M.O. Roland, J. A. Quayle, and P. G. Shekelle (1999), "The effect of panel membership and feedback on ratings in a two-round Delphi survey – results of a randomized controlled trial," **Medical Care**, 37(9), 964–968.

[26] S. W. Changchien and M. C. Lin (2005), "Design and implementation of a case-based reasoning system for marketing plans," Expert Systems with Applications, 28(1) pp. 43–53.

[27] V. Cavalli-Sforza and L. Ortolano (1984), "Delphi forecasts of land-use – transportation interactions," **Journal of Transportation Engineering**, 110(3), 324–39.

[28] W. C. Wedley, R. H. Jung and G. S. Merchant (1979), "Problem Solving the Delphi way," Journal of General Management. 5(1). 23.

A Distributed Storage Architecture based on a Hybrid Cloud Deployment Model

Emigdio M. Hernandez-Ramirez, Victor J. Sosa-Sosa, Ivan Lopez-Arevalo
Information Technology Laboratory
Center of Research and Advanced Studies of
the National Polytechnic Institute (CINVESTAV)
Ciudad Victoria, Mexico
(52) 834 107 0220 + 1114, 834 107 0241, 834 107 0220
ehernandez,vjsosa,ilopez{@tamps.cinvestav.mx}

***Abstract*—This paper presents a practical case of study that analyzes the behavior of a distributed storage architecture, which was developed on a hybrid cloud computing environment. Open source software was used for implementing this architecture. An elastic service can be supported by virtualization technologies that allow the architecture to increment and decrement resources on demand. Performance and resource consumption were evaluated applying different replication techniques, which offer several levels of data availability and fault tolerance. The obtained results help to identify the benefits and limitations that arise when a system based on this architecture is implemented. The prototype allows us to visualize some trade-offs when applying different replication techniques depending on the availability, efficiency, fault tolerance and privacy required by users.**

Keywords cloud computing, scalability, virtualization, high availability.

I. Introduction

To define the storage requirements for institutions or companies of any size has become a problem with no trivial solutions. This is mainly due to the very fast generation of digital information which behavior is very dynamic[1].

In this context, it is common that managers of storage resources, with the responsibility to make predictions about the resources that will be needed in the medium term, often face the following scenarios: a) predictions are below of real needs, in this case, there will be a problem of resources deficit; b) generation of an excessive expenditure on the purchase of storage resources, producing a complex administration and probably with resources that will not ever be used in the medium term. This situation makes it attractive the acquisition of storage services that implement an elastic concept, i.e., those having the ability to grow and being reduced on demand, wherein the costs of acquisition and management are relatively low.

Nowadays, this service model is called cloud computing. In this model, storage resources are provisioned on demand and are paid according to consumption.

Services deployment in a cloud computing environment can be implemented in basically three ways: private, public or hybrid. In the private option, resources belong to a company; this implies an initial strong investment for the company, because it is necessary to purchase a big amount of storage resources and assume the administration costs. In the public version, resources belonging to a third party, where costs are a function of the resources used. These costs include administration. Finally, hybrid version contains a mixture of both. The cloud computing model is mainly supported by the development of technologies such as virtualization and service-oriented architecture.

Distributed storage services over a cloud environment provide omnipresence and make it easier their deployment. This means that users can access their files from anywhere, while there exists an Internet connection and without requiring the installation of a special application (only it is needed a web browser).

Data availability, scalability, elastic service and pay only for consumption are very attractive characteristics found in the cloud service model. Virtualization is playing a very important role in the cloud computing. With this technology, it is possible to have facilities such as multiple execution environments, sandboxing, server consolidation, use of multiple operating systems, ease of software migration, among others.

Besides virtualization technologies, emergent tools for creating cloud computing environments that provide dynamic instantiation and release of virtual machines and software migration are also supporting the elastic service offered in this kind of computing model.

Although there are currently several proposals for cloud storage, such as Amazon S3 [10], RackSpace[3] or Google Storage[4], which provide high availability, fault tolerance and services and administration at low cost, there still are companies that do not feel confident to store their information in a third-party-owned environment. In these cases, such companies that would like to take advantages of cloud computing, would require to implement a private cloud solution. Unfortunately, this option often is beyond the scope of their budgets. This dilemma makes it attractive to think about a hybrid solution, in which companies or users in general can store some information using a private infrastructure, e.g. sensitive or most frequently used data, and store the rest of the information in a public cloud.

To evaluate different alternatives of implementation that might have this kind of companies was the main motivation of

the architecture presented in this paper. With the development and evaluation of a prototype of a storage service implemented on a hybrid cloud environment based on free software, we wanted to analyze the behavior of a service like this, taking mainly into account the low cost of the system implementation, the system efficiency, resource consumption and several levels of data privacy and availability by using different techniques of data replication.

The following list summarizes the contributions of this paper:

1) Proposal of a distributed storage architecture implemented on a hybrid cloud computing environment based on free distribution software.
2) Results of evaluations made at different settings applied to the infrastructure, offering several levels of data availability, fault tolerance and privacy, by means of implementing different replication mechanisms.
3) Proposal for an innovative replication mechanisms based on the Information Dispersion Algorithm [2], which was adapted to a hybrid cloud computing model.
4) A Prototype of a web distributed storage system that is supported by the architecture presented in this paper. This system was called DISOC (Distributed Storage on the Cloud) and represents our proof of concept.

The rest of the paper is organized as follows; section II includes related work. Section III describes the components that form the distributed storage architecture, which is supported by virtualization technologies. Section IV presents an evaluation of the tests and results obtained from the DISOC prototype and section V offers some final comments and conclusions.

II. Related work

Nowadays Amazon S3 is considered a pioneer of cloud storage solutions.It offers to its users different rates for storage, according to the amount of the stored data. These rates vary depending on the data availability required by users. Data availability is related to the replication technique that will be used in the Amazon infrastructure[10].

There exist also solutions that take advantages of public cloud storage using replication techniques that were originated in RAID, for example RACS[8], which is a proxy that is located between multiple cloud storage providers and customers. It is responsible for distributing data in a way that it provides an opportunity for clients to tolerate interruptions in a public cloud storage service or when the price for using the services is getting high. It uses replication in order to support those possible situations. RACS offers to its users an interface similar to Amazon S3, allowing operations such as PUT, GET, DELETE and LIST. Another proposal is HAIL[9], a cryptographic distributed system that allows file servers to provide a secure storage environment. HAIL supports the failure of any of the servers that make up the system, adding a degree of security to stored data using an approach based on the Reed Soloomon error correction codes.

Currently there are public cloud storage infrastructures such as Amazon S3[10], Rackspace[3], Google Storage[4] that are being used by distributed file systems such as Dropbox Dropbox[13], Wala[12], and ADrive[11], that allow users to store and share file. A common point in these infrastructures and applications is the use of public clouds. These services are being very useful for users wanting to have an unlimited storage space or to backup their data. However, the use of these type of solutions can be a challenging decision for a business environment. This is because some organizations have fear of storing sensitive data in a third party infrastructure or that the data could not be available at the time they were required. Our approach suggests creating a hybrid cloud storage environment (private + public), with low cost infrastructure, in which only part of the stored data are in the public environment, minimizing the likelihood of unauthorized access.

III. Infrastructure Description

In the previous section, it was mentioned that a small and medium businesses (SMB) could face some economical and technical challenges when trying to obtain the benefits of having their own cloud computing environment (private). Our proposal is trying to help with those challenges by designing and implementing an scalable and elastic distributed storage architecture based on a free and well known open source tools. This architecture is thought for combining private and public clouds by creating a hybrid cloud environment. For this purpose, tools such as KVM[5] and Xen[6], which are useful for creating virtual machines (VM), were evaluated. For managing a cloud environment is possible to use tools such as Open Nebula[15] and Eucalyptus[16].

The hard disks (HDs) integrated into the storage infrastructure are found in commercial computers (commodities). The use of this type of HDs makes this architecrure failure-prone. This was the reason why, we evaluate different replication mechanisms, which provide several levels of data availability and fault tolerance. Figure 1 shows the core components(a) included in the storage architecture (private cloud) and the distributed storage web system (DISOC) that is used as a proof of concept(b). It can be seen that the private cloud has an interface to a public cloud allowing a hybrid environment.

The core components of the architecture are the following:

- Virtual Machine (VM): In our current private cloud implementation, every core in a physical machine is ideally thought for running only one virtual machine. This situation can be changed depending on the level of workload. The open source tools KVM[5] and Xen[6] were evaluated to decide which one could offer a better performance in terms of virtual machine instantiation. Results of those tests are not included in this paper due to space limitation. KVM showed a slightly higher performance than XEN, reason why we chose KVM, similar results can be found at [17]. Each VM has a Linux operating system optimized to work in virtual environments, which requires a minimum consumption of disk space. The VM also includes an Apache web server, PHP and some basic tools that were used to build

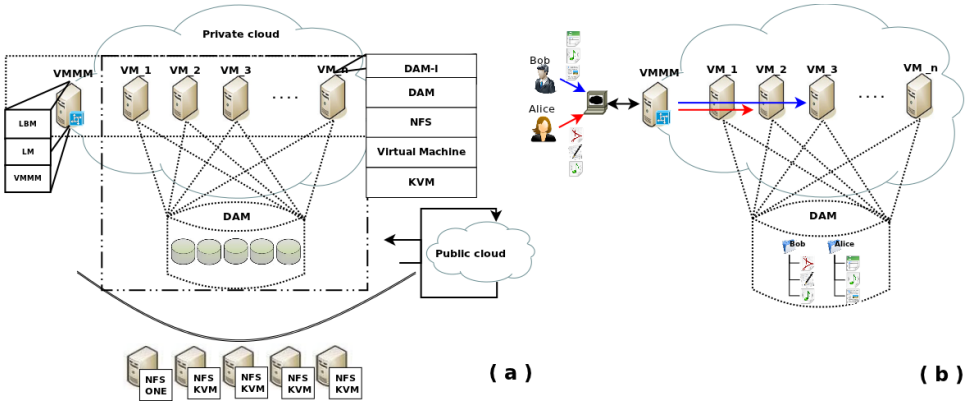


Fig. 1. Core components in a private/public cloud storage infrastructure.

the DISOC prototype. Every VM is able to access to a pool of disks through the DAM module.

- Virtual Machine Manager Module (VMMM). It has the function of dynamic instantiation and de-instantiation of virtual machine depending on the current load on the infrastructure. We evaluated two open source tools for managing virtual machines, Open Nebula(ONE)[15] and Eucalyptus[16]. We chose ONE, because it offers more simplicity in the installation/configuration and has more support and documentation available online.
- Data Access Module (DAM). In order to improve the speed of deployment of VMs and the storage service scalability, it was allocated a minimal physical disk space in every virtual machine (VM). The real disk space used by every VM was given by a Data Access Module Interface (DAM-I), which allows VMs to get access to disk space by means of a Data Access Module (DAM). The main function of DAM is to provide transparent access to the different disks that are part of the storage infrastructure. It allocates and retrieves individual files stored on different file servers. In this context, each VM has the notion of being interacting with a single disk. DAM is implemented over NFS and includes a file allocation algorithm that locates the whole file, or part of it, using a Round Robin policy that follows a sequential identification mechanism. This allocation scheme allows DAM to find the location of a file using a minimum of additional information (metadata). Since DAM is configurable, it is possible to evaluate the performance of the storage service according to several levels of availability, applying different replication techniques.
- Load Balancer Module (LBM). It is designed to distribute the load among different VMs instantiated on the physical servers that make up the private cloud. The LBM is configurable, so it is possible to define different balancing policies. The results presented in this paper consider a Round-Robin policy. LBM is the main gateway for the storage service. The NGinx web server[14] was adapted to become LBM, because this server can work as load balancer and has a low consumption of resources, essential point in a virtual environment with limited resources.
- Load Manager (LM). Basically, this module is responsible for monitoring the load that can occur in the private cloud. In general, it keeps track of the average response time per request in each VM. Exceeding a threshold (configurable), the manager informs the VMMM in order for it to deploy a new VM into the private cloud. LBM is also informed of a new VM that has to be considered in the load distribution. Likewise, when a low load threshold is reached, the VMMM will shut down a VM and the LBM will not consider it in the load balancing process in future requests.
- Distributed Storage On the Cloud (DISOC). It is a web-based file storage system that is used as a proof of concept of our architecture.

A. *Replication mechanisms*

High availability is one of the important features offered in a storage service deployed in the cloud. To accomplish this, the use of replication techniques is very common. DAM is the component that is configured to provide different levels of data availability. It currently includes the following replication policies: no replication, full replication, mirroring and IDA-based replication.

- No Replication. This replication policy represents the lowest level of data availability. With this scheme, only the original version of a file is stored in our disk pool, following a Round Robin allocation policy, depending on disk availability. This allocation method prevents files from being restricted to a single server, providing a minimal fault tolerance. Figure 2 (a) illustrates this allocation scheme using a pool of disks ($D_0 ... D_n$).
- Mirroring. This replication technique is a simple way to ensure higher availability, without high resource consumption. In this replication, every time a file is stored in a disk, DAM creates a copy and places it on a different disk. As shown in Figure 2(b), the distribution of files follows also a Round Robin policy, adding the copy of

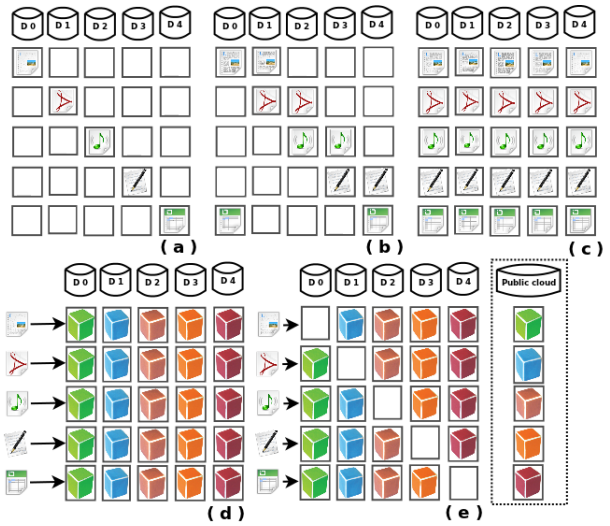


Fig. 2. Replication mechanisms

the file in the next available disk. The total number of bytes stored is $|F|*2$, where $|F|$ is the size of the original file.

- Total replication. Represents the highest data availability approach. In this technique, a copy of the file is stored in the total file servers available. It is also the strategy that requires the highest consumption of resources. The total sum of bytes stored is $|F| * n$, where n is the total number of file servers. As it is shown in Figure 2(c).
- IDA-based replication. In order to provide better data availability, with less impact on the consumption of resources, an alternative approach based on information dispersal techniques can be used. The Information Dispersal Algorithm (IDA)[2] is an example of this strategy. When it is required to store a file using IDA, a file of size $|F|$ is partitioned into n fragments of size $|F|/m$, where $m < n$. These fragments are distributed in n different disks. IDA only needs to obtain m fragments to reconstruct the original file. Under this scheme, even if $n-m$ disks failed, the file would still be recovered, that is why it is desirable that no more than $n - m$ file servers fail. IDA provides better fault tolerance than mirroring without needing to totally replicate the original file. In this prototype was evaluated IDA with $n = 5$ and $m = 3$ (this means a 60% of the original file is replicated). IDA seems attractive for being used in a hybrid cloud environment, since it is not necessary to save the entire file on a single file server (disk), so it could be possible to send k fragments of the file (where $k < m$) to a public cloud storage without revealing the content of the original file. As shown in Figure 2(d,e).

IV. Results

The evaluation scenario used to test our prototype of a infrastructure was built basically using 5 commercial PCs (commodities), which characteristics are shown in first section of table 1. This private cloud is able to be connected to a public cloud, allowing a hybrid cloud environment. The features of the VMs (for this test, there were only 5 VMs, each using one core) that were instantiated on the mentioned PCs are shown in section 2 of table I.

Physical machines				
	Cores	Memory	Hard disk	Network
1 pc	4	4 Gb	640 Gb	Ethernet 10/100
4 pc	2	2 Gb	250 Gb	Ethernet 10/100
Virtual machines				
5	1	1 Gb	1 Gb	Virtual
1	1	128 Mb	1 Gb	Virtual

TABLE I
Characteristics of the physical PCs and VMs used in the private cloud

In this evaluation, the access to a public storage cloud was emulated by connecting our private storage cloud with an external disk, located at a different network through a public internet connection. For the sake of simplicity (and keeping full control of the test) in this evaluation was not used a connection to the Amazon S3 public storage cloud.

Results obtained from this prototype are intended for evaluating: a) the impact of having an elastic service and, b) the behavior of the system when requiring several levels of data availability, applying different replication techniques.

A. The impact of having an elastic service

As a first step, it was evaluated the impact of having elasticity in the storage service versus a static service (without elasticity). In the elastic service, a new virtual machine is instantiated when a workload exceeds a defined threshold. The evaluation uses different workloads generated by Autobench[7]. The evaluation of the static service was useful for defining a benchmark that allows us to recognize the benefits obtained by an elastic service. In this context, it was compared the behavior of a single physical machine with a hard disk receiving an increasing workload versus applying the same workload on a set of virtual machines that were incrementally instantiated. For this test, the workload basically consists of a set of requests of a dynamically generated PHP web page. This web page emulates a processing time on the server by running a sorting algorithm (bubble type). Trying to emulate different levels of load on the server, it was defined a list containing different quantities of elements that had to be sorted. The results shown in Figure 3 represent the average response time a customer received when the load balancer only accessed to one physical machine (red line), and when the balancer accessed the same physical machine with 1 to 3 (blue line). It can be seen, at the beginning of the test, when the workload is low, how the response time offered by the static service (running only on one physical machine) is better, in some cases up to 4 or 5 orders of magnitude, compared to that obtained in the execution of the service

accessing to one virtual machine. In this test, a maximun response time of 30s was defined as the upper treshold for a new VM instantiation. It means that when the global system response time reaches 30 seconds, a new virtual machine will be instantiated and integrated into the storage service. It can be seen that the response time in the elastic service has some considerable falls during the test. This behavior is not occurring at the time of a new VM instantiation, but at time when the VM is included in the service by the load balancer. The instantiation and activation time of the new VM is between 60 and 90 seconds. At the time the workload increases, it will be necessary to instantiate another VM. The elastic service was always able to finish the workload offering an acceptable response time, while the static service collapsed and could not meet the total requests. Likewise, when this descending activity will be monitored until get running only one VM on all the infrastructure.

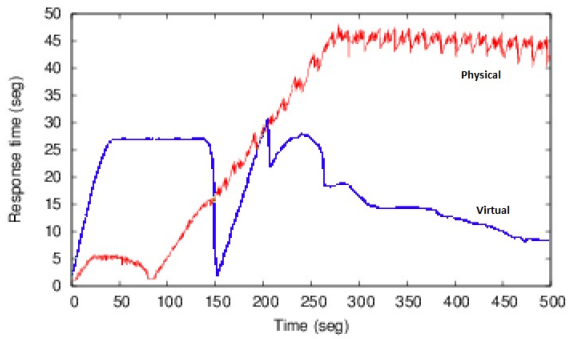


Fig. 3. Performance comparison between a fixed and elastic storage service

B. Data availability, evaluation of different replication policies

DAM component allows us to define the level of data availability required in the storage service. This can be done by applying different replication techniques. In this test, it was defined a benchmark that shows the benefits obtained of using a distributed storage system versus a centralized version.

For this evaluation, DAM was configurated for having access to a single disk. This test ran the storage system into one VM (emulating a centralized processing) with a single storage server (emulating centralized storage). The rest of the tests were always considered using a distributed processing (5 VMs) and distributed storage (5 disks that were distributed on different storage servers encapsulated by DAM). Since the replication with IDA policy is attractive to a hybrid cloud service, we compared its behavior in cases when it is only used on a private cloud and when it is also considering the use of a public cloud (hybrid model).

Two main metrics were taken for these experiments: 1) response time: it considers the time from when the user clicks on the button to upload or download a file, until the point when the file loading or downloading has finished, in this test, until the TCP connection is closed down. 2) service time: the time needed by DAM for locating a file (or part of it) and that the file is ready to be read by the system component that is requesting it.

The response time obtained for users during the uploading process is very similar independently of the replication technique that was used, except for the hybrid version of IDA. It can be seen that IDA was very affected when it involved the access to the external infrastructure (public cloud). The impact on hybrid IDA is given because some file fragments have to to be sent to/retrieve from the external infrastructure through a public internet connection. The main benefit of storing some file fragments in the external infrastructure is the fact of having more storage space available in the private cloud. It is important to remember that the number of fragments that are sent to the public infrastructure will never be greater than or equal to m, where m is the number of pieces required to build the original file.

As it is shown in Figure 4, the response time for downloading file using the hybrid version of IDA is also the most affected. Response times in the downloading process have similar behavior. For testing the behavior of this version of IDA, DAM was configurated to always obtain a fragment of a file from the the public cloud. It should be noted that this is not the typical case, because in a real scenario, the hybrid version of IDA only would obtain a fragment of a file from the public cloud in the cases when it was not able for DAM to obtain the m needed fragments from the private cloud, which means that more than n-m disks had failed (worst case).

Service times observed in Figure 4 suggest that again the higher consumption of time is due to the use of a public provider. It can be seen that the service time generated mainly by DAM is minimal compared to the total response time, independently of the replication technique that is used. The only exception of this is the hybrid version of IDA, which is being forced to get access to the public cloud.

V. Conclusion

This paper described the design, implementation and validation of a distributed storage architecture that takes into account a hybrid cloud model. It was introduced DAM, a simple mechanism for storage consolidation on a hybrid cloud environment, which is able to offer different levels of data availability based on users requirements. DAM uses a lightweight algorithm for file allocation, reducing the amount of metadata needed with a low resources consumption. Another point discussed was the real performance improvement obtained when using an elastic (virtualized) environment, instead of a physical environment. This will be true especially when the system is prone to receive big workloads. Finally, it is shown how the hybrid version of the IDA algorithm can be a viable solution for those SMB that want to obtain the benefits of cloud storage without exposing the content of all of their files in a third-party infrastructure.

Acknowledgments. This research was partially funded by project number 173455 from Fondo Mixto Conacyt-Gobierno del Estado de Tamaulipas.

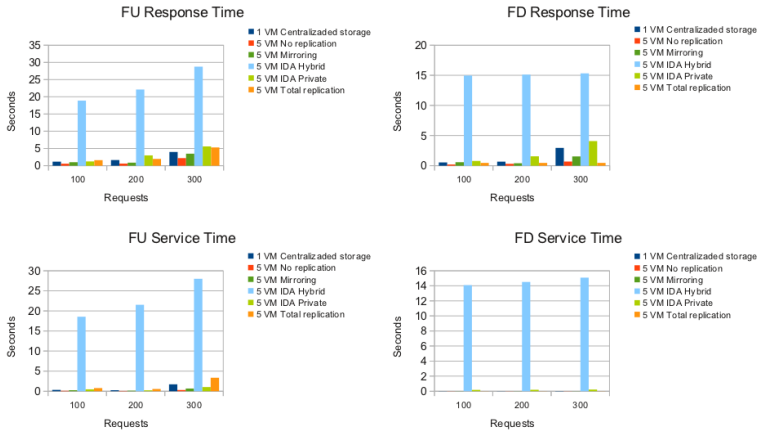


Fig. 4. Average response time and service time for file uploading(FU) and downloading(FD) using different replication techniques

REFERENCES

[1] John F. Gantz et al, *The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010*, An IDC White Paper - sponsored by EMC

[2] Michael O. Rabin, *Efficient dispersal of information for security, load balancing, and fault tolerance*, J. ACM 36, 2 (April 1989), 335-348.

[3] Rackspace Cloud Files, *http://www.rackspace.com/cloud/cloudhostingproducts/files*, August 2011

[4] Google Storage for Developers *http://code.google.com/apis/storage*, August 2011

[5] Kernel Based Virtual Machine, *http://www.linux-kvm.org* , August 2011

[6] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, and Andrew Warfield, *Xen and the art of virtualization*, Proceedings of the nineteenth ACM symposium on Operating systems principles (SOSP '03)

[7] Autobench, *http://www.xenoclast.org/autobench*, August 2011

[8] Abu-Libdeh, H et al, *RACS: a case for cloud storage diversity*, Proceedings of the 1st ACM Symposium on Cloud Computing, June 2010

[9] Bowers K. D. et al, *HAIL: a high-availability and integrity layer for cloud storage*, Proceedings of the 16th ACM Conference on Computer and Communications Security, November 2009.

[10] Amazon Simple Storage Service(S3), *http://aws.amazon.com/s3*, August 2011

[11] ADrive, Web storage, http://www.adrive.com,August 2011

[12] Wala, Secure online storage *http://www.wuala.com*, August 2011

[13] Dropbox *http://www.dropbox.com/features*, August 2011

[14] NGinx web server, *http://wiki.nginx.org*, August 2011

[15] OpenNebula home page, http://opennebula.org, August 2011

[16] Eucalyptus home page, http://www.eucalyptus.com, August 2011

[17] Comparative of xen and kvm available in: http://virt.kernelnewbies.org/XenVsKVM, August 2011

Comparison of Extended Fuzzy Logic Models of A-IFS and HLS: Detailed Analysis of Inclusion in the A-IFS of the Data Sets for Implication Operations

Xiaoyu HUANG
and
Tetsuhisa ODA

Aichi Institute of Technology
2-49-2 Jiyugaoka, Chikusa, Nagoya city, Aichi, prefecture, 464-0044, Japan

ABSTRACT

After the proposal of Zadeh's "Fuzzy Set Theory", several kinds of extended fuzzy set/logic models have been proposed. Some extended models treat any multi-dimensional fuzzy logic system. Typically, the models assume a pair of (*t, f*) for the truth-value of an ambiguous proposition **A**, while $t, f \in [0,1]$. The *t* means truthfulness and *f* means the falsity of the proposition **A**. Both Atanassov's "Intuitionistic Fuzzy Set" (A-IFS) model and Oda's Hyper Logic Space (HLS) model treat the two-dimensional logic space, but there are some differences regarding (1) the domain areas of definition and (2) the formulas of negation operation. For comparison of the two models, the authors investigated the following two conditions. Condition 1: Both propositions **A**=(*ta, fa*) and **B**=(*tb, fb*) are in the A-IFS area ($t+f \leq 1$), and both results of the implication operations by the two models return to the A-IFS area. Here, the HLS model can use both the A-IFS area and the contradiction area ($t+f > 1$) only as a calculating space. Condition 2: Though both propositions **A** and **B** are in the A-IFS area, the results of the implication operations by the HLS model are no longer in the A-IFS area but in the contradiction area. For the purpose of comparing the results, the following two methods are applied. Method 1: The resultant points (*t, f*) calculated by the implication are converted to the corresponding Interval Valued sets [*t, 1-f*] in one dimensional numerical truth-value space **V**. Method 2: As the contradicted area data is impossible to convert to the Interval Valued set, the resultant points from both models are integrated using one of Oda's formulas named $\boldsymbol{I_4}$. Since $\boldsymbol{I_4}$ is symmetrical to both data areas, it is proper to apply. Analyzing the results of both conditions, we concluded that the HLS model, especially the usage of the contradiction area, is useful for implication operation not only for treating the HLS data but also for treating the A-IFS data.

Keywords: Fuzzy Logic, Extended Fuzzy Logic, Intuitionistic Fuzzy Set, Hyper Logic Space, Fuzzy-set Concurrent Rating Method

1. INTRODUCTION

In 1965, L. A. Zadeh proposed the "Fuzzy Set Theory" for mathematically modeling a kind of ambiguous concept. After the proposal, many investigators including Zadeh himself followed and developed the theory from various points of view. In the early stage of this development, the Numerical Truth-value model was proposed. The Numerical Truth-value model assumes a real number value ***t***, ranging from 0 to 1, as a truth-value of a fuzzy proposition. Nowadays, the model is recognized to be representative of Fuzzy Logic (FL) models. Several kinds of extended Fuzzy Set/Logic models have since been developed. For example, the Interval Valued Fuzzy Set (IVF) model (See Fig. 1) is a well-known extended model. According to D. Dubois & H. Prade [1], the IVS includes most of the extended Fuzzy Set/Logic models.

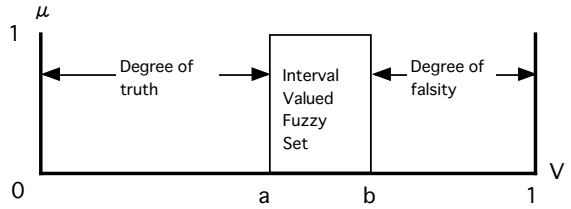


Figure1: Interval Valued Fuzzy Set (IVS)

In Bulgaria, K. Atanassov [2] proposed the Intuitionistic Fuzzy Set (IFS) model. IFS assumes not only the degree of membership function $\boldsymbol{\mu(x)}$ but also the degree of non-membership function $\boldsymbol{\nu(x)}$ of an ambiguous set. In Japan, the Hyper Logic Space (HLS) model was proposed by T. Oda [3]. The HLS model extended the numerical truth-value of the Fuzzy Logic in order to define the special indexes of the newly developed psychological measurement method namely Fuzzy-set Concurrent Rating (FCR) method. As HLS closely resembles IFS, the advantage of HLS over IFS has not been so clear, though HLS can also treat contradictory data sets while the IFS can not.

Among extended Fuzzy Set/Logic theories, K. Atanassov's IFS theory may be the most widely well-known model. (It is now also called the A-IFS, to distinguish it from another model with the same name proposed by G. Takeuchi and S. Chitani [4]. Takeuchi and Chitani's model has been called T-IFS.)

Apart from the European academic background, other models have been proposed in Japan, which assumed that *t* (truthfulness) and *f* (falsity) of a proposition are mutually independent as A-IFS assumed. One of the early models is Mukaidono and Kikuchi's "Between Fuzzy Logic" (BFL) model [5], in which they extended "Interval Valued Fuzzy Logic" by permitting the lower limit *a* to exceed the upper limit *b* of the interval value [*a, b*]. (See Fig. 2)

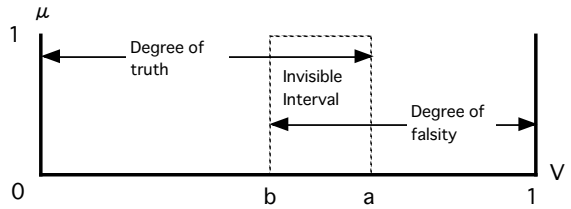


Figure 2: Between Fuzzy Logic (the case of invisible interval)

Against this background, the authors proposed the HLS model in which ***t*** and ***f*** are defined as perfectly independent, while A-IFS and other models have been assumed limited independencies of $t+f \leq 1$.

This study includes two analyses. Analysis 1 treats the condition that both results of the implication operations, derived by the models, of two points **A** and **B** in the A-IFS area return to the A-IFS area. In Analysis 2, another condition is investigated.

Throughout this paper, the effort of analysis is concentrated on uncovering the differences of the implication operation formulas between $\neg\mathbf{A}\vee\mathbf{B}$ by HLS and $\sim\mathbf{A}\vee\mathbf{B}$ by IFS.

This study is intended to show some advantages of the HLS model over the IFS model under the restriction of applying only A-IFS data sets.

Since the resultant point of the implication $\neg\mathbf{A}\vee\mathbf{B}$ can appear in the contradiction area, while the resultant point of $\sim\mathbf{A}\vee\mathbf{B}$ never spills out from the A-IFS area, the two models cannot be compared directly. So, the results are indirectly compared by calculating their integrated values. For the integration algorithm, the authors adopted the Combined Scoring Method 1 (***I4***) developed by T.Oda.

2. FCR-METHOD ,HLS MODEL AND OTHER CONCEPTS

In this chapter, the compendium of the HLS model, the FCR-method, and related concepts used in the latter part of this paper are introduced.

2.1 Hyper Logic Space

Hyper Logic Space (HLS) is the two-dimension fuzzy logic space **T** × **F.** Here, **T**=[0,1] means truthfulness while **F**=[0,1] means falsity. HLS has five special points in the coordinates:
0=(0,0) , **T**=(1,0) , **F**=(0,1) , **1**=(1,1) and **C**=(0.5,0.5).
The meanings of the points are as follows.

- **0**: empty point (Perfectly irrelevant)
- **T**: true point ("Truth" in Classical Logic.)
- **F**: false point ("False" in Classical Logic.)
- **1**: contradicting point (Perfectly contradicting)
- **C**: center

The points on the line t+f=1 compose the numerical truth-value space **V** of FL. The three points F, C and **T** appear on the line. The other points in HLS have not been treated in traditional FL. The area composed of the points t+f>1 is named "contradiction area", while the area composed of the points t+f<1 is named "irrelevance area". A-IFS has been treats only the area defined by t+f ≦ 1.

The logical operations of HLS are discussed in Chapter 3.

2.2 FCR-method

In 1993, T. Oda proposed a new technique to identify the fuzzy membership function of a concept by using three monopole scales. Then, the idea was developed to the Fuzzy-set Concurrent Rating method (FCR method) as a new technique for general psychological measurement [10]. The FCR method is designed to measure a subject's opinions or attitudes more naturally than traditional bi-polar rating scale methods. Since then, various new concepts and related algorithms have been developed [3,6-12].

The two-item type FCR method uses two independent rating scales for each question. (See Fig. 3)

The scales are used to measure positive and negative responses respectively. Thus, one question obtains a pair of responses and is represented as (*p,n*). By combining the pair, an integrated value is calculated. The integrated value is an alternative to a rated score value on a traditional bi-polar rating scale.

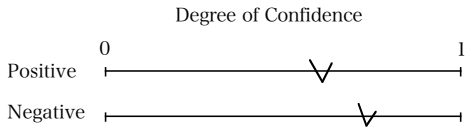


Figure 3: FCR-scale (two-item type)

Furthermore, possible contradictions in the responses or possible irrelevancies to the question can also be observed from the pair. In the FCR-method, the irrelevancy-contradiction index is very important for analyzing actual data. So, various kinds of formulas have been developed. But, the detailed explanations of the indexes are omitted here. Nowadays, $C_3=p+n-1$ is generally used as the irrelevancy-contradiction index in the FCR-method because of its simplicity and linearity. ($-C_3$ is identical to the "hesitation margin" introduced by P. Merin in her extended fuzzy logic/set model "Medative Fuzzy Logic" [13].) Furthermore, as a psychological measurement system, special instructions, that both scales should be independently rated, are needed when it is practically applied.

2.3 Integrated Value of the FCR-method

Assume the positive scale value ***p*** and negative scale value ***n*** of the FCR-scale are directly assigned to truthful ***t*** and falsity ***f*** respectively in HLS.

Fig.4 is illustrating the fundamental integration algorithms of the FCR-method, I_1, I_2 and I_3 has been proposed. In this figure, they are explained by using the projection lines pass through the observed point A.

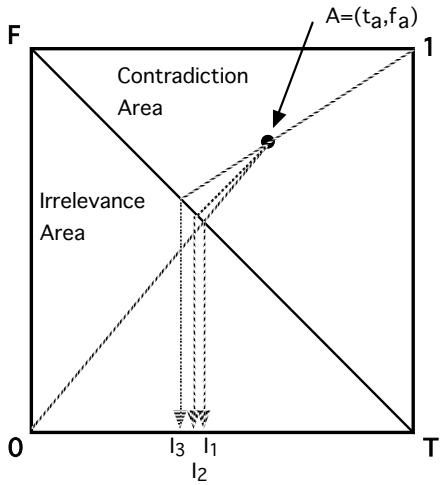


Figure 4: Fundamental integrated values illustrated in the HLS

Simple scoring method (*I1*): By assigning the score value ***1*** for ***t*** while **0** for ***f***, the weighted average score for the pair (*t, f*) is calculated by the following formula.

$$I_1 = \frac{t}{t+f} \quad if\ t+f \neq 0 \quad otherwise\ \ I_1 = 0.5 \qquad (1)$$

Reverse item averaging method (*I2*): Since the degree of falsity ***n*** can coincide with the degree of truthful ***t*** by negation, one of the basic integration formulas is defined by averaging ***t*** and ***1-n***.

In other word, the negation of the falsity (***1-f***) can be considered to be the alterative of the truthful ***t.***

$$I_2 = \frac{t+1-f}{2} \qquad (2)$$

Inverse scoring method (*I3*): By assuming that (1-t)

can be the alternative of n while (1-f) can be the alternative of t, the weighted average of the pair ((1-f), (1-t)) when the scores 1 and 0 are assumed respectively.

$$I_3 = \frac{1-f}{2-t-f} \quad if\ t+f \neq 2 \quad otherwise\ \ I_3 = 0.5 \tag{3}$$

By combining these integrated values, various combined scoring method were proposed and named I_4 to I_{11}.(See Table 1) [12]

Table 1. Combined Integration algorithms

Name of the algorithm	Definition	Eq.
Combined scoring method 1	$I4 = I1$ *if* $t + f \leq 1$ *otherwise* $I4=I3$	(4)
Combined scoring method 2	$I5 = I3$ *if* $t + f \leq 1$ *otherwise* $I5=I1$	(5)
Combined scoring method 3	$I6 = (I1 + I3) / 2$	(6)
Combined scoring method 4	$I7 = \sqrt{I_1 + I_3}$	(7)
Combined scoring method 5	$I8 = 2/(1/I1 + 1/I3)$	(8)
Combined scoring method 6	$I9 = (I1+I2 + I3) / 3$	(9)
Combined scoring method 7	$I10 = \sqrt[3]{I_1 \cdot I_2 \cdot I_3}$	(10)
Combined scoring method 8	$I11 = 3/(1/I1+1/I2 + 1/I3)$	(11)

2.4 Converting principle from a point to an interval

In this investigation, the well-known converting principle is introduced. According to the principle, one point of **A**=(*ta,fa*), which is in the two-dimensional fuzzy logic space **T** × **F**, which is identical to the HLS introduced here, is converted into the closed interval [*ta, 1-fa*] in the one dimensional numerical truth-value space **V**.

3. COMPARING HLS MODEL AND IFS MODEL

3.1 Common assumptions and definitions

Both models are closely resembles excepting the data area and the negation operations.

Assume both **A** and **B** are fuzzy propositions. The pairs of truth-value and false-value are referred to:

$$\mathbf{A} = (ta, fa),\ \mathbf{B} = (tb, fb) \text{ while } ta, tb, fa, fb \in [0,1] \tag{12}$$

The logical OR (∨) and the logical AND (∧) operations are just the same as follows.

$$A \vee B = (\max(t_a, t_b), \min(f_a, f_b)) \tag{13}$$

$$A \wedge B = (\min(t_a, t_b), \max(f_a, f_b)) \tag{14}$$

Since the fuzzy set operations are defined by the logical operations of the elements, the operations can be compared.

3.2 Differences of data area

Any data pair (***t, f***) of the A-IFS has the constraint of t + f ≦ 1 by the definition of the model. Meanwhile, t and f are completely independent with each other in HLS model.

3.3 Differences of negation operation

A negation operation in the HLS model is a natural extension of Zadeh's negation, and is so-called external negation of a proposition.

$$\neg\ \mathbf{A} = (1\text{-}ta\ ,1\text{-}fa) \tag{15}$$

Meanwhile, a negation operation in the A-IFS model is so-called internal negation; referred to exchange ***t*** and ***f.***

$$\sim \mathbf{A} = (fa, ta) \tag{16}$$

For distinguishing the negation operations of each model, the other symbol ~ is used for the negation operation of A-IFS.

The result of negation of A-IFS by (3) never protrudes out of the A-IFS areas.

3.4 Differences of implication operator

As a common definition for the implication operation (**A** → **B**) for both models, the "the negation of **A** or **B**" is adopted. But the definition of the negation operations are different depending the models, the algorithms are different.

The operations of each model are distinguished by the symbols of negation operation.

Hereafter, the hyper-logical space (i.e. [0,1]×[0,1] fuzzy logical space) is used for mathematical explanations and graphical presentations of both models.

The implication operations for HLS and A-IFS are as follows respectively.

$$\neg\mathbf{A} \vee \mathbf{B} = \{max(1\text{-}ta, tb), min(1\text{-}fa, fb)\} \tag{17}$$

$$\sim\mathbf{A} \vee \mathbf{B} = \{max(fa, tb), min(ta, fb)\} \tag{18}$$

4. ANALYSIS 1:

The case in which the result of HLS return to the IFS area.

s

In the HLS model, not only the data in the A-IFS area (t+f≦1), but also the data in the contradiction area (t + f> 1) can be treated while the A-IFS model can only treat the data in the irrelevance area (t+f<1) and the data in the numerical truth-value area (t+f=1). The authors compare the results obtained by each model to prove the superiority of the HLS model when the data A and B are both in the A-IFS area, because the A-IFS model is impossible to treat if a data is in the contradiction area. The constraints for A and B assumed here are as follows.

$$ta+fa \leqq 1,\ tb+fb \leqq 1 \tag{19}$$

In this chapter, first, under the conditions of (6), the conditions are clarified in which the result of implication operation ¬ **A** ∨ **B** by the HLS model returns into the A-IFS area. Then, the properties of both models' implication operators are compared.

4.1 Analysis of the result of inclusion operation by HLS model

The inclusion formula (Eq. 17) by the HLS model can be classified into four cases by its parameter values.

Case 1 : (1-ta≧tb) & (1-fa<fb),

$$\neg A \vee B = \{1\text{-}ta,\ 1\text{-}fa\} \tag{20}$$

----- coincident with ¬A

Case 2 : (1-ta≧tb) & (1-fa≧fb),

$$\neg A \vee B = \{1\text{-}ta,\ fb\} \tag{21}$$

Case 3 : (1-ta<tb) & (1-fa<fb),

$$\neg A \vee B = \{tb,\ 1\text{-}fa\} \tag{22}$$

Case 4 : (1-ta<tb) & (1-fa≧fb),

$$\neg A \vee B = \{tb,\ fb\} \tag{23}$$

----- coincident with B

(1) About the Case 1: Since the result of this case does not match with prerequisite condition (6), it is not treated in this chapter.

(2) About the Case 2: The restriction *(1-ta≧tb)&(1-fa≧fb)* can be transformed to

(ta≦1-tb)&(fa≦1-fb). (24)

The pentagonal area with hatching in Fig. 5 is showing the possible existence zone of the point **A** in which the point satisfying the restriction. In this figure, the hatched area means the relative location of **A** if the location of the point **B** was given.

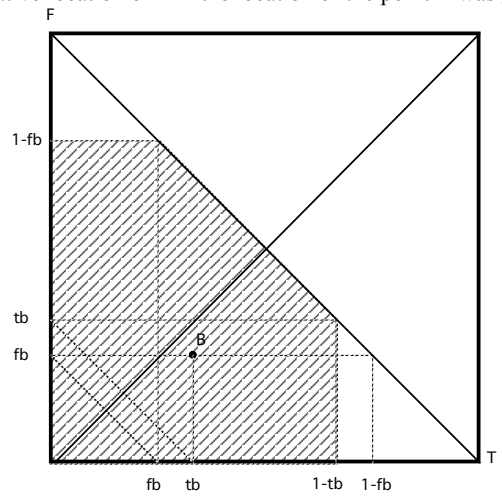


Figure 5: The zone in which point A satisfies the restriction of the Case 1

In Case 2, the condition in which the result of ¬**A**∨**B** returns to the A-IFS domain is $1 - ta + fb \leqq 1$. It means $ta \geqq fb$. Then, the result is shown as a trapezoidal area with hatching in the Fig. 6, as a relative location of **A** when **B** was given.

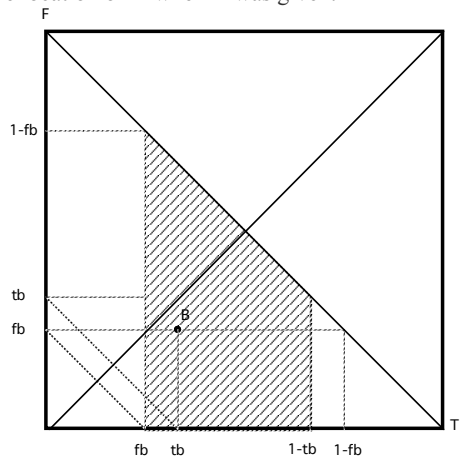


Figure 6: The zone that the point A satisfies the restriction of Case 2

(3) About Case 3: The condition (1-ta<tb)&(1-fa<fb) can be transformed to (ta>1-tb)&(fa>1-fb).
It is clear that the area which fulfilling the condition does not exist.
Proof:
From the condition,

$$(ta>1-tb)\&(fa>1-fb) \tag{25}$$

Then,

$$ta+fa>(1-tb)+(1-fb)$$

$$ta+fa>2-(tb+fb) \tag{26}$$

By the way, since **B** is a point in the A-IFS area,

$$tb+fb \leqq 1 \tag{27}$$

therefore $ta+fa>1$. (28)

This inequality is contradicting to the assumption that "**A** is existing in the A-IFS area".
Q.E.D.

(4) About Case 4: From the condition $(ta<tb)\&(1-fa \geqq fb)$, obtain

$$(ta>1-tb)\&(fa \leqq 1-fb). \tag{29}$$

As the result of the implication operation is just the same to the point **B**, it is always in the A-IFS area. In Fig. 7, the hatched triangle area is illustrating the possible existing zone of **A** satisfying the result of Case 4 as a relative location if **B** was given.

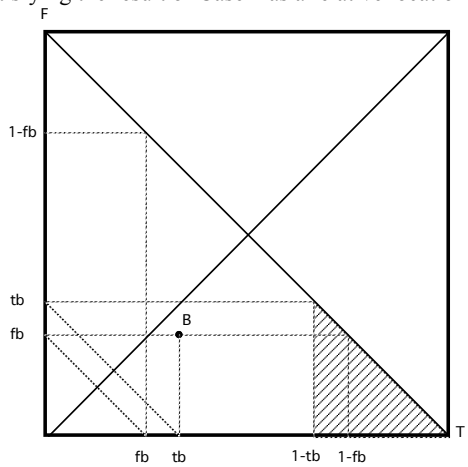


Figure 7: The zone that the point A satisfies the restriction of Case 4

Both ¬ **A** ∨ **B** and **B** fit into the IFS area without problems.

(5) Summarize the four cases: From Case 2, $1-fb \leqq ta < 1-tb$ therefore

$$\neg \mathbf{A} \vee \mathbf{B} = (1-ta, fb) \tag{30}$$

From Case 4, $ta \geqq 1-tb$ therefore

$$\neg \mathbf{A} \vee \mathbf{B} = (tb, fb) \tag{31}$$

In these cases, any result of the implication operation **A** → **B** returns into the A-IFS area.
In summary, the results of implication operation by the HLS model are shown as below.

$$\neg A \vee B = \begin{cases} (1-t_a, f_a) & if\ f_b \leq t_a < 1-t_b \\ (1-t_b, f_b) & if\ t_a \geq 1-t_b \end{cases} \tag{32}$$

4.2 Analysis of the results of implication operation by A-IFS model

The result of implication operation by A-IFS model is expressed as below.

$$\sim A \vee B = \left\{ \max(f_a, t_b), \min(t_a, f_b) \right\} \tag{33}$$

Eq. (33) can be divided into four cases by the conditions.
Case i :

$$if\ f_a \geq t_b\ and\ t_a \geq f_b\ then \sim A \vee B = (f_a, f_b) \tag{34}$$

Case ii :

$$if\ f_a \geq t_b\ and\ t_a < f_b\ then \sim A \vee B = (f_a, t_a) \tag{35}$$

Case iii :

$$if\ f_a < t_b\ and\ t_a \geq f_b\ then \sim A \vee B = (t_b, f_b) \tag{36}$$

Case iv :

$$if\ f_a < t_b\ and\ t_a < f_b\ then \sim A \vee B = (t_b, t_a) \quad (37)$$

These four cases are obtained through analyzing the relative position of the point **A** to the point ~ **B**. (See Fig. 8)
In these, only Case iii is special, because the result of the implication operation is just overlap with **B**.

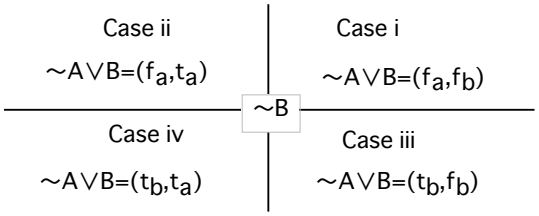


Figure 8: The four cases for analyzing the results of implication operation by the A-IFS model. The figure is showing the relative geometrical position of the point **A** to the point ∼**B** for each case.

4.3 Evaluation of the result of implication operations of each model

In summary, the results of implication operations of both IFS and HLS models are expressed together in one figure.

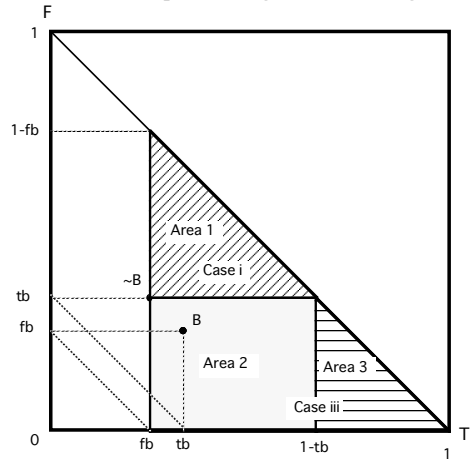


Figure 9: The results of implication operations by both IFS and HLS models

The areas to be analyzed are separated into 3 parts: i.e. Area 1 to Area 3.

Area 1: $\neg\mathbf{A}\vee\mathbf{B}=(1\text{-}ta, fb)$ -------- Interval = [1-ta,1-fb] (38)
$\sim\mathbf{A}\vee\mathbf{B}=(fa, fb)$ --------- Interval = [fa,1-fb] (39)
Area 2: $\neg\mathbf{A}\vee\mathbf{B}=(1\text{-}ta, fb)$ ---------Interval = [1-ta,1-fb] (40)
$\sim\mathbf{A}\vee\mathbf{B}=(fa, fb)$ ----------Interval = [fa,1-fb] (41)
Area 3: $\neg\mathbf{A}\vee\mathbf{B}=(1\text{-}ta, fb)$ ---------Interval = [1-ta,1-fb] (42)
$\sim\mathbf{A}\vee\mathbf{B}=(tb, fb)$ ----------Interval = [tb,1-fb] (43)

In the area 1, the right edges of the intervals have the same value *1-fb* for both models.
But, the left edges of the intervals have different values *1-ta* and *fa* respectively.
As the values *1-ta* and *fa* can vary under the restrictions $ta+fa\leq 1$ and $fa\geq tb$.
If and only if, the intervals are the same if *1-ta=fa*.
It is not known which model is better in this area.
In the area 2, the right edge of the interval is the same as bellow.
The left edge of ~ **A** ∨ **B** is a fixed value *tb*, ¬**A**∨**B** is *1-ta*, because Area 2 has a range of areas $fb \leqq ta \leqq 1\text{-}tb$, so transform $ta \leqq 1\text{-}tb$, obtain $tb\text{-}ta \geqq tb$. The interval of ¬ **A** ∨ **B** is narrow.
If *1-ta=tb*, then the range of the intervals are the same.
In the Area3, the results of the implication operations by each model are just the same point **B**.

5. ANALYSYS 2: The case that the result of the implication operation by the HLS model does not return to the A-IFS area.

In this case, the results of the implication operation belong to different areas by models. If the result of the implication by the HLS model is in the contradiction area, it is out of the framework of the A-IFS model, because it is out of the data area of A-IFS. Commonly considering, it is impossible to compare them directly. On the other hand, from the view point of FCR-method or HLS model, it is not so difficult to compare, because the resultant pair (t, f) will be integrated when the result is used in some inference system of application, since the paired data is not so easy to understand or use directly.
So, in this section, the results of operations are compared in one-dimensional criteria by calculating the integrated values. As to calculate the integrated value of the resultant points, the "combined scoring method 1" (symbol I_4) developed for the FCR-method is applied (Refer to Eq. 4).

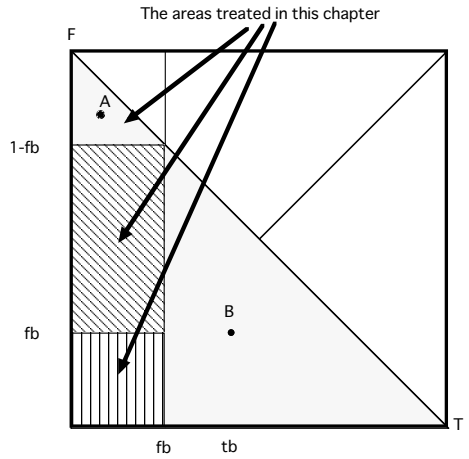


Figure 10: The areas where the results of the implication by HLS do not return to the A-IFS area.

5.1 Comparing the integrated values of the results calculated by each model

For example, when **A** locates at the upper left of **B**, substitute the result of ¬ **A** ∨ **B** in the formula I_3, and substitute the result of ~ **A** ∨ **B** in the formula I_1, then analyze the magnitude relation of the integrated values. The authors illustrate the result of the analysis. (See Fig.11)

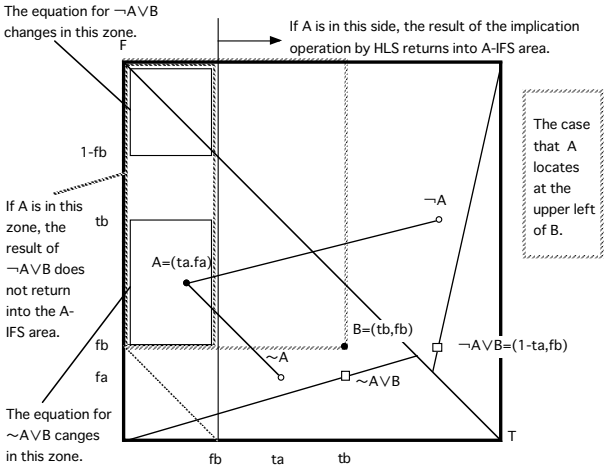


Figure 11: Summary of the results by both models
(The result by HLS does not return to the A-IFS area)

If the point **A** is in one of the three zones, the result of the implication operation does not return to the A-IFS area. The integrated value I_3 is equal to the value I_1 if the point **A** is in the top zone, while the values I_3 and I_1 are not consistent in the middle and the bottom zone. As illustrated in the Fig. 11, I_3 is greater than I_1. As seen as that, the integrated value of $\neg \mathbf{A} \vee \mathbf{B}$ is greater than $\sim\mathbf{A} \vee \mathbf{B}$, so the HLS is superior to the A-IFS in this case.

6. CONCLUSION

The HLS model can treat both contradiction area data and the A-IFS area while the A-IFS model precludes the contradiction area data by definition. The authors assumed that even if the original (observed) data sets are within the A-IFS area, if it is permitted to use the outside area of the definition of A-IFS, the result of the implication operation could be better than restricting the calculation space. Under such assumptions, this study analyzed the results of the implication operations by two extended fuzzy logic model, A-IFS and HLS, though both their data area is different. By classifying various cases, it is attempted to compare the results for testing the superiority of the models.

•For defining the implication equation $\mathbf{A} \rightarrow \mathbf{B}$, adopt

$$(\text{Not } \mathbf{A}) \vee (\mathbf{B}) . \qquad (44)$$

•For evaluating the result of implication, the same integration formula I_4 is applied.

(The I_4 can be used both the irrelevance area and the contradiction area by its symmetric feature. As it is a surjective function, it can be used for the inverse FCR-method proposed by E.Takahagi [9].)

Assuming that points **A** and **B** are in the A-IFS area, both of the resulting points appear inside and outside of A-IFS area were analyzed. It became clear that the result of HLS model is out of A-IFS area when a point **A** is at the left of a point **B**.

By setting three patterns for the results of implication are out of A-IFS area, followings became clear. In one pattern, the integrated value $I_3 = I_1$, in the other patterns, $I_3 > I_1$. By summarizing all of this investigation, it can be concluded that about the implications, the HLS model is showing superiority to the A-IFS model in most cases. But, in one case, Area 1 illustrated in the Fig. 9, A-IFS model can be superior to the HLS model depending on the values of the parameters of **A** and **B**.

According to the procedures of using the special integration algorithm, Analysis 2 could be seen as an attempt. But, at least Analysis 1 would be enough to demonstrate the usefulness of using the contradiction area for the logical calculation space.

In this paper, the only model compared to the A-IFS model is HLS. Regarding the definitions of the logical operations, logical OR and logical AND operations are the same in both models, but there can be better operations. For example, F. Smarandache [14] introduced the algebraic sum and algebraic product rules for the definitions as the logical OR and logical AND operations of his extended fuzzy/set logic model named "Newtorosophic Set/Logic". The definitions can be applied by using the method used in this paper. In the near future, such an investigation should be planned for exploring and constructing better fuzzy systems.

REFERENCES

[1] D. Dubois and H. Prade, "Interval-valued Fuzzy Sets, Probability Theory and Imprecise Probability", **Proceedings of EUSFLAT2005** (online).

[2] K. Atanassov, **Intuitionistic Fuzzy Set Theory and Applications**: Physica Verlag, 1999, pp.1-60.

[3] T. Oda, "Proposal of multi-dimensional multi-valued logic", **Journal of Japan Industrial Management Association**, Vol. 49, No. 3, 1998, pp.135-145.

[4] G. Takeuti and S. Titani, "Intuitionistic fuzzy logic and intuitionistic fuzzy set theory", **Journal of Symbolic Logic**, Vol. 49, 1984, pp.851-866.

[5] M. Mukaidono and H. Kikuchi, "Proposal of Interval Fuzzy Logic," **Journal of Japan Society for Fuzzy Systems**, Vol.2, No.2, 1990, pp.97-110

[6] T. Oda, A. Oba, R. Cheng, X. Huang, "Comparative Study of the Intuitionistic Fuzzy Set and the FCR-method", **Proceedings of Fuzzy System Symposium**, Vol. 23, 2007, WC1-4.

[7] T. Oda, J. Deng, T. Kimura and F. Hayashi: "Hyper Logic Space (HLS) Model and its Applications: Multidimensional Logic System as an Extended Fuzzy Logic System and a New Psychological Measurement Method", **Proceedings of the Third International Conference on Systems Science and Systems Engineering**, 1998 (Beijing), pp.128-135.

[8] J. Deng, T. Oda, M. Umano," Fuzzy Logical Operations in the Two-dimensional Hyper Logical Space Concerning the Fuzzy-set Concurrent Rating Method", **Journal of Japan Association for Management Systems**, Vol. 17, No. 2, 2001, pp.33-42.

[9] E. Takahagi, "Fuzzy Measures and Fuzzy-set Concurrent Rating Methods: Proposal of Inverse φs Transformation Method and Comparisons among Fuzzy-set Concurrent Rating Methods", **Journal of Japan Society for Fuzzy Theory and Intelligent Informatics**, Vol. 16, No. 1, 2004, pp.80-87.

[10] T. Oda, "Fundamental Characteristics of Fuzzy-set Concurrent Rating Method", **Journal of Japan Association for Management Systems**, Vol. 12, No. 1, 1995, pp.23-32.

[11] F. Hayashi and T. Oda, "The five-factor model of personality traits: An examination based on fuzzy-set theory", **The Japanese Journal of Psychology**, Vol. 66, No. 6, 1996, pp.401-408.

[12] Japan Society for Fuzzy Theory and Intelligent Informatics ed. , **Fuzzy and Soft-computing Hand-book**, Kyoritsu Shuppan, 2000.

[13] O. Montiel, O. Casillo, P. Melin, R. Sepulveda, "Human Evolutionary Model: A New Approach to Optimization", **Information Sciences,** 177, 2007, pp. 2075-2098.

[14] F. Smarandache, "Newtrosophic Set – A Generalization of the Intuitionistic Fuzzy Set", 2002 (online)

Management of Tracking Information of Digital Content for an Internet Inaccessible Environment

Seung-Won Lee, Choong-Bum Park, Eun-Ji You, Kyung-Min Park, Hoon Choi
Department of Computer Science and Engineering, Chungnam National University
Daejeon, Korea
{ seungwon, here4you, qutywing, km-park, hc }@cnu.ac.kr

Abstract

This paper describes a feature of mobile device middleware which distributes content files in a peer-to-peer manner using a wireless interface. The middleware manages the tracking information of digital content. It collects the information in a best-effort manner from devices which are not connected to the Internet. It then uploads the content tracking information to a server which is interested in the information for its business purpose. The paper introduces a method of managing and synchronizing the content tracking information. A simulation study was performed for verification of the proposed method and also for performance evaluation.

Keywords- mobile device peer-to-peer data transmission, middleware, content tracking information, synchronization

I. Introduction

Many of the software for PCs used today propagate through the Internet. With the agreement of the user, the usage information of the software and the system is collected and uploaded to a content server (CS) [1][2]. The usage information shows the preferences of the user and/or the popularity of software, and this information is used to increase the quality of the software [3][4][5]. Servers that provide content for public welfare and/or advertising purposes may want to obtain information regarding, for example, how their content is being propagated. The data gleaned from this process is called content tracking information (CTI). CTI may be managed for each content. It shows how many users obtained a certain content how many times. It may also show the migration path of the content, i.e., how the content is transferred to other devices in a network.

In some environments, the Internet is occasionally inaccessible for mobile devices when the surrounding wireless network condition becomes unreliable. Mobile devices in an area inaccessible to the Internet cannot download content from servers. Consequently, it is useful to exchange content files in a peer-to-peer manner between devices by WiFi or Bluetooth as shown in Figure 1 [6]. For this purpose, neighboring mobile devices are able to establish a MANET (Mobile Ad-hoc Network) and join and leave the MANET freely. When two devices in the same MANET are unable to transmit data directly because they are outside the signal transmission range, data communications is still possible by other devices in between by relaying the signal [7].

Not many studies have been reported about the CTI management, especially for mobile environment. In [5], a web server records URL of the downloader site whenever the downloader copies some content from the web server. DC Tracker [8] uses similar approach. In [9], a tracking technology for protecting unauthorized access of Internet content is investigated.

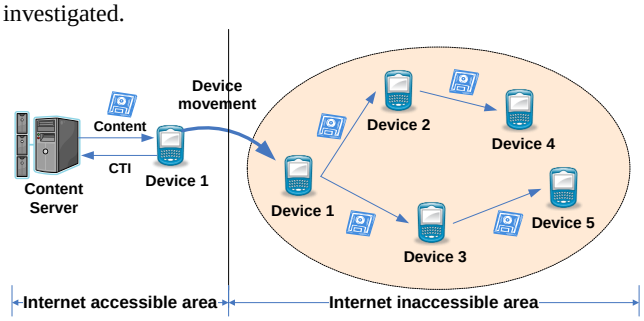


Figure 1. Propagation of content in a mobile environment

This paper proposes a method of collecting, generating and synchronizing CTI in a mobile computing environment with Internet *inaccessible* areas. By this method, content servers are able to collect CTI as much as possible from the devices which are not directly connected to the Internet. To the authors' knowledge, this is the first contribution on this subject.

Section 2 explains the requirements for managing a CTI and defines the format of a CTI, and Section 3 describes the CTI management method. In Section 4, a simulator to verify the proposed method is introduced along with its simulation results. Overhead of the proposed method is also evaluated. Finally concluding remarks are given in Section 5.

II. Content Tracking Information

Users of mobile devices which belong to the same MANET may have different interests; therefore, devices have different categories of content. In order to identify a neighboring device with the same category of content, each device must have a profile which includes user identification information, a list of content and the usage information of the content. By exchanging these profiles among mobile devices, a device can determine if it must send or receive content from a counterpart device. Therefore, the mobile devices participating in a MANET must periodically discover the existence of neighboring devices and then exchange profiles.

A mobile device must be able to transfer the content that it possesses to other devices participating in the MANET. Additionally, when particular content is transferred to a neighboring device, a new CTI must be generated and the generated CTI must be added to the CTI list, which is a collection

of CTIs. The copy of the updated CTI list is then transferred to the neighboring device along with the content.

CTI synchronization must occur so that each mobile device acquires as many CTIs as possible. Therefore, any device which is able to access the Internet can upload as many CTIs to the CS as possible. Detailed description about the synchronization is given in the next section.

A CTI represents a unit of path information that is stored with corresponding content to represent which device from which the content originated. The structure of the CTI is shown in Figure 2.

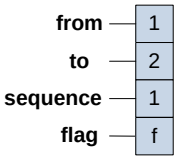


Figure 2. CTI format

The 'from' field contains the identifier of the device which provides the content, and the 'to' field shows the identification of the device which receives the content. The 'sequence' field is the number of times that the 'from' device has provided this content to other devices. The 'flag' indicates whether the CTI has been synchronized with the CTI in the CS.

The current structure of a CTI contains the minimum information needed to track only the migration path. Additional fields can be added to store other usage information such as the number of times that the content has been used or the location where a content transfer has taken place.

III. CTI MANAGEMENT

The CTIs which were previously uploaded to the CS should be deleted from the device to minimize the memory used by the CTIs. Finally, when the device returns to an Internet-inaccessible area from an Internet-accessible area after transmitting the CTI to the CS, the device synchronizes its CTI with the CTI of neighboring devices so that the CTI is also deleted from the neighboring devices. This will avoid the continuous expansion of memory that would otherwise be needed to store the CTI.

A. *CTI generation*

A CTI is generated whenever the content is sent to a neighboring device. The CTIs form the CTI list and then the list is sent to a neighboring device along with the content. The CTI list exists in every device as well as in each instance of the content.

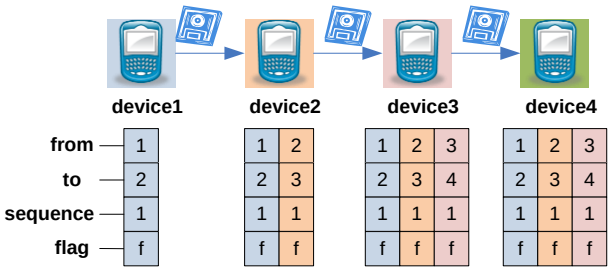


Figure 3. An example of a MANET with four devices

Figure 3 shows the process through which the generated CTI is added to the CTI list when the same content is distributed through device1, device2, device3 and device4, for example. In the figure, all of the values of the 'sequence' field are 1 as each device provided the content once.

The 'to' field of the CTI list of device1 in Figure 4 informs a user that device1 has provided this content in the order of device2, device4, and device6. According to the CTI list of device7, device7 received the content from device6, and device6 received this content from device1. Here, the value of the 'sequence' is 1 because this was the first time that device6 transferred the content.

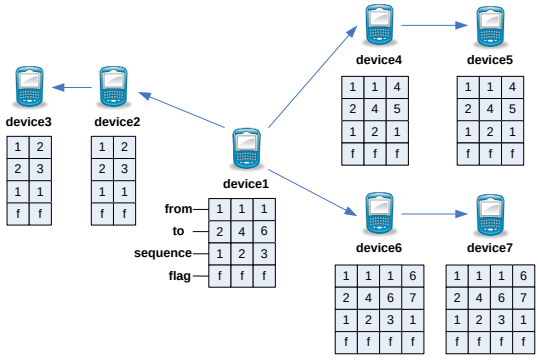


Figure 4. Device discovery and the CTI generation process

B. *CTI synchronization with CS*

If a certain device is in an Internet-accessible area, the device connects to the CS and uploads the CTI list. The CTI list, which was transferred to the CS in Figure 5, shows that there are three different values of 'from' (1, 4, and 6). When the transfer of the CTI list to the CS is complete, the corresponding device deletes all of the CTIs except for the CTI with the largest sequence value from each source (the 'from' field) from its CTI list. The values of the 'flag' of the remaining CTIs are then t (true) to indicate that the corresponding CTI and the CS are synchronized at this point.

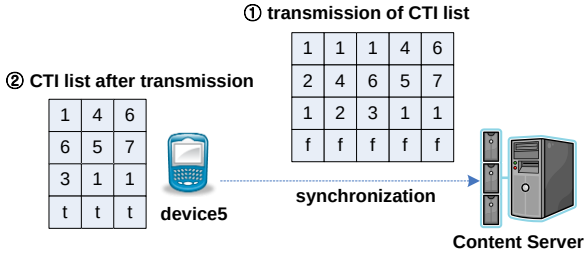


Figure 5. CTI synchronization with CS

How to utilize the CTI list for a CS is implementation dependent. A server may be interested in details such as who received the content from whom for each content and how many times a content was distributed in a day when we expand the CTI to augment a timestamp. Or it may be simply interested in how many devices use the content.

C. *CTI synchronization between devices*

When mobile devices with the same content can connect to each other, they synchronize the CTI list of all content they possess. This is done so that each device acquires as many CTIs as possible and so that all devices upload as many CTIs to the CS as possible when they access the Internet. Synchronization between mobile devices will result in a superset of the CTI lists

of each device. Here, the CTI list is grouped based on the value of the 'from' and 'sequence' fields.

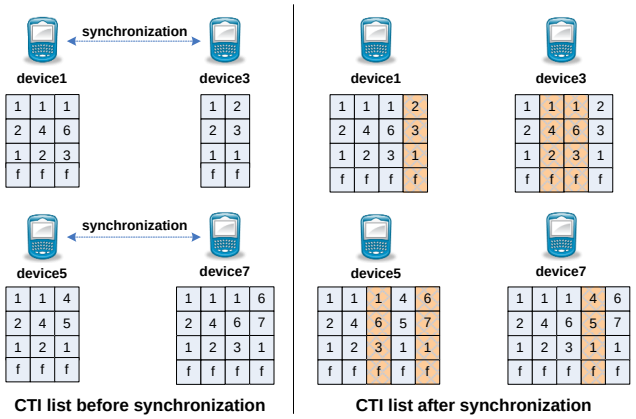


Figure 6. CTI synchronization between mobile devices

D. *CTI synchronization between devices – after synchronization with the CS*

When device5 in Figure 5 connects to other devices after synchronizing with the CS, it informs the neighboring devices that its CTI list has been synchronized with the CS and that the CTIs which have been uploaded to the CS need to be removed from the CTI list of the neighboring device. This procedure is depicted in Figure 7. Device11 checks the CTI list of device5. In part A in Figure 7, the value of 'from' is 1 and that of 'sequence' is 3. Additionally, the 'flag' is true; therefore, among the CTI elements with a 'from' value of 1 in the CTI list of device11, those with a 'sequence' value of less than 3 are deleted from the list, and the CTI with a 'sequence' value of 3 changes the 'flag' value to true.

The CTIs received from device11 are also added to the CTI list of device5, as shown in part B in Figure 7.

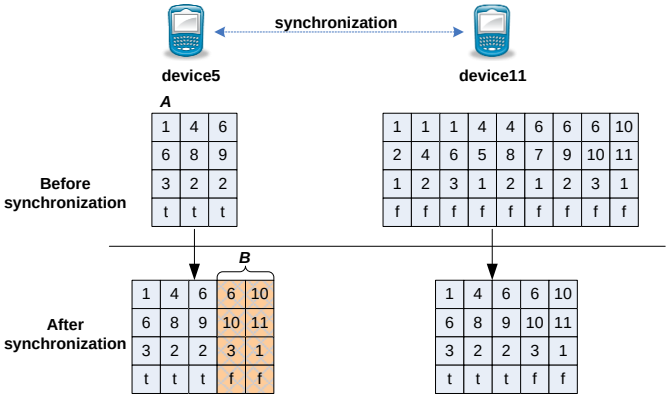


Figure 7. CTI synchronization between mobile devices after synchronization with the CS

With the synchronization algorithms described thus far, the effect of the synchronization between device5 and the CS will spread to neighboring devices, and other devices within the MANET will gain the benefits of the update without directly synchronizing with the CS.

IV. VERIFICATION and EVALUATION

A middleware which applies the previous method was implemented and tested on desktop computers. However a simulation was performed in order to verify logic of the proposed CTI management method with a reasonable number of mobile devices. A simulator was implemented and tests were carried out with a different number of mobile devices. The following functions can be configured in the simulator:

- Magnitude of MANET (number of fields)
- Position of the CS (Content Server)
- Number of mobile devices
- Printing of the CTI list of each mobile devices
- Content list provided by the CS
- Printing of the CTI list collected by the CS

The simulator provides a graphical user interface which shows the parameters above and a grid section. The small rectangle in the grid represents an area in which a mobile device can communicate with other mobile devices or servers through its wireless communication interface. When the simulation begins, all the device move to an adjacent grid (area) randomly at each simulation step, and each device attempts to locate neighboring devices in the same area. The devices located in the same area will discover each other and begin to communicate by their wireless interfaces, exchanging content and CTIs. A device will connect to the CS when it moves to the area of CS. It will then uploads its CTI list and download the content, if new content exists.

<Table 1> Example of collected CTI list

Information of device[0]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (2, 8, 2, false) (2, 6, 3, false) (6, 1, 1, false) (1, 0, 1, false)
Information of device[1]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (2, 8, 2, false) (2, 6, 3, false) (6, 1, 1, false) (1, 0, 1, false)
Information of device[2]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (2, 8, 2, false) (2, 6, 3, false)
Information of device[3]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (2, 8, 2, false) (2, 6, 3, false) (6, 1, 1, false) (6, 3, 2, false)
Information of device[4]
Content_0=> (4, 2, 1, false)
Information of device[5]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (5, 9, 1, false) (5, 7, 2, false)
Information of device[6]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (2, 8, 2, false) (2, 6, 3, false) (6, 1, 1, false) (6, 3, 2, false)
Information of device[7]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (5, 9, 1, false) (5, 7, 2, false)
Information of device[8]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (2, 8, 2, false)
Information of device[9]
Content_0=> (4, 2, 1, false) (2, 5, 1, false) (5, 9, 1, false)

The CS of the simulator collects the CTIs from mobile devices and keeps track of the migration path of each instance of content. Table 1 shows the information as collected by the CS;

this is comprised of the CTIs that were generated and managed by 10 different mobile devices. For example, the CTI list of device9 shows that device4 transmitted the content to device2, it then went from device2 to device5, and then from device5 to device9. Using this information, a content transfer path can be drawn. An example is shown in Figure 8.

Experiments have been performed with different number of fields and devices and the simulation confirmed that the proposed CTI list management and synchronization algorithms operate correctly.

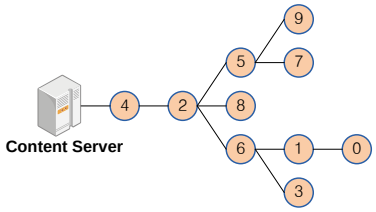


Figure 8. Content migration path

Performance evaluation of the CTI management method was also carried out by the simulator. It is assumed that a CS in an IAA has a digital content and, after one mobile device downloads the content from the CS, all the remaining mobile devices in this experiment obtain this content from other mobile device in a peer-to-peer manner. Number of simulation steps was measured from the moment a mobile device downloads the digital content until the moment CS collects all the CTIs regarding this content. Parameters of this simulation are as follows:

<Table 2> Simulation parameters

No. of Contents	No. of Fields	No. of Devices
1	400	5
1	400	10
1	400	15
1	400	20

Three different experiments were carried out. In the first one, labeled "CTIM", all the synchronization method explained in Section 3 were applied. In the second one, labeled "Test A", the synchronization method of Section 3 except the CTI synchronization between devices, CTI synchronization between devices after synchronization with the CS were applied. In the last experiment, labeled "Test B", the synchronization method of this paper was not applied. In Test B, each device has only one CTI which describes where this content is come from. CS needs to collect this CTI from each of the mobile devices of the experiment.

Figure 9 shows the results of simulation in terms of the number of simulation steps taken from the moment a mobile device downloads the digital content until the moment CS collects all the CTIs regarding this content. The values are mathematical mean of results of 50 simulation runs.

As the number of mobile devices increases, simulation steps of Test A becomes about 16 times longer than that of CTIM case. In Test B, it takes 10~50 times longer than CTIM case. It means that CS is able to collect tracking information of a digital content much sooner by having CTIM method, proposed in this paper. The reason CTIM works much faster is that any mobile device which moves into IAA and synchronizes with CS uploads tracking information as much as possible, including other devices' tracking information obtained by the method explained in Section 3.C.

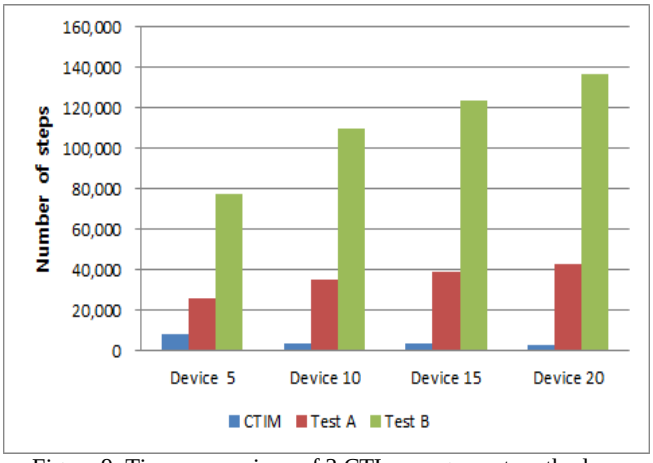


Figure 9. Time comparison of 3 CTI management methods

It is clear that the proposed method have advantage regarding the time. On the other hand, there is a disadvantage. Overhead of this method is that mobile devices need to store other devices' tracking information, resulting in additional storage consumption. However, as mentioned in Section 3.D, this additional space is released later. Figure 10 is a part of log message from the simulator and shows that how the additional space is minimized. When a device, device9 in this figure, synchronizes with other device (device 5 in this figure) which has 'flag' field of CTI set to 'true', it removes unnecessary CTIs from the CTI list. Considering that tracking information is a short text information, authors think this overhead is not critical to a modern electronic devices.

```
<  Before synchronization : Device[5] and Device[9] >
+device[5]'s CTI =
  -[content_0](7, 2, 2, true)(9, 6, 5, true)(1, 0, 1, true)(2, 4, 1, true) => [length:4]
  -Device Step :1090
+device[9]'s CTI =
  -[content_0](7, 9, 1, false)(7, 2, 2, false)(9, 1, 1, false)(9, 3, 2, false)(9, 8, 3, false)(9, 5, 4, false)(9, 6, 5, false)(1, 0, 1, false)(2, 4, 1, false) => [length:9]
  -Device Step :1353
<   After synchronization : Device[5] and Device[9] >
+device[5]'s CTI =
  -[content_0](7, 2, 2, true)(9, 6, 5, true)(1, 0, 1, true)(2, 4, 1, true) => [length:4]
  -Device Step :1090
+device[9]'s CTI =
  -[content_0](7, 2, 2, true)(9, 6, 5, true)(1, 0, 1, true)(2, 4, 1, true) => [length:4]
  -Device Step :1353
```

Figure 10, Reducing CTI list by the method of Section 3.D

V. CONCLUSION

In this paper, a method of generating and managing a CTI was proposed. Proposed CTI synchronization method was verified through a simulation.

One issue of this method might be scalability, i.e., a CTI list may become long and consumes large amount of memory space of mobile device. This problem can be avoided by limiting the maximum length of CTI list. When a CTI list reaches the

maximum, then, by depending on the implementation, the middleware may stop increasing the list in the case A or C of Section 3. It does not cause any problem for mobile devices or the CS because the proposed method works on best-effort basis, i.e., this method collects CTI from devices in an Internet inaccessible area on behalf of CS as much as possible, not necessarily all of the CTI. Without this method, the CS is not able to obtain content tracking information at all from devices which are not connected to the Internet.

ACKNOWLEDGEMENTS

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2011-C1090-1111-0008)

REFERENCES

[1] http://help.adobe.com/en_US/Dreamweaver/10.0_Using/WSc78c5058ca073340dcda9110b1f693f21-79cca.html - Accessed Aug. 20, 2009.

[2] Katchabaw M.J., Lutfiyya H.L., and Bauer M.A., "Usage-based Service Differentiation for End-To-End Quality of Service Management," 2003. Proceedings of the 2003 IEEE International Performance Computing and Communications Conference, pp.513-520, April. 2003.

[3] Jongyi Hong, Eui-Ho Suh and Junyong Kim, "Context-aware System for Proactive Personalized Service based on Context History," Expert Systems with Applications, Vol.36, No.4, pp.7448-7457, May. 2009.

[4] Yamato Y., Ohnishi H., and Sunaga H., "Study of Service Processing Agent for Context-Aware Service Coordination," Services Computing, 2008. SCC '08. IEEE International Conference on Volume 1, July. 2008.

[5] Tracer, http://www.trendbird.co.kr/1885 - Accessed Feb. 10, 2010.

[6] Choong-Bum Park, Byung-Sung Park, and Hoon Choi, "Service Management Middleware for Mobile Devices in a MANET Environment," The 13th IEEE International Symposium on Consumer Electronics, May. 2009.

[7] Chlamtac, M. Conti, and J.J.–N. Liu, "Mobile Ad Hoc Networking: Imperatives and Challenges," Ad Hoc Networks, Vol.1, No.1, pp.13–64, July 2003.

[8] Data Conversion Laboratory, DC Tracker (Digital Content Tracker), https://home.dclab.com/demo/dcdemo1.asp - Accessed Jan. 4, 2010.

[9] J-H. Hsiao,C-H. Li,C-Y. Chiu,J-H. Wang,C-S. Chen and L-F. Chien, "Effective Content Tracking for Digital Rights Management in Digital Libraries," Lecture Notes in Computer Science, Vo. 4172, pp.415-425, 2006.

Unemployment in USA mathematical modeling

Viktor V. Ivanov
Member Emeritus of AMS
United States of America

and

Valentina N. Korzhova, Malik F. Saleh
Management Information Systems Department, Prince Mohammad Bin Fahd University
Kingdom of Saudi Arabia

ABSTRACT

General mathematical theory of evolutionary system developed earlier is implemented to understand the unemployment in USA.

Certain ways to overcome this problem are based on investigation of the minimization of unemployment with regard to mathematical models of economics development.

Keywords: Evolutionary system; Mathematical model; Optimization problem; Unemployment problem; Work place

1. INTRODUCTION

A general mathematical theory of development of evolutionary systems (ES) and its various mathematical models (MM) start from the works [1], [2], and can be seen in monographs [1] - [3].

The present paper introduced additional restriction on MM so that the respective ES have the property of paying for themselves. We apply all this theory to understand reasons for too much unemployment in USA and suggest certain ways to correct that problem.

2. THE BASE MM OF ES

The basic minimal or simplest MM has the form

$$
\begin{aligned}
&m(t) = \int_{a(t)}^{t} \alpha(t,s)y(s)m(s)ds,\\
&0 \le y \le 1, 0 \le a(t) \le t, \alpha \ge 0,\\
&c(t) = \int_{a(t)}^{t} \beta(t,s)(1-y(s))m(s)ds, \beta \ge 0,\\
&R(t) = \int_{a(t)}^{t} m(s)ds, M(t) = \int_{0}^{t} m(s)ds,\\
&G(t) = M(t) - R(t),\\
&f(t) = m(t) + c(t), t \ge t^*.
\end{aligned} \tag{1}
$$

where $m(t)$ is the rate of creation of the first kind new generalized product (resource) quantity at the time instant t, which provides the fulfillment of the internal functions of ES, that is, restoration of itself and creation of the second kind product; $y(t)m(t)$ is a share of $m(t)$ for fulfillment of internal functions in the subsystem A of restoration and perfection of the system as a whole; (t, s) is the efficiency index for functioning of the subsystem A along the channel $(t, s)y(s)m(s) - m(t)$, i.e., the number of units of $m(t)$ created in the unit of time starting from the instant t per one unit of $y(s)m(s)$; $a(t)$ is a special temporal bound: the new product creating before $a(t)$ is never used at the instant t, but created after $a(t)$ is used entirely; $c(t)$ is the rate of creation of the second kind new generalized product quantity at the instant t, which provides the realization of the external functions of ES;$[1\mathrm{y(s)}]m(s)$ and

(t, s) are similar to ym and respectively but for the subsystem B of creation of the second kind product; $R(t)$ is the total quantity of the first kind product functioning at the instant t; $M(t)$ is the total quantity of the first kind product to be created during the time $t = 0$; $G(t)$ is the total quantity of the obsolete product at the instant t; $f(t)$ is the rate of the resource inflow from the outside ($m(t)$ and $c(t)$ are measured in the units of $f(t)$; t^* is the starting point for modeling; $[0, t^*]$ is the prehistory of ES, for which all the functions are given (their values will be noted by the same symbols but with the sign "*", e g, $m(t) = m^*(t)$, t $[0, t^*]$). It is obvious that all the relations (1) are faithful representations by definition. In a general case, the indices and depend on m, c, a, y, R, M, G, and f.

Anyone can see that (1) consists of 7 equalities and 7 inequalities connecting 14 values, namely: m, c, , y, ,$1 - y$, a, R, M, G, t, t^*, f, 0, all of which are nonnegative. Usually, α, β, y, f, and/or R are given, and the others are to be found. Even in the simplified formulation, MM (1) is the system of nonlinear functional relations, in which along with the nonlinear integral equation of the unusual form (the lower bound $a(t)$ can be unknown function) we have the system of functional inequalities.

We introduce here the additional restriction:

$$f(t) = m(t) + c(t) = k(t)c(t),$$
$$k(t) > 1, \quad (2)$$

where k is price of the unit c, meaning that all resources are obtained at the expense of the c price. The base simplest self-organized ES has the following MM:

$$\alpha'(t) = \int_{a(t)}^{t} \alpha(s)x(s)m(s)ds,$$
$$m(t) = \int_{a(t)}^{t} \alpha(s)y(s)m(s)ds,$$
$$c(t) = \int_{a(t)}^{t} \alpha(s)z(s)m(s)ds,$$
$$0 \le x, y, z \le 1, x + y + z = 1,$$
$$f(t) = \alpha'(t) + m(t) + c(t), t \ge t^* > 0, \quad (3)$$

where xm is a share of m for creation of new technology in the subsystem of ES.

Similar to (2), we introduce the restriction:

$$f(t) = \alpha'(t) + m(t) + c(t) =$$
$$k(t)c(t), k(t) > 1. \quad (4)$$

3. MORE COMPLICATED MM OF ES

The n-product MM, $n > 2$, can be formally written in the same form (1), where m, a, and c are the vector functions, and α, y, β, and $1 - y$ are the respective matrices (where the inequalities for the vectors and matrices are the same inequality for their appropriate components). The continuous MM can be described in the same form considering t and s as many-dimensional variables and examining the appropriate integrals as multivariate ones. The stochastic MM can be obtained by considering α, β, and f as functions of a random factor ω . The discrete MM can be represented in the same form if the integrals in (1) are understood in the sense of Stieltjes. The MM of ES (3), (4) can be generalized by the similar way.

4. OPTIMIZATION PROBLEMS

One of the important typical optimization problems for ES is maximization of the functional

$$I(y) = \int_{t^*}^{t} c(t)dt =$$
$$\int_{t^*}^{t} (\int_{a(t)}^{t} \beta(t, s)[1 - y(s)]m(s)ds)dt, \quad (5)$$

over y with regard to MM (1).

For the problems if ..., then ... and optimizations by x, y, we are in a need of frequent solutions of the Volterra-type equations considered above. In this connection, the respective effective numerical methods and software are very important

The first essential result on the properties of solutions of the problem (5) consisted qualitatively in that for "small" $T - t^*$ the desired $y(t)$ is minimally possible, but for "large" $T - t^*$ the desired $y(t)$may differ from the minimally possible on the larger initial part of the segment $[t^*, T]$. Only on the smaller final part of $[t^*; T]$ the desired $y(t)$ is minimally possible.

The notions "small" and "large" depend on the values of the functions α and β; namely, the greater the functions in question, the closer to t^* is the boundary between "small" and "large" segments. The result has obtained, in sequel, an important qualitative general interpretation or a law 1:

The record of an external function for any ES can be obtained only under the conditions of its sufficiently comfortable guarantee, that is, under the significant fraction of resources sent to internal needs of ES.

As to the same problem (5) and MM of ES of (3)- type, it was proven under certain conditions that the The following property or the law 2 takes place:

The record of an external function for any ES can be obtained only under the following priority of resource distribution: the highest priority has the subsystem C, then the subsystem A, and then subsystem B.

We consider here maximization of functional

$$R(t) = \int_{t^*}^{t} m(t)dt = \int_{t^*}^{t} (\int_{a(t)}^{t} \alpha(t,s)y(s)m(s)ds)dt, \quad (6)$$

over y and x, y with regard to MM (1)-(2) and (3)-(4).

5. INVESTIGATION OF THE PROBLEM (6), (1)-(2)

Using MM (1)-(2) and assuming y, α, β, and k are constants, we have

$$m(t) = \alpha y R(t), c(t) = \beta(1-y)R(t),$$
$$m(t) = (k-1)c(t),$$
$$t \geq t^* > 0, 0 \leq y \leq 1, \quad (7)$$

from where

$$R(t) = c(t)(\beta(1-y)) = \frac{(k-1)c(t)}{\alpha y},$$
$$y = \frac{(k-1)\beta}{\alpha + (k-1)\beta},$$
$$R(t) = \frac{(k-1)c(t)}{\alpha y} = \frac{c(t)(\alpha + (k-1)\beta)}{\alpha\beta},$$
$$R(t) = \frac{c(t)k}{\alpha}, \alpha = \beta. \quad (8)$$

It follows from (8) that unemployment the less the more new product with more price and less productivity cost. So, the modern slogan more goods and services made in USA can be corrected by more qualitative and requisite new goods and services (k is more) with less expenses (α is less) made in USA.

As to general case of MM (1)-(2), considering (6) with the replacement of m by $(k-1)c$, we come actually to the problem (5), and hence for maximization of work places (WP) number, we have to use the strategy in accordance to the law 1 above.

6. INVESTIGATION OF THE PROBLEM (6), (3)-(4)

Using MM (3)-(4) and assuming x, y, and k are constants, we have

$$\alpha'(t) = x\int_{t^*}^{t} \alpha(s)m(s)ds,$$
$$m(t) = y\int_{t^*}^{t} \alpha(s)m(s)ds,$$
$$c(t) = (1-x-y)\int_{t^*}^{t} \alpha(s)m(s)ds,$$
$$\alpha'(t) + m(t) = (k-1)c(t), t \geq t^* > 0, (9)$$

from where

$$R(t) = \int_{t^*}^{t} m(t)dt = \frac{y}{1-x-y}\int_{t^*}^{t} c(t)dt,$$
$$\frac{x+y}{1-x-y} = k-1, x+y = \frac{k-1}{k}$$
$$R(t) = yk\int_{t^*}^{t} c(t)dt. \quad (10)$$

It follows from (11) that $x + y \approx 1$, since k is rather large (it includes the prices of new WP and new technology). So the share $1 - x - y$ of WP in subsystem B is small, which is consistent with the law 2 above.

As to the general case of MM (3), (4), for maximization of the new products(5) on large period of the time $T - t^*$, we have (see [3]-[5]) the following relations:

$$x(t) = 1, t^* \leq t \leq \approx T - t^*,$$
$$y(t) = 1, T - t^* - (T - t^*)^{1/2}$$
$$\approx \leq t \leq \approx T - t^*,$$
$$x + y = 0, t \approx T - t^*. \quad (11)$$

7. CONCLUSION

In conclusion, we would like to emphasize that the main obstacles for realization of the maximization of new WP number by the ways above in practice can be production of namely requisite new products.

For ensuring this, special ES is needed to keep up with change of products needs and to make fast interaction with the science as ES to create respective new technology and with the education as ES to prepare new professional in accordance with respective new labor functions.

References

[1] V. Glushkov. *On a class of dynamic macroeconomic models*. Control systems and machines, 2, 1977.

[2] V. Glushkov and V. Ivanov. *Modeling of optimization of work places distribution between the branches of production A and B*. Kibernetika, 6, 1977.

[3] N. Hitronenko and Y. Yatsenko. *Mathematical Modeling in Economics, Ecology, and Environment*. KAP, 1999.

[4] V. Ivanov. *Model Development and Optimization*. KAP, 1999.

[5] V. Ivanov and N. Ivanova. *Mathematical Models of the Cell and Cell Associated Objects*. Elsevier, 2006.

Whitelist-based SIP Flooding Attack Detection Using a Bloom Filter

Ki Yeol Ryu, Ju Wan Kim, and Byeong-hee Roh

Division of Information and Computer Engineering, Ajou University
San 5 Wonchon-dong,Youngtong-Gu, Suwon 443-749, Korea
e-mail: {kryu, commind, bhroh}@ajou.ac.kr

ABSTRACT - With the nature of SIP with a text-based message format and its openness to the public Internet, it is exposed to a number of potential threats of Denial of Service (DoS) by flooding attacks. In this paper, we propose a whitelist-based SIP flooding attack detection schemes.[1]

Key Words: SIP, DDoS, Bloom Filter, SIP, Flooding Attack Detection

1. INTRODUCTION

Session Initiation Protocol (SIP)[1] has been widely adopted as the main signaling and session management protocol for most recent multimedia applications and systems such as VoIP, IP Multimedia Subsystem (IMS), and so on. With the nature of SIP with a text-based message format and its openness to the public Internet, it is exposed to a number of potential threats of Denial of Service (DoS) by flooding attacks. In SIP flooding attacks, attackers may generate massive malicious SIP request messages to a target SIP server in order to force the server to disconnect.

There have been threshold-based detection methods on SIP flooding attacks[2][3]. Since these approaches are based on certain threshold-values, their decisions may fail when normal traffic increases in normal situations. As another approaches to detect SIP flooding attacks, whitelist-based schemes[4][5] have been proposed. For the whitelists, the source addresses of users who successfully completed SIP registration processes by using REGISTER messages[5] or those who made legitimate sessions frequently[4] are used. With the whitelists, they can detect the flooding attacks by checking whether the number of mis-matched whitelist events occurs over a certain threshold. They also have some limitations to be used in large sized VoIP users' environments due to the heavy requirements to store huge whitelists and corresponding computational complexities to search a list. In addition, SIP supports various URIs, the use of only source addresses to construct the whitelist may be ineffective.

In this paper, we propose a whitelist-based SIP flooding attack detection schemes. For the whitelist, we use a Bloom filter approach to reduce the memory and the computational complexity. To maintain SIP session information using Bloom filter, the proposed method utilizes the three parameters such as source IP address, caller and callee's URIs.

2. PROPOSED METHOD

The SIP session setup is a three-way handshake with INVITE, 200 OK, and ACK as shown in Fig. 1. A calling SIP user agent (UA) transmits an INVITE request message to a callee UA to create a session through SIP proxy servers. Proxy servers receive and forward the INVITE message without disruption of the message. When the receiver UA receives the INVITE message, it sends a 200

This research was partially supported by the MKE, Korea, under the ITRC support program supervised by the NIPA (NIPA-2011-(C1090-1121-0011)). And, it was also supported by the Technology Innovation Program (Grant 10024119) funded by the MKE, Korea.

OK message to accept the session request. Then, the sender UA confirms the session setup by sending ACK, and the session setup is completed.

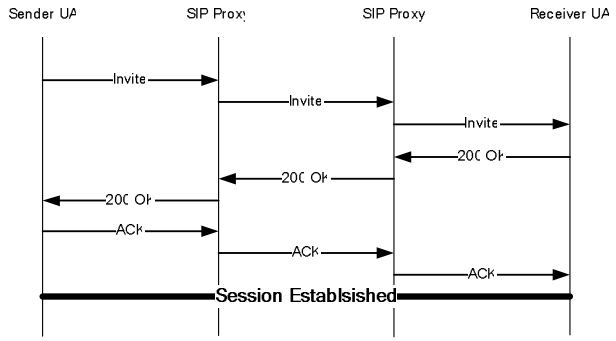


Fig. 1. Session establishment via an INVITE

2.1. Whitelist Construction

As shown in Fig. 1, the SIP server can acknowledge of a normal SIP session setup by monitoring the three-way handshake. In the proposed method, the normal session setup information is listed in the whitelist. For the whitelist, we use a Bloom filter[6] as follows.

It is assumed that there are k independent hash functions, $h_1, h_2, ..., h_k$ and a Bloom filter vector $\mathbf{V}$ of m bits, which is initially set to 0. We define a session string for each normal SIP session by three fields <source_IP_address, caller's URI, callee's URI>. Then, k independent hash functions are applied to the session string. The bit positions in vector $\mathbf{V}$ corresponding to hash function results are set to 1.

2.2. Attack Detection

With the whitelist provided by WM, SIP flooding attack symptoms can be detected easily as following. It is assumed that time is divided into a constant period of Δ, then Δ is a basic unit of attack measurement.

Let $\overline{M}_k$ be the non-membership counter, which is the number of incoming messages that are not members of the whitelist, during k-th time period, Δ_k. At the start of each time period, the counter is initialized to 0. The counter is calculated according to the procedure shown in Fig. 2. When an INVITE message is arrived, its session information string is formed, then Bloom filter membership test is applied to the string with the whitelist vector $\mathbf{V}$ provided by WM. $\mathbf{V}[h_j(b)]$ shown in Fig. 2 means the bit position in the vector $\mathbf{V}$ indicated by the hash function result of $h_j(b)$. If the membership test fails, then the counter is increased by 1.

Let R_k be the weighted average of $\overline{M}_k$ during k-th time period t_k. Then, we have

$$R_k = \alpha \cdot R_{k-1} + (1-\alpha) \cdot \overline{M}_k, k=0,1,2,\ldots \quad (1)$$

where α is the constant value for the weighted average between 0 and 1.

```
M̄_k =0;
while (during k-th time period)
  if (arrived message is INVITE) then
    get a string b from the session information of
the INVITE message;
for each hash function h_j  (j=1,2,…,k)
  if V[h_j(b)] != 1, then
    M̄_k ++;
    stop for-loop;
  endif
    endfor
endif
  endwhile
```

Fig. 2 Non-membership counter calculation

In order to determine the situation in which INVITE flooding attacks occur, we define the three states of NORMAL, ALERT, and ATTACK as in [3]. In NORMAL state, no attack is presumed. The ALERT state is a state in which an attack is in question but an attack has not yet been completely decided. In the ATTACK state, it is inferred that the SIP element is being attacked. The transition between those states is shown in Fig. 3. Let TH_{alarm} and TH_{attack} be the threshold values for the state detections, which are determined by an administrative policy of the service or network operators.

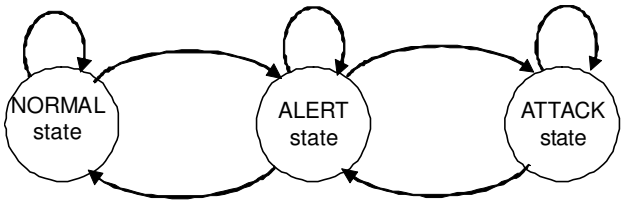


Fig. 3. Transition between states for attack detection

Let C_{ALERT} and C_{ATTACK} are threshold counter values to determine ALERT and ATTACK states, respectively. Then, the state are determined by the algorithm shown in Fig. 4

<variables>
count : counter for representing the degree of attack
$C_{\max}$: maximum of count
state : current state of the algorithm

<main algorithm>
Initially, k=0, count=0, state=NORMAL.
At each time period Δ_k,
- update the weighted average R_k

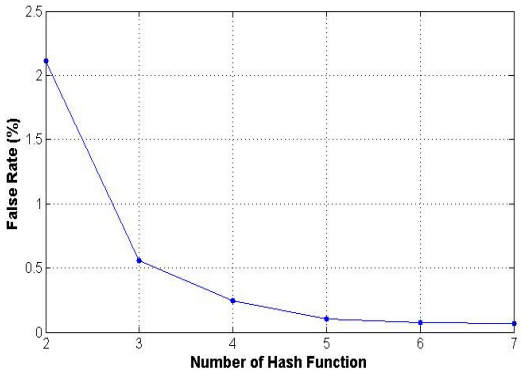
```
if (state==NORMAL)
    if ( R_k > TH_alarm )
        state=ALERT
    endif
elseif (state==ALERT)
    if ( R_k > TH_alarm )
        count++;
    else
        count --;
    endif
    if ( count ≤ C_ALERT )
        state=NORMAL
    elseif (count > C_ATTACK )
        state=ATTACK
    endif
else
    if ( R_k < TH_attack )
        count = MIN ( C_max , count++);
    else
        count --;
    endif
    if (count ≤ C_ATTACK )
        state=ALERT;
    endif
endif
```

Fig. 4. Attack Detection Algorithm

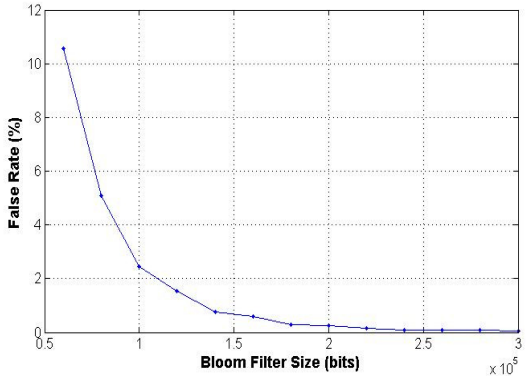
3. EXPERIMENTAL RESULTS

For the experiments to show the effectiveness of the proposed scheme, we used the OPNET simulation tool[7], which has a module for SIP simulation. We implemented a normal SIP traffic generation model from legitimate users on the OPNET by reference to SIPp, which is a free open source for traffic generation of the SIP protocol. And, we also implemented an attack SIP traffic generator by reference to INVITE Flooder, which is an open source to generate a flurry of SIP INVITE messages to a phone or proxy server.

We made a scenario to build a whitelist as follows. It is assumed that there are N registered UAs which can establish sessions through a common SIP proxy server. By having these all UAs established normal sessions each other, we obtained the whitelist, i.e. the Bloom filter vector **V**. Fig. 5. (a) shows a false-positive ratio for the attack detection when the Bloom filter vector size is fixed at 240,000 bits while the number of hash functions varies. And, Fig. 5. (b) also shows the false-positive ratio when the number of hash function is fixed at 5 while the Bloom filter vector size varies. From Fig. 5, to achieve the false-positive ratio less than 0.1%, k=5 and m=240,000 can be chosen when N is given 20,000. Likewise, the proposed method can detect SIP flooding attacks well, and can design the Bloom filter parameters according to the result of Fig. 5.



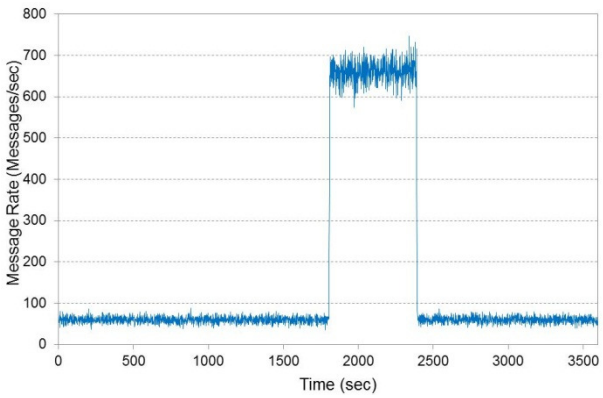
(a) varying k (m=240,000 fixed)



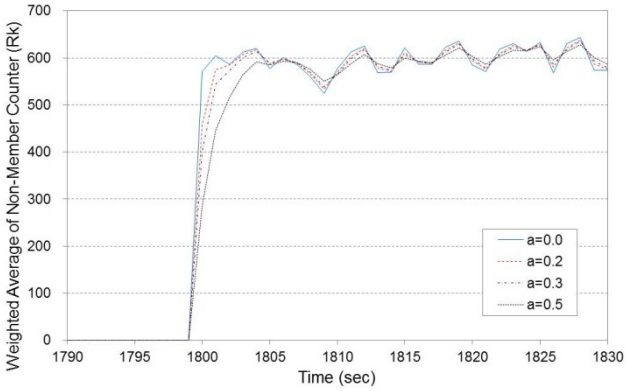
(b) varying m (k=5 fixed)

Fig. 5 False-positive ratios

Fig. 6 shows the sequence of R_k , which is the weighted average of non-membership counter, $\overline{M}_k$, for the case where the bulk-style attack sequence is added to the normal one. The aggregate traffic sequence is shown in Fig. 6 (a). As we can see from Fig. 6(b), the attack symptom can be effectively detected with the sequence of R_k . However, the time required for the decision whether there exists an attack symptom depends on the measurement time interval (Δ) and the weighted average constant (α). They can be determined in advance by an administrative policy of the service or network operators based on the analysis from the measurement of actual traffic flows.



(a) taffic sequence with a bulk-attack



(b) weighted average sequence

Fig. 6 R_k for bulk-style attack sequence

4. CONCLUSIONS

VoIP is one of the most crucial communication services for human life. However, SIP-based services are exposed to a number of potential threats of Denial of Service (DoS) by flooding attacks. In this paper, we proposed the whitelist-based detection schemes against SIP flooding attacks. To build the whitelist, we use a Bloom filter approach to reduce the memory and the computational complexity. With the Bloom filter, the proposed scheme requires a computational complexity of O(1) to process each message, while other whitelist-based approaches such as proposed in [5] have O(N). It is shown that the proposed method detects SIP flooding attacks with a very low false-positive ratio. We expect that the proposed method can contribute to provide secure SIP-based services and applications.

REFERENCES

[1] J.Rosenberg, H. Schulzrinne, G. Cvamarillo, A. Johnston, J. Peterson, R. Spark, M. Handley, E. Schooler, “SIP : Session Initiation Protocol,” IETF RFC 3261, June 2002

[2] M.A. Akbar, Z. Tariq, M. Farooq, "A comparative study of anomaly detection algorithms for detection of SIP flooding in IMS,” IEEE IMSAA’2008, Dec. 2008

[3] J. Ryu, B. Roh, K. Ryu, “Detection of SIP Flooding Attacks based on the Upper Bound of the Possible Number of SIP Messages," KSII Transactions on Internet and Information Systems, Vol.3, No.5, pp.423-574, Oct. 2009

[4] C.V. Zhou, C. Leckie, K. Ramamohanarao, "Protecting

SIP server from CPU-based DoS attacks using history-based IP filtering," IEEE Communications Letters, Vol.13, No.10, pp.800-802, October 2009

[5] Chen, E.Y.; Itoh, M.; , "A Whitelist Approach to Protect SIP Servers from Flooding Attacks," IEEE CQR'2010, June 2010

[6] L. Fan, P. Cao, J. Almeida, A. Z. Broder, "Summary Cache: A Scalable Wide-Area Web Cache Sharing Protocol," IEEE/ACM Tr. Networking, Vol. 8, No. 3, pp.281-293, June 2000

[7] OPNET Modeler, http://www.opnet.com

Simulation-Based Performance Evaluation of Predictive-Hashing Based Multicast Authentication Protocol[1]

Seonho Choi
Department of Computer Science, Bowie State University, Bowie, MD 20715, U.S.A.
and
Hyeonsang Eom
School of Computer Science and Engineering, Seoul National University, Seoul, South Korea
and
Edward Jung
School of Computing and SE, Southern Polytechnic University, Marietta, GA 30067

Abstract

A predictive-hashing based Denial-of-Service (DoS) resistant multicast authentication protocol was proposed based upon predictive-hashing, one-way key chain, erasure codes, and distillation codes techniques [4, 5]. It was claimed that this new scheme should be more resistant to various types of DoS attacks, and its worst-case resource requirements were derived in terms of coarse-level system parameters including CPU times for signature verification and erasure/distillation decoding operations, attack levels, etc. To show the effectiveness of our approach and to analyze exact resource requirements in various attack scenarios with different parameter settings, we designed and implemented an attack simulator which is platform-independent. Various attack scenarios may be created with different attack types and parameters against a receiver equipped with the predictive-hashing based protocol. The design of the simulator is explained, and the simulation results are presented with detailed resource usage statistics. In addition, resistance level to various types of DoS attacks is formulated with a newly defined resistance metric. By comparing these results to those from another approach, PRABS [8], we show that the resistance level of our protocol is greatly enhanced even in the presence of many attack streams.

Key Words*: denial of service, network protocol, authentication, multicast, resource requirement, cryptographic hashing, simulation*

1. Introduction

For real-time streaming applications, a new multicast authentication protocol was proposed which shows a higher level of resistance to Denial-of-Service (DoS) attacks [4, 5]. It was claimed that the resource (CPU, buffer, and network bandwidth) usage level can be greatly reduced by utilizing predictive hashing (PH) technique.

Our scheme is based on a block-based approach where a real-time data stream is divided into blocks of packets. In the predictive-hashing approach, packets in one block contain messages along with predictive authentication information required for authenticating the next block messages. Preliminary analysis on worst-case resource requirements conducted in our previous work [5] indicates that this new scheme consumes much less CPU and buffer space than one of the recently proposed denial-of-service (DoS) resistant multicast authentication schemes, pollution resistant authenticated block streams (PRABS) [8], and that its resistance to DoS attacks is greatly enhanced.

However, in the previous analysis, we derived various formulae, for estimating upper bounds on the resource requirements, in terms of coarse-level system parameters such as CPU times needed for performing erasure decoding, distillation decoding, signature verifications, etc. In addition, worst-case scenarios were assumed in analyses, which may yield too pessimistic estimation results that may not happen in real attack situations. Also, the analysis didn't provide a way to estimate the resource requirements for different attack types and/or different parameter settings. For example, by examining resource requirement changes for different block sizes (with the same bandwidth maintained for the data stream), we may have more insight on which block size we need to choose. The use of a smaller block size may lead to reduced resource usage compared to the use of a bigger block size. Also, using a smaller block size may loosen security condition under which it may be guaranteed that the worst DoS attack type cannot be launched, which the system designer may prefer. However, using a smaller block size means that the receivers will be more susceptible to messages losses due to bursty packet losses (from network congestion and/or from attack streams). Hence, the system designer needs to take these factors into consideration when determining values of the system parameters. Finally, to evaluate resistance level of our protocol against other approaches, a formal metric should be devised. By using a simulator we developed, it becomes possible to quantify resistance levels on different platforms.

We designed and implemented a simulator for the predictive-hashing based multicast authentication protocol. Multiple attack streams along with an authentic data stream may be generated and launched against a virtual receiver which is equipped with the predictive-hashing based protocol. Different attack types may be used in generating such attack streams with various system parameters such as block size, redundancy level for erasure encoding, message size, loss rate, packet (or block) period, and simulation duration, etc. The packets generated by the stream generators will be written to multiple files, and the system

[1] This work was supported by ARO grant 48575-RT-ISP.

parameters used by the stream generators are written to a separate file. These files are read by the virtual receiver process later and the packets will be fed as inputs to the packet processing object (called decoder) which performs authentication operations. One feature of this simulator is that it is platform independent, and the timing parameters such as block period or packet period may be specified in absolute time and will be enforced in any platform where the simulator is running. This simulator is written in Java.

By using this simulator, we may be able to obtain accurate information on resource usages by the protocol in a variety of attack scenarios with different parameter settings. The simulator also outputs detailed information on how much resources are used by each component in the protocol implementation. The type of DoS attack that may be launched varies depending upon the level of attack complexity. The attacker may simply generate packets randomly and launch an attack. Or, he/she may intercept authentic packets somewhere in the network and modify some fields in the packet and launch them against receivers. Alternatively, the attacker may intercept packets at one location in the network, reuse/modify some fields in the packets, and relay them to another location in the network, launching an attack with those relayed packets. This type of attack is named in this paper as a strong relay attack (SRA) and it will be addressed later on the effects of this type of attack and how to avoid such attack scenarios in PH-based approach. These various attack types may be specified as one of the input parameter to our simulator, and attack streams matching these specifications will be generated. Also, by using the simulator, we may formally define the concept of DoS resistance level (DRL) for different attack types as the number of attack streams the receiver may tolerate in terms of authentication throughput and memory requirement.

Preliminary information is given in Section 2. Section 3 provides a brief explanation on predictive-hashing based authentication protocol. Section 4 describes approaches we have taken in the design and implementation of the simulator. In Section 5, detailed simulation results are explained. Conclusions are presented in Section 6.

2. Preliminaries

Block: The original data (or message) stream at the sender side is divided into blocks, and each block data is packetized into the same number of packets. Each packet contains message portion (from original data stream) and additional information related to authentication may be attached. Block period, *p*, corresponds to a duration of time during which block packets are generated by the sender.

Erasure codes: This is one of the forward error correction (FEC) techniques to recover lost packets during transmission [6]. The encoder redundantly encodes information into a set of symbols. If the decoder (receiver) receives sufficiently many symbols, it can reconstruct the original information. An *(n, t)* erasure encoder generates a set of *n* symbols from the input. The decoder can recover all the original data as long as *n-t* symbols are available. *t* is named as a redundancy level.

Strong Threat Model: In this model, there is no limitation on attacker's capability:

- Packets may be eavesdropped, deleted, and spoofed (and sent).
- More than one of these attack operations may be combined to launch more powerful attacks against receivers. For example, packets may be first eavesdropped and deleted (blocked) so that receivers may not receive them. Based on the eavesdropped contents, newly spoofed packets may be sent immediately to the receivers. This attack scenario assumes that attackers can combine all of the above three operations at the same time.

Denial of service attacks

As in [6], an *attack level*, *f,* is introduced and used in this paper which defines the ratio of the bandwidth of injected invalid traffic to the bandwidth of valid traffic. For example, if an adversary injects 10,000 bytes of invalid data in one unit time while the sender is sending 1,000 bytes in a unit time, then the attack level is 10.

One-way accumulators

We can build a secure set membership operation by using one-way accumulators [1, 2]. There are several one-way accumulator schemes based on different cryptographic techniques. In distillation codes, Merkel hash trees [3] are used as one-way accumulators. When Merkel hash trees serve as one-way accumulators, the size of witnesses grows logarithmically with the size of the accumulated set.

3. Predictive Hashing Based Approach

We developed a new mechanism, which is based on *Predictive Hashing* (PH) and *One-way Key Chain* (OKC), to significantly reduce resource requirements at a receiver even in the presence of DoS attack packets flowing in. The basic idea of predictive hashing is that each block of packets conveys authentication information that will be used to authenticate the next block packets instead of sending authentication information within the same block as in previous approaches [6, 9, 10]. The PH technique allows receivers to save significant amount of buffer space since only authentication-related portions from each packet needs to be saved for future packet authentication, while the message portions of arrived packets are processed (or authenticated) immediately upon receipt. However, in our scheme, the sender needs to keep the message portions from two consecutive blocks in its buffer to calculate PH.

One-way key chain technique is already used in other contexts such as in one-time password [8], TESLA [11], etc. In our approach, the sender obtains a hash chain by applying hash operations recursively to some seed value, and obtained key values are assigned to blocks in backward order of their generation times. The sender uses the assigned key to calculate Message Authentication Codes (MAC) images of the prediction hashes/signature information for the next block, and attach them (along with other authentication related information) to the current

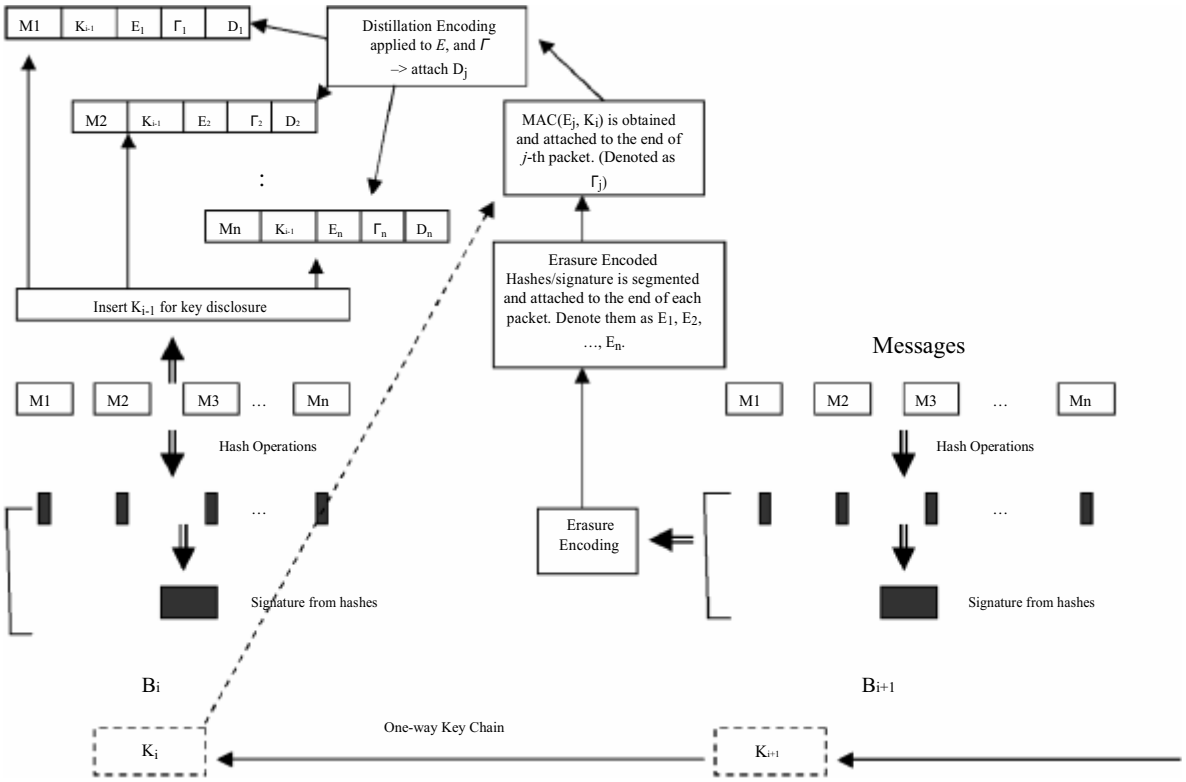


Figure 1. Overview of our PH-based scheme at a sender

block packets. Also, each block packet reveals the key used in the previous block to let the receivers use it in authenticating the previous block packets (or partitions) without applying erasure decoding and signature verifications in most of the cases. These mechanisms are combined with erasure codes and distillation codes to develop a multicast authentication protocol which is very resistant to Denial-of-Service attacks and resource-efficient. Figure 1 shows an overview of our approach at sender side. The receiver side operation is the reverse of the process shown in Figure 1.

PH Decoding Algorithm at Receiver

The receiver side algorithm is presented in Figure 2 in detail [4, 5].

4. Simulator Design and Implementation

The following are the goals of our simulator design:

- accurate measurement of resource usages at a receiver where the simulator is running: whatever computing platform the simulator is running, it should provide accurate measurements on resource usages (CPU time, memory, and bandwidth) as if it were acting as a receiver in a real network. To achieve this goal, the simulator is implemented in Java with a capability to adjust its execution scenario based upon timing parameters such as block period (or packet inter-arrival times).
- platform independence: this is achieved by implementing the simulator in Java. Also, specified timing parameter such as packet inter-arrival time will be enforced regardless of the computing platform.
- support for a variety of DoS attack types and other system parameters to be specified as input to the simulator: attack packet streams may be generated with various attack types including simple relay-attack and strong relay-attack types by attack stream generators. Also, other system parameter values, such as block size(n), redundancy level (t), message size in each packet, loss rate, and the number of blocks to be simulated, may be specified to the simulator.
- formulation and estimation of DoS resistance: the resistance level to various attack types is formulated as a number of attack streams that may be tolerated without affecting packet authentication throughput. That is, a threshold on the number of attack streams will be found beyond which packet authentication delays will increase indefinitely.

Taking these goals into consideration, we designed and implemented a simulator in Java, and carried out extensive simulations. Figure 3 shows the overall architecture of the simulator. The receiver side decoding is performed following the algorithm given in Figure 2.

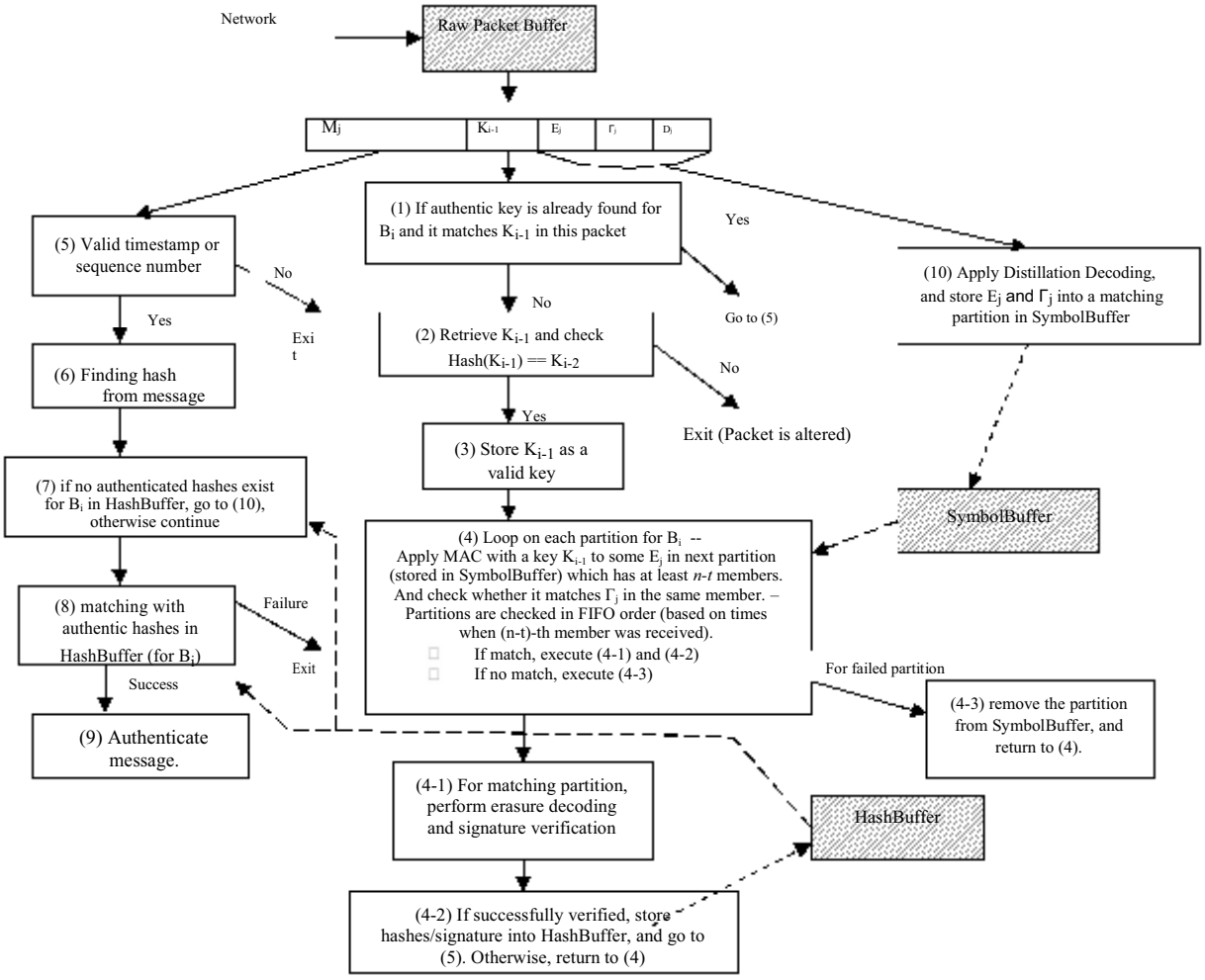


Figure 2. Detailed algorithm at the receiver

The reason why authentic and attack packets are stored into multiple files is to allow all the resources to be devoted later for running the receiver side decoding routine. Another alternative approach might have been to let packet generators run concurrently along with the receiver side decoding process. But, this approach would not permit us to measure exact resistance levels and resource usages as if the computer were wholly used for processing received packets as in real situation. We also assume that the overheads resulting from file access is insignificant to the simulator performance compared to the cases where the packets are received from the network.

5. Simulation Results

We conducted extensive simulations by running the simulator on a PC with Pentium 4 CPU (1.7 GHz) and 1GB of memory.

CPU Time Requirements

In Table 1 we show resource usage statistics in terms of CPU times with an attack level f=0 (i.e., no attack stream was introduced). The simulator was run for 100 block periods where a block period was 1.5 second. We used system parameter values n=32, t=16, loss rate =0.01, and message size = 1024 bytes.

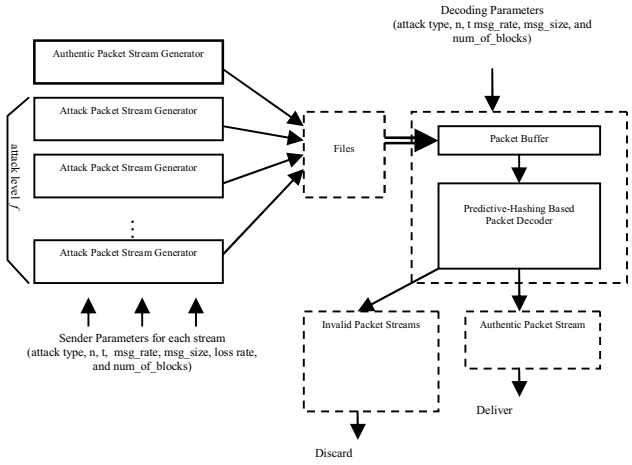


Figure 3. Simulator Architecture

algorithm steps (Figure 2)	Cumulative execution time (ms)	percentage
(1)	0	0.0
(2)	0	0.0
(3)	0	0.0
(4-1)	0	0.0
(4-2)	31	3.03
(4-3)	0	0.0
(4-Erasure Decoding)	0	0.0
(4-Signature Verification)	838	81.8
(5)	0	0.0
(6)	30	2.93
(7,8,9)	16	1.56
(10)	109	10.64
Total	1024	100%

Table 1. Cumulative CPU times and percentages of CPU times spent for each algorithm step when there was no attack stream.

Most of the CPU time is spent for the signature verification step (81.8%), followed by the distillation decoding step (10.64%). The average signature verification operation with data size of around 1200 bytes and signature size of 46 bytes took about 23-26 ms.

Figure 4 shows changes on CPU usage percentages among different algorithm steps with varying values of attack level. As is shown, more CPU time will be used for step (10), distillation decoding operation, while the percentage for step (4), signature verification operation, decreases as the attack level increases. This is due to the fact that only one signature verification operation is needed regardless of attack level in the PH-based approach, while the percentages of CPU times spent for the other steps will increase for higher attack levels. Note that this would not be true in PRABS [8] and signature verification cost will still dominate for higher attack levels.

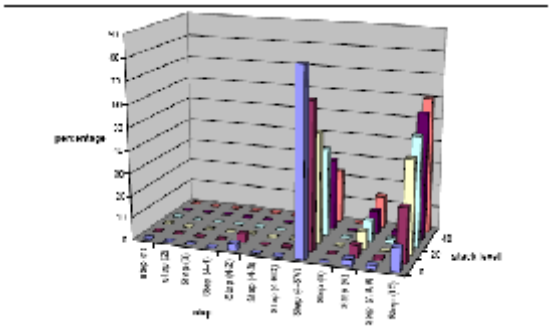


Figure 4. CPU time percentage changes with varying attack levels (from 0 through 50) in the presence of simple relay-attack streams

Resistance Level

For two different attack types, we ran the simulator to find threshold points (in terms of attack level) beyond which inter-packet authorization delay becomes steadily bigger than inter-packet arrival time. The ratios of inter-packet authentication delays to inter-packet arrival times are

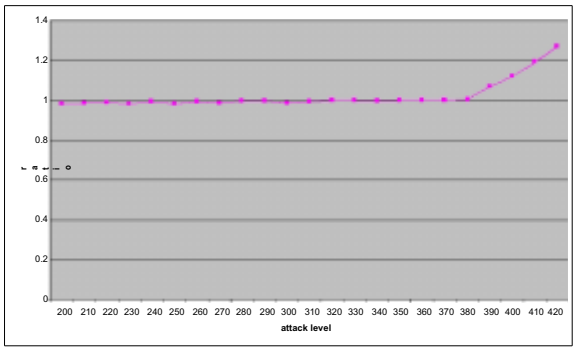


Figure 5. Ratios of inter-packet authentication delays to inter-packet arrival times with simple relay-attack streams. Resistance level is 390. With the PRABS approach, the resistance level becomes 26 as is shown in Figure 6.

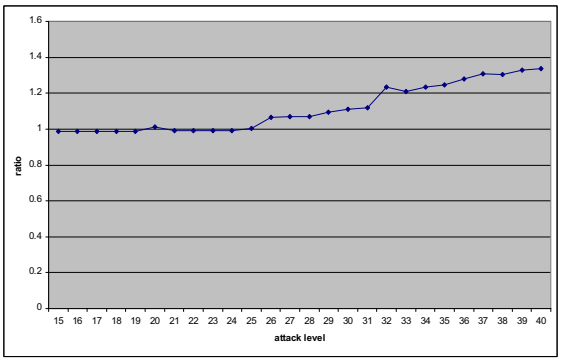


Figure 6. Ratios of inter-packet authentication delays to inter-packet arrival times with strong relay-attack streams. Resistance level is 26

shown in Figure 5 for simple relay attacks with different attack levels, where erasure-encoded symbol values in authentic packets are arbitrarily modified to generate attack packets. If this ratio is greater than 1.0, it means that the resistance level is reached. The simulations for other simple-relay attacks (such as modifying the key values or distillation code values) showed similar results due to the fact that, for any kind of simple relay-attacks, only one signature verification operation is needed in the PH-based approach.

Figure 6 shows the ratios of inter-packet authentication delays to inter-packet arrival times for strong relay-attacks when the same system parameter settings are used as in Figure 5. This figure may also be considered as the one showing performance gains we may achieve by using the PH-based approach compared to the PRABS approach when simple relay-attacks are launched. This is because the resource consumption in PRABS will increase in proportion to the number of attack streams regardless of attack types. In PRABS, the same number of signature verification operations is needed as the number of streams that a receiver receives.

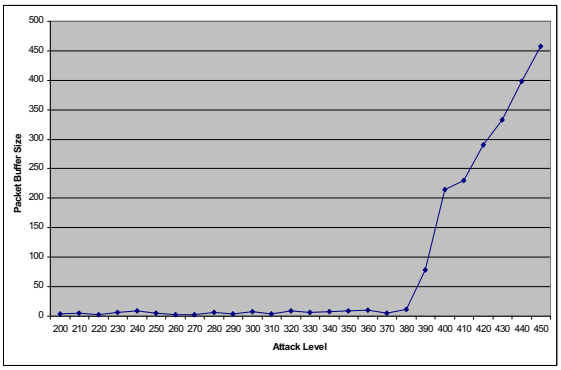


Figure 7. Memory requirement in terms of packet buffer size in the PH-based approach. The buffer size begins increasing around resistance level 390. With PRABS, the buffer size begins increasing around attack level = 26 as is shown in Figure 8.

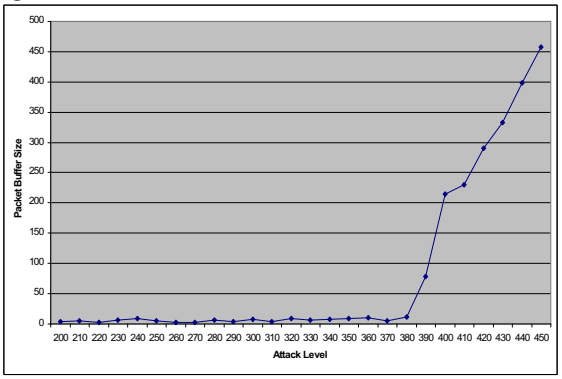


Figure 8. Memory requirement with strong relay-attack streams. It begins increasing near resistance level=26. Packet Buffer Size is measured in terms of the maximum number of packets stored in the packet buffer at any time

Memory Requirements

The memory requirements in the presence of simple relay-attacks streams and strong relay-attack streams are shown in Figure 7 and 8. Note that the major memory requirement comes from the packet buffer where the incoming packets are stored while the receiver is processing packets. The memory requirements for other buffers, such as Symbol Buffer and Hash Buffer, are negligible compared to that for the packet buffer due to the size difference, and are not shown here due to space limitation.

6. Conclusion

We designed and implemented a simulator based upon a new PH-based multicast authentication protocol. This simulator may be used on any computing platform to measure exact resistance level to various types of DoS attacks in different parameter settings. This tool may also be used to determine optimal system parameter values such as block period (p), block size (n), redundancy level (t), etc. This simulator was used to derive detailed resource usage information, and to measure the resistance levels against different types of DoS attacks on a selected computing platform. The result shows that our PH-based protocol outperforms other protocols (including PRABS which is outperformed by 15 times) in terms of resistance to DoS attacks.

References

[1] D. Adkins, K. Lakshminarayanan, A. Perrig, and I. Stoica. Taming IP packet flooding attacks. In *Proceedings of Workshop on Hot Topics in Networks (HotNets-II)*, Nov. 2003.

[2] N. Baric and B. Pfitzmann. Collision-free accumulators and fail-stop signature schemes without trees. In *Advances in Cryptology --EUROCRYPT '97*, volume 1233 of *Lecture Notes in Computer Science*, pages 480–494, 1997.

[3] J. Benaloh and M. de Mare. One way accumulators: A decentralized alternative to digital signatures. In *Advances in Cryptology – EUROCRYPT '93*, volume 765 of *Lecture Notes in Computer Science*, pages 274–285, 1993.

[4] Seonho Choi, "Denial-of-Service Resistant Multicast Authentication Protocol with Prediction Hashing and One-way Key Chain," ism, pp. 701- 706, In Proceedings of the Seventh IEEE International Symposium on Multimedia (ISM'05), 2005.

[5] Seonho Choi and Yanggon Kim, "Resource Requirement Analysis for a Predictive-Hashing Based Multicast Authentication Protocol," In Proceedings for EUC Workshops, pages 302-311, IFIP, August 2006.

[6] M. Goodrich, R. Tamassia, and J. Hasic. An efficient dynamic and distributed cryptographic accumulator. In *Proceedings of Information Security Conference (ISC 2002)*, volume 2433 of *Lecture Notes in Computer Science*, pages 372–388, 2002.

[7] C. Karlof, N. Sastry, Y. Li, A. Perrig, and J. Tygar, Distillation codes and applications to DoS resistant multicast authentication, in Proc. 11th Network and Distributed Systems Security Symposium (NDSS), San Diego, CA, Feb. 2004.

[8] Leslie Lamport, "*Password Authentication with Insecure Communication*", Communications of the ACM 24.11 (November 1981), 770-772

[9] A. Pannetrat and R. Molva. Efficient multicast packet authentication. In *Proceedings of the Symposium on Network and Distributed System Security Symposium (NDSS 2003)*. Internet Society, Feb. 2003.

[10] J. M. Park, E. Chong, and H. J. Siegel. Efficient multicast packet authentication using erasure codes. *ACM Transactions on Information and System Security (TISSEC)*, 6(2):258–285, May 2003.

[11] A. Perrig, R. Canetti, D. Song, and J. D. Tygar. Efficient and secure source authentication for multicast. In *Proceedings of the Symposium on Network and Distributed Systems Security (NDSS 2001)*, pages 35–46. Internet Society, Feb. 2001.

B2C Website Design and Customers' Affective Commitment: Exploring the Relationship

Jean Éthier
Département des Systèmes d'information, Université de Sherbrooke
Sherbrooke, Québec, Canada

Harold Boeck
Département de Marketing, Université de Sherbrooke
Sherbrooke, Québec, Canada

Geneviève Pellerin
Sherbrooke, Québec, Canada

Keywords: Customer commitment, Website design, purchasing process, cognitive appraisal.

1. INTRODUCTION

An abundance of recommendations can be found on how to design effective B2C websites [12, 8, 5, 14]. However, most of them do not take into account consumers' affective reactions to website use, even though these reactions influence key behaviors such as time spent on the website, purchasing, impulse buying, repurchasing, loyalty and commitment towards the website [1, 6, 11, 4, 3].

Based on the information systems and consumer behavior literature, this ongoing study explores the impact of four interface design features on consumers' affective commitment to websites. It proposes a research model based on a recent empirical study [7] that found that website interface features such as structure information and navigation are key elements impacting cognitive appraisals, which lead to affective reactions.

2. RESEARCH MODEL

The research model predicts several relationships between the constructs (Figure 1). The first four hypotheses predict that a positive assessment of each interface feature will positively influence the evaluation level of the online purchasing process. The fifth hypothesis predicts that a favorable evaluation of the online purchasing process will have a positive impact on commitment to the website.

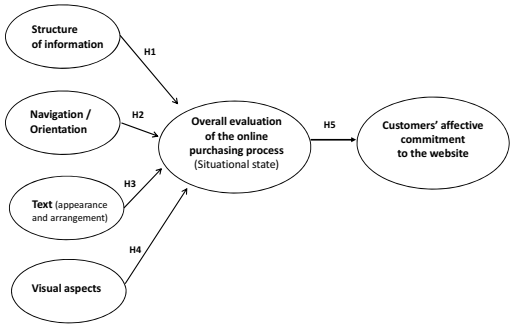


Figure 1. Research model

Each variable of the research model is theoretically justified and is measured with items identified in previous empirical research. The four website design features (structure of information, navigation/orientation, text, and visual aspects) come from the IBM design guidelines proposed for the development of high-usability websites [10] and from Hong and Moriai's [9] design principles. As a specific referent for affective reactions, the evaluation of the online purchase experience variable originates in [7] adaptation of a key cognitive appraisal identified by the appraisal theory of emotions [15]. Affective commitment to the website, which refers to the consumer's desire to continue the relationship with the website in the future, comes from Casalo et al.'s [2] study integrating other variables such as trust, perceived reputation and satisfaction. According to these authors, committed customers share four important characteristics: higher purchase intentions, better resistance to counter-persuasion, willingness to pay premium prices, and eagerness to recommend the website to others.

3. METHODOLOGY

A pre-test has already been conducted with 35 consumers to test the validity of each construct. Internal validity is confirmed for each variable, as the Cronbach's alpha coefficients are all within the threshold (> 0.70) suggested by Nunnally [13]: structure of information: 0.82; navigation/orientation: 0.77; text appearance and arrangement: 0.88; visual aspects: 0.94; evaluation of the online purchase experience: 0.94; and affective commitment to the website: 0.78.

The research model is currently being tested with data collected from a survey tracking purchasing experiences on a Canadian deal-of-the-day website that features discount coupons for different products and services usable at local or national companies. For two weeks, any consumer purchasing products or services on the website was prompted to answer an online questionnaire. The final results including hypotheses testing will be presented exclusively at the conference.

4. CONTRIBUTIONS

The expected contributions of this research project, both theoretical and practical, are numerous. First, the existence of an affective reaction emerging from the purchasing process on a B2C website should be confirmed. Second, the research model establishing antecedents for this affective reaction (website interface features and evaluation experience as a cognitive referent) is expected to be validated. Third, practitioners will be reminded of the importance of developing usable websites that integrate not only rational behavior but also customers' affective commitment to the purchasing experience offered by the website. Fourth, a set of guidelines for website design will be developed, derived from the items of the four website interface features.

5. REFERENCES

[1] T. Adelaar, S. Chang, K.M. Lancendorfer, B. Lee, M. Morimoto, "Effects of media formats on emotions and impulse buying intent", **Journal of Information Technology,** Vol. 18, No. 4, 2003, pp. 247-266.

[2] L.V. Casalo, C. Flavián, M. Guinalíu, "The influence of satisfaction, perceived reputation and trust on a consumer's commitment to a website", **Journal of Marketing Communications,** Vol. 13, No. 1, 2007, pp. 1-17.

[3] C.M. Chiu, C.-C. Chang, H.-L. Cheng, Y.H. Fang, "Determinants of customer repurchase intention in online shopping", **Online Information Review,** Vol. 33, No. 4, 2009, pp. 761-784.

[4] K. De Wulf, N. Schillewaert, S. Muylle, D. Rangarajan, (2006), "The role of pleasure in web site success". **Information and Management**, Vol. 43, No. 4, 2006, pp. 434-446.

[5] D.K.V. Duyne, J.A. Landay, J.I. Hong, **The design of sites: Patterns for creating winning web sites (2nd edition),** Upper Saddle River, N.J: Prentice Hall, 2006.

[6] S.A. Eroglu, K.A. Machleit, L.M. Davis, "Empirical testing of a model of online store atmospherics and shopper responses", **Psychology and Marketing,** Vol. 20, No. 2, 2003, pp. 139-150.

[7] J. Éthier, P. Hadaya, J. Talbot, J. Cadieux, "Interface design and emotions experienced on B2C web sites: Empirical testing of a research model", **Computers in Human Behavior**, Vol. 24, No. 6, 2008, pp. 2771-2791.

[8] S. Hassan, F. Li, "Evaluating the usability and content usefulness of web sites: A benchmarking approach", **Journal of Electronic Commerce in Organizations,** Vol. 3, No. 2, 2005, pp. 46-67.

[9] S. Hong, M. Moriai, **Evaluation criteria for the design of commercial Web sites,** Atlanta, Georgia: Department of Computer Information Systems, 1997.

[10] IBM, **Web design guidelines**, 2000, https://www-01.ibm.com/software/ucd/ucd.html. Access date: 6/11/2001.

[11] V. Mummalaneni, "An empirical investigation of web site characteristics, consumer emotional states and on-line shopping behaviors", **Journal of Business Research**, Vol. 58, No. 4, 2005, pp. 526-532.

[12] J. Nielsen, **Designing web usability**, Indianapolis, IN: New Riders Publishing, 2000.

[13] J.C. Nunnally, **Psychometric theory**, New York: Harper Business, 1991.

[14] R. Otaiza, C. Rusu, S. Roncagliolo, « Evaluating the usability of transactional web sites", **Third International Conference on Advances in Computer-Human Interactions**, 2010, pp. 32-37.

[15] I.J. Roseman, A.A. Antoniou, P.E. Jose, "Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory" **Cognition and Emotion**, Vol. 10, No. 3, pp. 241-277.

Near-Field Coupling Communication Technology For Human-Area Networking

Ryoji Nagai, Taku Kobase, Tatsuya Kusunoki, Hitoshi Shimasaki, and Yuichi Kado
Department of Electronics, Kyoto Institute of Technology,
Matsugasaki Sakyo-ku, Kyoto, Japan

and

Mitsuru Shinagawa
Faculty of Science and Engineering, Hosei University
Koganei-shi, Tokyo, Japan

ABSTRACT

We propose a human-area networking technology that uses the surface of the human body as a data transmission path and uses near-field coupling transceivers. This technology aims to achieve a 'touch and connect" form of communication and a new concept of "touch the world" by using a quasi electrostatic field signal that propagates along the surface of the human body. This paper explains the principles underlying near-field coupling communication. Special attention has been paid to common-mode noise since our communication system is strongly susceptible to this. We designed and made a common-mode choke coil and a transformer to act as common-mode noise filters to suppress common-mode noise. Moreover, we describe how we evaluated the quality of communication using a phantom model with the same electrical properties as the human body and present the experimental results for the packet error rate (PER) as a function of the signal to noise ratio (SNR) both with the common-mode choke coil or the transformer and without them. Finally, we found that our system achieved a PER of less than 10^{-2} in general office rooms using raised floors, which corresponded to the quality of communication demanded by communication services in ordinary office spaces.

Keywords: near-field coupling communication, human-area networking, common-mode noise, quasi electrostatic field, packet error rate, and signal to noise ratio.

1. INTRODUCTION

Wireless body area networks around the human body are expected to play an important role in various areas of application, such as in the remote monitoring of health, sports training, interactive gaming, sharing of personal information, secure authentication, train ticket wickets, and medical information systems [1]. Body-channel communication (BCC) technologies have recently been actively reported [2]–[6]. However, these communication technologies are only composed of transceivers (TRXs) on the human body (wearable TRXs). We propose human-area networking based on near-field coupling communication (NFCC), which consists of both wearable TRXs and those embedded in environments or in equipment that broaden the areas to which BCC can be applied [7]–[9]. We aimed at achieving the concept of "touch the network", which is a novel idea to access networks and exchange data by simply stepping on the floor. Typical examples of this concept for ticket wickets and Internet access systems are outlined in Fig. 1. When people carry wearable TRXs in their pockets, they can access networks through embedded TRXs by simply passing through ticket wickets. User IDs are then authenticated and fares are calculated and deducted. There is also a photograph that demonstrates our concept in Fig. 2. The person in this scenario has a wearable TRX attached to his body/clothes and he is accessing his favorite Web page by simply stepping on an embedded electrode while he is sitting on a chair. As the proposed communication system using embedded TRXs is able to connect networks all over the world, this system can be applied to a wide range of applications. However, embedded TRXs are more strongly susceptible to environmental noise from earth grounding, AC power, and equipment connected to networks than wearable TRXs. The quality of reception attained by embedded TRXs is worse for this reason. We found that the quality of communication was improved by implementing common-mode noise filters in embedded TRXs. The embedded electrodes were floated above a concrete floor because there is wiring under floors in real offices. We measured the signal to noise ratio (SNR) by taking this situation into consideration and demonstrated how much the packet error rate (PER) could be ensured in general office rooms.

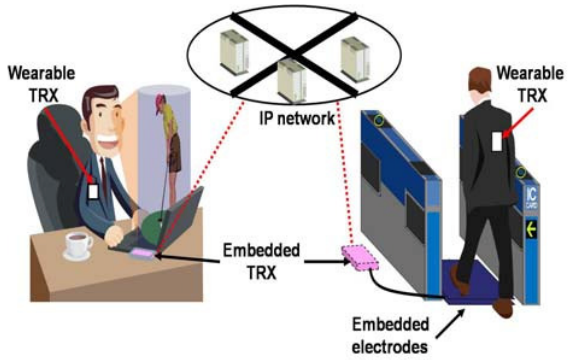


Fig. 1. Scenario for practical use.

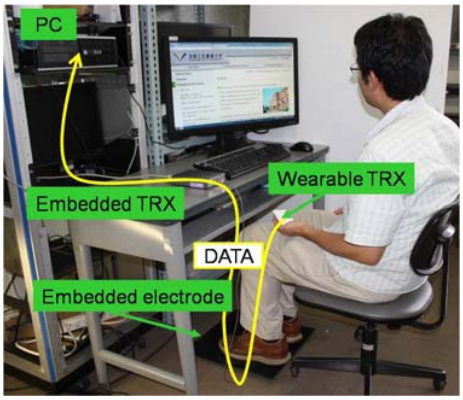


Fig. 2. Photograph of demonstration scenario.

2. Communication Model

The communication model for the NFCC system is shown in Fig. 3. The NFCC consists of two types of TRXs. The first is a wearable TRX that can be carried in jacket breast pockets or trouser pockets. The second is an embedded TRX that can be embedded in walls, desktop PCs, and wickets. When modulated signals are applied to a pair of parallel electrodes implemented in a wearable TRX, a quasi electrostatic field is generated near the electrodes. An electrical field signal is induced on the human body through a mechanism for near-field coupling. The signal loop is composed of two types of paths. The first is a forward path and the second is a return path. The forward path is a route from the electrode of the wearable TRX on the body side (signal electrode) to the upper electrode through the human body's surface. The return path is also a route from the lower electrode to the electrode on the wearable TRX on the side opposite the body (ground electrode) through earth grounding.

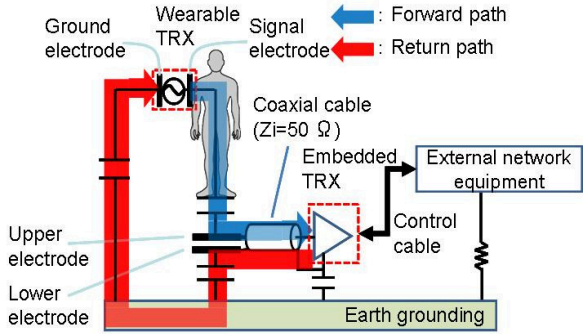


Fig. 3. Communication model for NFCC technology.

Communication where the wearable TRX transmits a signal and the embedded TRX receives it is called an up link. In contrast, communication where the embedded TRX transmits a signal and the wearable TRX receives it is called a down link.

We focused on common-mode noise as a critical factor that degraded the quality of transmission in NFCC systems. The intrusion route for noise is outlined in Fig. 4. The embedded TRX for the communication system is strongly coupled to earth grounding through an AC-power line and external network equipment. As a result, a common-mode noise loop is formed.

Next, we will describe a method of improving the quality of communication by implementing a filter to suppress common-mode noise and our evaluation of how the quality of communication varies with SNR in an ordinary office space.

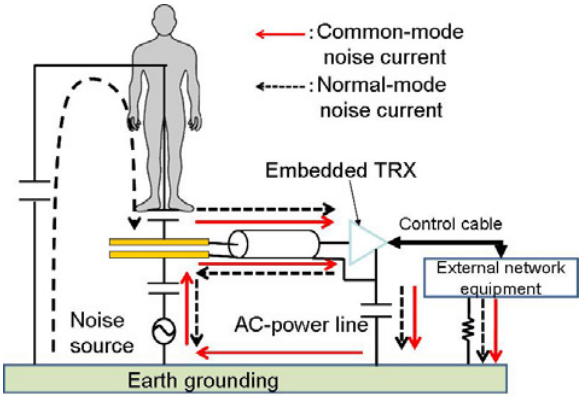
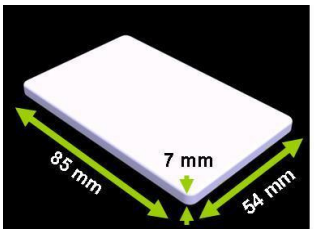
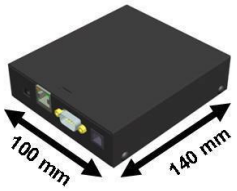


Fig. 4. Intrusion route for noise.

3. TRX Configuration

A card-type prototype wearable TRX and an embedded TRX that can be installed in environments, such as doors and floors, are shown in Fig. 5. The prototype uses a 6.75-MHz carrier frequency with binary phase shift keying (BPSK) modulation, and achieves a transmission rate of 420 kbps. The wearable TRX has a pair of parallel electrodes. It can operate for approximately one year on a single CR3032 button-type lithium-ion battery. The embedded TRX has an SMA connector acting as the signal input or output port and an RS232C serial port acting as the interface with external devices. It is driven by AC-power. In the example of rail ticket wickets that was described earlier, the card-type TRX can be carried in trouser pockets, transmitting ID information, and achieving communication with the embedded TRX built into the floor.





Wearable TRX		Embedded TRX	
Supply voltage	3.0 V	Supply voltage	6.0 V
Bitrate	420 kbps	Bitrate	420 kbps
Carrier Frequency	6.75 MHz	Carrier Frequency	6.75 MHz
		Interface	RS232C

Fig. 5. Configuration and basic specifications for TRX.

4. EXPERIMENTS

Evaluation of system

We measured the SNR for the up link as a function of the distance between the floor and the embedded electrodes, as shown in Fig. 6. As the distance between the floor acting as earth grounding and the embedded electrodes increases, the capacitance (C_L) between the floor and the lower electrode decreases. The received signal level decreases as the lower electrode is away from the floor. Raised floors in general office rooms are used to install wired communication networks or AC-power lines. We measured the SNR to ensure that our NFCC could be used on raised floors. To find what effect distance had on the SNR, we changed the distance with spacers made of foamed polystyrene. The embedded electrodes were connected to a spectrum analyzer and a person 1.76-m tall who wore shoes stood on the embedded electrodes. He wore the wearable TRX on his body. We measured the received signal power and the noise power. The distance between the person and the embedded electrodes was maintained.

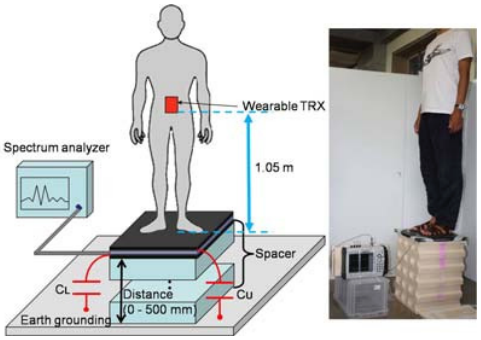


Fig. 6. System for measuring SNR.

Fig. 7 is a schematic of the experimental system. We used a phantom with the same electrical properties as the human body to ensure the experiments could be reproduced. The phantom was a rectangular solid filled with a gel material that absorbed water. Since there were spaces between the wearable TRX and human body in practical use, we placed an attenuator on the top surface of the phantom so that we could adjust the signal power. The wearable TRX was placed on the attenuator. The embedded TRX was connected to the embedded electrodes (350-mm-sq.) and the noise generator was connected to the noise electrodes (350-mm-sq.), both with a coaxial cable. We inserted the common-mode choke coil or the transformer between the embedded TRX and the embedded electrodes depending on the experiment. The embedded TRX was connected to a desktop PC with an RS232C cable. The noise electrodes were placed under the embedded electrodes. A rubber sheet, which was 5 x 350 x 350 mm, was inserted between the phantom and the embedded electrodes and between the embedded electrodes and the noise electrodes. The noise generator and the embedded TRX were driven by AC power. We held an attenuator with a thickness of 200 mm in the experiments, and we measured the quality of communication for the up link as a function of the SNR using a white noise generator.

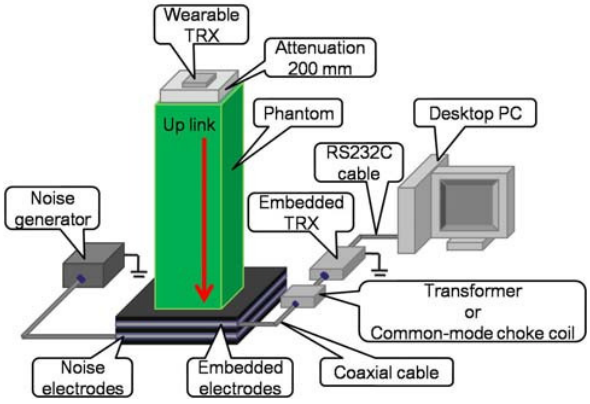


Fig. 7. System for measuring PER.

Results

The received signal and the noise power for the up link and the capacitance (C_L) as a function of a distance between the floor and the embedded electrodes are plotted in Fig. 8. The noise power increased by 3 dB when a person stood on the embedded electrodes. Although the SNR changed according to the distance, it remained at more than 23.9 dB. When the distance approached 0 m, the signal power increased, the noise power decreased, and the SNR was maximum. This is because the return path was enhanced due to increase in the value of C_L. The floating capacitance (C_U) between the upper electrode and the floor was comparable to C_L when the embedded electrodes approached the floor. Consequently, the balance between the impedance for the signal line and that for the ground signal line with respect to the earth grounding was better. As a result, the normal mode noise current was suppressed.

We designed and fabricated a common-mode choke coil and a transformer to suppress common-mode noise. The characteristics of these filters to suppress normal and common-mode noise are plotted in Fig. 9. Because the common-mode choke coil had high impedance for common-mode current, common-mode noise current was suppressed. As the transformer isolated the circuit for the embedded TRX from the embedded electrodes, common-mode noise current was suppressed.

The PER characteristics as a function of the SNR for the up link are plotted in Fig. 10. The total length of a packet was 22 bytes. Each packet consists of address, data, command, etc. We can see the SNR was improved by 2.5 dB at a PER of 10^{-2} when the transformer was used between the embedded TRX and the embedded electrodes. The SNR also improved by 5.0 dB at a PER of 10^{-2} when the common-mode choke coil was used. These results demonstrated that our system achieved a PER of less than 10^{-2}, which corresponds to the quality of communication demanded by communication services in office rooms with an SNR of more than 23.9 dB.

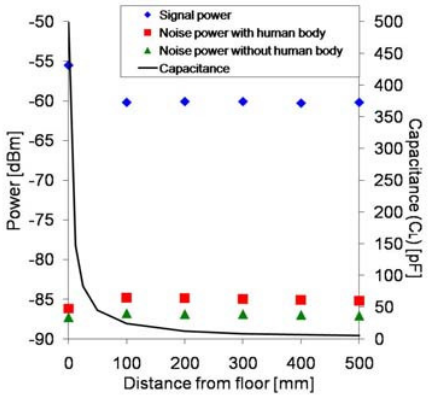


Fig. 8. Received signal, noise power, and capacitance characteristics.

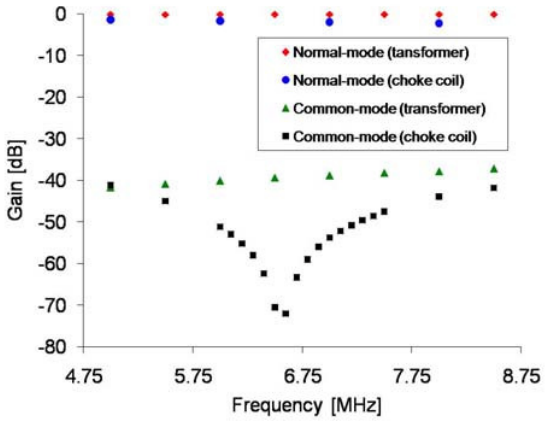


Fig. 9. Characteristics of filters to suppress normal and common-mode noise.

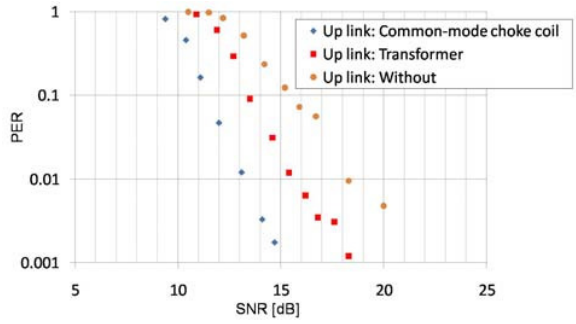


Fig. 10. PER characteristics as a function of SNR for up link.

5. Conclusion

We proposed a human-area networking technology using near-field coupling transceivers. We focused on the fact that embedded TRXs were strongly susceptible to common-mode noise in this work and made a common-mode choke coil and a transformer that acted as common-mode noise filters. We measured the PER of the up link as a function of the SNR both with the common-mode choke coil or the transformer and without them. Moreover, we measured the SNR as a function of the distance between the floor as earth grounding and embedded electrodes. As a result, our system could achieve a PER of less than 10^{-2} for the up link in general office rooms using a raised floor.

6. Acknowledgement

Part of this work was supported by a Grant-in-Aid for Scientific Research (A) 23246073 from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

7. References

[1] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and Victor C. M. Leung, "Body Area Networks: A Survey," Mobile Networks and Applications, Vol. 16, No. 2, pp. 171-193, Spring 2011.

[2] T. G. Zimmerman, "Personal Area Networks: Near-field intrabody communication," IBM Syst. J. , Vol. 35, no. 3-4, pp. 609-617, 1996.

[3] N. Cho, J. Yoo, S. J. Song, J. Lee, S. Jeon, and H. J. Yoo, "The Human Body Characteristics as a Signal Transmission Medium for Intrabody Communication," IEEE Trans. Microwave Theory and Techniques, Vol. 55, pp. 1080-1086, May 2007.

[4] S. J. Song, N. Cho, S. Kim, J. Yoo, S. Choi, and H. J. Yoo,"A 0.9V 2.6mW Body-Coupled Scalable PHY Transceiver for Body Sensor Applications," IEEE ISSCC, pp. 366--367, Feb., 2008.

[5] A. Fazzi, S. Ouzonov, and J. v. d. Homberg,"A 2.75mW Wideband Correlation -Based Transceiver for Body-Coupled Communication," IEEE ISSCC, pp. 204--205, Feb 2009.

[6] J. Bae, K. Song, H. Lee, H. Cho, L. Yan, H. J. Yoo, "A 0.24nJ/b Wireless Body-Area-Network Transceiver with Scalable Double -FSK Modulation," IEEE ISSCC, pp. 34-35, Feb 2011.

[7] M. Shinagawa, M. Fukumoto, K. Ochiai, and H. Kyuragi, "A near-field-sensing transceiver for intra-body communication based on the electrio-optic effect," IEEE Trans. Instrum. Meas., vol. 53, no. 6, pp.1533-1538, 2004.

[8] Y. Kado, "Human-Area Network Technology as a Universal Interface," Symposium on VLSI Circuit Digest of Technical Papers, pp. 102-105, 2009.

[9] Y. Kado and M. Shinagawa, "AC Electric Field Communication for Human-Area Networking," IEICE TRANS. ELECTRON, Vol. E93-C, pp. 234-243, MAR., 2011

Performance Analysis of VoIP over WiMAX

Humberto Véjar Polanco[1], Ernesto E. Quiroz M.[2], Juan J. Tapia A.[2]
[1] Universidad Tecnológica de Tijuana (UTT)
Carr. libre Tijuana-Tecate km.10, Fracc. El Refugio Quintas-C.,Tijuana,B.C., 22253
Tel. +52 (664) 969-4700, Fax. +52 (664) 969-4700
humberto.vejar@uttijuana.edu.mx
[2] Centro de Investigación y Desarrollo de Tecnología Digital (CITEDI-IPN)
Avenida del Parque 1310, Mesa de Otay, Tijuana, B. C., 22510
Tel. +52 (664) 623-1344, Fax +52 (664) 623-1388
{eequiroz, jjtapia}@citedi.mx

ABSTRACT

The evolution of WiMAX by introducing Quality of Service provisioning for voice, and uplink/downlink rates in the range of 35/75 MHz respectively, have made it a strong contender to the fourth generation mobile technology LTE (Long Term Evolution). In spite of the constant increase in multimedia traffic carried by mobile networks, voice remains one of the most important sources of revenue to provide sustainability for service providers. This confers great importance to the knowledge of VoIP performance over WiMAX. However efficient transport is hampered by the small size of the frames of digitized voice, and the added volume of WiMAX and IP system's information necessary for VoIP transport. In this context, in this paper, a computational module that provides performance figures of merit of the capability of WiMAX for VoIP transportation is presented.

KEYWORDS: *WiMAX, VoIP, 4G, Cell phone Communications.*

1. INTRODUCTION

The demand for wireless broadband services is growing rapidly worldwide, in some places even beyond the capability of operators to provide it. According to the Federal Communications Commission of the United States, the use of smart phones has grown nearly 700% in the United States in the last four years [1]. Furthermore, the volume of traffic in AT&T's mobile network has increased by 5,000 % in the last 3 years [2]. One of the drivers of these digital consumption leaps is WiMAX (Worldwide Interoperability of Microwave Access), which is a WMAN (Wireless Metropolitan Area Network) type network, based on the family of broadband wireless access technologies IEEE 802.16. WiMAX is a strong contender to existing last mile access technologies: Cable, DSL (Digital Subscriber Line). [3], as well as LTE (Long Term Evolution: 4th generation UMTS cellular systems).
The mobile WiMAX air interfaces use OFDMA (Orthogonal Frequency Division Multiple Access) to improve multi-path interference in NLOS (Non-Line of Sight) and LOS (Line of Sight) environments, with a range of up to 50 km.
Operates with asymmetric duplex transmission of 75 Mbps on the downlink and 35 Mbps uplink, provides high spectral efficiency (up to 2 bps/Hz), multi-channeling, and advanced MIMO (Multiple Input-Multiple Output) antenna technology.
To ensure worldwide application, WiMAX can use unlicensed and licensed spectrum, with variable bandwidth channels (12.5, 1.5, and 1.75 MHz multiples) up to a maximum of 20 MHz. 802.16 system access remains effective even in the presence of multiple connections per terminal, multiple levels of QoS (Quality of Service) per terminal, and a large number of users sharing the medium by statistical multiplexing.
The provision of QoS for real time services has also been integrated (2006, Release 1.0), with the aim to make it competitive with LTE systems.
As a technology for broadband provision, WiMAX is applicable to both: subscribers in dense urban areas, and for scattered rural communities, and can be used as a backhaul for Wi-Fi cellular clusters.
These characteristics have induced that beginning with the first commercial network in Korea in 2006, until September 2010, 592 WiMAX networks are operating in 149 countries, serving 13 million subscribers [4], number estimated to grow to 18 million in 2011 [5].
Based on the above, this work presents a computational module for WiMAX performance analysis in the transport of VoIP (Voice over IP: Voice over IP).

2. MOBILE WiMAX

WiMAX air interface technology is based on the standard IEEE 802.16 [6]. In particular, the current Mobile WiMAX technology derives from the IEEE 802.16e amendment approved by the IEEE in December 2005, which specifies the OFDMA air interface and provides support for mobility [7].

2.1. Mobile WiMAX Release 1.0

The Mobile WiMAX System Profile Release 1.0 [8] was developed in early 2006. It belongs to the family of WiMAX Forum standards, and was adopted by the ITU as the 6th air interface of the IMT-2000 family [9]. Support for the allocation of flexible bandwidth and integration of multiple types of QoS in WiMAX network, enables the provision of high-speed Internet access, VoIP, video sessions, multimedia chat and mobile entertainment. WiMAX issued a certification program to ensure interoperability of products from different manufacturers, achieving the first stamps of approval WiMAX Forum Certified for the 2.3 GHz spectrum in April 2008 and later for the 2.5 GHz spectrum.

The Mobile WiMAX Release 1.0 Profile is based on the IEEE air interface (Std. 802.16-2004, 802.16-2004, Cor. 1-2005, 802.16e-2005, 802.16-2004, Cor. 2) and WiMAX Forum's network specifications. Figure 1 shows the five sub-profiles and their components.

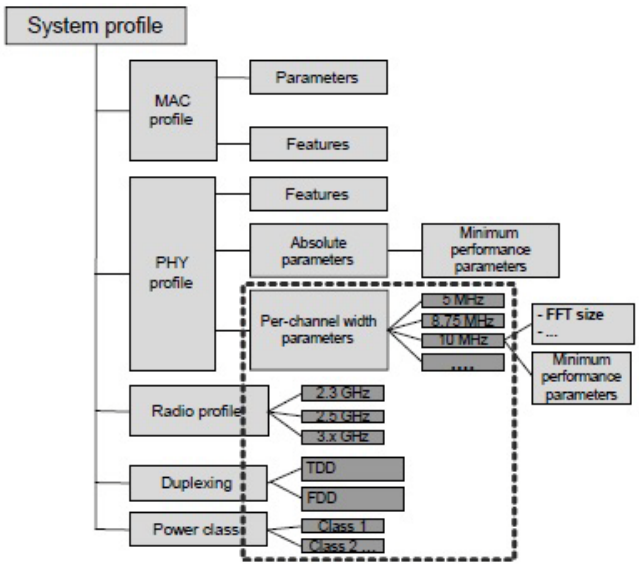


Figure 1.- Structure of Mobile WiMAX system profile[7].

2.2. Mobile WiMAX Release 1.5

WiMAX Forum works in the short-term migration to the profile called Release 1.5, which includes the following improvements:

FDD/HFDD efficient operation. FDD/HFDD (Half-duplex Frequency Division Duplex) operations optimization is based on dividing the 802.16 frame into partitions to be used by two different groups of mobiles having separate control channels, such as Uplinks MAPs and Downlink MAPs (downlink/uplink mapping), the fast feedback channel and HARQ ACK channel. This solution enables reuse of chipsets designed for version 1.0 (TDD) without compromising system performance to address FDD markets worldwide.

New bands. New classes of bands to provide a solution to the FDD transmission mode.

Improved MIMO. Closed loop MIMO operation and Beamforming (BF) further enhance the performance and coverage beyond version 1.0, which contains only open-loop MIMO capacity and some BF.

Improved MAC performance (specially improved VoIP capacity). Version 1.0 is highly optimized for data communications such as TCP/IP. The nature of data traffic implies transmission in "bursts". To adequately address this demand, Release 1.0 technology uses the mechanism of Downlink and Uplink MAP's, control messages transmitted in each frame, i.e. every 5 ms. While this is perfect for bursty traffic, support for the flow of data (VoIP, video) needs further optimization.

The idea of optimization is to use persistent allocation so that a simple MAP message provides information on the allocation of periodic resources matching the needs of a specific flow.

Extended/Improved network characteristics. Most extensions are related to Mesh Base Stations (MBS) Multicast and Broadcast Service. Release 1.5 extensions provide more flexible allocation of MBS zones, which are suitable also for small cells (micro and pico). Another attractive part of Release 1.5 is the set of support functions to Location Based Services (LBS).

Bluetooth coexistence in the same mobile. To provide a more efficient support for WiMAX terminals having additional WLAN (Wireless Local Area Network) interfaces and/or PAN (Personal Area Network), the latter based on Bluetooth [7].

3. TDD, VoIP and Codecs framing

In order to analyze VoIP handling by WiMAX in Releases 1.0 and 1.5, the way digital information is generated in the voice codecs is explained, and the format in which this information is organized to transport Internet and WiMAX. Internet adds control information from the protocols Real-Time Transport Protocol (RTP), User Datagram Protocol (UDP) and IP, these headers are compressed using the Robust Header Compression (ROHC) scheme. The WiMAX system then performs its own formatting to TDD frames.

3.1. Voice coding

Mobile WiMAX does not specify a preferred or base voice encoder. In this paper the AMR and ITU-T G.719 voice codec specifications are applied to carry on the performance analysis.

AMR Speech Encoder. The AMR speech codec is one of the standards adopted by the 3GPP for digitization of voice [10]. It is a variable rate encoder, which through an optimized link-adaptation mechanism, selects the best rate according to channel conditions and capacity. Every 20 ms produces one of 14 possible modes (Table 1, 3rd column), where each mode corresponds to a particular bit rate. The lower bit rate is used to transmit background noise during speech absence periods, and is known as Silence Indicator (SID).

Table 1. AMR Data Rate

Mode	Total speach bits	Channel
AMR 4.75	95	FD/HD
AMR 5.15	103	FD/HD
AMR 5.90	118	FD/HD
AMR 6.70	134	FD/HD
AMR 7.40	148	FD/HD
AMR 7.95	159	FD/HD
AMR 10.20	204	FD
AMR 12.20	244	FD

FD.-Full Duplex HD.-Half Duplex

In this work we use a simplified On-Off model of the AMR speech codec. During periods of active conversation (Figure 2), the highest bit rate (244 b/20 ms) is used, and during periods of inactivity, the SID rate is used (56 b/160 ms).

Header	Table of Contents	AMR Voice Frame	Octet + Alignment
8 bits	8 bits	244 bits	4 bits

Figure 2.-Encodec AMR voice packet fields (33 bytes)

G.719 speech codec. ITU-T specification G.719 describes a low complexity audio codec based on transformation, which operates at a 48 KHz sampling frequency and offers a complete audio bandwidth from 20 Hz to 20 kHz [11].

The coder processes 16-bit PCM linear input signals in 20 ms. frames with a 40 ms average delay. G.719 allows for any rate between 32 *Kbit/s* and 88 *Kbit/s* in increments of 4 *Kbit/s*, and 88 *Kbit/s* to 128 *Kbit/s* in 8 *Kbit/s* steps. One byte of the Table of Contents (ToC) is added at the beginning of each frame of compressed audio, along with the frame's length information. Figure 3 shows the format of the G.719 data packet @ 32 *Kbps*.

Table of Contents 1 byte	G.719 Voice Frame 80 bytes

Figure 3.- G.719 Packet Format on Mode 1

3.2. VoIP over IP

For real-time applications transport streams like Voice over IP (VoIP: Voice over IP) and video, packets are typically transported using the protocol stack RTP/UDP/IP (Real-Time Transport Protocol/User Datagram Protocol/Internet Protocol) [3]. Each protocol has an associated header, adding up 320 bits (Figure 4) or 40 bytes.

Payload
RTP 12 bytes
UDP 8 bytes
IP 20 bytes

Figure 4.- IP Packet Format and the transport headers

This is a huge expense compared to the VoIP packet payload. To reduce the burden of the protocol header, wireless systems use a technique known as Robust Header Compression (ROHC) [3], whereby the header is reduced to 32 bits (Figure 5). Thus VoIP packets for AMR and ITU-T G719 are reduced to two fields, as shown in Figures 5 and 6.

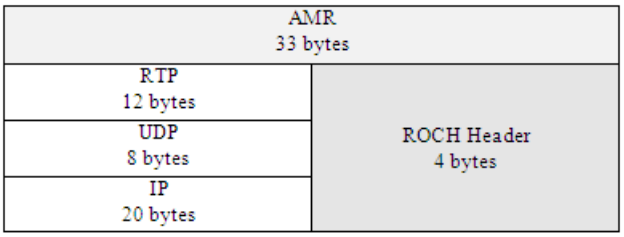


Figure 5.- AMR IP Packet Format and its transport headers

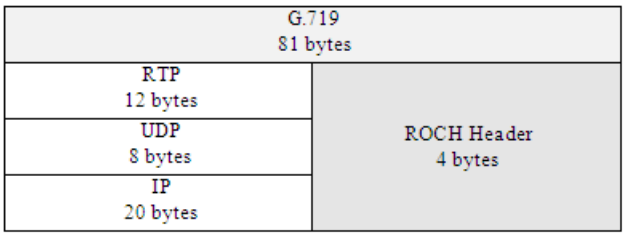


Figure 6.- G.719 IP Packet Format and its transport headers

3.3. MAC Frames

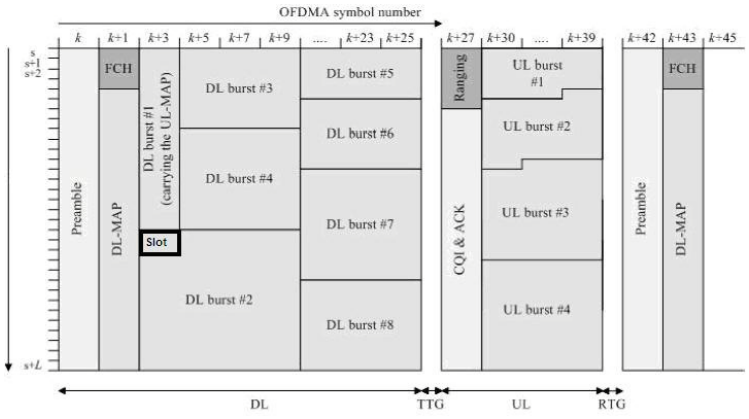
At the Medium Access Control (MAC) layer, WiMAX organizes information in Packet Data Units (PDU), as shown in Figure 7. Each MAC frame begins with a fixed length MAC header. This header may be followed by the load of the MAC PDU (MPDU). An MPDU may contain a field for Cyclic Redundancy Check (CRC).

MAC Header 6 bytes	Payload (Optional)	CRC (optional) 4 bytes

Figure 7.- MAC Frame General Format (Std. IEEE 802.16-2004 [12]) .

This way we can determine that the size of an active voice packet for the AMR codec is 47 bytes, and for an inactive voice packet is 21 bytes. As for G.719 (with a rate of 32 kbps) is of 95 bytes.

Frame structure and allocation of Mobile WiMAX traffic. The 802.16e standard provides different physical layer modes and configurations of radio channel [13]. In the TDD mode, the data mapping is done in two dimensions: time and subcarriers, where an OFDMA frame has a duration of 5 ms subframe comprising a downstream (downlink: DL) subframe and an uplink (UpLink: UL) as shown in Figure 8 [3].



courtesy of WiMAXFORUM

DL.- DownLink UL.- UpLink

TTG.- Transmission Time Interval RTG.- Receive/transmit Transition Gap

Figure 8.- Structure of 802.16 TDD Frame

The number of PDUs that can be accommodated in these sub frames depends on the modulation scheme, coding rate and channel quality factor (G).

4. COMPUTATIONAL MODULE

The application was developed on the Microsoft Visual Studio 2008 with the C # programming language [14]. The main assumptions adopted are as follows:

- Use the OFDM WiMAX TDD frame with duration of 5 ms, bandwidth of 5 and 10 MHz and the PUSC permutation mode (Partial Usage of the subchannels).
- Apply the operating characteristics of the voice codecs G.719 and AMR to generate voice traffic.
- To calculate the VoIP traffic load in both encoders applies a simplified On-Off model (on-off).
- For AMR, during periods of active conversation, it uses the highest bit rate, this is 244 bits/20 ms, and during periods of inactivity, a rate of 56 bits / 160 ms.
- For G.719, during periods of active conversation, using the highest bit rate, this is 160 bytes/20 ms, and during periods of inactivity rate 80 bytes/20 ms.

Figure 9 shows the capture screen and output of the module for the calculation of operating parameters and efficiency in the transport of VoIP.
The blank fields offer choices, these are detailed below: a) bandwidth (5 or 10 MHz), b) G parameter (1 / 4, 1 / 8, 1 / 16, 1 / 32), c) modulation schemes (BPSK, QPSK, 16QAM and 64QAM) d) Coding rate (1 / 2 CTC, 3 / 4 CTC 2 / 3 CTC, 5 / 6 CTC), e) Number of retransmissions (1, 2, 4), f) voice decoder (AMR, G.719), g) Detection of voice (detection or no detection), h) data rate voice packets (12.2 kbps, 10.2 kbps,7.95 kbps, 7.40 kbps, 6.70 kbps, 5.90 kbps, 5.15 kbps, 4.75 kbps and 1.80 kbps for the AMR codec, 32 kbps and 64 kbps for the G.719 codec).
Pressing the Calculate button, the module calculates: a) The operating parameters of WiMAX, b) length (bytes) of a package of WiMAX Voice, c) the number of slots and OFDM symbols used by the voice packet, d) gross rate (bits / sec.) transmission of voice, e) the rate for IP transport, f) rate of speech (without header), g) percent Efficiency, h) the percentage of occupation of a WiMAX frame.

The expressions for calculating the above data can be founded in [15].

Efficiency calculations

The most important formulas from the perspective of performance analysis are:

- Efficiency of WiMAX VoIP package is:
  Efficiency = Packet Size / Voice Packet size adapted for WiMAX

- Occupation of a VoIP PDU TDD frame is:
  % Occupancy = Packet Size adapted for WiMAX / TDD Down Link Frame Size

- Maximum number of simultaneous calls a single TDD frame can support is:
  No. calls per frame = 1 / Occupancy Rate

Substituting the appropriate values in the above formulas, we can calculate the efficiency, the occupancy rate and the number of calls per frame for both AMR to G.719

5. PERFORMANCE ANALYSIS OF VOIP OVER WIMAX

Because WiMAX was originally conceived as a means of transporting data, WiMAX versions 1.0 and 1.5, show improvements in the treatment of VoIP, which is particularly important for several reasons: (a) The voice is a time service actual maximum tolerance delay of 200 msec. (b) Vocoder digital frames are very small (about 264 bits), compared to data services, leading to proportionally larger control headers compared to the payload.

The module "VoIP_over_WiMAX" calculates and displays the operating parameters according to Figure 8. To compute the performance of WiMAX systems for voice transportation, we focus on the following figures of merit.

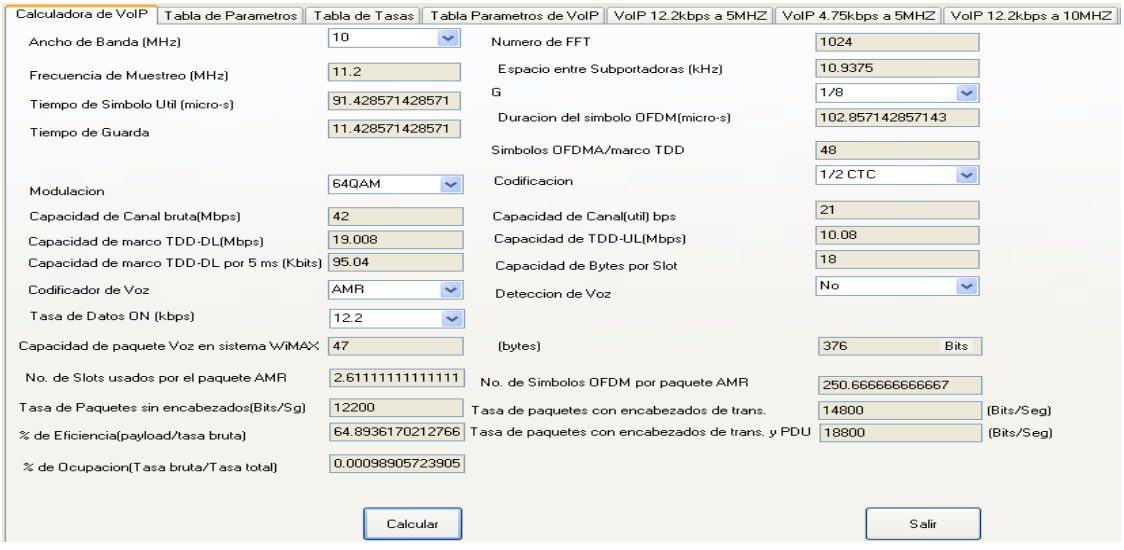


Figure 9.-Screen to enter the operating parameters of WiMAX

- Frame occupation percentage (FOP): The ratio of the number of bits in a PDU (VoIP packet with WiMAX headers, IP headers and payload) to the total number of bits of a TDD frame.
- Packetization efficiency percentage (PEP): The ratio of the payload length (bits) to the PDU's total number of bits.
- Net VoIP Rate (NVR): Is the total number of voice and control bits sent in a one second period by the WiMAX system.
- Number of calls per frame (NCPF): Maximum capacity of VoIP calls a a TDD-DL frame can accommodate, if it were to transports only VoIP traffic. Obtained by dividing the total capacity of a TDD-DL frame by the FOP.

In the tables below FOP, PEP, NVR and NCPF calculations are presented according to the operational characteristics of 1.0 and 1.5 Releases, and AMR/ G.719 voice codecs.
Given a bandwidth of 5 MHz for TDD-DL transmission, channel coding rate and G parameter are fixed. Table 2 lists the total capacities in Mbps and the resulting FOP when the modulation scheme is changed.

In the AMR case, it is clear that when upping the levels of modulation, there is a better occupancy of spectrum capacity, which means that the same VoIP rate adopts lower FOP figures. Results vary from 1.287% with BPSK to 0.214% with 64QAM. Similar behavior is observed for AMR in version 1.5, and G.719 in Releases 1.0 and 1.5.
On the other hand, the change in version 1.0 to 1.5 provides a slight gain by reducing the FOP in both AMR and G.719.
Since G.719 is a full band audio encoder (up to 20 KHz), while AMR is a narrowband codec (up to 4 KHz), the coded frames of the former are longer, and therefore have a greater FOP.
.

Table 3. Percentage efficiency of bundling and VoIP Average Gross Rate

	AMR		G.719	
	1.0	1.5	1.0	1.5
PEP (%)	32.35	35.04	41.414	42.68
TBPV (bps)	11,490	10,050	23,010	21,570

PEP and NVR parameters are independent of the modulation scheme, and are shown in Table 3 for the same conditions as in Table 2.

The change from 1.0 to 1.5 in terms of transport efficiency is insignificant in both versions, as we can see comparing 32.35 vs. 35.04 and 41.41 vs. 42.68.

As explained above, the improvement lies in that in 1.5 the initial packet carries WiMAX headers that identify the flow, and is omitted in all successive packets.

NVR values of Table 4 consider an activity/silence relationship of 60/40. It is found that Release 1.5 decreases 1,440 bps the data volume transmitted by both the AMR and the G.719 encoders compared to Release 1.0.

Maintaining the same encoding, bandwidth and G conditions, Tables 4 and 5 provide estimates of the number of calls per TDD frame for AMR and G.719 ITU-T, respectively.

Table 4. Number of calls per TDD frame with the AMR encoder

		1.0		1.5	
Modulation	Data Rate (Mbps)	% Occup.	NLLM	% Occup.	NLLM
BPSK	3.168	1.287	77	1.188	84
QPSK	6.336	0.643	155	0.594	168
16QAM	12.672	0.321	311	0.297	336
64QAM	19.008	0.214	467	0.198	505

Viewed from the perspective of call capacity, the TDD-DL frame could accommodate 77 simultaneous conversations in the AMR lower level of modulation case, up to 467 in the highest level. In this last comparison it is necessary to mention the fact that capacity is being measured in a single frame, which is not equivalent to the total capacity, as voice frames are generated with a 20 ms periodicity and TDD frames every 5 ms.

Table 5. Number of calls per TDD frame with the G.719 encoder

		1.0		1.5	
Modulation	Data Rate (Mbps)	% Occup.	NLLM	% Occup.	NLLM
BPSK	3.168	2.5	40	2.425	41
QPSK	6.336	1.25	80	1.212	82.5
16QAM	12.672	0.625	160	0.606	165
64QAM	19.008	0.417	239	0.404	247

The increase in the number of G.719 calls is less notable than in the AMR case, which is attributable to the denser digital volume of G.719 compared to that of AMR.

Table 2. Occupancy rate of the TDD-DL frame for a VoIP package

				AMR[1]		G.719[2]	
				1.0	1.5	1.0	1.5
Modulation	Encoding	G	Data Rate (Mbps)	POM (%)	POM (%)	POM (%)	POM (%)
BPSK	½	1/8	3.168	1.287	1.188	2.5	2.425
QPSK	½	1/8	6.336	0.643	0.594	1.25	1.212
16QAM	½	1/8	12.672	0.321	0.297	0.625	0.606
64QAM	½	1/8	19.008	0.214	0.198	0.417	0.404

1. AMR @ 12.2 Kbps 2. G.719 @ 32 Kbps

6. CONCLUSIONS

- The VoIP_over_WiMAX module calculates the performance in the transport of VoIP for any combination of settings (bandwidth, modulation technique, channel condition, etc.) considered in the WiMAX standards 1.0 and 1.5. Includes AMR and G.719 ITU-T coders parameters. And the possibility of integrating other codec formats if necessary.
- The transport efficiency of VoIP over WiMAX 1.0 is 32.32, which is slightly improved by Release 1.5 to 35.04. A similar improvement is obtained in the case of the G.719 coder. The figures allow us to affirm that WiMAX is not efficient in terms of the relation payload to control headers. This situation can be improved with the addition of more voice frames in a single PDU, with the risk of increasing the BER in case of packet loss.
- The capacity calculation of calls that can be transported within the same TDD-DL frame shows the enormous flexibility and capability of WiMAX, since its base number is 77 and can grow up to 467 (with conditions indicated in Section V).
- Issues to consider in the improvement of this tool include the calculation of loading and efficiency of VoIP sessions at cell level. Consider the effect of VoIP packet loss and eventual re-transmission.

Acknowledgments. This work was supported by the IPN through grant SIP-IPN 2010-0060 and COFAA Exclusivity Scholarship.

7. REFERENCES

[1] Kelly Riddell, Amy Thomson, "IPhone Network Jams Open Market for Time Warner Cable", *Bloomberg Business week*, March 8, 2010.

[2] Federal Communications Commission, "Mobile Broadband: the Benefits of Additional Spectrum", *OBI Technical Paper Series*, No. 6, October 2010, pp. 7.

[3] Mo-Han Fong and Robert Novak, Sean McBeath, Roshni Srinivasan, "Improved VoIP Capacity in Mobile WiMAX Systems Using Persistent Resource Allocation", *Nortel Networks, Huawei Technologies, Intel Corporation*, *IEEE Communications Magazine*, October 2008.

[4] Maravedis, *4G Digest*, Volumen 6, No. 12, February 9, 2011.

[5] Maravedis, *4G Digest*, Volumen 6, No. 10, January 12, 2011.

[6] IEEE, "IEEE Standard for Local and Metropolitan Area Networks, Air Interface for Fixed Broadband Wireless Access Systems", *IEEE 802.16-2004*, October 2004.

[7] Marcos Katz, Frank Fitzek, Eds., "WiMAX Evolution: Emerging Technologies and Applications", *John Wiley & Sons, Ltd.*, ISBN: 978-0-470-69680-4, 2009, pp. 3.

[8] WiMAX Forum™, "Mobile System Profile Release 1.0", Approved Specification Revision 1.4.0, 2 Mayo 2007.

[9] International Telecommunication Union, Press Release, "ITU Radiocommunication Assembly approves new developments for its 3G standards", http://www.itu.int/newsroom/press_releases/2007/30.html, October 19, 2007.

[10] AMR speech CODEC; General description. (Release 8). 3GPP, TS 26.071 V8.0.0, December 2008.

[11] ITU-T, "G.719: Low-complexity, full-band audio coding for high-quality, conversational applications", *ITU-T,* June 2008.

[12] IEEE, IEEE Std. 802.16-2004, 2004.

[13] Fan Wang, Amitava Ghosh, Chandy Sankaran, Philip J. Fleming, Frank Hsieh, Stanley J. Benes, "Mobile WiMAX systems: Performance and Evolution", *Networks Advanced Technologies, Motorola Inc., IEEE Communications Magazine*, October 2008.

[14] Charles Petzold, "Programming Microsoft Windows with C#", *Microsoft Press*, 2002.

[15] Loutfi Nuaymi, "WiMAX: Technology for Broadband Wireless Access", *John Wiley & Sons*, ISBN:9780470028087, 2007

Proceedings of the International Conference on Information and Communication Technologies and Applications (ICTA 2011), and the International Conference on Design and Modeling in Science, Education, and Technology (DeMset 2011)

Scheduling Active Nodes of Clusters in WSNs to Minimize Energy

Zixiang Wang, Senlin Zhang, Meikang Qiu and Meiqin Liu

***Abstract*— Minimizing energy is a challenge problem in *Wireless Sensor Networks* (WSNs). Aggregation Nodes (AGNs) in the WSNs implement the preliminary data fusion and packets relay. Since the packets are transmitted to the closest AGNs at first, the AGNs divide the networks into several clusters. Sensors are usually uniformly deployed in the sensing area, and the distances between sensor nodes and AGNs are different. Under a certain fusion performance constraint, not all sensor nodes need to be active. In this paper, the knowledge of energy balancing and parameter estimation is employed to reduce energy consumption. We propose an algorithm to schedule the active sensor nodes for each cluster in WSNs. Both the number of active sensor nodes and data length can be obtained through our algorithm. Our method properly assigns the active sensors with different distances to the AGN of a cluster, and the total energy consumption of the cluster is reduced employing our scheme. Experimental results demonstrate the effectiveness of our approach.**

***Index Terms*— WSN, energy, MSE, active nodes, probability**

I. Introduction

Wireless Sensor Networks (WSNs) develop very fast during these years. The networks consist of multiple sensors are introduced to military, environment monitoring, health, and many other areas. With the development of digital electronics, the sensor nodes in WSNs are smaller in size, more accurate in detection and faster in data processing. The sensors in WSNs are embedded with batteries. In many situations, the batteries are impossible to be replaced. Today energy harvesting technologies [1] that enable sensors convert ambient energy to electric energy is emerging. But due to some external limitation, energy harvesting cannot provide stable energy. Therefore, minimizing energy consumption is an important problem in the applications of WSNs.

Reducing energy consumption can be achieved in many aspects. The authors in [2] optimized energy arrangement regarding the architecture design of a sensory node controller on the hardware. The system usually does not require all the sensor nodes to be active. Some researches focus on the management of sensor state while satisfying the performance index. The schemes of the sleeping sensor nodes of a cluster-based sensor networks are proposed in [3], [4]. In [5], authors gave solution to balance energy consumption in data-gathering networks. The authors in [6], [7] worked on sensor networks with multiple states. Based on the tradeoff between energy, deadline and reward, tasks can be rotated to run relatively critical applications while meeting energy and time constraints [6]. By scheduling sensor nodes with some different active modes and a sleeping mode, energy consumption can be reduced [7]. In the work of [8], authors discussed the energy minimization on an in-line topology, and calculated the optimal transmission distance between two sensor nodes.

It is known that the energy consumption in transmitting a packet experiences a path loss [9], which is proportional to the α-th power ($\alpha > 1$) of transmission distance. The multihop and *Aggregation Nodes* (AGNs) are usually employed in WSNs to reduce total energy consumption. In that scheme, sensor nodes transmit the data to the nearest AGN at first. The ANGs fuse the data and relay it to the next AGN. The covered region of an AGN can be treated as a cluster and the AGN is the clusterhead. The cost to transmit a certain packet to the AGNs is different due to the sensor's location. The approaches proposed early [3], [5] implemented the energy balancing in WSNs, which scheduled sensors to achieve energy balancing and maximize the lifetime of WSNs. But these researches with energy balancing usually did not take reward into consideration. In the signal process of WSNs, *Mean Square Error* (MSE) is the most important reward. Approaches that minimize energy consumption in WSNs with MSE constraint [10], [11] are proposed to obtain optimal data length and number of active sensor nodes. Yet the approaches did not give us the solution to schedule sensors with different distances to the clusterhead. In a WSN system, sensors located in different areas need to be assigned in a balanced way. Therefore, a scheme that minimizes energy consumption and considers sensor's location at the same time is significant.

In this paper, we propose an algorithm to schedule active nodes in one cluster considering the sensor's location. Our target is to minimize energy consumption with both MSE constraint and energy balancing. In long distance transmission, the data transmitted are quantized to several bits in order to save the transmission energy. To reach an ideal fusion result and implement an accurate parameter estimation, the data need to have more bits. But longer data length increases the cost in transmission. Sometimes that more sensor nodes provide data with less bits gives better fusion quality. The algorithm we propose in this paper will give us the optimal data length and number of active nodes. We schedule the active sensor nodes with different transmission distances. The sensors of the same cluster will exhaust in a balance way employing our scheme. The contributions of this paper are as the follows:

- First, we obtain the optimal transmission data length and number of active sensor nodes with numerical solutions.

This work was supported in part by the NSFC under Grants 61071061 and 60874050, the Program for NCET in University under Grant NCET-10-0692, the Zhejiang Provincial NSF of China under Grants R1100234 and Z1090423, the Research Project of ZPED under Grant Z200909334, the Fundamental Research Funds for the Central Universities under Grant 2009QNA4012, the Fund of Aeronautics Science under Grant 20102076002, and the ASFC under Grant 20102076002.

- Second, we schedule active sensor nodes to balance the energy consumption of every sensor distributed in the same cluster.
- Third, Our approach can reduce the total energy consumption by scheduling active sensor nodes.

The paper is organized as follow: In Section II, we address the problem to be solved. We proposed an algorithm to obtain the optimal active sensor nodes number and data length with MSE constraint of a simple network structure in Section III. In Section IV, we extend the type of WSN and schedule the active sensors. In Section V, we give the experimental results. The conclusion is shown in Section VI.

II. Basic Concepts and Models

The problem we discuss in this article should satisfy the following assumptions:

- All the sensors in the WSN are homogeneous. They have same detection probability, process speed, memory size, and consume equal energy in the same mode.
- The constants included in this paper are always the same.
- In the process of data acquisition, the actual value is corrupted by additive noise. The noises are zero mean, and with same variance σ^2.
- The packets relay is implemented by the AGNs, so the sensor nodes will send packets directly to the AGNs.

There are two main parts energy consumption in WSNs: circuit energy and transmission energy. Circuit energy is the energy used to sustain the normal function of a sensor. The energy cost in data transmission is the transmission energy. In our work, we assume the sensor nodes have two modes: active and sleeping. The energy dissipation of sensor nodes in sleeping mode is quite little. Therefore, we neglect the energy consumption in sleeping mode. The functions a sensor needs to accomplish are detection, quantization and transmission if active.

With the observed value z_k, the fusion center in WSN needs to make an accurate estimation on the parameter θ we are curious about. As we assumed, all the sensors have the same detection probability, then more active sensor nodes gives better fusion result in the fusion center [12]. According to the accuracy requirement, the observed value should be quantized to more bits data [10]. Assume the time slot T is the time transmitting a packet and one packet contains L bits quantized data. If the observed value of sensor S_k is quantized to L_k bits, the energy consumption E_k every time slot T can be presented as follow [13]

$$E_k = ca_k(2^{L_k} - 1) + E_a, \tag{1}$$

where $a_k = d_k^{\alpha}$, c is constant during transmission, E_a denotes the circuit energy, d_k is the distance between two hops, and α is the path loss exponent. The total energy consumption per time slot is

$$E_{total} = \sum_{S_k \in S_r} E_k = \sum_{S_k \in S_r} (ca_k(2^{L_k} - 1) + E_a), \tag{2}$$

where S_r represents the sensor set whose sensor elements are required in the data fusion.

From Eq. (2), we know both elements in S_r and transmission data length L have influence on the total energy E_{total}. Hence, the problem is that under the fusion quality constraint, the data length and active nodes number should be obtained in order to achieve minimum cost. The energy discrepancy due to the transmission distance cause the imbalance of sensor nodes' lifetime. So we also need to determine the number of active nodes with different transmission distance to balance energy dissipation.

III. Scheduling for the Circle-Based Cluster

In this section, we consider a special kind of cluster, whose distance between sensor nodes and the clusterhead (AGN) are the same. All sensors are on a circle of a certain radius R centered on the AGN as shown in Fig. 1. We call it circle-based cluster. From Eq. (2), we know

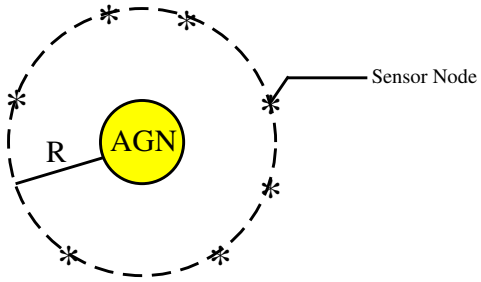


Fig. 1. This is a very simple WSN, the distances between all sensors and fusion center are same.

energy dissipation of all active sensor nodes is same in the circle-based cluster.

Raw observed value is impractical to store in the local memory of sensor nodes, and directly transmit observed data is high energy cost. Quantization is required to prolong the WSN's lifetime and improve the system's performance. Suppose the observed value z_k is bounded to [-W, W], and we uniformly divide [-W, W] into 2^{L_k} intervals. Then z_k is round to the nearest endpoint of the 2^{L_k} intervals. Constraint by the embedded chip in sensors, the data length cannot be too long. The number of quantization bits has an upper bound. Quantization brings quantization error, so the quantization bits of each node should not be too small. Then there is a range of the quantization bits: $L_{min} \le L_k \le L_{max}$. We employ quasi-BLUE estimator [10] to obtain an estimated value $\hat{\theta}$ in the fusion center. The MSE of quasi-BLUE estimator is

$$E(\hat{\theta} - \theta)^2 = \Big(\sum_{S_k \in S_r} \frac{1}{\sigma^2 + \delta_k^2} \Big)^{-1}, \tag{3}$$

where $\delta_k^2 = \frac{W^2}{(2^{L_k}-1)^2}$.

Suppose D_r is the MSE required, i.e.,

$$\Big(\sum_{S_k \in S_r} \frac{1}{\sigma^2 + \delta_k^2} \Big)^{-1} = D_r. \tag{4}$$

Then we have

$$L = log_2\Big(\frac{W}{\sqrt{n_a D_r - \sigma^2}} + 1 \Big), \tag{5}$$

$$n_a \geq \frac{\sigma^2}{D_r}, \tag{6}$$

where L denotes the data length (L bits), n_a denotes the number of active sensor nodes. Insert Eq. (5) into Eq. (2), the energy dissipation is

$$E_{total}^{D_r} = \frac{cd^{\alpha} n_a W}{\sqrt{n_a D_r - \sigma^2}} + n_a E_a. \tag{7}$$

$$n_{aopt} = \arg\min_{n \in N^+} [E_{total}^{D_r}]. \tag{8}$$

Thus, $L_{opt} = log_2(\frac{W}{\sqrt{n_{aopt} D_r - \sigma^2}} + 1)$.

Lemma 1: The function $E_{total}^{D_r} = \frac{n_a W}{\sqrt{n_a D_r - \sigma^2}} + n_a E_a$ has only one extreme point.

Proof:

$$E_{total}^{D_r} = \frac{cR^{\alpha} n_a W}{\sqrt{n_a D_r - \sigma^2}} + n_a E_a$$
$$= \frac{cR^{\alpha} W}{\sqrt{-\frac{1}{n_a^2}\sigma^2 + \frac{1}{n_a} D_r}} + n_a E_a \tag{9}$$

Obviously, $\frac{cR^{\alpha} W}{\sqrt{-\frac{1}{n_a^2}\sigma^2 + \frac{1}{n_a} D_r}}$ $(n \in N^+)$ has only one extreme point. $n_a E_a$ is a monotonic increasing linear function. Therefore, $E_{total}^{D_r}$ has only one extreme point. ■

Because the active sensor node number n_a is discrete, we cannot calculate n_{aopt} through derivation. From Lemma 1, we know there exists an extreme point of $E_{total}^{D_r}$. So we design an algorithm to search the optimal value n_{aopt} in a loop way. The total energy $E_{total}^{D_r}$ is depended on the active nodes number n_a and data length L. There is a linear relationship between $E_{total}^{D_r}$ and n_a. While the relationship between $E_{total}^{D_r}$ and L is exponential. To accelerate the searching process, we choose L as our variable. The algorithm is shown in Algorithm 1. We set the initial value $L = 1$. The corresponding energy can be calculated with Eq. (7). We gradually increase L and calculate the corresponding $E_{total}^{D_r}$ until the energy consumption is larger than the last time. Then we find the optimal data length L_{opt}. The corresponding optimal active nodes n_{aopt} can be obtained with L_{opt}.

Algorithm 1 Optimal Data Length Algorithm

Require: Observed value bound W, MSE constraint D_r and noise variance σ^2;
Ensure: Optimal data length L_{opt} and optimal number of active nodes n_{aopt};
1: Initialize $L = 1$;
2: $n = \lfloor \frac{\sigma^2 + \frac{W^2}{(2^L - 1)^2}}{D_r} \rfloor$; % function $\lfloor * \rfloor$ represents the nearest interger
3: Calculate $E_{total}^{D_r}$ with Eq. (7);
4: do;
5: $E_{total}^{D_r} 0 = E_{total}^{D_r}$;
6: $L = L + 1$;
7: Calculate $E_{total}^{D_r}$;
8: while $(E_{total}^{D_r} < E_{total}^{D_r} 0)$;
9: $L_{opt} = L - 1$;
10: if $L_{opt} < L_{min}$;
11: $L_{opt} = L_{min}$;
12: else if $L_{opt} > L_{max}$;
13: $L_{opt} = L_{max}$;
14: $n_{aopt} = \lfloor \frac{\sigma^2 + \frac{W^2}{(2^{L_{opt}} - 1)^2}}{D_r} \rfloor$;
15: Output results L_{opt} and n_{aopt}.

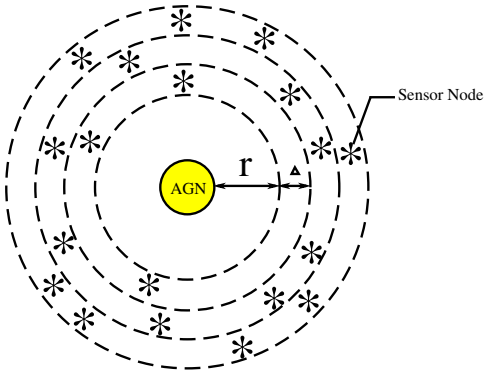


Fig. 2. The Multiple-Corona Structure Cluster

IV. Active Sensor Assignment for Different Distances

The cluster type discussed in Section III is a very special topology which is not common in our daily life. The sensor nodes usually are randomly deployed over an area. Although the shapes of different clusters are different, their structure can be represented as multiple-corona type as shown in Fig. 2. We reckon sensor nodes in coronas with same width centered at the AGN no matter what cluster shape of a WSN is. The energy dissipation of one sensor $E_k = ca_k(2^{L_{opt}} - 1) + E_a$ is influenced by the distance between sensor node and AGN. To maintain the coverage of WSNs, we need to balance the energy consumption of sensors. If active sensor nodes are randomly assigned, the outside sensors will exhaust more quickly. The scheme that make every sensor in the same cluster exhaust almost at the same time is the optimal. In Section III, we find the optimal solution of n_{aopt}. In this section, we assign n_{aopt} sensors to different coronas. Assume the sensor nodes are uniformly distributed on the coverage area, hence, the node density of the cluster is

$$\rho = \frac{N}{\pi R^2}, \tag{10}$$

where N is the number of sensor nodes, R is the radius of circle and ρ is the node density. We divide the cluster into many coronas as shown in Fig. 2. We define the sensors in the corona with inner radius r centered on the AGN belong to the set S^r, $S^r \subseteq S$, where S is the set comprising all sensor nodes. The width of each corona is a small value Δ. We denote r_{min} the inner radius of the innermost corona. The number of sensor sets is $n_s = \frac{R - r_{min}}{\Delta}$. Our target is to determine the active sensor node number n^r of each S^r. The number N_{S^r} of sensor nodes of S^r is

$$N_{S^r} = \rho \iint r d\theta dr = 2\pi\rho \int_r^{r+\Delta} r dr$$
$$= \pi\rho(\Delta^2 + 2r\Delta). \tag{11}$$

The energy when all sensors in S^r are active is

$$E_r = \rho \iint [cr^{\alpha}(2^{L_r}-1)r + E_a r] d\theta dr$$
$$= 2\pi\rho[\tfrac{1}{\alpha+2}c(2^{L_r}-1)r^{\alpha+2} + \tfrac{1}{2}E_a r^2]|_r^{r+\Delta}$$
$$= 2\pi\rho\{\xi[(r+\Delta)^{\alpha+2} - r^{\alpha+2}] + E_a(\Delta^2+2r\Delta)\}, \quad (12)$$

where $\xi = \frac{1}{\alpha+2}c(2^{L_r}-1)$. Then, the average energy consumption of each sensor in S^r is

$$\bar{E}_r = E_r / N_{S^r}$$
$$= \frac{2\{\xi[(r+\Delta)^{\alpha+2} - r^{\alpha+2}] + E_a(\Delta^2+2r\Delta)\}}{\Delta^2+2r\Delta}. \quad (13)$$

We assume the total energy of the battery embedded in a sensor is $\bar{E}$. Therefore, the total energy of the sensor set S^r is $\bar{E}N_{S^r}$. That means the sensors of S^r can provide energy of transmitting $\frac{\bar{E}N_{S^r}}{\bar{E}_r}$ packets. Suppose there are two sensor sets S^{r_1} and S^{r_2}. If the lifetime of two sets is almost the same, then the following equation is satisfied

$$\frac{1}{N^{r_1}}\frac{\bar{E}N_{S^{r_1}}}{\bar{E}_{r_1}} = \frac{1}{N^{r_2}}\frac{\bar{E}N_{S^{r_2}}}{\bar{E}_{r_2}}, \quad (14)$$

where N^{r_1} is the packets transmitted from S^{r_1} per unit time, N^{r_2} is packets transmitted from S^{r_2} per unit time. We call N^r $(r = r_1, r_2, r_3, ...)$ the packet transmission rate. There is a relationship between two packet transmission rates of sensor sets in a cluster. The ratio can be calculated by Eq. (15). For simplification, We define $\beta = N^{r_{min}}$, $N_\beta = N_{S^r}|_{r=r_{min}}$, $E_\beta = \bar{E}_r|_{r=r_{min}}$. Thus, the general form of packet transmission rate can be denoted as

$$N^r = \frac{N_{S^r}}{N_\beta}\frac{E_\beta}{\bar{E}_r}\beta. \quad (15)$$

By scheduling the active sensor nodes, we can balance the energy of sensors with different distances to the AGN. The active sensor nodes proportion allotted to different sensor sets can be obtained based on N^r.

Lemma 2: If we denote $n_{aopt}|_{d=r}$ the optimal active nodes number as the distance d between the nodes and the AGN is r, then $\forall r > 0$, $\exists \Delta > 0$, the following equation satisfies

$$n_{aopt}|_{d=r} = n_{aopt}|_{d=r+\Delta}.$$

Proof: Assume the MSE constraint is D_r, when $d = r$,

$$n_{aopt}|_{d=r} = sol\{minE^{Dr}_{total}|_{d=r}\} \quad (16)$$

$\forall \varepsilon > 0$, $\exists \Delta_1 > 0$,

$$|\, n_{aopt}|_{d=r+\Delta_1} - n_{aopt}|_{d=r}\,| = 1, \quad (17)$$
$$|\, n_{aopt}|_{d=r+\Delta_1-\varepsilon} - n_{aopt}|_{d=r}\,| = 0. \quad (18)$$

Thus, $\exists \Delta \in (0, \Delta_1)$,

$$n_{aopt}|_{d=r} = n_{aopt}|_{d=r+\Delta}. \quad (19)$$

■

Lemma 2 shows n_{aopt} and L_{opt} are same for all sensors in a sensor set if we choose proper Δ. In general , small Δ will be suitable. The maximum valid Δ can be obtained by Algorithm 2.

If we want to balance the sensor energy dissipation of different area, the active sensor nodes should be scheduled in a special probability way rather than random scheme. The probability P_{S^r} a sensor set be chosen by the AGN is

$$P_{S^r} = \frac{\frac{N_{S^r}E_\beta}{\bar{E}_r N_\beta}\beta}{\sum_{S^{r_k}\in S}\frac{N_{S^{r_k}}E_\beta}{\bar{E}_{r_k}N_\beta}\beta} = \frac{\frac{N_{S^r}E_\beta}{\bar{E}_r N_\beta}}{\sum_{S^{r_k}\in S}\frac{N_{S^{r_k}}E_\beta}{\bar{E}_{r_k}N_\beta}}. \quad (20)$$

We call P_{S^r} the activation probability.

Different sensor sets have different n_{aopt} and L_{opt}. n_{aopt} and L_{opt} are depended on the internal radius of a sensor set. With the given parameter W, σ and the MSE constraint Dr, we can calculate the n_{aopt} and L_{opt} of each sensor set. We denote $n_{aopt}(S^r)$ the n_{aopt} of sensor set S^r, $L_{opt}(S^r)$ the L_{opt} of sensor set S^r. With the activation probability P_{S^r}, we can schedule all the sensors in a cluster, and assign the active nodes to each sensor set properly. The sensor activation of the same sensor set is random. We propose the *Active Nodes Scheduling* (ANS) algorithm to schedule the overall sensors for a cluster. The ANS algorithm is shown in Algorithm 3.

Algorithm 2 Δ Acquisition Algorithm

Require: Minimum radius r_{min} and maximum radius r_{max};
Ensure: Corona width Δ;
1: Initialize $\Delta = 1$;
2: Calculate number of total coronas, $n_{corona} = \frac{r_{max}-r_{min}}{\Delta} + 1$;
3: Calculate L_{opt} of every inner radius with Algorithm 1;
4: for $(n = 0, n < n_{corona}, n++)$
5: if $(L_{opt}|_{r_{min}+n\Delta} - L_{opt}|_{r_{min}+(n+1)\Delta} > 1)$;
6: $\Delta = \Delta - d$, d is a small positive number;
7: Jump to Step 2;
8: Output Δ.

Algorithm 3 Active Nodes Scheduling (ANS) Algorithm

Require: Observed value bound W, MSE constraint D_r and noise variance σ^2;
Minimum radius r_{min} and maximum radius r_{max}.
Ensure: Optimal active nodes assignment;
1: Calculate Δ employing Algorithm 2;
2: Divide all sensors into $\lfloor \frac{r_{max}-r_{min}}{\Delta} \rfloor + 1$ sensor sets;
3: Obtain n_{aopt} and L_{opt} of each sensor set by Algorithm 1;
4: Calculate P_{S^r};
5: Determine the active sensors proportion of each sensor set with P_{S^r}.

V. Experimental Results

In this section, we present the experimental results for the algorithm proposed in the paper. In all simulations, we set the path loss exponent $\alpha = 2$ and the circuit energy $E_a = 1$ unit every time slot. First, we simulate the solution to find the optimal data length of Circle-Based cluster, then we show the result after the scheduling active sensor nodes of each corona with ANS algorithm.

A. Simulation for Circle-Based Cluster

According to Eq. (7), the number of active sensor nodes depends on the observed value bounds W, MSE constraints D_r, noise variance σ^2 and the transmission distance R. In all the following simulations, we set the transmission distance $R = 2$. We simulate our algorithm with different observed value bounds W, different MSE constraints D_r and noise variance σ^2. Our target is searching optimal data length and number of active sensor nodes.

We set $D_r = 0.5$, $\sigma = 1$ and $W = 4, 8, 16$, and 32. The optimal data length L_{opt} obtained by our algorithm are 2, 3, 4, and 5 respectively. The energy consumption is shown in Fig. 3. We set $W = 16$, $\sigma = 1$ and $D_r = 0.1, 0.2, 0.5$, and

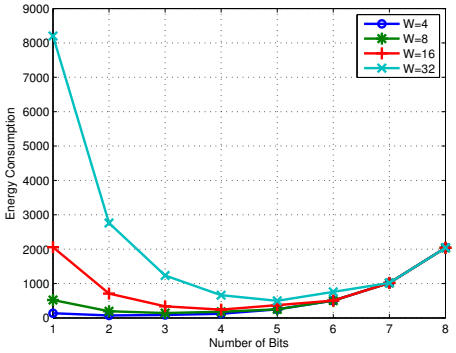


Fig. 3. The Energy Consumption of Different bounds W

1. The optimal data length L_{opt} obtained by our algorithm are 4, 4, 4, and 4. The energy consumption is shown in Fig. 4. We set $W = 16$, $D_r = 10$ and $\sigma = 0.5, 1, 2$,

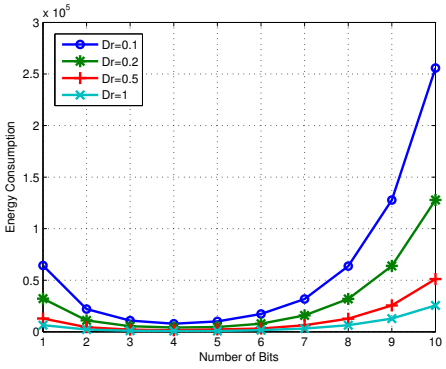


Fig. 4. The Energy Consumption of Different MSE Constraints D_r

and 5. The optimal data length L_{opt} obtained by our algorithm are 5, 4, 3, and 2. The energy consumption is shown in Fig. 5. The experimental results demonstrate our algorithm is effective in obtaining the optimal solution of data length L. The corresponding number of active nodes can be calculated by Eq. (4).

B. Simulation for Active Nodes Assignment

We divide a cluster into five coronas S_1, S_2, S_3, S_4 and S_5 from inside to outside. Assume the internal radius of the corona closest to the AGN $r_{min} = 1$. Through our algorithm, the corona width $\Delta = 1$ is obtained.

Assume there are 110 sensors in the WSN. If the sensors are uniformly distributed in the coverage of the WSN, there are about 10 sensors in S_1, 16 sensors in S_2, 22 sensors in S_3, 28 sensors in S_4 and 34 sensors in S_5. Assume battery embedded sensor can provide 10000 units energy. We assume $Dr = 0.5$, $W = 16$, $\sigma = 1$. The

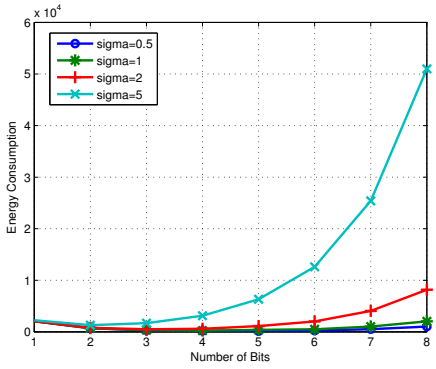


Fig. 5. The Energy Consumption of Different Noise Variance σ^2

optimal data length L_{opt} obtained by our algorithm is 4 bits.

If we employ random scheme to schedule the active sensor nodes, the total energy consumption is more and active nodes will be assigned in an imbalance way. Assume the AGN activates sensor nodes 10, 100, 500, and 1000 times, respectively, and the number of rest sensors is given in Table I. We can find more sensors

TABLE I
REST SENSORS

Activated 10 times						
Sensor Set	S_1	S_2	S_3	S_4	S_5	Sum
Random Scheme	9	16	22	28	34	109
ANS Algorithm	9	16	22	28	34	109
Activated 100 times						
Sensor Set	S_1	S_2	S_3	S_4	S_5	Sum
Random Scheme	9	16	22	28	26	101
ANS Algorithm	9	16	22	27	34	108
Activated 500 times						
Sensor Set	S_1	S_2	S_3	S_4	S_5	Sum
Random Scheme	9	15	17	6	4	51
ANS Algorithm	6	14	19	21	29	89
Activated 1000 times						
Sensor Set	S_1	S_2	S_3	S_4	S_5	Sum
Random Scheme	9	13	3	0	0	25
ANS Algorithm	4	4	3	7	7	32

alive employing our algorithm with the same condition, and our approach can help prolong the lifetime of sensors. We record the energy consumption every 10 time slots, and plot the their value each moments. The energy consumption of our algorithm and random scheme is shown in Fig. 6. If we schedule active nodes with our approach, the total energy consumption will be less.

In the Table I, there are more sensors rest in the outside sensor sets employing ANS algorithm. This is important to balance the energy consumption for each cluster in the WSN. Because outside corona is larger than the inside, a good schedule need to maintain more sensors in the outside sets. The contrast can be seen in Fig. 7 and Figure 8. We assume the AGN activates sensors randomly from the five sensor sets for 1000 times and repeat the process 50 times. The rest sensors distribution is shown in Fig. 7. From Fig. 7, we can clearly discover that the sensors outside are sparser than that of the inside. We repeat the

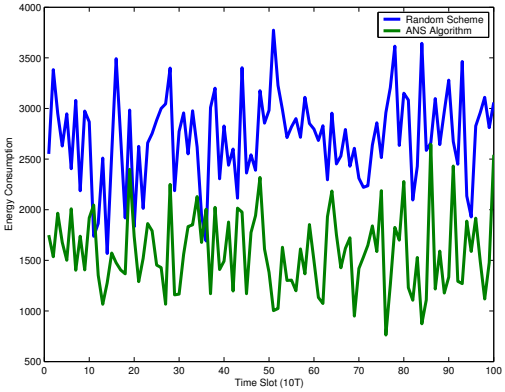


Fig. 6. The blue line is the energy consumption of random scheme, while the green line is the energy consumption of our algorithm.

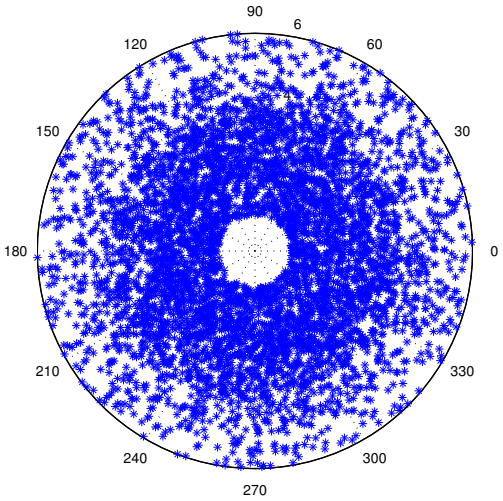


Fig. 7. The blue star in the circle means the sensors alive after the AGN randomly activated sensors in the WSN for 1000 times.

simulation with the algorithm we proposed, the result is shown in Fig. 8. The sensors still alive are almost uniformly distributed in the sensing area, and we can see the sensors remain in the WSN are more in quantity compared to that in Fig. 7.

VI. Conclusion

In this paper, we focused on the energy-saving problem for every cluster in WSNs. Our goal was to minimize the total energy with a given MSE constraint. For the networks whose sensors were uniformly distributed, the imbalance of energy dissipation caused by the transmission distance was considered. We proposed an algorithm to obtain the optimal number of active sensor nodes and transmission data length in the circle-based cluster. To the more general clusters, we gave a solution to schedule the active sensors with different distances to the AGNs. Our approach balanced the energy consumption compared to the conventional random scheme. The algorithm we proposed to schedule sensors can assign the sensors in

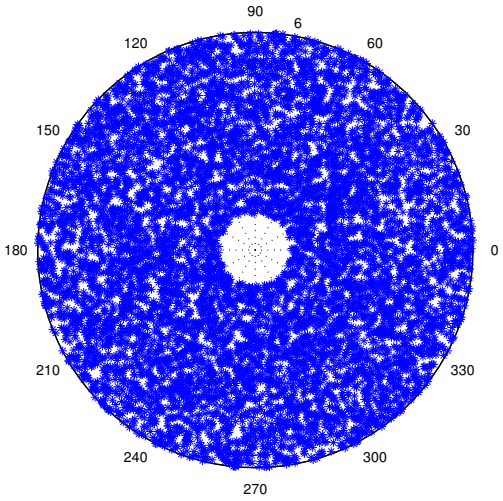


Fig. 8. The blue star in the circle means the sensors alive after the AGN activated sensors in the WSN with our scheme for 1000 times.

a near-optimal way. The experimental results proved that our algorithm was efficient.

References

[1] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials*, pp. 11–19, 2010.

[2] R. Gao and Z. Fan, "Architectural design of a sensory node controller for optimized energy utilization in sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 2, pp. 415–428, 2006.

[3] J. Deng, Y. Han, W. Heinzelman, and P. Varshney, "Scheduling sleeping nodes in high density cluster-based sensor networks," *Mobile Networks and Applications*, vol. 10, no. 6, pp. 825–835, 2005.

[4] Y. Wang, K. Huang, C. Lin, and C. Hung, "The optimal sleep control for wireless sensor networks," in *2009 Joint Conferences on Pervasive Computing*, 2009, pp. 95–102.

[5] H. Zhang and H. Shen, "Balancing energy consumption to maximize network lifetime in data-gathering sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 10, pp. 1526–1539, October 2009.

[6] C. Rusu, R. Melhem, D. Mosse, "Maximizing the system value while satisfying time and energy constraints," *IBM Journal of Research and Development*, vol. 47, no. 5/6, pp. 689–702, September/November 2003.

[7] M. Perillo and W. Heinzelman, "Optimal sensor management under energy and reliability constraints," in *Wireless Communications and Networking*, 2003, pp. 1621–1626.

[8] F. Shebli, I. Dayoub, A. M'foubat, A. Rivenq, and J. Rouvaen, "Minimizing energy consumption within wireless sensors networks using optimal transmission range between nodes," in *2007 IEEE International Conference on Signal Processing and Communications*, November 2007, pp. 24–27.

[9] S. Cui, A. J. Goldsmith, and A. Bahai, "Joint modulation and multiple access optimization under energy constraints," in *Global Telecommunications Conference*, December 2004, pp. 151–155.

[10] J. Xiao and Z. Luo, "Universal decentralized estimation in an inhomogeneous sensing environment," *IEEE Transation on Information Theory*, vol. 51, no. 10, pp. 3564–3575, October 2005.

[11] J. Li, G. AlRegib, "Distributed estimation in energy-constrained wireless sensor networks," *IEEE Transactons on Signal Processing*, vol. 57, no. 10, pp. 3746–3758, October 2009.

[12] M. Barkat, *Signal Detection and Estimation*. Artech House, 2005.

[13] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, September 2005.

Applying Scientific Research Skills to Teaching in a New Domain: A Case Study

Annie HUI
Northern Virginia Community College
6901 Sudley Road, Manassas, VA 20109

ABSTRACT

This paper presents a case study of a new instructor's research on learning how to teach in a new domain. Two components are identified as essential to good teaching. They are a good teaching methodology, and a good standard on the materials to be taught. With the goal to stimulate exploration rather than unidirectional information transfer, a teaching methodology that encourages collaborative learning is drafted out. This methodology is implemented on a 200-level computer organization course to test for its effectiveness. The course materials are then evaluated with reference to existing courses on the same subject.

Keywords: teaching method, collaborative learning

1. INTRODUCTION

When an instructor is teaching in a field of her own expertise, she has a large pool of knowledge to draw from. Typically, such an instructor knows the subject in far greater depth and breadth than her students do. This wide gap, together with her general skills in discovering and reasoning about such knowledge, allows the instructor to define a clear scope for the students to explore. Therefore, the students develop a sense of security. They believe that the instructor is able to guide them during their exploration. Perhaps more implicitly, they also believe that the instructor is able to address most of their questions regarding the subject. In other words, the instructor serves not only as a facilitator of learning, but also as a handy source of expert knowledge.

This is not the case when the instructor is newly introduced to the field herself. In this situation, this knowledge gap between the instructor and the students may only be marginal. The instructor and the students may often find themselves navigating together in territories that are uncharted to the both of them.

With this being an inevitable situation for anyone who ventures into a new field, the challenge is then: *How to make good use of this situation?* More specifically, the challenge for the instructor is how to make good use of the "I-don't-knows" while keeping the students engaged in class and guaranteeing that the students' learning outcomes meet an objective standard. It is worth noting that while this situation is only a transitional experience for educators, it is an everyday experience for researchers. Therefore, some skills that a researcher possesses to manage her day-to-day navigation in the uncharted territories of knowledge may be useful for a new educator.

This paper presents a systematic attempt to meet this challenge. The instructor has identified two important components in the making of a good course. The first is the effectiveness of the teaching methodology. The second is the quality of the materials being taught. Both components will be examined in this paper in the context of a 200-level computer organization course with a given syllabus. The remainder of this paper is organized as follows. Section 2 outlines the methodology used to design the specific course. Section 3 reports the outcomes of a one-semester experiment of applying this methodology to the course. Section 4 discusses the effectiveness of this methodology based on the findings of the experiment. Section 5 examines the quality of course with reference to a survey of 21 existing courses on the same subject. Section 6 concludes this work with a discussion of its implications.

2. METHODOLOGY

The fundamentals

In brainstorming an approach to learn collaboratively, the following points are found to be important:

- The need for a minimal set of clearly defined knowledge
- Encouragement on brainstorming and technical reading
- A constant refinement of the subject roadmap
- The ability to address the difficult questions raised by the students

Each of these points is elaborated below.

The need for a minimal set of clearly defined knowledge: The quality of the students' education is important. It is necessary to define a minimal set of knowledge that the students must master. This set of knowledge must be specific and measurable, instead of general and un-testable. For example, "to know the fetch-decode-execute cycle" is general. "To describe the fetch-decode-execute cycle through a trace of how the data is moved among the registers during the execution of one instruction" is specific. The students are expected to acquire this set of basic knowledge either through classroom instruction, or through self-reading. Towards this ends, it is helpful for an instructor new to a subject to select a textbook.

Encouragement on brainstorming and technical reading: Brainstorming is the process, in which a problem is explored in multiple perspectives, ideas are proposed, and alternative solutions are examined. While this process is easy to conduct in a small group discussion, to conduct effective brainstorming in a class size of 20 to 25 requires some careful preparation. One useful technique for increasing the depth of class preparation is to assign students with an open-ended, challenge

problem in a homework which is due before the class discussion. Additionally, class preparation may include the assignment of a scholarly or technical paper for the students to read and comment on. Such a reading exercise not only sharpens the focus of an in-class discussion, but also develops the technical reading skills of the students.

Constant refinement of the subject roadmap: Coursework requires a well-defined scope. No matter how exciting a topic may be, a class simply does not have the liberty to run wild on its discussion at the expense of other topics. Therefore, it is important for the instructor to help the students develop a roadmap on the subject. The purpose of this subject roadmap is to put topics into perspectives and to establish connections among them. As the class makes progress in its exploration of the subject, this map needs to be constantly refined.

Ability to address the difficult questions raised by the students: It has been observed that students are stimulated to pursue a subject with greater fervor, when their questions, especially the challenging ones, are taken seriously by the instructor and appropriately addressed. This may require extra reading, thinking, and writing for the instructor, but this approach has been shown to be worthwhile and to result in students that are more engaged in their study. This in turn suggests that the student is more concerned about being respected, taken seriously, and commended for his diligence than being "spoon fed" a default answer, or worse yet, ignored.

Design of the course

This section presents a discussion on how the fundamental ideas are being implemented into the course structure. The subject matter is first divided into different levels of difficulty. The educational instruments are then applied to expound knowledge at each level of difficulty.

Levels of difficulty defined

For course design purpose, the instructor classifies the knowledge of computer organization into three levels of difficulty: basic, advanced, and research.

Basic knowledge refers to materials that are so essential to the field that any person educated in this field is expected to know in depth. Examples of basic knowledge include data representations, the instruction execution cycle, and the mechanism of multi-level memory access.

Knowledge at the advanced level is the ability to reason, analyze, and invent using the basic knowledge. A specific example of advanced knowledge would be the ability to calculate how a change to the length of the exponent field may affect the range and/or the precision of a floating point representation. The ability to propose a design that expands the function of an instruction set architecture is another example of advanced knowledge.

The research level covers all areas beyond the advanced level. Topics in the research level are characterized by open-endedness.

Educational instruments

The course consists of weekly lectures, weekly homework assignments, tests, and projects. This structure has been given to the instructor as a guideline. Within this framework, the instructor is to define the purposes of each of these educational instruments and to design each instrument accordingly.

Lectures

Lectures are the primary means to communicate all the knowledge at the basic level, and to stimulate thinking at the advanced level. In addition, lectures are places to build the subject roadmap. To prevent dryness and boredom, students are invited to work out the details of the basic knowledge in class. Advanced topics are introduced during lectures, and are to be investigated in depth in the homework assignments. Latest hot topics, such as Watson the supercomputer, are also covered during lectures.

Homework assignments

Homework assignments serve three purposes: first, to reinforce basic knowledge through practice, second, to promote self-learning through guided investigations of difficult problems, and lastly, to give pointers for further research. Problems at the basic level are typically straightforward. Problems at the advanced level are essentially the instructor's formulation of a difficult problem in such a way that a constructive and definitive conclusion can be drawn. The design of problems at the advanced level requires the most effort on the side of the instructor. It begins with an open problem that intrigues the instructor. The investigation process may involve breaking down the problem into sub-problems and considering alternative approaches. When a satisfactory conclusion can be drawn to close the problem, the instructor may then transform her own investigation process into a collection of questions, hints, and additional resources for the students to consider when they examine the problem themselves. Research problems are open questions, which have yet to be formulated. Students are expected to develop their own outline of the problem based on resources such as technical and scholarly papers. Homework assignment questions at the basic level and the advanced level are mandatory, while research problems are optional. Students receive bonus credits for their contributions of any non-trivial findings on the research problems.

Projects

Projects are large-scale problems that need more than one week of effort on the side of the students. A mandatory project requires students to solve a non-trivial problem using a simple assembly language. At the basic level, the student is expected to be able to think at the machine level. At the advanced level, the student is expected to explore the limitations of such a low-level language and to consider ways to overcome them.

Students are offered an optional opportunity to perform independent research on a topic of their own interest. The goal is to expose them to professional and scholarly activities early in their college education. The research project involves a review of between 10 and 30 pieces of scholarly literature, a written proposal of the research topic based on this review, a formal report, a presentation, and a peer evaluation. The students are given the guidelines and the expectations on various types of research (such as surveys, innovative proposals, in-depth analyses, and implementations). The formal

report of the project is to comply with a format adopted by such communities as the IEEE or the ACM.

Tests

Tests serve both as an evaluation tool and as a means to reinforce students' understanding of the subject matter. Test questions cover both the basic and the advanced levels and are meant to test students' command of the subject under an extremely tight time constraint. Students may use any books, devices, and internet resources available to them, but are forbidden from communicating with other humans.

3. RESULTS AND ANALYSIS

This section reports students' performance of a 200-level computer organization class that is taught based on the methodology outlined in this paper.

All students enrolled in this course are expected to have already received the equivalent of two semesters of computer science education covering programming techniques, data structures, algorithm analysis and calculus. Prior knowledge on computer hardware, assembly language programming or operating systems is not required. However, it has been observed that such prior knowledge greatly affects how much the students respond to the intellectual stimuli presented to them. Therefore, students' experience is a key factor to consider in this analysis. The students are placed into one of three categories: computer science majors (CS), non computer science majors (non-CS), and working professionals (PRO). A person is considered as a working professional if he/she has over 10 years of work experience in the computer science, information technology, or engineering field.

The class size is 23, composing of 16 CS majors, 4 non-CS majors, and 3 professionals. The average performance of each group over a scale of 100 is shown in Table 1.

Performance over 100 points	Experience			Class average
	CS	Non-CS	PRO	
Total score	82.25	70.69	97.70	82.25

Table 1: Average score over the whole course

The students' performance is analyzed for their knowledge at each level of difficulty. On a scale of 100 points, 55 points are allocated for basic knowledge, and 45 points are for advanced knowledge. Research level work is given bonus credits on top of the 100 points. The 50-point passing mark corresponds to 90.9% (=50/55) of the set of basic knowledge that a person educated in computer organization ought to have. At the basic level, the class average of 48.97 is very close to the 50-point mark. Both CS majors and professionals perform within just 1 point of the 50-point mark. Non-CS majors' average score at this level is 45.95, which is 4 points below the 50-point mark.

At the advanced level, the class average score is 26.55 out of 45 points. The difference in performance is noticeable. As a group, the professionals' average is 10 points (~38%) above the class average, while the non-CS majors' average is 5 points (~19%) below.

At the research level, the differences among the three groups are even wider. The class average bonus score for homework is 5.89. The professionals' average is 4 points (~70%) above the class average. The non-CS majors' average is about 3 points (~47%) below the class average. The details of these data are reported in Chart 1. (Note: The sum of each group's basic, advanced and research level homework average shown in Chart 1, is equal to the group's average shown in Table 1, subtracted by the group's optional research project average. The optional research project scores are excluded from this calculation because only 2 students have taken up such projects.)

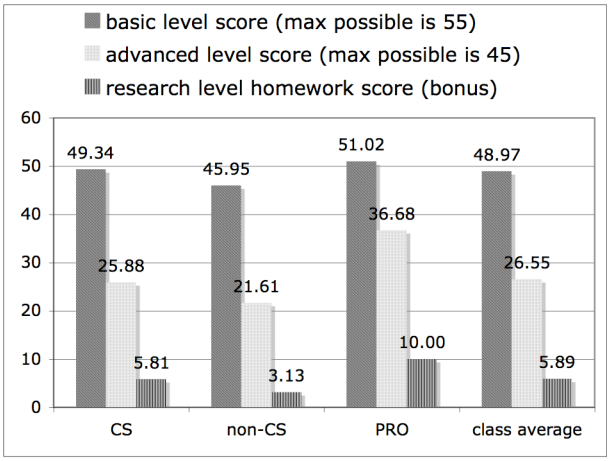


Chart 1: Performance at three levels of difficulty

The performance of the class is further analyzed based on each type of evaluation tool, namely, homework assignments, projects, and tests. These data are summarized in Tables 2 to 4. The analysis, which follows, provides some insights into the learning patterns of each group of students.

Homework performance	Experience			Class average
	CS	Non-CS	PRO	
Basic level score	20.98	16.95	20.98	20.28
Advanced level score	16.13	9.54	17.92	15.22
Research level bonus score	5.81	3.13	10.00	5.89

Table 2: Homework performance at three levels of difficulty

Based on the data in Table 2, it can be observed that the average homework scores of professionals and CS-majors are comparable at both the basic and the advanced levels of difficulty. Non-CS majors fare lower for homework problems. This could be because homework assignments tend to require details, such as the working out of the mechanism of the instruction execution cycle. Non-CS majors often forego such questions because they are very time-consuming.

Project performance	Experience			Class average
	CS	Non-CS	PRO	
Basic level score	3.94	3.75	3.33	3.83
Advanced level score	1.00	0.00	3.33	1.13
Number of research project attempts	2	0	0	Total: 2

Table 3: Project performance at three levels of difficulty

The data of Table 3 show that the performance of all groups of students is similar on the basic level project problem. Non-CS majors simply do not attempt the project problem at the advanced level. The professionals perform significantly better than CS majors on the project problem at the advanced level, but not at the basic level. The basic level average score of the professional group is lowered by some individuals in the group that have decided not to attempt the projects at all. It is worth noting that, while the professionals regularly attempt the optional bonus homework questions, they do not participate in the optional independent research projects. The two research projects completed at the end of this course are attempted by CS majors. This is likely because the homework challenges are short enough to be completed on a weekly basis. In contrast, an independent research project is too time consuming for them.

Test performance	Experience			Class average
	CS	Non-CS	PRO	
Basic level score	24.43	25.25	26.70	24.87
Advanced level score	8.75	12.08	15.43	10.20

Table 4: Test performance at two levels of difficulty

Tests are made intentionally challenging. Students are told to attempt all the test questions, which add up to a total of 75 points. Realistically, they are only expected to be able to complete 45 points worth of the questions within the tight time constraint. The remaining 30 points are counted as bonus credits (which no one has managed to receive.) Test performance is summarized in Table 4. It can be seen that all three groups of students perform well on test questions that are at the basic level of difficulty. A surprising phenomenon is that the non-CS majors fare much better than the CS majors on the test questions at the advanced level. A careful review on the students' works and track records reveals the following behavioral patterns:

1. Non-CS majors, who under-perform on weekly homework assignments, use tests as a means to catch up on their grades, while CS majors who have performed well may forego challenging problems towards the end of the course so as to redirect their time to other subjects.
2. Non-CS majors are often majors in mathematics or physics. They tend to have an advantage over CS majors, on short mathematically challenging problems that involve system analysis.
3. CS majors, who regularly perform well on homework assignments, get bogged down by complex test problems which the non-CS majors simply avoid because of the tight time constraint.

4. DISCUSSIONS

The effectiveness of the methodology is evaluated by three criteria. First, do the students meet the requirement of mastering the basic set of knowledge of the subject? Second, do the students demonstrate an advanced level of understanding of the subject? Such understanding is characterized by the ability to reason. Third, are the students motivated to explore unknown territories of the subject, beyond the comfort zone of basic knowledge? We now investigate each of these criteria.

The average score on basic knowledge is close to the 50-point marker, which corresponds to knowing 90.9% of basic knowledge. The sub-group with the lowest average score for basic knowledge is only 4 points off. Therefore, it can be safely concluded that the first criterion is nearly met. The student performance on advanced knowledge varies based on the students' interest and prior knowledge of the subject. This is expected. Non-CS students have significantly lower performance than the other two groups on the advanced level. It remains to be investigated how to improve the performance of this category in the future. To various degrees, students from all three categories respond positively to the optional research questions in homework. As effort in this level is optional, a positive response can indicate that, either students are interested in the topic and the topic is manageable within their ability, or students need to make up for their grades and are seizing every opportunity available to them, or a mixture of both reasons. Overall, the study shows that the methodology by itself seems to be effective.

5. EXISTING COURSES

In previous sections, we have discussed a teaching methodology that has shown to be effective when practiced on a 200-level computer organization course. This section evaluates the quality of the course. In the absence of a universal test, we use the choice of textbook as a gauge of difficulty, with the assumption that students passing a course are expected to be able to solve the basic level problems provided by the course's textbooks. The amount of knowledge that a student may gain from a course has to be built upon whatever prior knowledge the student already has. Therefore, the course prerequisites are a second factor to consider in evaluating its level of difficulty.

A survey is done on the courses of computer organization and computer architecture, offered by universities and colleges in the US between 2009 and 2011. The purpose of this survey is two-fold. First, it helps the new instructor set objective and realistic standards on the minimal set of knowledge that the students are expected to master at their level. Second, it gives the instructor some perspective on the trends in the field.

The 21 courses selected for this study satisfy at least two of the three criteria listed below:

- Either "computer organization" or "computer architecture" appears in the course title;
- The course materials, including syllabus, schedule, lecture notes, assignments and references, are fully available on the internet and require no restricted access;
- The course description, level, syllabus and prerequisites are comparable to our course.

These 21 courses (sorted by the names of the institutions) are:

CSCI140: Assembly Language & Computer Organization, at Clark University (ClarkU)

CS251: Computer Organization, at Dickinson College (Dickinson)

CS465: Computer Systems Architecture, at George Mason University (GMU)

CS314: Computer Organization, at Mississippi College (MC)

CSC234: Computer Organization & Assembly Language, at North Carolina State University (NCSU)

CS200: Computer Organization, at Northern Arizona University (NAU)
CS147: Computer Architecture, at San Jose State University (SJSU)
SE320: Computer Organization & Architecture, at Stony Brook University (Sunysb)
ISE390: Introduction to Computer Organization, at Stony Brook University (Sunysb)
CSCI305: Computer Organization & Programming, at The Citadel (Citadel)
CS61C: Machine Structures, at University of California Berkeley (Berkeley)
CS152: Computer Architecture and Engineering, at University of California Berkeley (Berkeley)
CSE141: Introduction to Computer Architecture, at University of California San Diego (UCSD)
CMSC313: Assembly Language & Computer Organization, at University of Maryland Baltimore College (UMBC)
CMSC411: Computer Architecture, at University of Maryland Baltimore College (UMBC)
CMSC611: Advanced Computer Architecture, at University of Maryland Baltimore College (UMBC)
CMSC411: Computer Systems Architecture, at University of Maryland College Park (UMD)
ITCS3182: Computer Organization & Architecture, at University of North Carolina Charlotte (UNCC)
ESE534: Computer Organization, at University of Pennsylvania (UPenn)
EE382N: Microarchitecture, at University of Texas at Austin (UTexas)
ITCS6810: Computer Architecture, at University of Utah (Utah)

The course codes generally reflect the academic departments and the levels of the courses, with the exception of *CS61C*, which is a lower-division undergraduate level course offered at Berkeley. For this study, we consider it as 100-level. For course codes with 4 digits, the most significant digit is an indicator of the course level.

Choices of Textbooks

Seven institutions offer a computer organization or architecture course as an introductory course (that is, at levels 100-200) in their curriculums. These courses may or may not be followed up at a higher level. Out of these seven courses, three adopt [7] as their main textbook. Two of the remaining four courses use [8], one uses [9], and one uses the instructor's own notes. Additionally, of the seven introductory level courses, three cover microcontrollers, assembly languages or systems internals. They use [1],[5] or [6] as their second textbook.

In the sample set, there are eight courses at the 300 level. The most popular textbook for them is [8], adopted by three courses. One course uses [3], one uses [4], one uses [7]. One uses the instructor's own notes. Some courses at this level cover primarily assembly languages or systems internals. They use [2], [5] or [6], either solely or in conjunction with [8].

Three courses are offered at the 400 level. Two of them use [8] while one uses [4]. Three of the courses are graduate level (500-600) courses, and all of them use [4] as their main textbook.

The choices of textbooks used by a course provide some good estimate on the difficulty of the course materials. The three most popular textbooks in computer organization and architecture are [7] for levels 100-200, [8] for levels 300-400, and [4] for levels 500 and above.

Null and Lobur's [7] has 14 chapters and it provides a broad introduction at an easily understandable manner. The notable contribution of this textbook is the MARIE instruction set architecture, which consists of 13 instructions, and is defined on an accumulator-based CPU. The MARIE assembly language is sufficiently complex to handle non-trivial integer arithmetic problems such as fibonacci number computation. The limitations of the MARIE architecture include the lack of support for recursive function and instruction-level pipelining. The courses we surveyed that adopt [7] use mainly the first 7 chapters. Our course also adopts this book as the textbook and covers the same chapters.

Patterson and Hennessy's [8] gives a full coverage of the field based on the MIPS system developed by one of the authors. The book examines a fully functional academic version of the MIPS machine and its assembly language. The MIPS is a RISC architecture and is designed for efficient support of a 5-stage pipeline. The authors offer free access to their lecture notes that illustrate the functions of MIPS and the 5-stage pipeline. These presentations are in such great details that many educators simply use these slides without any modifications. The book consists of 7 main chapters and 5 chapters of appendices. In this survey, all of the courses that adopt this book cover at least 6 of the main chapters to various degrees of details. Our course adopts this book as a reference text.

Hennessy and Patterson's [4] investigates performance issues of systems. The materials of [4] are built upon the foundation of [8], It starts with Amdahl's law as the key measure of system performance. Various architectural issues are examined, including instruction-level parallelism and thread-level parallelism. The book also covers memory hierarchy and GPU. This book is suitable for use at an advanced level.

Stalling's [9] was a classic textbook during the time when the x86 was the state-of-the-art architecture. In many ways, Null and Lobur's [7] seems to have come out of the philosophy of [9]. However, with RISC gradually phasing out CISC, Patterson and Hennessy's [8] has taken over the academic leadership of the field in the past 5 years.

Significance of Background Knowledge

All of the 100- and 200-level courses surveyed require at least some basic knowledge of computer science, and a programming language, such as Java. Three of the courses at this level also require, as a prerequisite, one semester of training in discrete mathematics, digital logic, C/C++ or assembly language such as MIPS. Those courses that do not have such prerequisites cover some of these topics within their syllabi. Our course fits into this latter category. Students are required to have two semesters of computing education as their background knowledge. They learn digital circuit design and assembly language programming as a part of our course.

All the 300-level courses assume that students have at least the equivalent of 200-level data structure knowledge (for CS students) or electrical engineering knowledge (for EE students). In addition, familiarity with C/C++, assembly languages, and/or some basic computer organization knowledge, is a requirement.

Courses at level 400 require students to have prior knowledge in very specific subjects, such as programming language organization, or systems programming at level 300. Graduate level courses all require students to have background experience that is equivalent to a 400-level computer organization course.

The following table reports a count of the number of courses at level-*x* that require background knowledge of subject *y*, where *x* is the column entry and *y* the row entry of the table.

Level / Experience (as prerequisites)	100	200	300	400	500	600
Java or Computing 1	1	2				
Computing 2	1	2				
Data structure 2			3			
Discrete mathematics	1		1			
C/C++			3			
Assembly language	1		1	1		
Digital logic	1		1	1		
Circuit design			2	1		
Basic computer organization			1	1	1	
Programming language organization				1		
Systems programming				1		
300-level Electrical engineering			1	1	1	
400-level computer organization						2

Table 5: Prerequisites required by computer organization or architecture courses at various levels

Observations of Special Interest

Berkeley introduces the C language, the MIPS assembly language, and the 5-stage pipeline among other essential architectural concepts in *CS61C*. One main objective of this course is to get the freshmen familiarized with the concept of machine parallelism at multiple levels. *CS61C* then provides students with a strong background for *CS152* and a few other upper-division systems courses including Operating Systems, Programming Languages and Compilers, and Computer Security.

UMBC offers computer organization and architecture courses only at an advanced level, beginning at *CMSC313*, which introduces architectural concepts through an in-depth investigation of the internals of C and assembly language. This course is followed up by *CMSC411* and *CMSC611*.

The *EE382N* course offered at UTexas provides an interesting electrical engineering perspective on computer organization. The course explains some of the shortcomings of the multi-core processors.

This survey reveals that there is no standard syllabus for this subject. Each institution offers some courses that cover this subject to various degrees of breadth and depth based on the institution's curriculum. In reading this survey, the reader is reminded to avoid the fallacy of measuring the quality of a course solely by its level of difficulty. A truly successful education is one that takes into consideration the individuals' backgrounds, and stimulates the individuals to achieve learning objectives that are challenging but reachable. In this context, our course has met the standard expectations at its level.

6. CONCLUSION AND FUTURE WORK

We conclude this paper with a discussion on the implications of the findings in this preliminary teaching experiment. The students in this study generally respond well to a teaching method that provides them with a basic set of knowledge and stimulates them to make further explorations by themselves. This teaching method works exceptionally well for a beginner instructor who is actively engaged in the same learning process with the students. As an instructor grows in her own knowledge, the gap between what the instructor knows and what the students know will widen. An experienced instructor may need to take more conscious effort to stimulate the students towards discovery-based deep learning, instead of reducing her teaching to a one-directional information transfer. More insights may be obtained to compare this methodology with those adopted by experienced educators, in order to learn how they have achieved this goal. And this is a worthy endeavor for an instructor, especially of this age. For many have observed that ours is an age of information explosion, in which "greater access to knowledge is not equivalent to greater knowledge; an ever-increasing plethora of facts and data is not the same as wisdom; breadth of knowledge is not the same as depth of knowledge; and multitasking is not the same as complexity."[10]

7. REFERENCES

[1] M. P. Bates, **PIC Microcontrollers: An Introduction To Microelectronics**, Newnes, 2ed., 2004.

[2] R. Bryant and D. R. O'Hallaron, **Computer Systems: A Programmer's Perspective**, Addision Wesley, 2ed., 2010.

[3] I. Englander, **Architecture of Computer Hardware, Systems Software, and Networking; An IT approach**, Wiley, 4ed., 2009.

[4] J. L. Hennessy and D. Patterson, **Computer Architecture: A Quantitative Approach**, Morgan Kaufmann, 4ed., 2006.

[5] K. R. Irvine, **Assembly Language for Intel-based Computers**, Prentice Hall, 5ed., 2006.

[6] B. W. Kernighan and D. M. Ritchie, **The C Programming Language**, Prentice Hall, 2ed., 2010.

[7] L. Null and J. Lobur, **The Essentials of Computer Organization and Architecture**, Jones&Bartlett Learning, 2ed. 2006.

[8] D. Patterson and J. L. Hennessy, **Computer Organization and Design, The hardware/software interface**, Morgan Kaufmann, 4ed., 2008.

[9] W. Stallings, **Computer Organization and Architecture: Designing for Performance**, Prentice Hall, 8ed., 2009.

[10] W.T. Lukeman, An amazon review on Nicholas Carr's book **The Shallows**, link: http://www.amazon.com/Shallows-What-Internet-Doing-Brains/dp/0393072223/ref=cm_cr_dp_orig_subj

Image toolbox for CMOS image sensors simulations in Cadence ADE

David Navarro, Zhenfu Feng, Vijayaragavan Viswanathan, Laurent Carrel, Ian O'Connor
Université de Lyon; Institut des Nanotechnologies de Lyon INL-UMR5270, CNRS, Ecole Centrale de Lyon, Ecully, F-69134, France

***Abstract*— This paper presents a toolbox we have developed in order to help analog designers for image sensors simulations. Such matrix structures, composed of millions of pixels are too difficult to handle manually, especially in terms of input generation and output analysis. A graphical toolbox has been developed in Cadence Analog Design Environment (ADE) to overcome these problems. That toolbox has been completely written in Cadence SKILL language. It permits to manipulate images as input, and to automatically generate images at output, in order to check the quality of electronic design. Low level aspects can also be analyzed at system (image) level.**

Image sensor, electronic design, APS, simulation, Cadence, Analog Design Environment.

1. INTRODUCTION

Image sensors in standard CMOS technology are now a well established alternative to the CCD image sensors technology. Indeed, process maturation, integration possibilities and low power consumption make CMOS image sensor widespread. Moreover, recent 3D technologies focused researchers and industry on new image sensor architectures and 3D floorplanning [1] [2].

CMOS image sensors are electronic systems that are composed of analog blocks (mainly a pixel matrix and a noise reduction block –CDS: Correlated Double Sampling-), digital blocks (such as controller and decoders), and mixed blocks (ADC: Analog to Digital Converter) [3]. Fig. 1 details a basic CMOS image sensor floorplan.

2. PROBLEMS IN IMAGE SENSOR SIMULATIONS AND STATE OF ART

A major problem in these matrix structures is density: it is difficult to make system-level analysis because of the multiplicity of inputs and outputs. Indeed, in a classical pixel, a 3-transistors active pixel, also called 3-T APS [3], shown in Fig.2, 3 inputs are necessary. A "reset" signal –for initialization-, a "select" signal –for reading- and luminosity are required. The two digital control signals, "reset" and "select", have precise timings. These signals are common for each line, and have to respect an interlaced sequence. In a m x n matrix, M lines have to be considered to generate these signals. The major problem comes in fact from the third input: m x n luminosity inputs have to be considered.

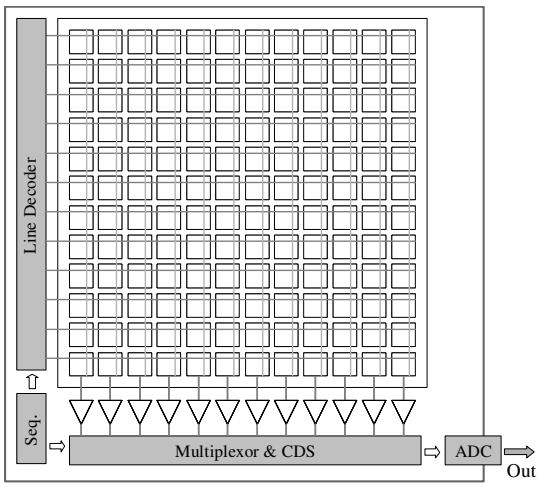


Figure 1. classical CMOS image sensor floorplan

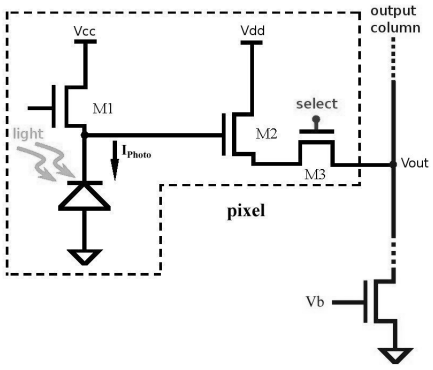


Figure 2. classical CMOS 3-T Active Pixel Sensor (APS)

Luminosity is often set as a constant value, because anyway classical pixels integrate current over an exposure time (also called integration time). Depending on the photodiode model, the luminosity is set with a current source that emulates the photocurrent if a simple equivalent schematic is drawn, or luminosity can be processed if model is of upper level (for example in Verilog-A or VHDL-AMS language [4]). Of course, it is mandatory to simulate

different values on pixel to realistically simulate the sensor characteristics. The problem is how to set easily thousands or millions of design variables. The same problem exists at sensor output: it is difficult to manage and analyze millions of analog values. In classical CMOS image sensors, outputs are voltages, sometimes ADC output.

As a consequence, designers rarely simulate full imagers at a time. Blocks are validated separately, and small matrixes are used to validate a global simulation. Test-chips (ASICs) are also required to properly characterize pixels characteristics. This toolbox is proposed to help designers running image sensor – level simulations.

Available image sensor simulators are TCAD and focus on physical and FDTD simulations [5] [6]. On the other hand, image processing toolbox exist [7]. ECAD image sensor simulators are also missing.

Some high-level models –for example VHDL-AMS or MATLAB- have been developed [5], but it appears that the gap between a real analog structure and a high-level model make it uneasy to use within an optimization work. Experience showed that analog designers trust better in a level 53 spice model than in a third-party high-level model, and it is a drawback to use several simulation platforms. Moreover, computer power calculation has rapidly increased past ten years, and computer clusterization is now a widespread solution, so simulation time can be lowered. Moreover, designers cannot do without post-layout simulations that are easily accessible in the classical micro-electronic design flow.

To answer the above mentioned problems, we propose in this paper a graphical toolbox that is dedicated to electronic designers, as it is integrated in Cadence Analog Design Environment (ADE). Novelty is to propose an easy way to manage many (millions) parameters, This toolbox is a first step of a new kind of image simulator that is briefly presented in conclusion.

Paragraph II details the input-output mechanism we have developed, and paragraph III details the graphical user interface, and test-case results.

3. INPUT AND OUTPUT PROCESS

In order to handle all the input and output signals, a high-level mechanism has been set. It permits to read an image as input, and generates an image as output. As Fig. 3 shows, several steps are required.

We consider an image that has the same resolution as the image sensor. In that way, a pixel in the input image will be converted into luminosity or a photocurrent value. Then, that input value is set to the pixel input. A first skill processing function converts image values into lux, watts or amperes, and then a second one assigns each value to each pixel. Design variables in ADE are used to make that mapping. Before assignment, an automatic creation of thousands or millions of design variables is done. Names are also fixed by software, for example lum_0_0 for the first pixel, lum_3_200 for 201th pixel of fourth line. Then, the simulation that was configured in ADE is run. Classical ADE output, for a classical CMOS 3T pixel, is shown in Fig. 4.

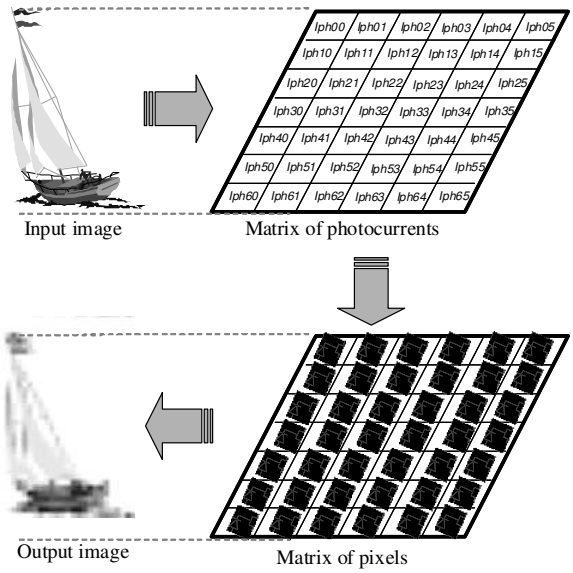


Figure 3. Image design flow detail

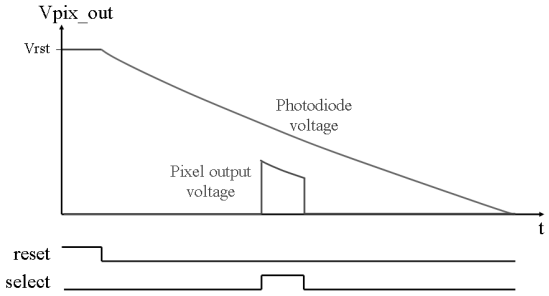


Figure 4. classical output voltage of a CMOS 3T APS

During a "select" window, the analog voltage at photodiode is followed to the pixel output, with a well-known voltage drop due to pixel area minimization constraints (all transistors are N-type like the photodiode to avoid P-Well to N-Diff distance rule). During that window, the analog voltage is read by blocks that are located out of the matrix: noise reduction blocks (CDS: Correlated Double Sampling), then ADC. CDS block is a noise reduction block. To cancel temporal noise, it samples the pixel output voltage twice: once during the reset state to measure the maximal and starting voltage, and once at select time. Sampling times –times that are used to read and store values- are automatically calculated, according to generators properties

in schematic. Circuit simulation is also different if we consider a matrix of pixel or a complete or almost complete image sensor (pixels matrix, CDS, ADC).

If desired, via a checkbox, the simulation tool can also automatically make the double sampling to give more accurate results in a pixel matrix simulation. In that case, two simulations are configured and run: a first one measures the maximal pixel output voltage at reset, and a second is exactly the one configured by designer in ADE.

As a result, pixel output voltage can be a raw voltage value or a simulated double sampled one, as illustrated in Fig. 5. In case of automatic double measurement, two simulations are run. The original netlist is modified in order to turn on the select transistor, so Voh can be read at reset state. It is effectively sampled 1µs before reset signal falling edge. Considering classical timings in such circuits, pixel output voltage will have a stable and final value. This timing is a parameter that can be changed though "setup" function. Then, the user-defined simulation is run, and Vos signal is sampled at the middle time of select window.

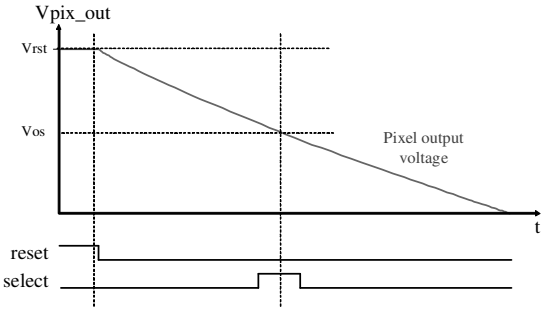


Figure 5. Automatic timing to read voltages values at pixel output: on user request, simulator can output Vos or (Vrst – Vos) based data.

At end of simulation, Cadence Spectre output files are processed in order to display results. It is mandatory to detail a classical output simulation to best explain the internal result processing. The output voltage, computed with one of the two previously explained manners, is converted in grey level. Three solutions can also be configured: raw, absolute or relative coding.

- Raw coding is a basic reading of Vos output signal voltage and supply voltage Vcc is set as maximal digital code (255 in 8-bit or 1023 in 10-bit). Equation 1 gives a calculation example for 8-bit resolution. This solution can be used as a debug mode for designer, in order to check the raw simulation output compared to classical (manual) simulation.

$$\text{Raw grey-value} = 255 \times \frac{V_{os}}{V_{cc}} \qquad (1)$$

- Absolute coding consists in setting the maximal pixel output (Vrst) as maximal digital code. Saturation, ie 0v, is set as the minimal value. Equation 2 gives the equivalent calculation for 8-bit resolution. This calculation gives an equivalent result as a classical CDS block would do.

$$\text{Absolute coding grey-value} = 255 \times \frac{V_{os}}{V_{rst}} \qquad (2)$$

- Relative coding consists in finding maximal and minimal pixel output voltages (respectively Vmax and Vmin) in the matrix, and to set them to maximal and minimal codes. Equation 3 gives the equivalent calculation for 8-bit resolution. It is what an image signal processor would do to optimize dynamic range.

$$\text{Relative coding grey-value} = 255 \times \frac{V_{os}}{V_{max} - V_{min}} \qquad (3)$$

These three solutions permit to optimize hardware and software in image sensor: hardware blocks such as CDS and ADC, and software for image signal processor at output. Hardware block that can be hierarchically optimized are at pixel, matrix, matrix and CDS, and matrix and CDS and ADC levels.

4. USER INTERFACE, TEST-CASE RESULTS

The graphical user interface is showed in Fig. 6. It is opened via an "imager" menu in ADE.

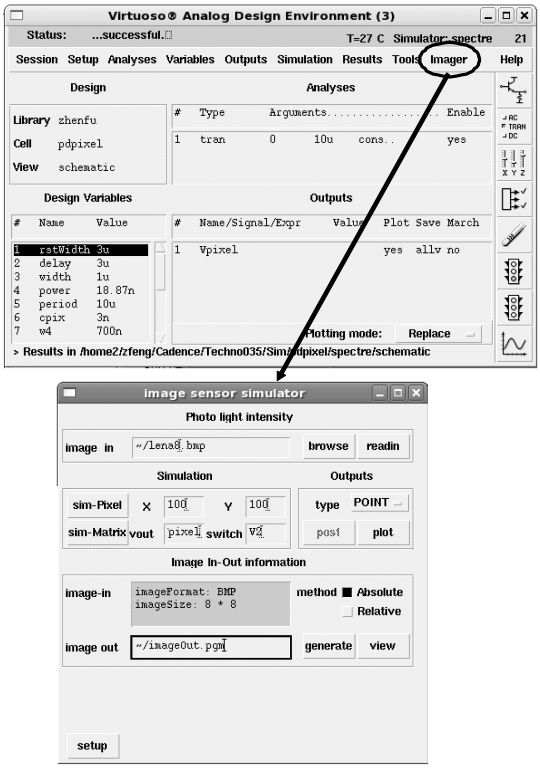


Figure 6. Toolbox graphical user interface

User can select the input image (in uncompressed BMP or PGM formats). Clicking the read-in button imports image and converts color or grey pixels values into light information (light power or photocurrent according to the

photodiode model that is used). This conversion is done by considering silicon sensitivity to light.

As analysis setup is launched from a single pixel schematic, it is possible to select a single pixel simulation (sim-Pixel), or a matrix simulation (sim-Matrix). For a single pixel simulation, coordinates of pixel within the matrix (pixel at location 100, 100 in Fig. 6) are selected. For a m x n matrix simulation, the same pixel is simulated m x n times with respective m x n stimuli. As a matter of fact, output node of the pixel (vout) as to be defined. Vout is set as "pixel" value in Fig. 6. It is to notice that many pixel structures can be simulated, since simulation and toolbox can run on any schematic. Moreover, this toolbox could be used for any matrix simulation, like memories.

As test example, we have considered an input image, a classical 3T pixel, and several configurations. As we can observe, from an input image, Fig. 7, it is possible to automatically set the matrix data (light) input that is applied on electronic structure, to simulate electronic circuit with Cadence Spectre, and then to monitor output images according to the chosen reading mode configuration (Fig. 8 to Fig. 10). Raw or absolute coding will be used to characterize low level characteristics of pixels matrix, or to study a system composed of matrix and Correlated Double Sampling block, eventually with Analog to Digital Converter (ADC).

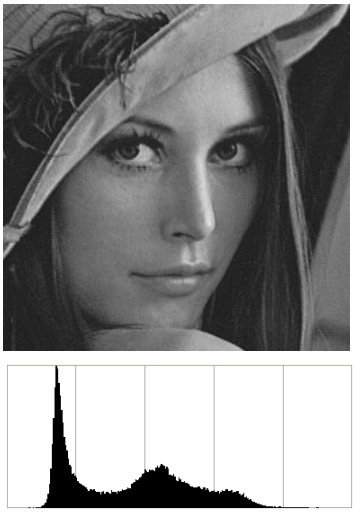


Figure 7. Input image and histogram

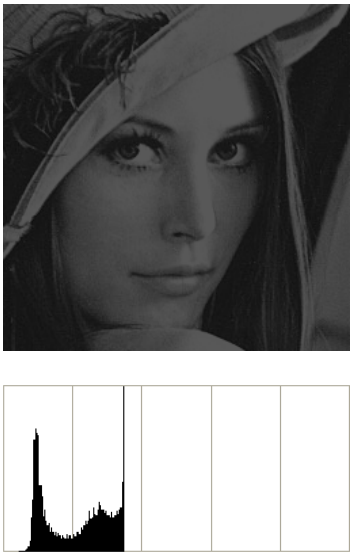


Figure 8. Raw output image and histogram, Wfollower = 1 µm

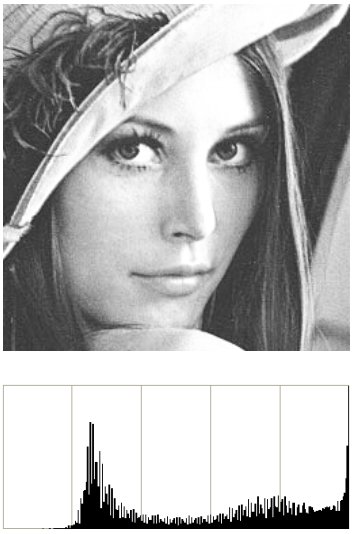


Figure 9. Absolute coding output image and histogram, Wfollower = 1 µm

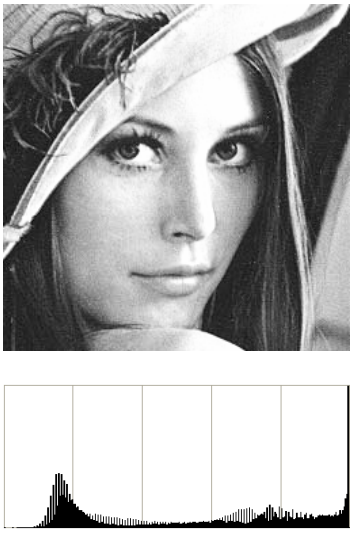


Figure 10. Relative coding output image and histogram, Wfollower = 1 µm

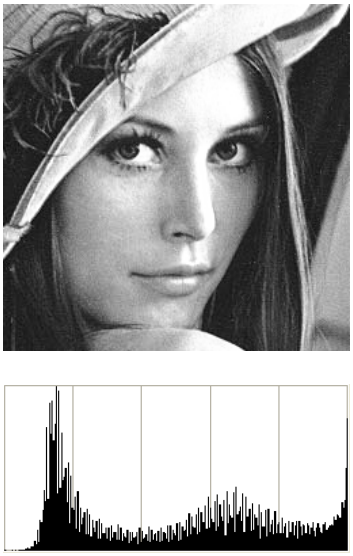


Figure 11. Relative coding output image and histogram, Wfollower = 15 µm

As double sampling is done in CDS block, it is indeed useless to simulate an image sensor that comprises such a block with the relative coding option. Relative coding is useful to characterize a standalone pixels matrix, or specific smart image [8] sensors that don't embed CDS. Following blocks parameters, such as amplifiers, CDS and ADC, can also be studied and dimensioned. We can observe that contrast and mean values differ from input to output images, and it is more visible on histograms. Histograms, that present grey values versus number of pixels, clearly show that transfer function of sensor is not linear. Moreover, Fig. 11 shows the transistor size impact on output image quality. Electronic impact on image sensor quality can also be clearly studied.

5. CONCLUSION

This paper presented a new toolbox for CMOS image sensor simulations in Cadence Analog Design Environment. It is written in SKILL language, and it permits to input automatically an input image, fitting each image pixel on each pixel of the electronic sensor under study. Light stimuli is calculated and applied on each. Output images permit to check the quality of analog design in the pixel. Low level aspects of the sensor can be analyzed at system (image) level. This toolbox is the first point of a dedicated fast simulator we are developing. It will enable several mega-pixels simulation within a few minutes, by studying a single pixel in a parametrical way, and by projecting result on the matrix. Variability analysis on this simulator will also be available in order to support the matrix aspect, and in order to monitor low-level (physical) impacts on output image.

REFERENCES

[1] [1] Knickerbocker, J.U. et al. "3-D Silicon Integration and Silicon Packaging Technology Using Silicon Through-Vias", IEEE Journal of Solid-State Circuits, vol. 41, no. 8, pp.1718-1725, August 2006.

[2] Topol et.al., "Three-dimensional integrated circuits", IBM Journal Research and Development,. Vol. 50 no. 4/5, pp. 491-506, July/September 2006.

[3] E.R. Fossum, "CMOS Image Sensors: Electronic Camera-On-A-Chip", IEEE Trans. on Electronic Devices, Vol 44, N° 10, October 1997.

[4] D. Navarro, D. Ramat, F. Mieyeville, I. O'Connor, F. Gaffiot, L. Carrel, "VHDL & VHDL-AMS modeling and simulation of a CMOS imager IP", Forum on specification & Design Languages, Lausanne, Switzerland, September 2005.

[5] Silvaco, "CMOS Image Sensor Simulation - ATLAS", www.silvaco.com/content/kbase/CIS_april2010.pdf

[6] CrossLight Inc, "3D Simulation of CMOS Image Sensor", www.crosslight.com/applications/crosslight_3DCIS3.pdf

[7] Matworks, "MATLAB Image Processing Toolbox", http://www.mathworks.com/products/image

[8] J. Kramer, R. Sarpeshkar, C. Koch, "Pulse-Based Analog Velocity Sensors", IEEE Trans. On Circuits and Systems II : Analog and Digital Signal Processing, Vol 44, pp 86-101, 1997

Modeling in Service Innovation: 10 Propositions

Jukka OJASALO
PhD, Professor
Laurea University of Applied Sciences
02650 Espoo, Finland
jukka.ojasalo@laurea.fi

ABSTRACT

The purpose of this paper is to examine theoretical grounds of modeling and service innovation, and to provide propositions for using modeling in service innovation. This conceptual article is based on an extensive literature analysis on modeling and service innovation. First, this article discusses the general principles of the modeling of business processes and systems. Then, it discusses service blueprinting, which is a specific type of a business process modeling tool developed for services. After that, it explains the special characteristics of innovation management in services. Next, as research implication, it provides ten propositions for using modeling in service innovation. Then, it draws the final conclusions. The ten propositions for using modeling in service innovation relate to identifying the problem to be modeled, the psychology of individuals and organizations, avoiding pitfalls, taking advantage of value co-creation, modeling experience, designing and modeling all elements of service experience, basing the model development in deep customer understanding, fostering creativity, thinking different and radically new service business models, and establishing collaborative networks.

Keywords: Modeling, Business model, Business process modeling, Innovation management, Service design

1. MODELING BUSINESS PROCESSES AND SYSTEMS

This section discusses the principles of modeling at general level in organizations and problem solving. Modeling has a long history and a large number of applications. It has been widely used, for example, in the area of organizational and business process development, as well as in information systems and services design [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Kettinger, Teng and Kuha, for example, conducted a study in which they examined altogether 25 methodologies, 72 techniques, and 102 tools in business process reengineering [12]

A model is an external and explicit representation of part of reality as seen by the people wish to use that model to understand, to change, to manage, and to control that part of reality in some way or other [13]. Model is a statement of a problem, characterized by a set of inputs, a set of outputs, and relations between them [14]. Models are used for exploring possible consequences of actions before they take them, which can be called "reflection before action" [15].

According to Pidd, (p. 119) "..a model is a convenient world in which one can attempt things without the possible dire consequences of action in the real world. In this sense models become tools for thinking. This thinking might relate to one-time events.. Or thinking might concern occasional events.. Alternatively, the thinking might concern routine events.. We also use models as tools for thinking when we try to understand a complex system, even if we contemplate no immediate action."[16]

A business process is a collection of activities that takes one or more kinds of input and creates an output that is of a value to the customer [17]. It is defined as structured, measured sets of activities designed to produce a specified output for a particular customer or market [18]. It refers to a set of related tasks performed to achieve a defined business outcome [19]. It is network of activities and buffers through which the flow units have to pass in order to be transformed from inputs to outputs [20].

Denna, Perry, and Jasperson brought forward three basic types of business processes: (1) acquisition/payment, (2) conversation, and (3) sales/collection [21]. The acquisition/payment process includes the activities of acquiring goods and services needed by the organization to perform its functions. The conversion process refers to activities of transforming goods or services from raw material to finished products. The sales/collection process includes activities of attracting customers, delivering goods or services, and collecting payments for delivered goods and services.

Curtis, Kellner, and Over identified four most common perspectives to process models: functional, behavioral, organizational, and informational [22]. The functional perspective illustrates a process by showing what activities are being performed and which data flows are needed to link these activities. The behavioral perspective illustrates a process in terms of when activities are being performed and how they are performed. It uses, for example feedback loops, iterations and triggers. The organizational perspective illustrates a process by showing where and by whom activities are being performed. The informational perspective illustrates a process by showing the entities being produced or manipulated by the process. Entities refer to documents, data, or products.

Lue and Tung proposed a framework for selecting business process modeling methods [23]. Their framework is described is the following. The framework consists of modeling objectives, perspectives of modeling methods, and characteristics of modeling methods, as follows. *(1) Objectives of process modeling* include three alternatives: communication, analysis, and control. *(a) Communication*. The primary objective of modeling may be

facilitating communication related to modeling. Process designers need to describe existing and improved processes. They have to agree upon a common representation among themselves. The need to share their knowledge of business processes with other employees. Simplicity and clarity may be the most desired features of a modeling for the communication purpose. *(b) Analysis.* Another objective of modeling may be analyzing and improving existing processes. Identifying the best process requires generating alternative representations, simulating process behaviors, and measuring process performance. *(c) Control.* Managing and monitoring a business process may also be the objective of modeling. Since there are several interrelated processes in the organization, there is need to control process operations, manage process relationships, and audit performance. Modeling methods of automated procedures, multi-level process descriptions, and other sophisticated modeling tools can used to achieve this objective. The second main element in Lue and Tung's (ibid.) framework relates to perspectives of modeling. *(2) Perspectives of modeling methods* consist of the object perspective, activity perspective, and role perspective. *(a) Object perspective*. This perspective emphasizes what is being done. The objects that are being manipulated in the process are followed in the modeling. These objects can data, documents, or physical goods. Data flow diagram (DFD) is an example of the object perspective approach. *(b) Activity perspective*. This perspective is about how things are done. The modeling methods focus on representation on the activities being performed and relationships between activities. Integrated definition of function modeling IDEF0 (see e.g. Kim and Jang [24]) is an example the activity perspective. *(c) Role perspective.* The role perspective focuses on who does what. A business process is modeled by representing roles and relationships between roles. The role activity diagrams (RAD) are an example of role perspective methods. The third main element in Lue and Tung's (ibid.) framework relates to characteristics of modeling methods. *(3) Characteristics of modeling methods* include formality, scalability, enactability, and ease of use. *(a) Formality.* This refers to question: how formal or precise are the languages and notations of the modeling method? Some methods have a set of well defined notations and require formal semantics to be strictly followed, while others only have a set of guidelines. Formal methods may be well positioned to provide a more precise representation of a process and have the benefits of well-developed properties for advanced analysis. However, they may also be less flexible in terms of modeling ambiguous processes and human involvement. *(b) Scalability*. This relates to question: how large and complex a business process can the modeling method represent? Some methods can handle large processes and offer mechanisms that support multi-level representations, while others are best suited for modeling processes that are relatively small in size. *(c) Enactability*. This relates to the question: does the modeling method support automated enaction and process manipulation? Some modeling methods only allow process designers to depict a process in a static state, while others also provide automated tools for process simulation and analysis. *(d) Ease of use*. This relates to question: how difficult is the modeling method for process designers and other non-technical employees to understand and use? Some methods use simple and easy-to-understand notations such as arrows and boxes, while others utilize more complicated mathematical symbols and formulae.[25] According to Martin and McClure, a good model should provide a good basis for communication, be capable of subdivision, and have a consistent notation [26].

Willemain examined professional modelers and reported on following findings related to models, modeling process, and modelers [27].The qualities of an effective model, in decreasing order importance, are (1) validity, (2) usability, (3) value to client, (4) feasibility, and (5) aptness for client's problem. The relevant qualities of an effective modeling process are (1) problem context, for example discovering the real problem, (2) model assessment, for example validation and verification, (3) model structure, for example selection of key variables and elaboration of submodels, (4) model realization, for example prototyping and data collection. The important qualities of a modeler include: (1) the modeler's mindset, for example creativity, sensitivity to client, and persistence, (2) nontechnical expertise, for example communication and teamwork skills, (3) OR/MS (Operations Research/Management Sciences) expertise, and (4) subject matter expertise.

Pidd put forward six simple principles of modeling [28]. *(1) Model simple, think complicated.* Models are simple representations of a complex world. Models should be easy to understand, at least in outline form, and should be easy to manipulate and control. Relatively simple model can support complicated analysis. However, a simple model does not have to be a small model. *(2) Be parsimonious, start small, and add.* It is impossible to know in advance how complicated the model should be. The principle of parsimony in modeling means that one should develop models gradually, starting with simple assumptions and adding complications only if necessary. Rather than attempting to build a final model from scratch in one effort, one can make initial assumptions that are known to be too simple, but allow proceeding in the modeling. Then, one will refine the initial far-too-simple model over time until it is good enough and fits for its intended purpose. One should deliberately develop a series of models, each more complex than its processors. The modeler build models that are too simple and, when their limitations become too obvious, throws them away and builds another to overcome some of the limitations. Through a series of prototypes, the modeler gradually ends up to a model that fits the original purpose. *(3) Divide and conquer, avoid megamodels*. Developing a set of small (interrelated) models id often most useful when a large model is needed. According to Raiffa (p. 7), "Beware of general purpose, grandiose models that try to incorporate practically everything. Such models are difficult to validate, to interpret, to calibrate statistically, and, most importantly to explain. You may be better off not with one big model but with a set of simpler models" [29]. *(4) Use metaphors, analogies, and similarities.* Modelers can seek an analogy with some other system or an association with some earlier work. The modeler relies on his own or somebody else's previous experience. The idea is to search for previous well-developed logical structures similar to the problem at hand. Analogies are most useful in the early stages of modeling. *(5) Do not fall in love with data*. Some people assume, that because a model is a representation of some system, examination of data from that system will reveal all they need to construct the model. Such an assumption may be a mistake, even though exploratory data analysis is useful. The availability of user-friendly software packages for data analysis may also make people to imagine that modeling is primarily data analysis, preferably with lots of data. However, modeling should drive data collection, not the other way round. One should first

think about the type of model that might be needed before attempting large-scale data collection. *(6) Modeling may feel like muddling through*. Model building is not a linear process which moves from step 1 to step 2 to step 3 and so on. A pretence that model building is a rational process may create various problems, particularly for beginners.[30]

Willemain suggested four ways for teaching and improving modeling capability. Firstly, *don't forget craft skills*. "Soft" qualities in modeling were emphasized more than "hard" qualities. "Soft" qualities include creativity, teamwork, and communication skills, while "hard" qualities cover technical knowledge, subject matter knowledge, and OR/MS knowledge. Secondly, *don't forget model assessment*. Effective models are valid and usable. Thirdly, *don't forget the client*. Working and interacting with the clients is important in order to understand the context of modeling problem and to assess the model. Fourthly, don't forget wisdom. In addition to understand equations and algorithms, it is very important to open up discussion about important issues of less technical nature, and to do less talking and more listening.[31]

2. SERVICE BLUEPRINT

This section discusses service blueprinting, which is a specific type of business process modeling approach developed for services. As referred earlier, there a large number of methods for modeling systems and business processes. From all available concepts, services blueprint, introduced by Shostack [32], is perhaps the most well known and popular in the service design context [33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45]. Compared to other to other process-oriented design techniques and tools, service blueprints are first and foremost customer-focused approach, allowing firms to visualize the service processes from their customers' perspective [46].

Service blueprinting is a mapping technique for visualizing service systems. It is a holistic method of seeing in snapshot all relevant resources, actors, and activities involved in the service delivery process, which is essentially a dynamic and living phenomenon. A service blueprint documents all process steps and point of divergence in a specific service. This documentation is carried to whatever level of detail that is needed to distinguish between any two competing services.[47, 48] A service blueprint is a map or picture that portrays the service system so that the different people involved in providing it can understand and deal with it objectively, regardless of their roles or their individual points of view. It visually displays the service by simultaneously representing the process of delivery, the points of customer contact, the roles of customers and employees, and the visible elements of the service. A service blueprint visually breaks a service down into its logical components and depicts the steps and tasks in the process, the methods by which the tasks are executed, and the evidence of the service as the customer experiences it. Blueprinting is a particularly powerful technique in the services context, since services are essentially customer experiences rather than objects or technologies.[49]

Service blueprinting offers several benefits. A service blueprint shows time in diagrammatic form, all the main functions of the service, all possible fail points and processes to correct those, and the relationships between the front and back offices [50]. The advantages of service blueprints also relate to providing a platform of innovation; recognizing roles and interdependencies among functions, people, and organizations; facilitating both strategic and tactical innovations; transferring and storing innovation and service knowledge; designing customer interaction from the customer's point of view; suggesting critical points for measurement and feedback in the service process; clarifying competitive positioning; understanding the ideal customer experience; and education of service design [51, 52, 53].

The main elements of a service blueprint are *customer actions, onstage contact employee actions* (actions visible to the customer), *backstage contact employee actions* (actions invisible to the customer), *support processes, and physical evidence*. Service blueprint also includes the *line of interaction, line of visibility*, and *line of internal interaction*. These lines dive the map into different zones where the actions of customers, contact employees, and support personnel are placed. [54, 55, 56, 57, 58]

The process of building a service blueprint includes the following steps: (1) identify the service process to be blueprinted, (2) identify the customer or customer segment experiencing the service, (3) map the service process from the customer's point of view, (4) map contact employee actions and/or technology actions, (5) link contact activities to needed support functions, and (6) add physical evidence of service at each customer action step.[59, 60]

3. SPECIAL CHRACTERISTICS OF SERVICE INNOVATION

Goods and services have four basic differences. Services are intangible, heterogeneous, perishable, and they are produced and consumed simultaneously. These distinctive characteristics have various implications to services management.[61]

The characteristics that distinguish goods and services have several implications for service innovation. De Brentani [62, p.102] summarized these implication drawing on several sources [63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75] as follows.

Intangibility. Intangibility often requires successful use of tangible evidence to help explain or portray the service. It can simplify and shorten the new product development process for services. Intangibility often allows for quick reactions to changed customer needs. However, it also risks haphazardness in service design. Due to intangibility sustainable advantage and a proprietary position are difficult to achieve in service business, for example through patent. This results in a proliferation of similar services and diminished incentive for the innovator to invest time and resources in truly pioneering efforts in service development.

Simultaneous production and consumption. Services are often produced and consumed in the presence of customers. Consequently, for new services, production and delivery are central facets of what customers purchase, and thus the successful service design is likely to require involvement from many different functional specialties within the firm.

Heterogeneity. Production and delivery depend on company personnel. Consequently, consumption can vary at each purchase occasion. On the one hand, this offers opportunities to better satisfy client needs through service customization. On the other, this can also be viewed as a lack of consistency in the service product, in other words, poor service quality, perceptions of unreliability and heightened customer risk.

Perishability. Services cannot be preproduced and inventoried. Thus, opportunities to produce and sell the service are irrevocably

lost when demand is either above or below the firm's capacity. Service firms can incur high costs when facilities and staff are idle during low demand periods, and lost revenue during peak demand periods. The challenge for new service development is in designing alternate service level offerings. The challenge is to design alternate service level offerings, for example "full" versus "limited" service offerings for low and peak demand periods. Indeed, developing a countercyclical line of services may also be the objective of service innovation.[76 , p.102]

The models describing the innovation process in the services context typically suggest that the innovation process is a sequence of phases during which the initial ideas for a new service are refined into a deliverable service in the market place. Ojasalo [77] proposed the following general phases for an innovation process of services: (a) strategy development related to the new service, (b) generating and screening of ideas and selecting one for development, (c) business analysis; including markets, internal conditions, and profitability, (d) development of service concept, (e) testing, (f) launch, and monitoring and modifying the service.

4. 10 PROPOSITIONS FOR USING MODELLING IN SERVICE INNOVATION

Based on the earlier literature review 10 propositions for using modeling in management of service innovation are presented next.

1. *Identify the problem to be modeled*. A problem can be understood as a gap between existing and desired condition. A problem has certain basic dimensions, which should be examined: content, location, owner, absolute and relative magnitude, and time perspective. For example, in which part of the service process does the problem exist? In which department or customer segment of the services company does the problem exist? Is this problem perceived by the front line employees or top executives of the service company? Who resist solving and modeling the problem? How much revenue do we lose due to the problem, measured in absolutely in dollars and relative of the total sales? When did the problem appear the first time? How often does the problem appear? Is there a trend with its development? [78]
2. *Consider the psychology of individuals and organizations.* Different individuals and organizations may be involved in the problem in hand, and its modeling. They may have the very different interest or perceptions of the situation. Different backgrounds, experience, knowledge, skills, attitudes, and motives of people involved affect both the modeling process and outcome. For example, if potential customers of the new services being developed and modeled are competitors, then it may be reasonable not to involve them in the process at the same time. The modeler's former experience in modeling may often be an advantage. On the other hand, it may sometimes make the modeling effort biased. Moreover, the organizational culture, political issues, power games, and potential tendency to turn a blind eye to problems may all affect the modeling.[79]
3. *Avoid pitfalls*. Modelers should try to avoid certain basic pitfalls. One should not mix symptoms and the root cause of the problem. The root cause of the problem may cause a variety of symptoms. Symptoms disappear only by identifying and solving the underlying root cause problem. Also, the modeler should not have too strong an opinion of the nature of the problem in advance. Moreover, one should not make modeling effort just from one perspective, like for example from the perspective of the middle management. In addition, one should not forget that the problem is perceived differently among different stakeholders. Customers of the services company may have totally different opinion from the view of the top management. Moreover, the modeler should make sure that the modeling project does not fall by the wayside. Sometimes there is big risk for this, since time, money, personnel, information, and other resources needed in the modeling are scarce. Finally, one should not make a careless objective setting at the beginning of the project since this will cause a great waste of resources in the later phases of the modeling.[80]
4. *Take advantage of value co-creation*. The meaning of value and the value creation process are rapidly shifting from a supplier company-centric view to customer experiences and joint value co-creation. In value co-creation, customers engage in the process of both defining and creating value [81]. When value is co-created, the supplier contribution is a value proposition that can support customer's value creation processes, and the customer contribution is the value actualization.[82] Co-creation means joint creation of value by the provider and the customer.[83] Value is co-created in learning together, and dialog operates as an active interactive process of learning. Customers are in a proactive role. They are involved at every stage of service development. An active dialog improves identifying customers' latent needs and wants. Customers may also be directly involved to develop new value propositions, i.e. co-design is one form of co-creation.[84] In value co-creation, new levels of access and transparency are needed. Focus of quality is on customer-company interactions and co-creation experiences.[85]
5. *Model experience.* Many models focus on mapping and displaying visible and observable reality. This approach functions well in the context of technical systems and machines. However, when modeling human and social systems, such as organizations, this is clearly not enough. In the case of service organizations, it is paramount to focus on subjective experience of individuals. To the large extent, services are produced by humans, not by machines. The production of services does not take place in isolation of customers, but instead in interaction with them. Thus, the subjective total experience of both service customers and front line employees is essential. Modeling activities, materials and data flows, organizational structures, etc. is important, but the experience of paying customers is the most important. The criteria for good customer experience should always come from the customers themselves. When a physical good, such as a camera, does not function, it can easily be observed and measured by the quality control. However, subjective customer experience is much more difficult to see and

observe. It requires specific techniques and deep customer understanding.

6. *Design and model all elements of service experience.* If customer experience is the central element of modeling in the context of services innovation, what is driving it? What is customer's service experience composed of? The following design principles are useful to remember.[86] (a) All features of design must share a common purpose. The fulfillment of each design aspect is essential for the design to be considered total. (b) Every design component and the overall design have a viable lifespan beyond which it will be rendered obsolete. (c) The value attributed to the design of a service is influenced by experience and is concerned with characteristics which the customer determines important. (d) The design process is depending on the identification, evaluation, prioritisation and selection of resources with regard to materials, time, labour, expertise and creativity. (e) The design outcome is the result of the synthesis among knowledge, resources and creativity. (f) The design process is ongoing in nature and has to be constantly improved. (g) A design strategy serves to manage change and to encourage it. (h) The success of design depends on its ability to relate to those responsible for its management, i.e. design must reflect the strategic position of a service provider. (i) Sound management is needed to integrate design within the business function. (j) Service design has to satisfy everybody, not just those for whom it is directly intended.
7. *Base the model development in a deep customer understanding.* In order to develop successful service business models, a deep customer understanding is needed. The traditional methods, such as surveys and personal interviews are not enough anymore. The modeler has to go beyond explicit knowledge and observable reality. It is vital to get access to tacit knowledge and understand customers' latent needs. Of course, the modeler needs to know what the customers of his business model say and think. But, in addition, more importantly, he or she needs to know what customers feel, imagine, and dream. This can be done with the help of new methods that stimulate and enable customers and users to construct in their minds and in their imagination the kind of world, the kind of service they would like to have. Such new methods have been developed in the recent years. They are typically different kinds of emotional tools, including various visual, playful, storytelling techniques that reveal dreams, fear, and excitement. With such methods users and customers are seen as creators rather than just informants. The assumption is that the modeler does not know in advance what he will get. The idea is to stimulate the customer to reveal his/her latent needs, and elaborate them further.
8. *Foster creativity.* Analytical, objective, logical, and systematic methods and skills are important for innovating new service models. In addition to skills, creativity has a crucial role in developing new breakthrough service models. The main characteristics of a creative culture are: (a) leadership by visionary, enthusiastic champions of change, (b) top management support and encouragement of creativity, both financial and psychological, (c) an effective communication system. Leaders share the business vision with their staff and empower them to optimize their potential in achieving the business goals, (d) flexibility towards new thinking and new behavior patterns; the creative organization readily adapts to change and proactively searches for new opportunities, (e) customer focus — the satisfaction of all customers, both internal and external — is the dominant prevailing ethos in innovative companies, (f) desire to look for ideas among competitors, customers, academe, suppliers, and even industries with a different focus, and (g) desire to exploit the talent of the entire organization, stressing the point that "a good idea does not care who has it".[87]
9. *Think different and radically new service business models.* The nature of service business is constantly changing. This requires the modeler to think outside the box. Increasingly, he or she has to imagine totally new business models. The following strategies can be useful in developing radically new business models in services. (a) Focusing on redefining certain core attributes of existing products in ways that fundamentally shift the financial model away from goods to services. (b) Focusing on leveraging data, information and knowledge gathered automatically or through customer interactions that reinforce existing products or services, or create new revenue-generating services. (c) Focusing on transforming existing products or services of creating entirely new value through internet or mobile delivery and experience. (d) Focusing on expanding the value of existing products or services by adding services that meet a broader set of customer needs or by introducing complementary service offerings.[88]
10. *Establish collaborative networks.* The meaning of collaborative networks is increasingly important in almost any type organization and function.[89, 90] Networking has great potential in innovation management as well.[91] The fundamental advantage of collaborative networks to modelers is that they enable fast and dynamic access to external resources needed in the model development effort. Growing all vital competences in-house is, in many cases, simply too expensive and time consuming. This is particularly true if the model development is takes place in a smaller company. By utilizing case specific expertise, information, and other resources from the network the modeler is able to conduct his or her task more efficiently and effectively.

5. CONCLUSIONS

This paper examined theoretical grounds of modeling and service innovation, and provided propositions for using modeling in service innovation. This article was based on an extensive literature analysis. It reviewed the literature on the general principles of the modeling of business processes and systems, service blueprinting, and special characteristics of service innovation. This conceptual article contributed to the literature by making ten propositions for using modeling in service innovation. These propositions related to identifying the problem to be modeled, the psychology of individuals and organizations, avoiding pitfalls, taking advantage of value co-creation, modeling

experience, designing and modeling all elements of service experience, basing the model development in deep customer understanding, fostering creativity, thinking different and radically new service business models, and establishing collaborative networks.

6. REFERENCES

[1] Will, H. J. 1975. Model Management Systems. In **Information Syst. and Organization Structure**, E. Grochla and N. Szyperski, Eds. W. Gruyter, Berlin, 468-482.

[2] Dolk, D. R. and Konsynski, B. 1985. Model Management in Organizations. **Information and Manag.** 9, 1, 35-47.

[3] Applegate, L. M., Konsynski, B. R. and Nunamaker, J. F. Model Management Systems: Design for Decision Support. **Decision Support Systems**. 2, 81-91.

[4] Geoffrion, A. M. 1987. An introduction to structured modeling. **Management Science**. 33, 5, 547-588.

[5] Raghu, T. S., Jayaraman, B., and Rao, H. R. 2004. Toward an integration of agent- and activity-centric approaches in organizational process modeling: incorporating incentive mechanisms**. Information Systems Res**. 15, 4, 316-335.

[6] Danesh, A. and Kock, N. 2005. An experimental study of process representation approaches and their impact on perceived modeling quality and redesign success. **Business Process Management Journal.** 11, 6, 724-35.

[7] Sun, S. X., Zhao, J. L., Nunamaker, J. F, and Sheng, O. R. L. 2006. Formulating the data-flow perspective for business process management. **Inf. Sys. Res**. 17, 4, 374-91.

[8] Damij, N. 2007. Business process modelling using diagrammatic and tabular techniques. **Business Process Management Journal**. 13, 1, 70-90.

[9] Frye, D.W. and Gulledge, T.R. 2007. End-to-end business process scenarios. **Ind. Mgmt. & Data Syst**. 107, 6, 749-61.

[10] Turetken, O. and Schuff, D. 2007. The impact of context-aware fisheye models on understanding business processes**:.** ***Information and Management***. 44, 40-52.

[11] Wegmann, A. and Le, L.-S. Regev, G., and Wood, B. 2007. Enterprise modeling using the foundation concepts of the RM-ODP ISO/ITU standard. **Information Systems and e-Business Management.** 5, 397-413.

[12] Kettinger, W.J., Teng, J.T.C., and Guha, S. 1997. Business process change: a study of methodologies, techniques and tools. **MIS Quarterly**. 21, 55-80.

[13] Pidd, M. 1999. Just Modeling Through: A Rough Guide to Modeling. **Interfaces.** 2 (March-April 1999), 118-132.

[14] Wright, G., G. P., Chaturvedi, A. R., Mookerjee, R. V., and Garrod, S. 1998. Integrated modeling environments in organizations: an empirical study. **Information Systems Research**. 9, 1, (March 1998), 64-84.

[15] Boothroyd, H. A. 1978. **Articulate Intervention**. Taylor & Francisand, London.

[16] Pidd, M. 1999. Just Modeling Through: A Rough..

[17] Hammer, M. 1990. Reengineering work: don't automate. Obliterate. **Harvard Business Review**. 68, 4, 104-12.

[18] Davenport, T .H. 1993. **Process Innovation: Reengineering Work through Information Technology**. Harvard Business School Press, Boston, MA.

[19] Davenport, T.H. and Short, J.E. 1990. The new industrial engineering: information technology and business process redesign. **Sloan Management Review**. 31, 4, 11-27.

[20] Laguna, M. and Marklund, J. 2005. **Business Process Modelling, Simulation, and Design**. Pearson Education, Inc., Upper Saddle River, NJ.

[21] Denna, E.L., Perry, L.T. and Jasperson, J. 1995. Reengineering and REAL business process modeling. In **Business Process Change: Reengineering Concepts, Methods, and Technologies** V. Grover and W. J. Kettinger, Eds. Idea Group Publishing, London, 350-75.

[22] Curtis, B., Kellner, M.I. and Over, J. 1992. Process modeling. **Communications of the ACM**, 35, 9, 75-90.

[23] Luo, W. and Tung, Y. A. 1999. A framework for selecting business process modeling methods. **Industrial Management & Data Systems.** 99, 7, 312-319.

[24] Kim, S.-H. and Jang, K.-J. 2002. Design performance analysis and IDEF0 for enterprise modeling in BPR. **Int. Journal of Production Economies**. 76, 121-133.

[25] Luo, W. and Tung, Y. A. 1999. A framework for..

[26] Martin, J. and McClure, C. 1985. **Diagramming Techniques for Analysts.**. Englewood Cliffs, Prentice-Hall.

[27] Willemain, T. R. 1994. Insights on modeling from a dozen experts. **Oper. Res**. 42, 2, (March-April 1994), 213-222.

[28] Pidd, M. 1999. Just Modeling Through: A Rough..

[29] Raiffa, H. 1982. **Policy analysis: a checklist of concerns**. PP-82-2. Int. Inst. for Appl. Syst. Anal. Laxenburg, Austr.

[30] Pidd, M. 1999. Just Modeling Through: A Rough..

[31] Willemain, T. R. 1994. Insights on modeling..

[32] Shostack, G. L. 1984. Designing services that deliver. **Harvard Business Review**. 62, (Jan-Feb 1984), 133-139.

[33] Shostack, G. L. 1982. How to Design a Service? **European Journal of Marketing**. 16, (Jan-Feb 1982), 49-63.

[34] Shostack, G. L. 1984. Designing services that deliver.

[35] Shostack, G. L. 1987. Service design in operating environment. In **Developing new services**, W. R George and C. E. Marshall, Eds. AMA. Chicago, 27-43.

[36] Shostack, G. L. 1987. Service positioning through structural change. **J. of Marketing**. 51, (Jan 1987), 34-43.

[37] Kingman-Brundage, J. 1989. The ABC's of service system blueprinting. In **Designing a Winning Serv. Strat**., M. J. Bitner and L. A. Crosby, Eds. AMA, Chicago.

[38] Kingman-Brundage, J. 1993. Service mapping: gaining a concrete perspective on service system design. In **The service quality handbook**. E. S. Eberhard and W. F. Christopher, Eds. Amacon, New York, 148–63.

[39] Kingman-Brundage J. 1995. Service mapping: back to basics. In **Understanding servs. manag**. W. J. Glynn and J. G. Barnes, Eds. John Wiley & Sons, Chichester, 119–42.

[40] Kingman-Brundage J, and George, W. R. 1996. **Using service logic to achieve optimal team functioning**. QUIS5, International Service Quality Assoc., New York, 13– 24.

[41] Kingman-Brundage J, George W. R, Bowen, D. E. 1995. ''Service logic'': achieving system integration. *Int.* **Journal of Service Industry Management**. 6, 4, 20–39.

[42] Gummesson, E. and Kingman-Brundage, J. 1991. Service design and quality: applying service blueprinting and service mapping to railroad services. In **Quality Management in Services**. P. Kunst, and J. Lemmink, Eds. Van Gorcum, Netherlands.

[43] Fleiss, S. and Kleinaltenkamp, M. 2004. Blueprinting the service company. Managing services processes efficiently. **Journal of Business Research**. 57, 392-404.

[44] Bitner, M. J., Ostrom, A. L., and Morgan, F. N. 2008. Service Blueprinting. A practical technique.. for service innovation. **Calif. Manag. Rev**. 50, 3 (Spring 2008), 66-94.

[45] Johne, A. and Storey, C. 1998. New service development: a review of the literature and annotated bibliography. **European Journal of Marketing**. 32, 3/4, 184-251.

[46] Bitner, M. J. et al. 2008. Service Blueprinting. A practical

[47] Shostack, G. L. 1987. Service positioning through structural change. **J. of Marketing**. 51, (Jan 1987), 34-43.

[48] Shostack, G. L. 1984. Designing services that deliver.

[49] Zeithaml, V. A., Bitner, M. J., and Gremler, D. D. 2009. **Services Marketing. Integrating Customer Focus Across the Firm**. McGraw-Hill, New York.

[50] Johne, A. and Storey, C. 1998. New service development: a review of the literature and annotated bibliography. **European Journal of Marketing**. 32, 3/4, 184-251.

[51] Gummesson, E. and Kingman-Brundage, J. 1991. Service..

[52] Bitner, M. J. et al. 2008. Service Blueprinting. A practical

[53] Zeithaml, V. A., et al. 2009. **Services Marketing. Integ**..

[54] Fitzsimmons, J. A. and Fitzsimmons, M. J. 2006. **Service Management. Operations, Strategy, Information Technology**. McGraw-Hill, New York.

[55] Zeithaml, V. A., et al. 2009. **Services Marketing. Integ**..

[56] Bitner, M. J. 1993. Managing the Evidence of Service. In **The Service Quarterly Handbook**. E. E. Scheuing and F. Christopher, Eds. AMACOM: New York, 358-370.

[57] Kingman-Brundage, J. 1989. The ABC's of service system blueprinting. In **Designing a Winning Service Strategy**, M. J. Bitner and L. A. Crosby, Eds. AMA, Chicago.

[58] Bitner, M. J. et al. 2008. Service Blueprinting. A practical

[59] Zeithaml, V. A., et al. 2009. **Services Marketing. Integ**..

[60] Bitner, M. J. et al. 2008. Service Blueprinting. A practical

[61] Parasuraman, A., Zeithaml, V.A. and Berry, L.L. 1985. A Conceptual Model of Service Quality and It´s Implications for Future Research. **J. of Marketing**, 49, Fall, 41-50.

[62] De Brentani, U. 1995. New Industrial Service Development: Scenarios for Success and Failure., **J. of Business Research**, 32, 93-103.

[63] Berry, L.L. 1980. Services Marketing Is Different. **Business,** 30, May-June, 24-29.

[64] Shostack, G. L 1984. Designing Services that Deliver. **Harvard Business. Review**, 62, January-February, 133-139.

[65] Shostack, G. L. 1987. Service Positioning Through Structural Change. **J. of Marketing**, 51, January, 34-43.

[66] Easingwood, C. J. 1986. New Product Development for Service Companies. **J. of Product Innovation Management,** 3, December, 264-275.

[67] Wind, Y. J. 1982. **Product Policy: Concepts, Methods and Strategy.** Addison-Wesley, Reading, MA., 550-553.

[68] Gummesson, E. 1981. The Marketing of Professional Services: 25 Propositions, in **Marketing Services,** Donnelly, J.D. and George, E.R. eds., American Marketing Association, Chicago.

[69] Lynn, S. A. 1987. Identifying Buying Influences for a Professional Service: Implications for Marketing Efforts. **Industrial Marketing Management**, 16, 119-130.

[70] Grönroos, C. 1982. An Applied Service Marketing Theory. **Eur.J. Marketing**, 16, 30-41.

[71] Grönroos, C. 1990. **Service Management and Marketing: Managing the Moments of Truth in Service Competition,** Lexington Books, Lexington, MA.

[72] Jackson, R. W. and Cooper, P. D. 1988. Unique Aspects of Marketing Industrial Services. **Industrial Marketing Manag**. 17, 111-118.

[73] Maister, D. H. and Lovelock, C. H. 1982. Managing Facilitator Services. **Sloan Management Review**, 23, Summer, 19-31.

[74] Easingwood, C. J. and Mahajan, V. 1989. Positioning of Financial Services for Competitive Advantage. **J. of Product Innovation Management**, 6 August, 207-219.

[75] Levitt, T. 1976. The Industrialization of Services. **Harvard Business Review**, 48, September-October, 63-74.

[76] De Brentani, U. 1995. New Industrial Service Development: Scenarios for Success and Failure., **J. of Business Research**, 32, 93-103.

[77] Ojasalo, J. 2008. Innovation Management in Knowledge Intensive Services. **The Business Review, Cambridge**, 9, 2, 212-219.

[78] Kubr, M. (ed.) 1996. **Management Consulting. A Guide to Profession.** ILO, Geneva. 172-173.

[79] Büyükdamgaci, G. 2003. Process of Organizational Problem Definition: Hoe to Evaluate and Improve. Omega. **The International Journal of Management Science**. 13, 327-338.

[80] Kubr, M. (ed.) 1996. **Management Consulting. A Guide to Profession**. ILO, Geneva. 172-173.

[81] Ojasalo, K.2010. The Shift from Co-Production in Services to Value Co-Creation. **The Business Review, Cambridge,** 16, 1, 171-177.

[82] Gummesson, E. 2008. Extending the New Dominant Logic: From Customer Centricity to Balanced Centricity. **The Journal of the Academy of Marketing Science**. 36, 1, 15-17.

[83] Prahalad, C.K. and Ramaswamy, V. 2004. Co-creation experiences: The next practice in value creation. **Journal of Interactive Marketing**, 18, 3.

[84] Flint, D.J. & Mentzer, J.T. 2006. Striving for Integrated Value Chain Management Given a Service-Dominant Logic for Marketing. In Lusch & Vargo (eds.), **The Service-Dominant Logic of Marketing,** M.E. Sharpe, Inc.

[85] Prahalad, C.K. & Ramaswamy, V. 2004. Co-creating unique value with customers. **Strategy and leadership**, 32, 3, 4-9.

[86] Doyle, S. A. and Broadbridge, A. 1999. Differentiation by design: the importance of design in retailer repositioning and differentiation. **International Journal of Retail &**

Distribution Management, 27, 2, 72-82. Doyle and Broadbridge's research results were summarized by Mager, B. 2004. **Service Design – A Review**. Proma Print Kolen, p. 44. See also Zehrer, A. 2009. Service Experience and Service Design: Concepts and Applications in Tourism SMEs. **Managing Service Quality**, 19, 3, 332-349.

[87] Flynn, M., Dooley, L., O'Sullivan, D. and Cormican, K. 2003. Idea Management for Organisational Innovation. **International Journal of Innovation Management**, 7, 4, 417-442.

[88] Tekes. 2010. **The Future of Service Business Innovation**. Tekes Review 27, 2.

[89] Ojasalo, J. 2004. A Framework for Managing Cross-Cultural Differences in International Business Networks. **The Business Review, Cambridge,** 1, 2, 77-83.

[90] Ojasalo, J. 2004. Key Network Management. **Industrial Marketing Management,** 33, 3, 195-205.

[91] Ojasalo, J. 2008. Management of Innovation Networks — A Case Study of Different Approaches**. European Journal of Innovation Management,** 11, 1, 51-86.

Sequential Metamodeling Approach for Optimum Design of Contact Springs Used in Electrical Connectors

Kun-Nan CHEN
Department of Mechanical Engineering, Tungnan University
New Taipei City 222, Taiwan
knchen@umd.edu or knchen@mail.tnu.edu.tw

ABSTRACT

In this paper, optimum designs of a contact spring used in an electrical connector are achieved using a sequential metamodeling approach, with the objective function of the optimization problem being defined as the maximum von Mises stress in the contact spring. In order to ease the computational burden of the optimization process, the procedure is split into two stages, each with four design variables. This two-stepped scheme utilizes the Face-centered central composite experimental design concept, performs non-linear contact finite element analysis on every design point, builds response surface models with regression analysis, and uses the quadratic programming technique to optimize the approximated models.

Keywords: Electrical Connector, Contact Spring, Optimum Design, Metamodeling and Response Surface.

1. INTRODUCTION

With ever increasing demands for portable electronic devices, the reliability of their rechargeable power systems has become an important issue. A portable device connects to a battery or an electronic charger through an electrical connector. An electrical connector serves to couple two circuit devices in an electronic system. A basic electrical connector consists of four elements [1]: contact interface, contact finish, contact spring and connector housing. The contact interface can be categorized into two groups, i.e., the separable interfaces and permanent interfaces, while the contact spring performs three functions in a connector: supplying an electrical path between two subsystems, producing the normal contact force that establishes and maintains the separable interfaces, and permitting the formation of the permanent connections. A separable-interface connector may have many varieties and different shapes and sizes depending on a given set of requirements for a particular application. Figure 1 shows a common cell phone battery, several types of electrical connectors, and a connector embedded in a cell phone.

One of the most important factors affecting the reliability of a high-cycle electronic connector is their mechanical performance, which includes contact forces, deformation and stresses, etc., in the contact springs. Localized damage to the Au plating of a connector caused by a simple, manual coupling operation could lead to a high rate of functional failure [2]. Due to growing demands for smaller connectors with higher mechanical performance, a proper design of the contact spring to achieve the required mechanical performance is increasingly difficult. Weight *et al.* [3] modeled and optimized the contact spring of a constant force electrical connector (CFEC) used in a personal digital assistant docking station, with the ratio of the minimum force to maximum force calculated over the mechanism displacement as the objective function, which provides a good measure about how invariable the contact force of the mechanism truly is. Hsu *et al.* [4] parameterized the geometry of a contact spring pair of a board-to-board connector, and minimized the insertion force while kept the contact normal force and resulting stress within specified ranges. The mating of a contact pair usually involves nonlinear contact force and large deformation, and even plastic theories. Manninen *et al.* [5] and Deshpande and Subbarayan [6] studied the press-fit connector of a printed circuit board and a land grid array (LGA) connector, respectively, using nonlinear finite element (FE) analysis with plastic deformation consideration.

In this paper, the shape and size of a contact spring used in an electrical connector (a separable-interface connector) are analyzed to meet a particular set of constraints using the finite element method with a non-linear contact model and large deformation theories. Further, structural optimization on the contact spring is attained using a two-stepped, sequential metamodeling approach to simplify the optimization procedure. The structural optimization problem for the contact spring is solved to minimize the maximum von Mises stress occurred when the contact spring is engaged with the contact plate. The two-stepped procedure integrates the experimental design concept with a faced-center central composite design, non-linear contact finite element analysis on every design point, regression analysis for building response surface model, and optimization on the approximated model using the quadratic programming technique. Meanwhile, during the optimization procedure, design space reduction scheme is adopted to improve the accuracy. Finally, simulation of a contact plate approaching to and then moving away from the contact spring is rendered to examine the relation between the normal contact force in the contact spring and the traveling distance of the rigid contact plate.

Fig. 1. Cell phone battery and electrical connectors

2. RESPONSE SURFACE APPROXIMATION

The response surface methodology [7], which was originally intended as an empirical modeling approach, is a collection of procedures including design of experiments (DOE), model selection and fitting, and optimization on the fitted model. The methodology has long been expanded to include simulation modeling and approximations. In particular, RSM has been employed by many authors, e.g. [8-11], to solve design optimization problems, especially in the area of multidisciplinary design optimization. A response surface approximation (RSA), usually in the form of a simple polynomial function, can be built from DOE (with numerically simulated experiments) and model fitting. Once a polynomial RSA is created, the optimization on the function can be easily accomplished by most optimization techniques. The most attractive features of RSA are less number of repeated response evaluations and optimization without needing the sensitivity information. In recent years, applications of RSA or RSM-based design optimization in microelectronics have increased quite dramatically. In the present work, with the help of RSA, a minimum stress design, which minimizes the von Mises stress in the contact spring after connection, will be presented. The optimum design maximizes the reliability of the contact springs, as far as reducing the stress is concerned.

Response surface approximation plays a crucial role in RSM. A response surface is a functional expression for a relationship between a response and a set of dependent variables. A complex function (or a response) y can be approximated by a response surface approximation $\hat{y}$ with k independent variables (or factors) $x_1, x_2, \ldots, x_k$ as

$$y = \hat{y}(x_1, x_2, \cdots, x_k) + \varepsilon \tag{1}$$

where ε is the error between the approximated and the exact values of y. The approximating function $\hat{y}$ usually takes on the form of a polynomial whose coefficients can be determined by the least squares method using data from a chosen set (decided by design of experiments) of the independent variables and the resulting responses. For a second order polynomial expression, the approximating function $\hat{y}$ has the form

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j \tag{2}$$

where the βs are the regression coefficients to be determined, and there are $(k+1)(k+2)/2$ such coefficients.

A successful application of RSA is greatly dictated by a proper choice of sampling points in design space, i.e., design of experiments. A face-centered central composite design (FCCD) with its independent variables confined within certain upper and lower bounds belongs to a family of central composite designs, which are the most popular second-order designs. An FCCD consists of 2^k factorial points, $2k$ face-centered configurations, and one center point, for a total of 2^k+2k+1 design points. Figure 2 demonstrates two FCCDs with k=2 and k=3. When performing RSM-based design optimization, response surface approximations are often repeatedly executed, and since a good RSA result may only be valid within certain distance around the center design point, the design space can be reduced after each iteration. In general, either increasing the number of experimental trials (design points) or downsizing the design space can effectively enhance the accuracy of RSA. However, the added number of experimental runs can also significantly raise the experimental or computational cost. Therefore, the design space reduction method seems to be a preferable choice, and this technique is done by introducing a pair of move limits on every design variable. After each iteration, both limits in every pair are moved closer to the other by the same ratio, resulting in a much smaller design space centering at the optimizer obtained from the previous iteration. Then a new response surface is subsequently constructed and a new optimum sought. If some portion of a new design space exceeds the boundary of the previous one, the portion is excluded before performing DOE and RSA.

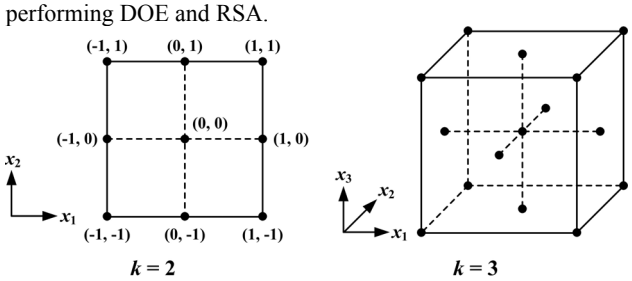


Fig. 2. Face-centered central composite designs

Following the construction of an RSA, the quality of the approximation can be assessed by some statistical testing functions. The coefficient of multiple determination R^2 is a measure of the amount of predictability for the response y by the approximating function $\hat{y}$, and the coefficient is defined as

$$R^2 = \frac{S_r}{S_t} = 1 - \frac{S_e}{S_t} \tag{3}$$

and

$$S_r = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 \tag{4}$$

$$S_e = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{5}$$

$$S_t = S_r + S_e \tag{6}$$

where S_r, S_e and S_t are called the sum of squares due to regression, sum of squares due to residual and total sum of squares, respectively; y_i is the ith observation, $\bar{y}$ the average value of all observations, $\hat{y}_i$ the response surface approximation evaluated at the ith set of independent variables, and N the total number of observations. R^2 takes on a value between 0 and 1. A larger value of R^2 does not necessarily indicate a closer fit of the approximation to the response since adding a variable will always raise the value of R^2. The adjusted coefficient of multiple determination, which will not increase if an added variable is not statistically significant and therefore is a better indicator than R^2, is defined as [7]

$$R_{adj}^2 = 1 - \frac{S_e/(N-p)}{S_t/(N-1)} = 1 - \left(\frac{N-1}{N-p}\right)(1 - R^2) \tag{7}$$

where p denotes the number of the regression coefficients.

3. PROBLEM DEFINITION AND FORMULATION

The contact spring of an electrical connector under study is shown in Fig. 2. Most part of the contact spring, except the contact head, is enclosed in the connector housing. During a connection process, a contact pad from the other subsystem moves in to touch the contact head and complete an electrical path between the two subsystems. As the contact pad continues to move in further, the spring is furthermore compressed, producing an increasing contact normal force. A greater contact

force has positive effects on contact electrical resistance and the mechanical stability of the interface, an increasing normal force leading to a decreasing contact resistance and to a better ability to withstand disturbances. However, countering these positive effects is the effect of rising contact force on the stresses in the spring. Higher stresses inevitably reduce the fatigue life of the contact spring, and even produce excessive plastic deformation that could result in a reliability problem on the connector. One of the objectives of this study is to minimize the von Mises stress in a shape design of the spring after the connector is mated. The contact spring is assumed to be made of beryllium copper, whose properties are listed in Table 1. A plastic hardening relation for the material, adopted from [6], is also assumed and given in Table 1.

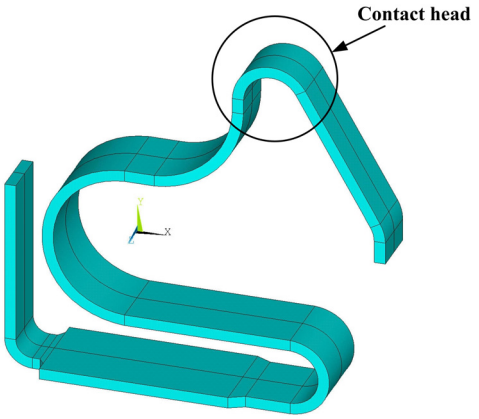


Fig. 2. Contact spring of an electrical connector

Table 1. Properties of beryllium copper

Young's modulus (GPa)	Poisson's ratio	Plastic hardening	
		Plastic strain	Stress (MPa)
113.85	0.3	0.0	621
		0.09454	759

The structural shape optimization of the contact spring requires repeated nonlinear elastic-plastic contact analysis, which could lead to a prohibitively high computation cost and a convergence difficulty if a conventional first-order mathematical programming technique is used to solve the 3-dimentional finite element model in this research. To ease the computational burden and reduce the risk of no convergence, a metamodeling scheme, the response surface methodology, is adopted to solve the problem.

Design Variables and Constants

Before commencing a finite element analysis, the structural model of the contact spring is parameterized. The geometrical parameters, shown in Figs. 3 and 4, establish the shape and size of the structure and are employed as inputs, i.e. design variables and constants, to the analysis and optimization. Figure 3 illustrates the geometrical parameters on the *x*-*y* plane for the contact spring, including design variables r_1-r_4 and d_1-d_3, constants C_{d1}-C_{d6}, C_{r1}-C_{r2} and t, and geometrical constraints H_h, V_h, V_t and H_g. Figure 4 shows the parameters on the *z* axis, which are the beam width parameters (in *z* axis) at various locations, consisting of design variables w_1-w_4 and constants C_{w1}-C_{w4}. The beam thickness t is set constant in consideration of manufacturability, and geometrical constraints H_h, V_h, V_t and H_g are required to satisfy either assembly or functionality restraints. The rest of the constants are set due to their less significant influence on the analysis results. Every cross section perpendicular to the centroidal axis of the bended and curved beam is assumed to have a rectangular shape. The *z* dimensions of the solid model are formed by linearly (in *z*) connecting adjacent beam width parameters. In order to reduce the complexity of the optimization process, the procedure is split into two stages. The first stage treats only r_1-r_4 and d_3 as the design variables and the rest as constants. By considering the geometrical constraints H_h, V_h and V_t, the parameters r_4, d_1 and d_2 may be linked to the other design variables, cutting down the total number independent variables in this stage to four (r_1-r_3 and d_3). When the first-round optimization is completed, the second stage will begin, based on the results obtained from the former, by setting four new design variables w_1-w_4 and the rest constant. The initial values of the design parameters and the constants are given in Table 2. The invariant C_{d5} may be regarded as a function of other constants by noticing the constraint *Hg*, and therefore it is removed from Table 2.

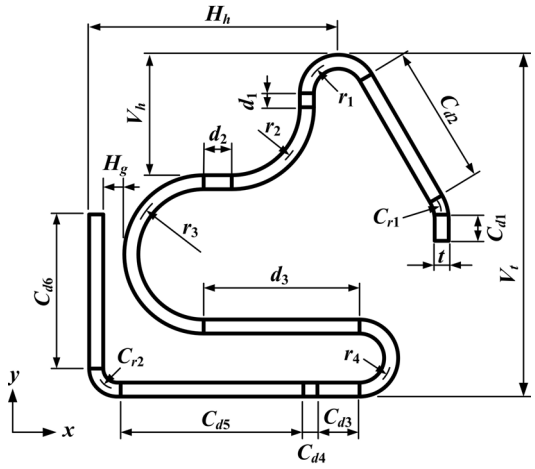


Fig. 3. Geometrical parameters on *x*-*y* plane for the contact spring

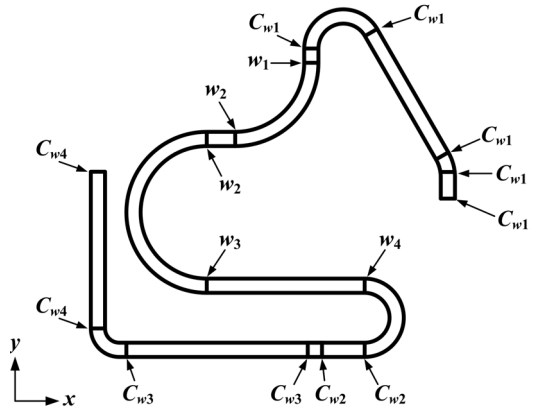


Fig. 4. Geometrical parameters on *z* axis for the contact spring

Table 2. Initial values for the geometrical parameters of the contact spring (all units in mm)

r_1	r_2	r_3	r_4	d_1	d_2	d_3
0.45	1.07	1.03	0.45	0.2	0.4	2.22
t	H_h	V_h	V_t	H_g	C_{d1}	C_{d2}
0.15	3.57	1.8	4.91	0.4	0.3	2.0
C_{d3}	C_{d4}	C_{d6}	C_{r1}	C_{r2}	w_1	w_2
0.6	0.2	2.2	0.45	0.3	0.7	1
w_3	w_4	C_{w1}	C_{w2}	C_{w3}	C_{w4}	
1	1	0.7	1	1.2	0.7	

Parameterized Finite Element Models

Finite element software ANSYS is used for computational modeling of the contact spring. Following a preliminary convergence analysis, the finite element mesh shown in Fig. 5 is proven satisfactory. The finite element model is consisted of 10,638 three dimensional elements, mostly bricks and a few tetrahedron elements. In addition, a finite element contact pair is also defined over portions of the contact pad and spring surfaces that may potentially engage contact, and a friction coefficient of 0.1 between the contact pair is assumed in the analysis. The contact pad, whose contacting surface is assumed to be rigid, has an initial position of 0.075 mm (half of the beam thickness) from above the contact head. Downward compression distances of 1 mm and 1.5 mm on the contact head are realized and studied by moving down the contact pad 1.075 mm and 1.575 mm, respectively.

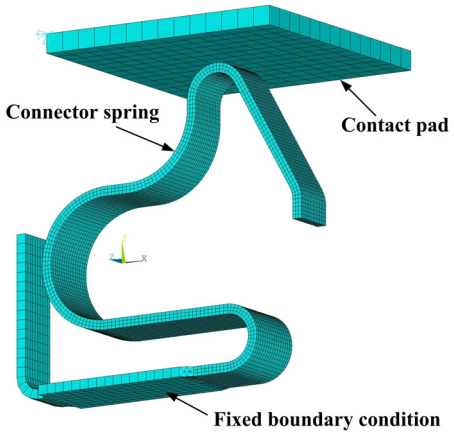


Fig. 5. Finite element mesh of the contact spring model

Optimization Problem Formulation

In this study, the objective function of the optimization problem is the maximum von Mises stress in the contact spring during the entire coupling and decoupling processes between the contact spring and contact pad. And this maximum stress is minimized in an attempt to increase the reliability of the spring and therefore the connector. Mathematically, the optimum shape design problem can be written as

$$Minimize \quad f(\boldsymbol{x}) = \sigma_{\max} \tag{8}$$

Subject to

$$x_i^{\mathrm{L}} \le x_i \le x_i^{\mathrm{U}}, \quad i = 1, 2, \cdots, k \tag{9}$$

where $\sigma_{\max}$ denotes the maximum von Mises stress, x_i are the design variables, and the superscripts U and L represent the upper and lower bounds, respectively. Also, the total number of the design variables is symbolized by k, which is equal to four in both stage one and two. For stage one, the lower and upper bounds for the design variables are: (0.3, 0.5) for r_1, (0.5, 1.2) for r_2, (0.5, 1.2) for r_3, and (1.5, 2.5) for d_3, and for stage two, they are: (0.5, 1.2) for w_1-w_4. The selections of the bounds are based on manufacturability and engineering judgment.

To approximate objective function $\sigma_{\max}$ by RSA, repeated finite element analyses are performed on all design configurations, after which a explicit functional relation, also known as the response surface, of $\sigma_{\max}$ with respect to the design variables is created by least squares curve fitting to a polynomial model. The minimization of the multi-variable polynomial function subjected to side constraints can be easily carried out by common optimization routines.

4. RESULTS AND DISCUSSION

A typical run of the nonlinear elastic-plastic contact analysis using ANSYS takes approximately 950 CPU seconds on a PC. In either stage of the optimization process, there are $N=2^k+2k+1=25$ sets ($k=4$) of design points and, therefore, 25 different analyses need to be performed to constitute one iteration. With the help of shrinking design space boundaries, it usually takes several iterations to achieve convergence. For a comparison purpose, FE analyses using the initial values given in Table 2 yield $\sigma_{\max}$=685.524 MPa for the case of 1 mm downward compression and $\sigma_{\max}$=741.917 MPa for 1.5 mm downward compression. The following subsections will present the optimization results of 1 mm and 1.5 mm downward compression on the contact head for both stages and also the analyses of contact normal forces during the contact engagement process.

Optimum Design after Stage One

For the case of 1 mm downward compression on the contact head, finite element analyses are executed repeatedly using 25 sets of input parameters for the first iteration in stage one, then a quadratic model is fitted. Statistical testing is performed on the fitted model, and gives an R^2 value of 0.992 and an adjusted R^2 value of 0.982, which represent a very good fit of the approximations to the responses. After 6 iterations, convergences are apparent. The final optimized result is checked by another ANSYS verification run, and it produces $\sigma_{\max}$=577.055 MPa at r_1=0.5 mm, r_2=1.2 mm, r_3=1.199 mm and d_3=1.5 mm. This optimum design has its parameters located either at their upper or lower bounds. Figure 6 shows the von Mises stress distribution of this optimum model for the case of 1 mm downward compression on the contact head after stage one optimization.

For the case of 1.5 mm downward compression, larger stresses and even permanent deformations are expected. The same procedure as for the 1 mm case but with a downward compression of 1.5 mm on the contact head is executed. The optimization process converges within 10 iterations. The verification run gives $\sigma_{\max}$=652.563 MPa at r_1=0.351 mm, r_2=0.825 mm, r_3=0.951 mm and d_3=1.921 mm, and the von Mises stress distribution of this optimum model is shown in Fig. 7.

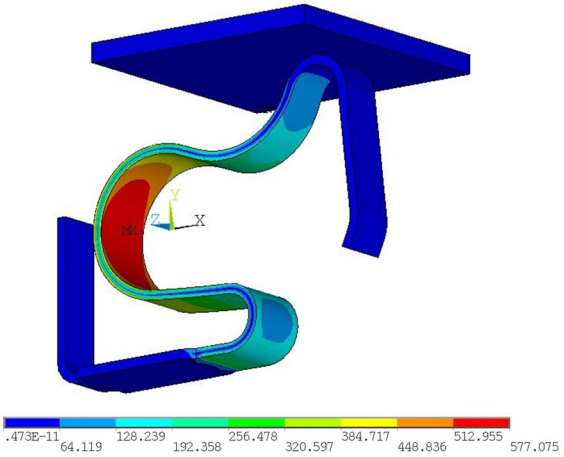


Fig. 6. Stress distribution of the optimum model for the 1 mm compression case after stage one optimization

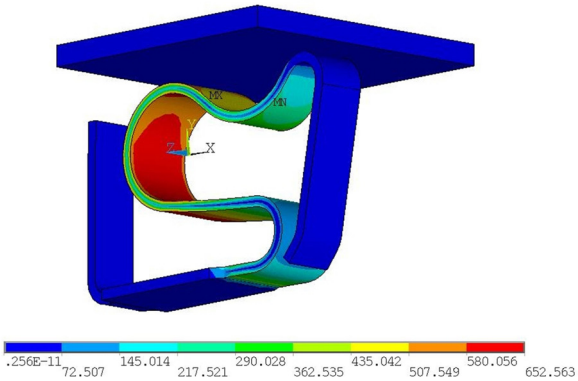


Fig. 7. Stress distribution of the optimum model for the 1.5 mm compression case after stage one optimization

Optimum Design after Stage Two

Based on the optimal parameters acquired from stage one, new design variables, w_1-w_4, are introduced to perform optimization in stage two. Again, a multi-variable quadratic function is matched using the stress data from FE analyses. The fitted model is then tested for its adequacy, and the testing reveals a satisfactory outcome for both 1 mm and 1.5 mm compression cases: adjusted R^2 values of 0.966 and 0.919 for the 1 mm case and the 1.5 mm case, respectively. Within 10 iterations, both cases are converged. The optimum model for the 1 mm case, situating at w_1=0.516 mm, w_2=1.196 mm, w_3=1.192 mm and w_4=0.502 mm, yields σ_{max}=511.156 MPa, shown in Fig. 8. Similarly, the optimum model for the 1.5 mm case, positioning at w_1=0.728 mm, w_2=0.500 mm, w_3=0.937 mm and w_4=0.815 mm, generates σ_{max}=628.926 MPa, shown in Fig. 9. These optimized responses represent significant improvements over those of the initial models (σ_{max}=685.524 MPa and 741.917 MPa). Notice that only one of the design parameters attains its optimal value at the boundary for the 1.5 mm case.

Contact Normal Force vs. Displacement

Beside the stresses in the contact spring, the contact normal force is another important factor dictating the performance of a spring. Figures 10 and 11 show the contact normal forces experienced by the optimum designs of the spring after stage one and stage two optimizations, respectively. When the contact head is compressed downward by approximately 1 mm, the force reaches the maximum at 663.83 mN for the stage-one optimum model, and at 718.50 mN for the stage-two optimum model. A further examination reveals that the spring suffers no plastic strain and undergoes no plastic deformation. If the contact pad is moved back to the initial position, the spring will recover completely to its original shape. When the compressing distance of the spring is extended to 1.5 mm, plastic deformations are clearly visible. Figures 12 and 13 display the force vs. displacement curves as the contact pad travels downward and back. The maximum forces are observed at 1459.29 mN for the stage-one optimum model and at 909.44 mN for the stage-two optimum model, both occurring near the 1.2 mm mark of distance traveled by the contact pad. When the contact pad further moves down, the yielding strength of the spring is exceeded and the normal force drops. As the contact pad continues to press downward to the farthest distance mark of 1.575 mm and then returns to its initial position, the spring endures a maximum plastic strain at the location coincident to that of the maximum stress. The plastic strains lead to a permanent deformation of -0.41 mm (downward) for the stage-one optimum model and -0.43 mm (downward) for the stage-two optimum model, at the tip of the contact head after the contact pad disengages the spring. These irreversible deformations may be significant enough to cause a reliability problem. One possibility to resolve the dilemma is to conduct a new optimization practice that seeks to minimize both the maximum stress and the maximum equivalent plastic strain simultaneously.

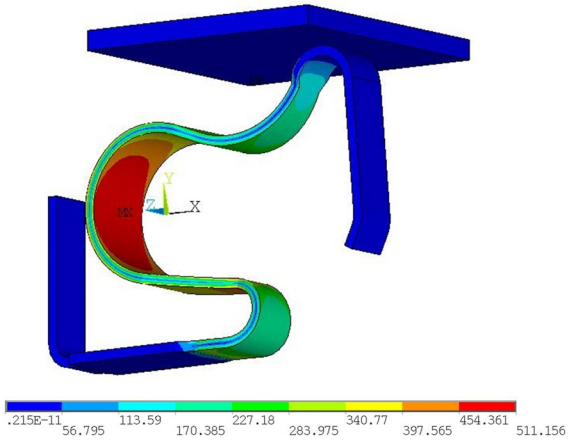


Fig. 8. Stress distribution of the optimum model for the 1 mm compression case after stage two optimization

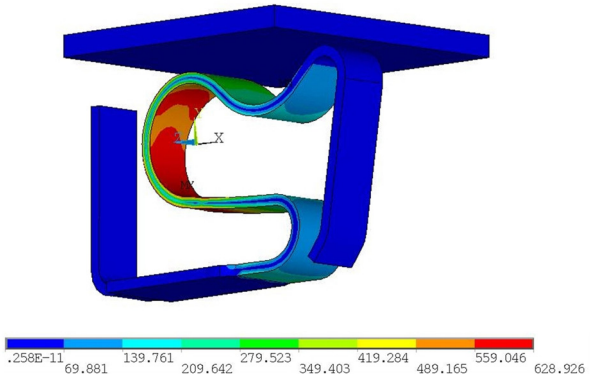


Fig. 9. Stress distribution of the optimum model for the 1.5 mm compression case after stage two optimization

5. CONCLUSIONS

Optimum designs of a contact spring used in an electrical connector have been achieved using a sequential metamodeling approach, with the objective function of the optimization problem being defined as the maximum von Mises stress in the contact spring. In order to ease the computational burden of the optimization process, the procedure was split into two stages, each with four design variables. This two-stepped scheme utilized the FCCD experimental design concept, performed non-linear contact finite element analysis on every design point, built response surface models with regression analysis, and used the quadratic programming technique to optimize the approximated models. Two cases were studied, one with the contact head of the spring pressed downward 1 mm, the other 1.5 mm. Both cases showed that the responses of the optimum models represented significant improvements over those of the initial models. However, simulation results revealed that significant permanent deformations were observed for the 1.5 mm case.

6. ACKNOWLEDGEMENT

This research work has been supported by the National Science Council of Taiwan, ROC under grant no. NSC 100-2918-I-236-001. The financial support is gratefully appreciated.

7. REFERENCES

[1] R.S. Mroczkowski, **Electronic Connector Handbook**, New York: McGraw-Hill, 1998, ch. 1, pp. 1.2-1.10.

[2] P. Arrowsmith, P. Kapadia, A. Hawleyc and R. Sodhid, "Investigation of a connector electrical failure," **Surface and Interface Analysis**, Vol. 43(1-2), 2011, pp. 600-603.

[3] B.L. Weight, C.A. Mattson, S.P. Magleby and L.L. Howell, "Configuration selection, modeling, and preliminary testing in support of constant force electrical connectors," **Journal of Electronic Packaging**, Vol. 129, 2007, pp. 236-246.

[4] Y.-L. Hsu, Y.-C. Hsu and M.-S. Hsu, "Shape optimal design of contact springs of electronic connectors," **Journal of Electronic Packaging**, Vol. 124, 2002, pp. 178-183.

[5] T. Manninen, K. Kanervo, A. Revuelta, J. Larkiola and A. S. Korhonen, "Plastic deformation of solderless press-Fit connectors," **Material Science and Engineering A,** Vol. 460-461, 2007, pp. 633-637.

[6] A. Deshpande and G. Subbarayan, "LGA connectors: an automated design technique for a shrinking design space," **Journal of Electronic Packaging**, Vol. 122, 2000, pp. 247-254.

[7] R. H. Myers and D. C. Montgomery, **Response Surface Methodology: Process and Product Optimization Using Designed Experiments,** New York: John Wiley & Sons, Inc., 1995.

[8] H. Agarwal and J. E. Renaud, "Reliability based design optimization using response surfaces in application to multidisciplinary systems," *Engineering Optimization*, Vol. 36(3), 2004, pp. 291-311.

[9] F. van Keulen and K. Vervenne, "Gradient-enhanced response surface building," **Struct Multidisc Optim**, Vol. 27, 2004, pp. 337-351.

[10] O. Yeniay, R. Unal and R. A. Lepsch, "Using dual response surfaces to reduce variability in launch vehicle design: a case study," **Reliability Engineering & System Safety**, Vol. 91, 2006, pp. 407-412.

[11] G. Steenackers, R. Versluys, M. Runacresb and P. Guillaumea, "Reliability-based design optimization of computation-intensive models making use of response surface models," **Quality and Reliability Engineering International**, Vol. 27(4), 2011, pp. 555-568.

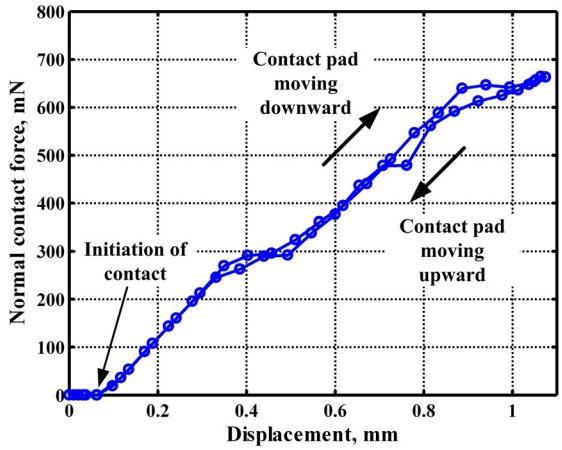


Fig. 10. Contact normal force experienced by the optimum model for the 1 mm case after stage one optimization

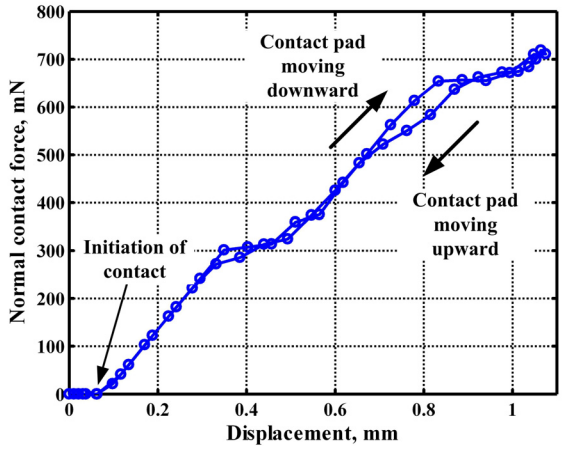


Fig. 11. Contact normal force experienced by the optimum model for the 1 mm case after stage two optimization

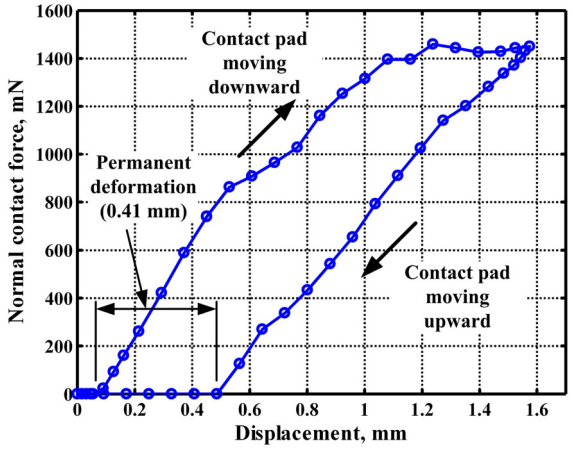


Fig. 12. Contact normal force experienced by the optimum model for the 1.5 mm case after stage one optimization

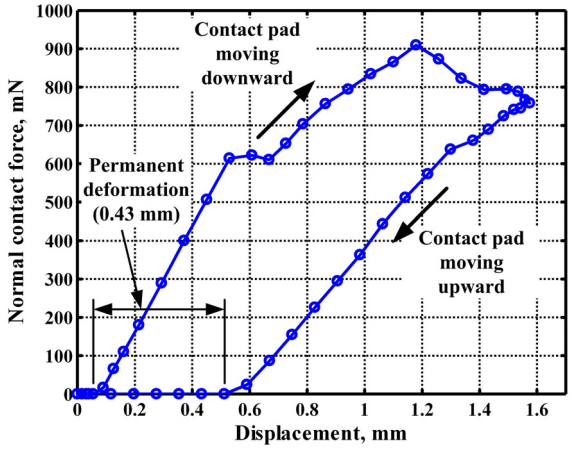


Fig. 13. Contact normal force experienced by the optimum model for the 1.5 mm case after stage two optimization

Time Domain Modeling Of A Band-Notched Antenna For UWB Applications

S.MRIDULA, Binu PAUL, P.MYTHILI
Division of Electronics Engineering, School of Engineering
Cochin University of Science and Technology, Kochi – 682 022, India

and

P.MOHANAN
Centre for Research in Electromagnetica and Antennas, Department of Electronics
Cochin University of Science and Technology, Kochi – 682 022, India

ABSTRACT

The time domain modeling of a coplanar wave guide (CPW) fed band-notched antenna for UWB applications is presented. The annular ring antenna has a dimension of 36x36 mm^2 when printed on a substrate of dielectric constant 4.4 and thickness 1.6 mm. The uniplanar nature and compact structure of the antenna make it apt for modular design. The crescent shaped slot provides a notch in the 5.2-5.8 GHz frequency band to avoid interference with WLAN. The pulse distortion is insignificant in the operating band and is verified by the measured antenna performance with high signal fidelity and virtually steady group delay.

Keywords: Ultra wideband, UWB antenna, monopole antenna, and wireless communications.

1. INTRODUCTION

High data rate and excellent immunity to multi-path interference make Ultra-wideband (UWB) technology one of the most promising solutions for future short-range high-data wireless communication applications. The allocation of the frequency band from 3.1 to 10.6 GHz by FCC [1] with a –10 dB bandwidth greater than 500 MHz and a maximum equivalent isotropic radiated power spectral density of – 41.3 dBm/MHz for UWB radio applications presents an exciting opportunity to antenna designers. UWB reaps benefits of broad spectrum in terms of the bit rates it can handle. By Shannon's theorem, the channel capacity C is given by,

$$C = W.\log_2\left(1+\frac{S}{N}\right) \qquad (1)$$

where W is the bandwidth and S/N is the signal to noise ratio. It can be seen that the bit rate (capacity) can be easily increased by increasing the bandwidth instead of the power, given the linear – versus- logarithmic relationship. Range of operation of such systems are determined by the Friis formula,

$$d \propto \sqrt{\frac{Pt}{Pr}} \qquad (2)$$

d being the distance, Pt the transmit power and Pr the receive power. Equations (1) and (2) together suggest that it is more efficient to achieve higher capacity by increasing bandwidth instead of power, while it is equally difficult to achieve a longer range. Thus, UWB primarily is a high-bit, short-range system.

UWB technology is a derivative of the time hopping spread spectrum (THSS) technique, a multiple access technology particularly suited for the transmission of extremely narrow pulses. It has been standardized in IEEE 802.15.3a as a technology for Wireless Personal Area Networks (WPANs). The challenges in UWB antenna design are bandwidth enhancement, size miniaturisation, gain and radiation pattern optimization.

Monopole antennas are used in communication systems at a wide range of frequencies. Electrical properties of these antennas are dependent upon the geometry of both the monopole element and the ground plane. The monopole element is either electrically short with length much less than a quarter-wavelength or near-resonant with length approximately a quarter-wavelength. This element can be thin with length-to-radius ratio much greater than 10^4 or thick with length-to-radius ratio of 10^1 -10^4. In addition, the ground-plane dimensions may vary from a fraction of a wavelength to many wavelengths. Traditionally, a monopole geometry consists of a vertical cylindrical element at the center of a perfectly conducting, infinitely thin, circular ground plane in free space. Electrical characteristics of such antennas are primarily a function of only three parameters; the element length, element radius, and the ground-plane radius, when each is normalized to t

he excitation wavelength. Radiation pattern of such antennas are uniform in the azimuthal direction. UWB monopole antennas fall in to volumetric and non-volumetric categories based on their structures. Non-volumetric UWB antennas are microstrip planar structures evolved from the volumetric structures, with different matching techniques to improve the bandwidth ratio without loss of the radiation pattern properties. A number of traditional broadband antennas, such as self-complementary spiral antenna, bi-conical antenna, log-periodic Yagi-Uda antenna [2], etc., were developed for UWB radio systems in the past. However, most of these antennas may be too bulky to be applicable in compact UWB communication equipments, such as handsets, PC cards, personal digital assistants (PDAs) and so on. In order to reduce system complexity and cost, it is necessary to develop miniature, light weight, low cost UWB antennas. Many efforts have been made to design such antennas. The fundamental design practice to realize ultra wide bandwidth is to match multiple resonances by suitable techniques [3-8]. Antenna design for UWB systems calls for special care, for if the surface currents on different parts of the antenna undergo significant time delays before summed up at the antenna terminal or transmitted as a free wave, signal dispersion may result [9].

The UWB printed monopole antenna consists of a monopole patch and a ground plane, both printed on the same or opposite side of a substrate, while a microstrip line or CPW is located in the middle of the ground plane to feed the monopole patch. Compared with the ultra-wideband metal-plate monopole antenna, the UWB printed monopole antenna does not need a perpendicular ground plane. Therefore, it is of smaller volume and is suitable for integrating with monolithic microwave integrated circuits (MMIC). To broaden the bandwidth of this kind of antennas, a number of monopole shapes have been developed, such as heart-shape, U-shape, circular-shape and elliptical-shape etc. The UWB printed monopoles are more suitable for smaller portable devices where volume constraint is a significant factor.

Due to the co-allocation of the UWB frequency band with frequency bands reserved for narrowband wireless technologies, there is a need to provide filtering in those bands to avoid interference from or causing interference to narrowband devices. So the use of a band stop filter becomes necessary. Several antennas have been reported in literature aiming at size reduction, bandwidth enhancement and WLAN interference avoidance [10-14]. Planar monopole and dipole antennas show good promise for use in UWB systems. Coplanar waveguide (CPW) fed antennas have the advantage of a balanced structure, since the feed lines and the radiating structure are on the same side of the substrate. [15-16].

CPW fed slot antennas are also very good candidates for UWB applications. The antennas discussed in [17] use a large slot for bandwidth enhancement and L or T shapes for size reduction. A CPW fed tapered ring slot antenna which can achieve a relatively large bandwidth is introduced in [18]. The wide band slot antenna [19] uses a large aperture and a modified microstrip feed to create multiple resonances. In another technique, a rotated slot is proposed [20] wherein two modes of close resonances are excited by a microstrip feed line. A tapered slot feeding structure is used to transform the guided waves to free space waves in [21]. In [22], a microstrip fed triangular slot antenna with a double T shaped tuning stub is introduced. The double T shaped stub is fully positioned within the slot region on the opposite side of the triangular slot. But the antenna has large dimension of 55x65mm2 with limited bandwidth of 3.3GHz.

The uniplanar nature and compact structure of the CPW fed annular ring antenna presented in this paper make it apt for modular design. The crescent shaped slot inserted into the UWB antenna aims at rejecting the 5.15-5.825GHz band limited by IEEE 802.11a and HiperLAN/2.

2. ANTENNA GEOMETRY

The structure comprises of a slotted annular ring shaped monopole antenna fed by a 50Ω CPW with a serrated ground plane as shown in Fig.1. The antenna is printed on a substrate of $\varepsilon r = 4.4$, loss tangent $\tan\delta = 0.02$ and thickness h=1.6 mm.

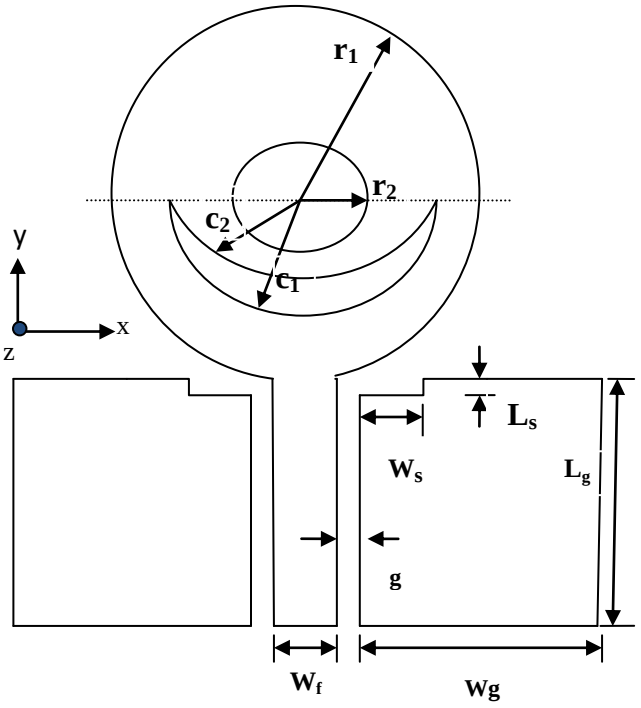


Fig.1 Antenna Geometry (all dimensions in mm)
Lg=15, Wg=16, g=0.35,Wf=3, Ls=1.2,Ws= 3
r1=11, r2=2.3, c1=6.5, c2=6.2

The strip width (W_f) and gap (g) of the Coplanar Waveguide (CPW) feed are derived using standard design equations for 50Ω input impedance [23]. The dimensions are optimized for ultra wideband performance after exhaustive simulation using Ansoft HFSS V.12.The accuracy of the antenna dimension is very critical at microwave frequencies. Therefore photolithographic technique is used to fabricate the antenna geometry. Photolithography is the process of transferring geometrical shapes from a photo-mask to a surface.

3. FREQUENCY DOMAIN RESULTS

Fig.2. illustrates the reflection characteristics of the antenna, measured using HP 8510C Vector Network analyzer. The antenna exhibits 2:1 VSWR bandwidth from 2.9 GHz to 17.4 GHz, with a notch in the 4.8 GHz – 5.8 GHz band. The antenna is developed from a conventional CPW fed disc antenna of radius r_1. The inner disc of radius r_2 inserted into the disc results in an annular ring antenna, shifting the lower edge of the resonant band from 3.26 GHz to 3.09 GHz, thus catering to the UWB requirement from 3.1 to 10.6 GHz. The crescent shaped slot of dimensions c_1,c_2 introduces a notch in the reflection characteristics. The serrations in the ground plane are responsible for fine tuning and precise positioning of the notch.

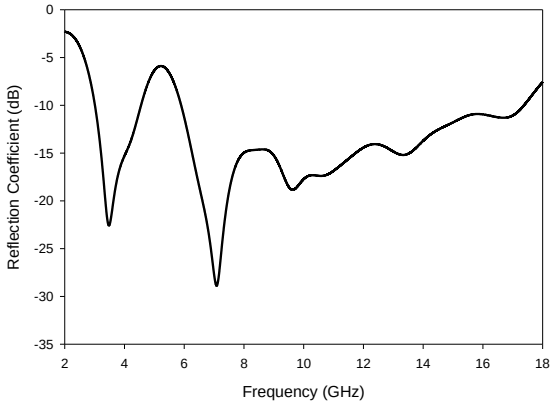


Fig.2.Reflection characteristics of the antenna

The current distribution in the antenna at different resonant frequencies in the operating band is illustrated in Fig.3. The bi-directional currents in the crescent shaped slot at 5.6 GHz [Fig.3(c)] account for the notch in the reflection characteristics. Typical measured radiation patterns of the antenna at 3.5 GHz and 7.1 GHz are shown in Fig.4. The antenna is linearly polarized along Y direction with good cross polar isolation in the entire band of operation. The antenna exhibits an average gain of 0.9dBi in the operating band. These characteristics confirm the suitability of the antenna for UWB operations.

4. TIME DOMAIN RESULTS

Good frequency domain performance does not necessarily ensure satisfactory time domain behavior. Linear phase delay or constant group delay is a mandatory requirement for an UWB antenna. A flat group delay is required so that the high and low-frequency signal components arrive at the receiver simultaneously. To study the time domain behaviour, two identical prototypes of the antenna are used as a transmitter – receiver system [24]. As shown in Fig.5, the measured group delay remains almost constant with variation less than 2 nanosecond for the face to face orientation. Similar results are obtained for the side by side and back-to-back orientations. This indicates a good time domain performance of the antenna throughout the operating band, barring the notch band.

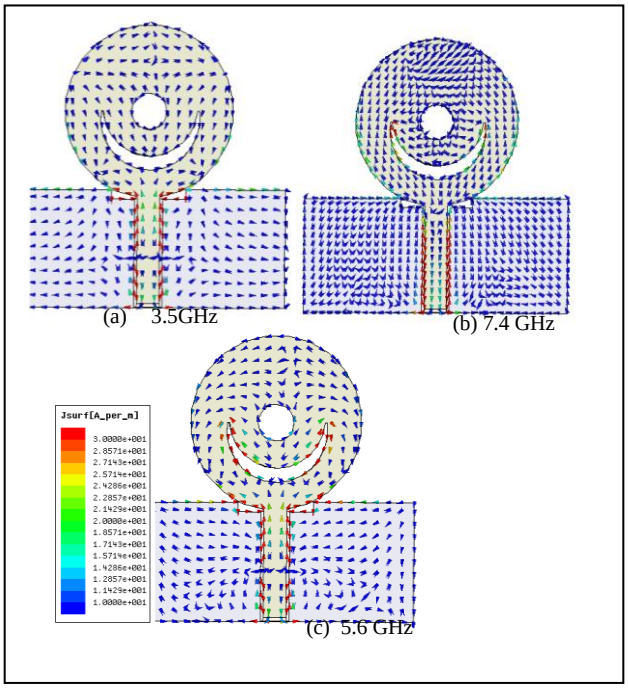


Fig.3. Current distribution at various frequencies in the operating band of the antenna

Transient response of the antenna is studied by modeling the antenna by its transfer function. For this, the transmission coefficient S_{21} is measured using HP8510C Network analyzer in the frequency domain for the face-to-face and side-by-side orientations placing the antennas at a distance R=10cm. From the S_{21} values of the UWB antenna system thus measured, the transfer function of the system is computed as follows.

$$H(\omega) = \sqrt{\left(\frac{2\pi R c \cdot S_{21}(\omega) \cdot e^{\frac{j\omega R}{c}}}{j\omega} \right)} \quad (3)$$

Where c is the free space velocity and R is the distance between the two antennas. This transfer function is multiplied with the spectrum of the input signal, which is chosen as a fourth order Rayleigh Pulse given by

$$S_i(t) = \frac{(16x^4 - 48x^2 + 12)e^{-x^2}}{\sigma^4} \; ; \quad where \;\; x = \frac{t-1}{\sigma} \;\; , \sigma \; is \; the \; pulse \; width \quad (4)$$

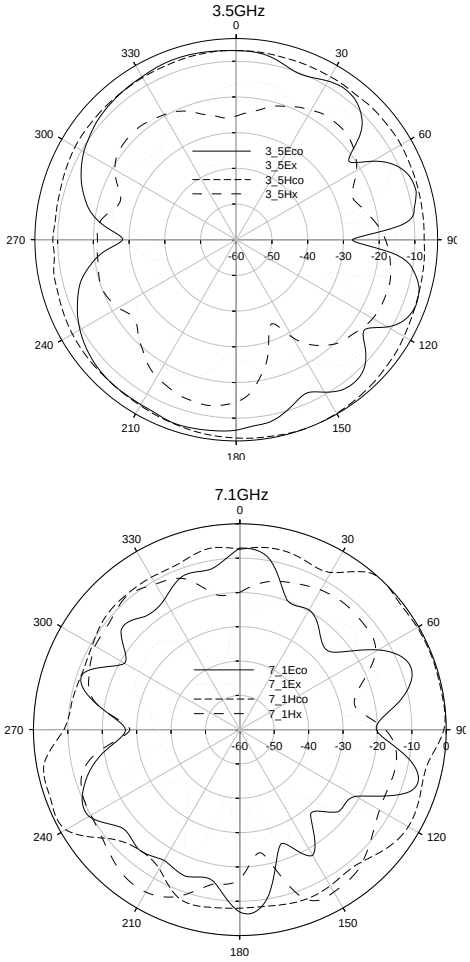


Fig.4 Measured Radiation Pattern

The inverse FFT of the product of H(ω) and the spectrum of the input signal gives the waveform at the receiver. The transmitted and received wave forms for the face-to-face and side-by-side orientations of the antenna are shown in Fig.6. It is evident that the received pulses are almost identical.

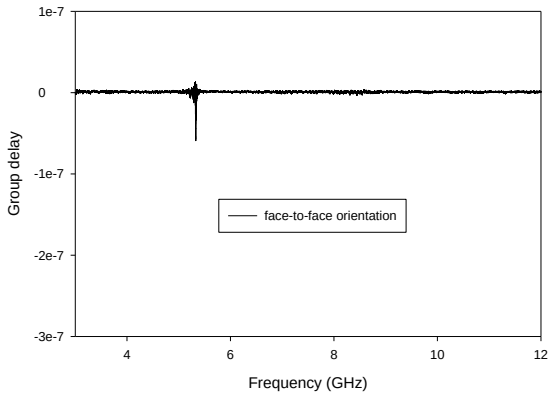


Fig.5. Measured group delay of the antenna (nsec)

In UWB systems it is very important to characterize the transient behavior of the radio propagation channel, specifically for impulse radio systems. Pulse fidelity involves the autocorrelation of two different time domain waveforms and compares the shape of the pulses disregarding the amplitude and the time delay. A low fidelity between transmitted and received pulse means that the dis- tortion of the received pulses is high and hence loss of system information is high [25]. The fidelity factor between transmitted and received signals in Tx/Rx setups between two identical antennas in different orientations are calculated for the fourth order Rayleigh pulse [Fig.7].

$$F(\theta,\varphi) = \underbrace{max}_{\tau} \left[\frac{\int_{-\alpha}^{\alpha} S_t(t) S_r(t+\tau,\theta,\varphi) dt}{\sqrt{\int_{-\alpha}^{\alpha} {S_t}^2(t) dt \int_{-\alpha}^{\alpha} {S_r}^2(t,\theta,\varphi) dt}} \right] \quad (5)$$

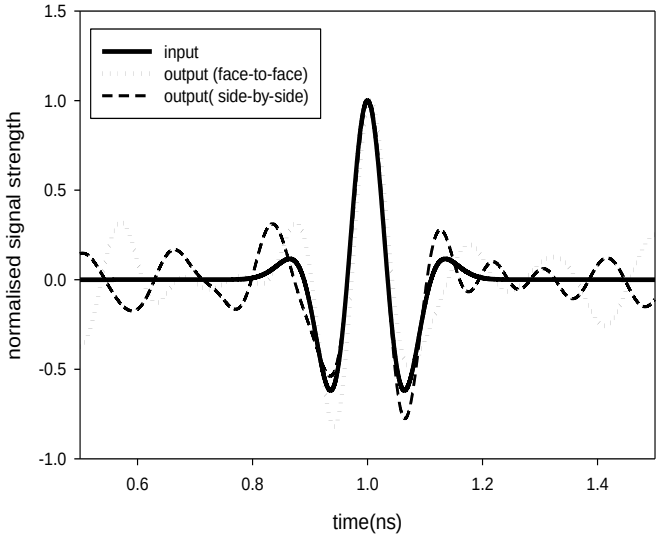


Fig.6 Transmitted and Received Pulse for different orientations of the antenna

It is clear from the figure that fidelity factor is greater than 0.9 for τ=50ps, where τ is the pulse width fidelity factor. These values for the fidelity factor show that the antenna imposes negligible effects on the transmitted pulses.

According to FCC regulations, UWB systems must comply with stringent EIRP limits in the frequency band of operation. EIRP is the amount of power that would have to be emitted by an isotropic antenna to produce the peak power density of the antenna under test. To obtain EIRP, we use similar transmit and receive antennas and total frequency response of the system H(ω) is calculated as

$$EIRP = S_i(f)\sqrt{H(f)} \cdot \frac{4\pi r f}{c} \quad (6)$$

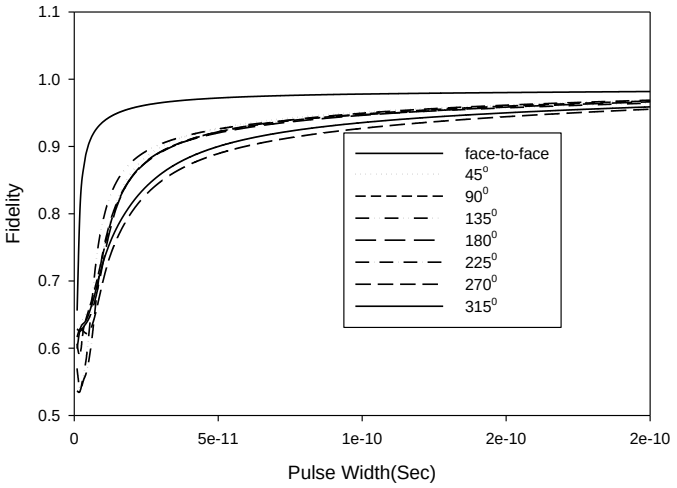


Fig.7 Fidelity of the antenna in different orientations

Fig.8 shows the measured EIRP emission level of the antenna excited with a fourth order Rayleigh pulse with pulse width factor τ =50ps. As it is clear from the figure, EIRP of the antenna satisfies the FCC masks for the entire UWB band.

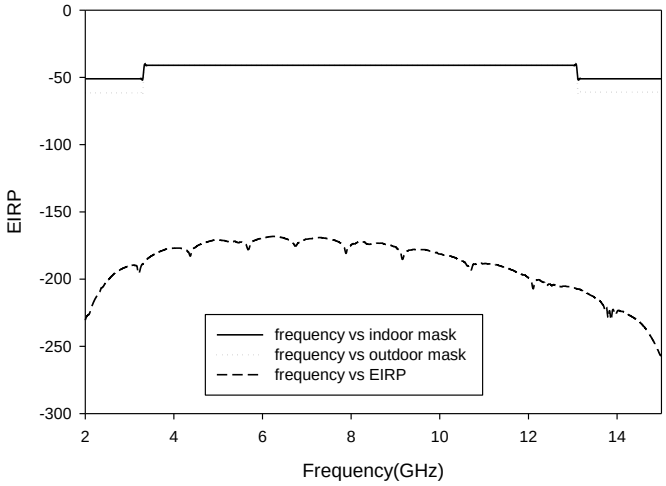


Fig.8 Measured EIRP of the antenna

5. CONCLUSIONS

The time domain modeling of a compact UWB wideband monopole antenna with band-rejection characteristics is presented. The prototype offers -10dB impedance band from 2.9 GHz to 17.4 GHz, with an overall size of 36mm x 36mm, catering to the UWB frequency requirement. Furthermore, the crescent shaped slot inserted into the radiator rejects the 5.2 - 5.8 GHz WLAN band. Broad impedance bandwidth, stable radiation patterns, reasonable gain and excellent time domain characteristics are the main attractions of this antenna.

6. ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support by the AICTE, Govt. of India under the scheme RPS(C) File.No: 8023/BOR/RID/RPS-12/2008-09 dt 30.10.2008. They are also thankful to C.M.Nijas, Research scholar, CREMA, Department of Electronics, CUSAT for the help rendered in fabrication and measurement.

7. REFERENCES

[1] Federal Communications Commissission, First report and order, revision of part 15 of Commission's rule regarding UWB transmission system **FCC02-48**, April 2002.

[2] Kraus, J.D.: **'Antennas'** McGraw-Hill, 2nd edn., Ch. 15, 1988.

[3] S. I. Latif, L. Shafai, and S. K. Sharma, "Bandwidth enhancement and size reduction of microstrip slot antennas," **IEEE Trans. Antennas Propag.,** Vol. 53, No. 3, Mar. 2005, pp. 994–1003.

[4] N. Behdad and K. Sarabandi, "A multiresonant single-element widebandslot antenna," **IEEE Trans. Antennas Propag.,** Vol. 3, No. 1, Jan. 2004, pp.5–8.

[5] J. Y. Jan and J. W. Su, "Bandwidth enhancement of a printed wide-slot antenna with a rotated slot," **IEEE Trans. Antennas Propag.,** Vol. 53, No. 6, Jun. 2005, pp. 2111–2114.

[6] T. G. Ma and C. H. Tseng, "An ultra wideband coplanar waveguide-fed tapered ring slot antenna," **IEEE Trans. Antennas Propag.,** Vol. 54, No.4, Apr. 2006, pp. 1105–1111.

[7] T. G. Ma and S. K. Jeng, "Planar miniature tapered-slot-fed annular slot antennas for ultrawide-band radios," **IEEE Trans. Antennas Propag.,** Vol. 53, No. 3, Mar. 2005, pp. 1194–1202.

[8] E. S. Angelopoulos, A. Z. Anastopoulos, D. I. Kaklamani, A. A. Alexandridis, F. Lazarakis, and K. Dangakis, "Circular and elliptical CPW-fed slot and microstrip-fed antennas for ultrawideband applications," **IEEE Antennas Wireless Propag. Lett.,** Vol. 5, No. 3, Jun. 2006, pp.294–297.

[9] K. Siwiak and D. McKeown, **Ultra-Wideband Radio Technology**.New York: Wiley, 2005, pp. 97–111.

[10] J. Liang, C. C. Chiau and C. G. Parini, "Study of Printed Circular Monopole Antenna for UWB Systems," **IEEE Trans. Antennas Propag.,** Vol. 53, No. 11, November 2005, pp. 3500-3504.

[11] Pengcheng Li, Jianxin Liang and Xiadong Chen, "Study of printedelliptical/circular slot antennas for

ultrawideband applications antenna **IEEE Trans. Antennas Propag.,** Vol. 54, No. 6, June 2006, pp. 1670-1675.

[12] Q. Wu, R. Jin, J. Geng, and J. Lao, "Ultra-wideband rectangular disk monopole antenna with notched ground," **Electron. Lett.,** Vol. 43, No. 11, May 2007pp. 1100–1101.

[13] Wen-Shan Chan,and Kuang-Yuan Ku,"Bandwidth enhancement of open slot antenna for UWB applications," **Microwave and Optical Technology Letters,** Vol. 50, No. 2, February 2008, pp.438-439.

[14] M. Ojaroudi, C. Ghobadi, and J. Nourinia, "Small square monopole antenna with inverted T-shaped notch in the ground plane for UWB application," **IEEE Antennas Wireless Propag. Lett.,** Vol. 8, Jul. 2009, pp. 728–731.

[15] J. Liang, L. Gu, C.C. Chiau, X. Chen and C.G. Parini, "Study of CPW-fed circular disc monopole antenna for ultra wideband applications," **IEE Proc.-Microw. Antennas Propag.,** Vol. 152, No. 6, December 2005, pp.520-526.

[16] Xinan Qu,1 Shun-Shi Zhong,1 and Wei Wang, "Study of the band-notch function for a UWB circular disc monopole antenna," **Microwave and Optical Technology Letters,** Vol. 48, No. 8, August 2006, pp.1667-1670.

[17] S. I. Latif, L. Shafai, and S. K. Sharma, Bandwidth enhancement and size reduction of microstrip slot antennas, **IEEE Trans.Antennas Propag.,** Vol**.** 53, 2005, pp. 994-1003.

[18] T.G.Ma and C.H. Tseng, An ultra wide band coplanar waveguide-fed tapered ring slot antenna, **IEEE Trans. Antennas Propag., Vol.** 54, 2006, pp. 1105-1111.

[19] N.Behdad and K.Sarabandi, A multiresonant single element wide-band slot antenna, **IEEE Trans.Antennas Propag.,** Vol. 53, 2005, pp. 994-1003.

[20] J.Y.Jan and J.W.Su,Band width enhancement of a printed wide slot antenna with a rotated slot, **IEEE Trans.Antennas Propag.,** Vol. 53, 2005, pp. 2111-2114.

[21] T.G.Ma and S.K.Jeng, Planar miniature tapered slot fed annular slot antennas for ultra wide band radios, **IEEE Trans.Antennas Propag.,** Vol. 53, 2005, 1194-1202.

[22] JoongHan Yoon, Triangular slot antenna with a double T shaped tuning stub for wide band operation, **Microwave and Optical Technology letters.,** Vol. 49, 2007, pp. 2123-2128.

[23] R.Garg,P.Bhartia,I.Bahl and A.Ittipiboon, **Microstrip Antenna Design Handbook**.Norwood,MA:Artech House, 2001.

[24] Y.Duroc,A.Ghiotto,T.P.Vuong and S.Tedjini, UWB Antennas: Systems With Transfer Function and Impulse Response, **IEEE Trans.Antennas Propag.,** Vol. 55, 2007, pp. 1449-1451.

[25] A. Mehdipour, K. Mohammadpour-Aghdam and R. Faraji- Dana, "Complete dispersion analysis of vivaldi antenna for ultra wideband applications" **Progress In Electromagnetic Research,** PIER 77, 2007.

Efficiency of Electric Power Utilities Using Data Envelopment Analysis: An Application to Practical Comprehension

Katsumi Nishimori, Kazuki Sakuragi
Tottori University, Japan
nisimori@ele.tottori-u.ac.jp

Abstract

We have investigated an efficiency of electric power utilities of the United States and Japanese electric companies using data envelopment analysis. The analysis can be newly constructed with an effective graphical expression of the efficiency in a diagram of two-input one-output models. We show that it is very useful to graphically illustrate the assessment of efficiencies due to the analysis based on time series data with several examples and discussion.

1. Introduction

Japan's earthquake gave Tsunami and nuclear crisis (11 March, 2011) as a nuclear power plant accident caused by the earthquake. In promoting energy conservation and efficient use of energy worldwide, in particular, an energy strategy has become very important for electric power management. As an assessment using the measure model for efficient activity of the industry, data envelopment analysis (DEA) is widely used due to a nonparametric data treatment [1]. Input and output data obtained from the activities are evaluated with the DEA efficiency that depends on each industry's activity. The analysis explicitly gives us the improvement points in the activity data. The DEA has been developed from CCR (Charnes, Cooper and Rhodes) analysis to BCC (Banker, Charnes, and Cooper) analysis [1]. Using them to measure the performance of decision making units (DMU) of the power industries, more precise assessments are carried out to the efficient use of energy in the electric power management.

The electric power companies in Japan are vertically integrated as structures of electricity business unlike those in Europe and the United States (US). Each company of Japan makes a monopoly-type business in the respective regions and also has several functions, such as generation, transmission, distribution and sales. Overall efficiency of the electric power managements has localized effects of the regions in comparison with Japanese power companies.

In this report, the US data is employed for the entire US power industry of EIA [2] and Japanese data of power industry as a whole entity are the data of averaged 9 Japanese electric power companies [3] without Okinawa power company.

We have compared the efficiencies of electric power utilities of US and Japan power industries by DEA method. The efficient use of the power equipments is investigated by the results of CCR and BCC analyses with time series data from 1998 to 2009.

DEA technique measures a relative efficiency between business entities based on the data provided as a lot of plural input data elements and plural output products [4]. However, it is difficult to individually pick up the effect from many inputs and outputs. Therefore, 2-input 1-output data are adopted to maximize the efficiency for simplicity. The efficiency has to be evaluated by choosing the input and output data heuristically. When the optimal DEA assessment can be done, we can make an improvement clear by the difference between efficiency and inefficiency results in CCR and BCC analyses.

Here, we describe a brief outline of model (CCR and BCC analyses) of DEA and compare the energy-use efficiency of the electricity business industry using each DEA analysis by 2-input 1-output type expression with time series data. Especially we show that it is very useful to graphically illustrate the assessment of its efficiency due to our analysis with the time series data by several examples and discussion.

2. DEA method

We describe the DEA method briefly. There are two methods of CCR and BCC [1]. At first we can apply the theory of minimization of the LP (linear programming) method to the theory of CCR maximization about the mathematical procedure of the

CCR analysis of DEA method using the dual transformation [1]. That is;

$$\begin{aligned} &[\text{LP minimization}] \qquad \min \quad \theta \\ &\text{s.t.} \qquad \theta\, \mathbf{x}_0 - X\lambda \geqq \mathbf{0} \\ &\qquad\qquad \mathbf{y}_0 - Y\lambda \leqq \mathbf{0} \\ &\qquad\qquad\qquad \lambda \geqq \mathbf{0} \end{aligned} \qquad \cdots\cdots (1) .$$

Here, x_0 is each DMU's input data and y_0 is output data. θ is the objective function in the LP minimization operation of LP method. X and Y are virtual input and virtual output vectors, respectively. λ represents the weight vector of nonnegative values.

BCC model is developed to extend the CCR model to variable returns to scale. The eq. (1) of CCR is rewritten on the weight-vector λ of the LP algorithm, by adding the constraints $\mathbf{e}^T\lambda = \Sigma\lambda_j = 1$ as follows;

$$\begin{aligned} &\min \quad \theta \\ &\text{s.t.} \quad \theta\, \mathbf{x}_0 - X\lambda \geqq 0 \\ &\qquad \mathbf{y}_0 - Y\lambda \leqq 0 \\ &\qquad \mathbf{e}^T\lambda = 1, \quad \lambda \geqq 0 \end{aligned} \qquad \cdots (2)$$

The application examples are shown in the next section using these analyses.

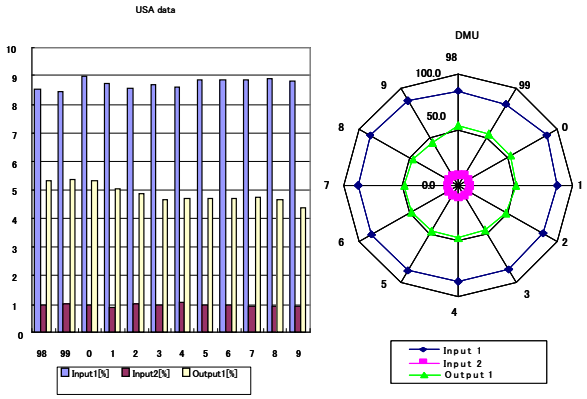
3. Results of electric power industries

Here, DEA analysis is applied to measure efficient utility that shows inefficient and efficient DMU comparison of US and Japanese electricity industries. We describe the results using graphic illustration to the assessment of its efficiency due to the analysis with the time series data.

3.1 The US electric power industry

DEA analysis of the electric power industry in the US is carried out by the previous reports [5, 6]. We show the recent analysis by the use of time-series data for the year 1998 up to 2009. Each DMU data is collected from the time series data of the annual indices [2]. The DMU data is shown in Fig. 1. The graph (a) is a bar graph which is displayed every year. The graph (b) is a radar chart of the graph (a) to compare each other in the later.

Input 1 is the ratio of operating expenses to total sales fee (%; O.E.: expense) and the input 2 is the electric energy loss (%; E.L.: energy loss) which is similar to the definition of Vaninsky in the reference [5]. The output is capacity utilization factor (%; C.U.: capacity utilization) for the efficiency of the US electric power industry [5].



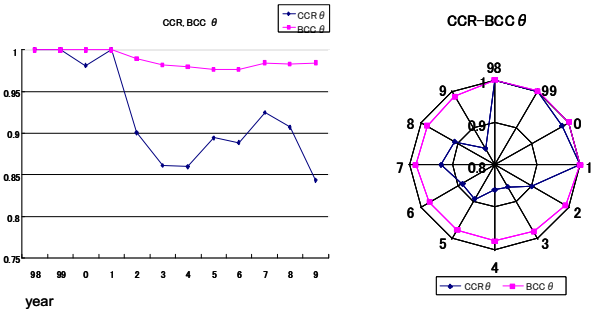
(a) (b)

Fig. 1. (a) Actual DMU data from 1998 to 2009 of whole electric power industry in the United States. (b) Radar charts of (a).

The actual data are indicated for 12 years from 1998 to 2009. The energy loss every year seems small from the changes of actual data, but the loss power is a huge amount just across whole the electric power business.

The electric energy we use is produced from the other primary energies of mechanical, chemical, thermal and nuclear energy etc. The electricity energy is very convenient for the use, but it gives rise to the energy loss due to the energy conversion, long distance transmission and distribution to industries and societies.

The efficiency consideration for electric energy use is very important in energy management. Using the above data, CCR and BCC efficiency analyses are carried out. Fig. 2 shows the results. The BCC results from 1998 to 2004 are coincident with those of Vaninsky [5].



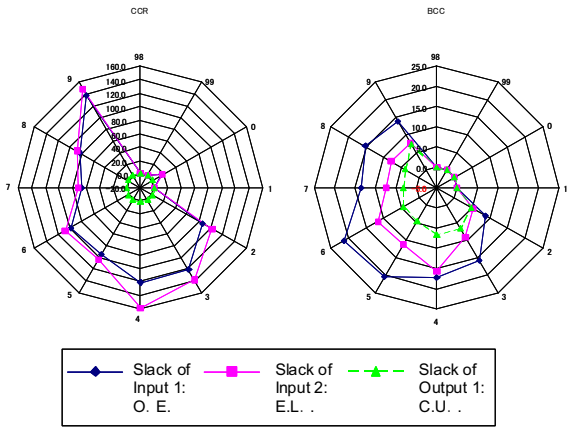
(a) (b)

Fig.2. (a) is the results of CCR and BCC efficiency analyses and (b) shows the radar chart of result (a) for whole electric US industry.

The radar chart of Fig.2 (b) expresses the results in a compact area. We can immediately read the difference between CCR and BCC efficiency plots at the years from 1998 to 2009, because the time sequence is according to clock wise direction in a radar chart.

The efficiency differences can be recognized as the distortion of a polygonal shape from a circle for a period from 2001 to 2009. The point of "1" shows the year 2001 for example. Also those of other years are similar to this expression. The radar chart emphasizes that the BCC efficiency is greater than the CCR efficiency by using the distortion of the shapes.

The differences are investigated in detail using the slack analysis [7] with respect to the actual inputs and output. Fig.3 shows the differences (i.e. slacks) between the virtual DMU (θ=1.0) and the actual DMU results of BCC and CCR, respectively. When the difference is zero, the DMU becomes efficient. The efficient DMU points are concentrated to the center of a circle.



(a) (b)

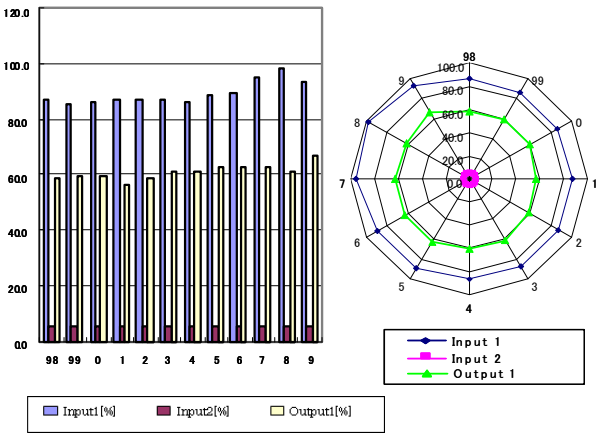
Fig.3. Radar charts of (a) CCR and (b) BCC inefficiency (slacks) analyses of US electric industries.

In the CCR result of Fig. 3(a) the slack points for both inputs 1 (expense) and 2 (energy loss) make the polygonal shapes, which inflate at the lower left for a period from 2002 to 2009. The respective slack points show the inefficiencies. Contrary to this shape, the CCR inefficiency curve of the above mentioned Fig. 2(b) shrinks to the circle center for the period. US has a lot of continuing economic crises corresponding to the period, such as the terrorist attacks in 2001, bankruptcy of Enron, the attack on Iraq in 2003, a steep rise of oil prices from 2004 and the collapse of Lehman Brothers in 2008. The slacks of BCC analysis of Fig. 3(b) also indicate the similar result weakly. While they are numerically small. The result of the US is called as the Type 1.

3.2 Averaged Japanese power industry

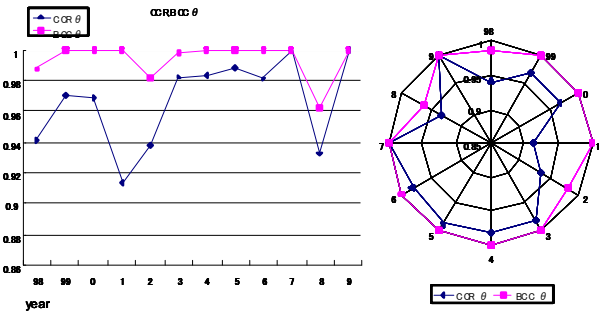
Next, our analysis is applied to Japanese electric power industries to compare with the above US industries.

Japanese data of power industry as a whole entity are obtained from averaged 9 Japanese electric power companies [3] without Okinawa Power Company. The actual time series DMU data are shown in Fig. 4.



(a) (b)

Fig. 4. (a) Time series DMU data averaged for 9 electric companies in Japan.
(b) Radar charts of time series DMU data of (a).

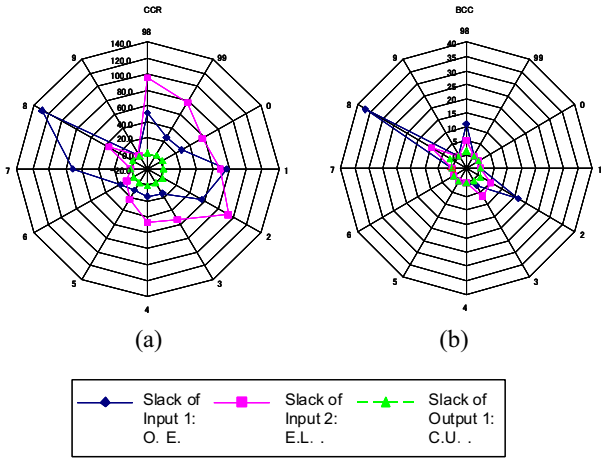


(a) (b)

Fig. 5. (a) CCR and BCC efficiency results. (b) Radar charts of averaged 9 electric companies of Japan.

There is a big difference of the values between the input 1 (expense) and 2 (energy loss) data as shown in Fig. 4. In addition the annual change of those values is small, however the DEA calculation was carried out without hindrance. The CCR and BCC efficiencies are shown in Fig. 5(a) and the radar charts (b), respectively.

In contrast with above US results, the rapid recover appears in the CCR efficiencies of averaged Japan power industry, even after 2001 terrorist attacks or the collapse of Lehman Brothers (LB) in 2008 worldwide.



(a) (b)

Fig.6. Radar charts of (a) CCR and (b) BCC inefficiency (slacks) analyses of averaged Japanese 9 electric companies.

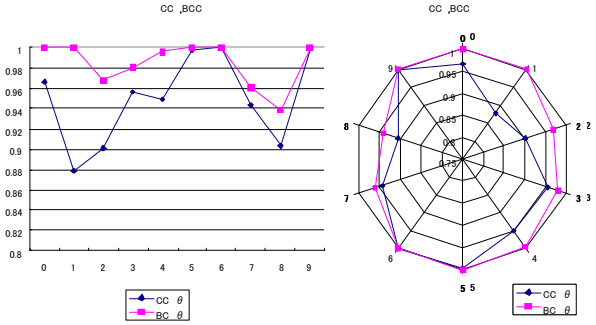
Figure 6 shows the annual results using CCR and BCC analyses for the 9 averaged Japan industries. Fig. 6(a) and (b) show the radar charts of the CCR and BCC slacks, respectively. In the CCR slacks results of Fig. 6(a), the curve of input 2 (energy loss) is clearly separated from that of input 1 (expense). This implies that the operating expense has a trade-off relation with the energy loss in Japan.

In addition, the influence of Lehman shock of 2008 is remarkably seen for the BCC slack peak than the CCR slack peak. While the effective use of electricity is indicated in other years. The result of the averaged Japan is called as the Type 2.

3.3 Tokyo electric power company.

In order to compare with the US and averaged Japan power industries, further analyses are investigated for several major power companies in Japan,.

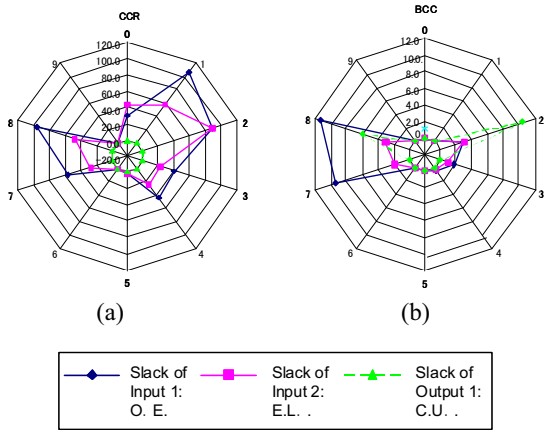
First, we can see the case of Tokyo electric power company (TEPCO). Figure 7 shows the results of the annual CCR and BCC efficiencies in (a) and the radar chart of (a) is shown in (b). The results are similar to those of the average Japan industry.



(a) (b)

Fig. 7. The line graph (a) is the CCR and BCC results of TEPCO. The (b) shows the radar charts of (a).

Figure 8 indicates the radar charts of (a) CCR and (b) BCC slack analyses of TEPCO. In Fig. 8(a), the CCR slacks for the energy loss (input 2) have similar deviations to those for the operating expense (input 1) in the period from 2002 to 2004. The former however is larger than the latter in the CCR result of Fig. 6(a) of averaged Japan industry. This means that the electric transmission and distribution network of TEPCO is well run over Kanto region than that of whole Japan. TEPCO suppresses the energy losses owing to the well developed power grid. In the BCC result of Fig. 8(b) the scale of coordinates is multiplied by 10. Therefore the effects of the slacks are less than the CCR result of Fig. 8(a) by 10 times. The effect due



(a) (b)

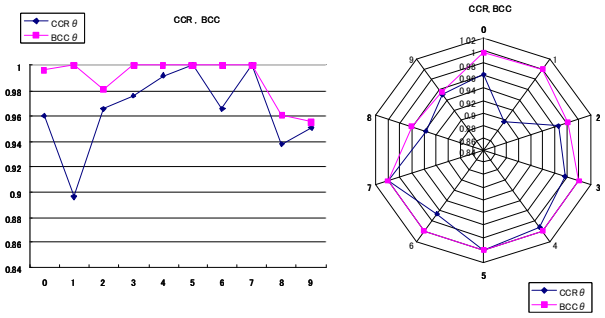
Fig.8. Radar charts of (a) CCR and (b) BCC un-efficiency (slacks) analyses of TEPCO.

to 2001 terrorist attacks or the collapse of Lehman Brothers (LB) in 2008 can however be seen remarkably. The result of TEPCO is called as the Type 3. Similar results are obtained for the Kansai electric power Co. (KEPCO) and Chubu electric power Co. (CEPCO), which are the major companies of Japan power industry. TEPCO, KEPCO and CEPCO supply 1/3, 1/6 and 1/6 of the whole Japanese electric power generation every year, respectively.

As we have seen, the above results of the three power companies make the most shape of that of the averaged 9 Japanese companies. However the slacks of energy losses are more prominent in the latter than the former. To see the difference in detail, let us investigate other local electric power companies of Japan in the following.

3.4 Kyushu electric power company.

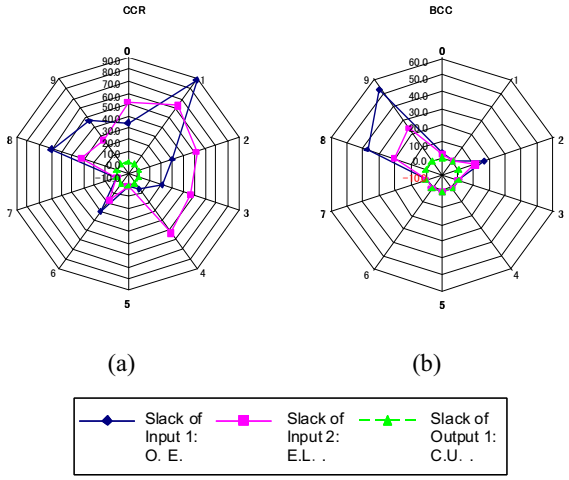
The CCR and BCC efficiency results of Kyushu electric power company (Kyuden) resembles those of TEPCO, as shown in Fig. 9. However the recover at 2009 of LB shock in 2008 is weaker than that of TEPCO.



(a) (b)

Fig. 9. (a) CCR and BCC efficiency results, and (b) the radar charts of Kyuden.

Figure 10 shows the slack results by (a) CCR and (b) BCC analyses of Kyuden. In the CCR result we can find that the slacks (inefficiencies) of input 2 (E.L.: energy losses) are beyond those of input 1 (O.E.: operating expense ratio) at the years from 2002 to 2004. We have ever seen the same feature in Fig. 6(a) of the averaged Japanese industry. This means that Kyuden has one of the significant influences to the CCR inefficiency result of the averaged Japanese power industry. However the Japanese major 3 power companies do not show it clearly. Since Kyuden provides the electric power over the complex topographical area with a lot of islands, it seems that the energy losses considerably increase. The result of Kyuden is called as the Type 4.

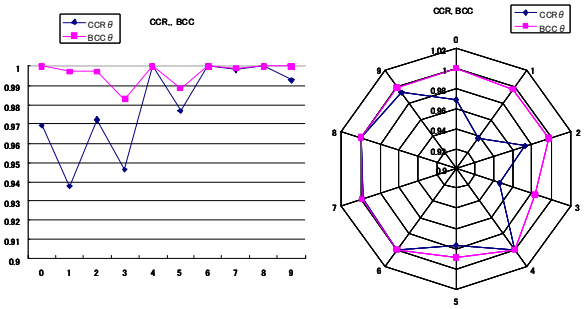


(a) (b)

Fig.10. Radar charts of (a) CCR and (b) BCC inefficiency (slacks) analyses of Kyuden.

3.5 Chugoku electric power company.

Finally, we describe a different case of Chugoku electric power company (Chuden) among Japanese power companies. The business area of Chuden is located in the west Japan about 800 km distant from Tokyo. Figure 11 shows the results of CCR and BCC efficiency analyses.



(a) (b)

Fig.11. (a) CCR and BCC results and (b) the radar charts of Chuden.

The influence of Lehman shock at 2008 disappears and both of CCR and BCC results indicate that the efficiencies are efficient (θ = 1.0) from 2006 to 2008 in the graphs. The closed curve of the CCR efficiency

is strongly biased to the left. This leads to the formation of polygons with bulges in the top right corner, as shown in CCR result of Fig. 12(a). The results of Fig. 12 are derived from the CCR and BCC slack analyses. Especially, as the BCC slack analysis has sharp peaks in Fig. 12(b), the efficiencies can be recovered in the short interval of year by year. This reflects the business behavior of Chuden as a monopoly electric supplier and producer over the localized small service area. The result of Chuden is called as the Type 5.

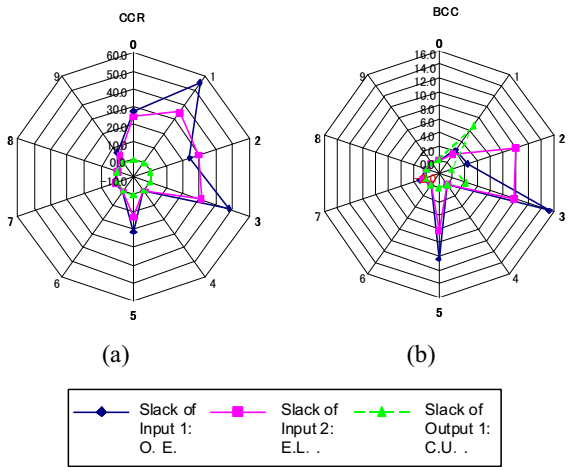


Fig.12. Radar charts of (a) CCR and (b) BCC inefficiency (slacks) of Chuden.

DEA analyses have been carried out about the power industries in Japan and the US during recent one decade. The respective analyses based on radar chart scheme are able to pick up the practical business features of power industries, which show the effective use of electric energy with the above examples. In measuring the DEA efficiency of power industries, we used the 2-input 1-output diagrams using time series data. The radar chart depends on how to choose the data. However, when we select suitable data sets, useful results can be easily obtained from the clear features of radar charts.

4. Conclusion

We show a new analysis and expression of DEA method to support improvement to practical efficiencies of the US and Japanese power industries. A comparison of the results can be summarized as follows: First, it is very useful to distinguish CCR and BCC efficiencies which can be visually recognized with feature shapes of radar charts. Second, the improvement points are directly indicated by DEA slack analysis on the radar charts with 2-input 1-output diagrams using time series data. Third, it is easy to classify the practical business behaviors of power industries with both of efficiency and slack analyses owing to the respective radar chart features. Therefore we can propose the practical analysis as a new powerful tool for DEA methods.

5. References

[1] A.Charnes et al.: “Measuring the efficiency of decision making units”, **European Journal of Operational Research**, Vol. 2, 1978, pp. 429-444.

[2] http://www.eia.doe.gov - a website of the Energy Information Administration of the US, a section of the US Department of Energy (DOE).

[3] Federation of Electric Power Companies: http:// www.fepc.or.jp/, http://www.energia.co.jp/index.html.

[4] W.D. Cook and L.M. Seiford: “Data envelopment analysis (DEA) – Thirty years on”, **European Journal of Operational Research**, Vol.192, 2009, pp.1-17.

[5] A. Vaninsky: ”Efficiency of electric power generation in the United States: Analysis and forecast based on data envelopment analysis”, **Energy Economics**, Vol. 28, 2006, pp. 326-338.

[6] K. Tone and M. Tsutsui: “Decomposition of cost efficiency and its application to Japanese-US electric utility comparisons”, **Scio-Economic Planning Sci.**, **14**, 2007, pp. 91-106.

[7] P.V. Geymueller: “Static versus dynamic DEA in electricity regulation: the case of US transmission system operators”, **Central European Journal of Operations Research**, Vol.17, 2009, No.4, pp. 397-413.

FUNCTIONAL MAPPING

Donald V. Poochigian
Department of Philosophy and Religion, University of North Dakota
Grand Forks, North Dakota 58202, United States

ABSTRACT

Space is either a continuous Newtonian surface, or a discontinuous Leibnizian field. A Newtonian surface is a dimensionally extended geometric point. A Leibnizian field is dimensionally unextended geometric points. Mapping traces point to point. There being one point in Newtonian space, point cannot be traced to point. There being disassociated points in Leibnizian space, point cannot be traced to point. Sketching is discontinuous sequencing of point to point. Mapping is continuous relation of point to point. Elements being discontinuous, identity of one with another is nominal. Elements being continuous, identity of one with another is real. A uni-dimensionally extended point being disassociated from other points in a Leibnizian field, there are points in mapping space. A uni-dimensionally extended point being associated with another point on a Newtonian surface, there is tracing in mapping space. Extension of a non-dimensional Leibnizian point into a uni-dimensional Newtonian line converts a Leibnizian point into a Newtonian surface. Tracing point to point, extension of a non-dimensional Leibnizian point into a uni-dimensional Newtonian line constitutes mapping.

FUNCTIONAL MAPPING

1. ENIGMA

Mapping is axiomatic, a tracing of theorem derivation from axioms. Within an axiom system, axioms identify constituents and constituent sequencing constituting mapping from axiom to theorem. Limited by the constituent and constituent sequencing constraints of an axiom system, mapping delineates tautological transformation of constant elements from axiom to theorem, and/or analogical transmutation of inconstant elements from axiom to theorem.

Any mapping identified within an axiomatic system constitutes truth within that axiomatic system. Any mapping unidentified within an axiomatic system constitutes falsity within that axiomatic system. Assigned truth and falsity of primitive constituents is axiomatic. Only the truth and falsity of mapped constituents is derivative.

The enigma of mapping. Now is an enigma. Mapping occurs within either Newtonian or Leibnizian space. Something is always between any two things within Newtonian space. Nothing is always between any two things within Leibnizian space. Newtonian space is an infinitely extended point. Leibnizian space is an infinity of unextended points. It is impossible to map a point to a point within Newtonian space because transited points are indiscernible. It is impossible to map a point to a point within Leibnizian space because transitional points are indiscernible.

Solution to the enigma of mapping. False is impossible in Newtonian space, and true is impossible in Leibnizian space. Resolution is a compromise, converting a Leibnizian point into a limited Newtonian space. Hereby, a Leibnizian point *a* is extended in one dimension to another Leibnizian point *b*, composing a line *ab*. False is wherever the line formed by extension of Leibnizian point to point is unextended.

Provided is a proof, Frege asserting,

> The method of proof is a syntactic method for proving validity, it depends only on the geometric shapes of the signs in the argument forms and the arguments that are substitution instances of them.[1]

However, "accepting the truth-value of a sentence as constituting what it means,"[2] when same "geometric shapes" can have different "truth-value," then, "How on earth there can be a definition where there is no question about connexions between sign and thing signified by it is a puzzle."[3]

"Connexions between sign and thing signified by it is a puzzle" when discontinuous. A discontinuum from *a* to *b* constitutes semiotic, and its identity constitutes intuition. There being more than one element immediately constituent to every limiting element, transition from *a* to *b* is unnecessary. Unnecessary, each limiting element presents a concealed contradiction.

"Connexions between sign and thing signified by it is [not] a puzzle" when continuous. A continuum from *a* to *b* constitutes logic, and its identity constitutes reason. There being no more than one element immediately constituent to every limiting element, transition from *a* to *b* is necessary. Necessary, each limiting element presents no concealed contradiction.

A continuous one dimensional extension, *ab* remains the same point. Here *a* demarcates one limit of the extended point, and *b* demarcates the other limit. Arbitrary, however, is whether *a* becomes *b*, or *b*

becomes *a*. Whichever or neither, Leibnizian space converts into Newtonian space by this means.

Proof being a path mapped from point *a* to point *b*, a difficulty arises. A path mapped from point *a* to point *b* requires, "a whole series of . . . other perspectives . . . such that . . . the space which consists in relations between perspectives [*a* and *b*] can be rendered continuous."[4] Provision of this "series of . . . other perspectives" requires segmentation of the space from point *a* to point *b*. Segments are identifiable only if distinguished by a limit. Indistinguishable without a limit, mutually unlimited segments are identifiable as a segment.

A limit is an unambiguous separation of segments. Constituents common to different segments, as occurs with Euler-Venn Diagrams, constitute an ambiguous separation of segments. Common to otherwise different segments, these are understandable as integrating the otherwise different segments into a single segment by their mutual relation to the uncommon constituents of the otherwise different segments. Now there is no topological "***system of neighborhoods***."[5]

A "***system of neighborhoods***" requires unambiguous separation of segments. Unambiguous, segments share no common member. Sharing no common member, there is no means of transiting from segment to segment. There being no means of transiting from segment to segment, there is no means of mapping *a* to *b*. There being no means of mapping *a* to *b*, there is no means of proof. There being no means of proof, there can be no proof.

Topological neighborhoods composed of unrelated points, "one's definition of 'logically rigorous' tends to boil down to 'it convinces me.'"[6] Proof becomes a subjective psychological, not objective logical, state of being. Contrastingly, objective logical proof constitutes either an injective transformation of *a* into *a*, or a surjective transmutation of *a* into *b*. Transformation reorders common elements; transmutation substitutes uncommon elements.

Whether reordering common elements, or substituting uncommon elements, inscribed is an unbroken continuum from *a* to *b*. Unbroken, *a* is extended in one dimension into *b*, each the limits of a continuous line. As the limits of a continuous line, *a* and *b* are aspects of the same thing, the point/line, not different things.

To what logic comes is different understanding of indifferent things, a proof constituting a particular sequence of conceptual alterations. Identity of *a* and *b* being inconstant over constant observation of *a* and *b*, identity is not observational, it is conceptual. Requisite for an explanation is distinguishing between abstraction and quality.

2. MAPPING

Mapping is transition from origin to conclusion within a space. Something being between any two things in a Newtonian space, a Newtonian space inscribes a continuous surface. Nothing being between any two things in a Leibnizian space, a Leibnizian space inscribes a discontinuous field. A continuum not being a discontinuum, and a discontinuum not being a continuum, neither is reducible into the other. Each requires an element not constituent of the other to account for their difference.

Thus, presuming space is either a continuous surface without origin or conclusion, or is a discontinuous field without transition, renders mapping puzzling. For any space to be sensationally distinguishable while conceptually indistinguishable is impossible. Therefore, all space is sensational and conceptual. For any space to be sensationally indistinguishable while conceptually distinguishable is impossible. Therefore, no space is conceptually basic.

After all, a concept is a continuum. Concepts are a discontinuum. Therefore, no concept is composed of concepts. Only when introducing an insensate substance can mapping be explained. Abstraction satisfies the requirements of such an insensate substance.

Ontologically, abstraction is substantively distinguished by spatiotemporal constancy, and quality by spatiotemporal inconstancy. Epistemologically, abstraction is experientially distinguished by sense, and quality by sensation. Mapping requires an ambiguously continuous and discontinuous qualitative space. Ambiguous as member of the set of all things constituent of the set, and member of the set of all things not constituent of the set, a set limit is concurrently representable as a set element and not a set element.

Resolution of ambiguity within such a space constitutes mapping. Separating and relating, abstraction constitutes identity. Abstract reidentity of the ambiguously continuous or discontinuous as the certainly continuous or discontinuous resolves ambiguity. Resolving ambiguity, abstract reidentity constitutes mapping.

A field constitutes points with shared limits composing a "neighborhood" of "close" points.[7] Sharing limits, the points compounding a field are ambiguous. Mapping is the resolution of this ambiguity.

The limit of a point is the collection of all ambiguous constituents of the point. Mapping is the resolution of the ambiguity of the limit of contiguous

points. Exclusive disjunction terminates mapping. Inclusive disjunction extends mapping.

Justification of mathematical mapping initiates with Ernst Zermelo's axiom of choice whereby, "for any set whose members are sets that are non-empty and mutually exclusive, there exists at least one set having exactly one element in common with each of the sets belonging to the original set."[8] Relevantly, "each of the sets belonging to the original set" can be understood as different worlds. Proceeding thus, identifying the "one element in common" can be the same "'natural' properties,"[9] or if not, the same "essential" property,[10] this latter constituting a theoretical entity.[11]

Assuming same "essential" property need not manifest same "'natural' properties" in all occurrent worlds integrates "'natural'" and "essential" properties. Doing so, each occurrence of an essential property in a different world is a different manifestation of it. Although exhibiting different "'natural' properties," "sameness" is exhibited by mapping "'natural' properties" in one manifestation to "'natural' properties" in another manifestation.

Illustratively, assuming an "injection," "surjection," or "bijection," how is $\boldsymbol{D} = \boldsymbol{T}$ known?[12] Ruth Barcan Marcus indicates $\boldsymbol{D}$ and $\boldsymbol{T}$ must appear the same.[13] But, considering "1+1=2," "1+1" and "2" do not appear the same. Apparent sameness being unnecessary, identity is by "the 'rule' which tells us what $f(x)$ is."[14] Now, definiens is determined by identity with corresponding axiomatic archetype, and definiendum is determined by identity with corresponding axiomatic archetype. Here "the 'rule'" constitutes the "one element in common with each of the sets [$\boldsymbol{D}$ and $\boldsymbol{T}$] belonging to the original set [$\boldsymbol{D} = \boldsymbol{T}$]." Although not appearing the same, $\boldsymbol{D}$ maps to $\boldsymbol{T}$ by tracing through the medium of "the 'rule.'"

IDENTITY. Analogical identity can be likeness to a description or an archetype, Bertrand Russell's conception of a propositional function illustrating likeness to a description.[15] Conformity to a rule being a relation, though, a proof is incomplete without conformity to a second rule identifying conformity with the first rule. And conformity with the second rule being a relation, a proof is incomplete without conformity to a third rule identifying conformity with the second rule, and so on, a proof being endless.

Relation is an unbroken path between two elements within a domain. If a broken path, as with topology, how is an element prior to the break known to be the same element subsequent to the break? Elements in different domains are proven related by tracing an unbroken path between them, incorporating both into a common domain. Proof is tracing such an unbroken path, mathematically constituting identifying a dense set.[16] It is material when physical, an unbroken path of matter between limits. It is mental when phenomenal or conceptual.

Phenomenal it is semiotic, an unbroken path of consciousness between limits.[17] Conceptual it is logical, an unbroken path of elements between limits. Semiotic is particular, not general; logic is general, not particular. Particular is neither an analogical archetype nor analogical identity. General is either an analogical archetype or analogical identity.

Experience composes qualitative and abstract elements. Qualitative experience is sensation, and abstract experience is sense. Sensation is extended in time and/or space, and sense is extended in neither. Grouped and ungrouped qualitative elements, whether sensate or imaginative, are indistinguishable. Therefore, distinguishing grouped and ungrouped qualitative elements is not qualitative. By elimination, it is abstraction.

Observationally abstraction is intentional. It can appear as an unextended sense of self-identity, or an extended sense of self-identity containing or contained within another identity. Quality is known by sensation; abstraction is known by sense. There is no sensation of a quality distinguishing it as fused or diffused elements. There is no constant sense of a quality differentiating it as fused or diffused elements. There is a sometimes sense of something differentiating it as fused or diffused elements. Therefore, inconstant abstraction distinguishes quality as whole or parts.

Space and time are properties of quality, but not of abstraction. A quality over space and time is distinguishable from the quality at a constituent space and/or time. A quality continuous over a spatial/temporal range differs at each spatial/temporal segment of the range. An abstraction continuous over a spatial/temporal range does not differ at each spatial/temporal segment of the range.

Because inconstant, sense of abstraction can be mistaken. Error is determined by with what other quality or qualities a sensed quality is analogically identified. Sense as fused or diffused elements is sense of analogical identity of whole with whole or part with part. Identity of elements with what is self-evidently whole distinguishes elements as fused; identity of an element of elements with what is self-evidently part distinguishes elements as diffused. Since what is self-evidently part can be mistaken in its own identity, error is determined by with what other quality or qualities this sensed quality is analogical identified, etc.

3. PROOF

Distinguishing "between a subject matter under study and discourse about the subject matter,"[18] between metamathematics and mathematics, David Hilbert concludes moving from "the subject matter" to the "subject matter understudy" requires "One [to] at all times be able to replace 'points, lines, planes' by 'tables, chairs, beermugs.'"[19] Kurt Gödel institutes this project by a process of mapping whose, "underlying idea is to find a 'model' (or 'interpretation') for the abstract postulates of a system, so that each postulate is converted into a true statement about the model."[20]

This proceeds by identifying,

> A function [consisting] of three things:
>
> (1) a domain ***D***,
> (2) a target ***T***,
> (3) a rule which, for *every* $x \in \boldsymbol{D}$,

specifies a *unique* element $f(x)$ of ***T***.

> Item (3) is the heart of the matter.
> It is important that $f(x)$ be *uniquely* defined, so that there is no ambiguity attached to it.[21]

Employing such a function, Gödel implements,

> a method for completely 'arithmetizing' the formal calculus. The method is essentially a set of directions for setting up a one-to-one correspondence between the expressions in the calculus and a certain subset of the integers. Once an expression is given, the Gödel number uniquely corresponding to it can be calculated.[22]

Proceeding thus,

> there is one persistent source of difficulty. . . . the axioms are interpreted by models composed of an infinite number of elements. This makes it impossible to encompass the models in a finite number of observations; hence the truth of the axioms themselves is subject to doubt. . . . Non-finite models, necessary for the interpretation of most postulate systems of mathematical significance, can be described only in general terms; and we cannot conclude as a matter of course that the descriptions are free from concealed contradictions.[23]

Such a "difficulty" concerns definition, not analogy, though. Indeed, "Non-finite models . . . can be described only in general terms," but by archetypal analogy, they can be identified at any level of scale. It is perhaps for this reason in defining number,

> The set-theorist Ernst Zermelo proposed that the number 0 is the empty set (ϕ) and for each number *n*, the successor of *n* is the singleton of *n*, so that 1 is $\{\phi\}$, 2 is $\{\{\phi\}\}$, 3 is $\{\{\{\phi\}\}\}$, etc. So every number except 0 has exactly one member.[24]

Even if postulate systems are described in particular terms by this method, though, they are still contradictory. Contradiction is not simply a pragmatic function of postulate systems, it is an intrinsic function.

Nothing in Hilbert and Gödel's mapping scheme is inconsistent with pairing every element of "the subject matter" ***D*** with one element of the "subject matter understudy" ***T***, and every element of ***T*** with one element of ***D***, and no element of ***D*** or ***T*** with more or less than one element of the other, establishing a "bijection" of "the subject matter" and the "subject matter understudy."[25] Doing so, "the situation is perfectly symmetrical; and if we turn all the arrows round we define another function . . . in the opposite direction." Proceeding thus, "the subject matter" is converted into the "subject matter understudy."

Achieving this, "the subject matter" $\boldsymbol{D}^1$ of which the current "subject matter" ***D*** constituted the "subject matter understudy," is convertible into the current "subject matter understudy" ***T***, and so on, until the universe is converted into the current "subject matter understudy" ***T***. Now every ***D*** is converted into ***T***. Of course achieving this, the process can be reversed "in the opposite direction," converting every ***T*** into ***D***.

Now whether the current "subject matter understudy" ***T*** is "the [current] subject matter" or not is ambiguous. Thus it is impossible in the process of mapping "the subject matter" ***D*** to the "subject matter understudy" ***T***, to specify "a *unique* element $f(x)$ of ***T*** . . . so that there is no ambiguity attached to it." Proof by mapping as Hilbert and Gödel propose is unnecessary, then. Ambiguity is resolved only by embracing it in circularity, conjoining otherwise separate linear conversions of "the subject matter" and the "subject matter understudy."

Topological transmutation of abstraction and quality is impossible because an infinite expansion of abstraction cannot produce a quality, and an infinite reduction of quality cannot produce an abstraction. Solution is intentional alteration, consciousness substituting for Russell's,

> Between two perceived perspectives which are similar, we can imagine a whole series of other perspectives, some at least unperceived, and such that between any two, however similar, there are others still more similar. In this way the space which consists in relations between perspectives can be rendered continuous, and (if we choose) three-dimensional.[26]

4. CONCLUSION

Mapping constitutes transiting from point to point in Leibnizian space. Leibnizian space constitutes a field of unrelated points. Being unrelated, it is impossible to transit from point to point. Therefore, mapping is impossible.

Mapping is possible by extending a point in Leibnizian space in one dimension to another point. Composed is a line, constituting an extended point. Initiating and concluding points of the line constitute aspects of the same line/point. Thus, they constitute the same point, not separate points, the space within which the extended line/point occurs remaining a field of unrelated points.

To assemble such a field requires identifying the limit of a domain as an exclusive disjunctive of all otherwise ambiguous constituents of the domain. A point transited at the limit of a domain is identified by inclusive disjunction, not exclusive disjunction, providing transiting of the extended point into the contiguous domain. All other points contiguous to the extended point are continued to be resolved exclusively, otherwise sustaining the perimeters of the transited domains. Similarly, all other points contiguous to the extended point are resolved exclusively, sustaining the perimeter of the transiting linearly extended point. Excised from the contiguous domains by this means, the extended point constitutes an autonomous self-contained domain, a set of one.

5. REFERENCES

[1] Alan Hausman, Howard Kahane, Paul Tidman, **Logic and Philosophy: A Modern Introduction**, 10th ed., Belmont, California: Thomson Wadsworth, 2007, p. 91.
[2] Gottlob Frege, "On Sense and Meaning," **Translations From the Philosophical Writings of Gottlob Frege**, 3rd ed., eds. Peter Geach and Max Black, Totowa, New Jersey: Rowman & Littlefield, 1980, p. 63.
[3] Gottlob Frege, "Translation of Parts of Frege's *Grundgesetze der Artihmetik*," trans. P. E. B. Jourdain and J. Stachelroth, in **Translations from the Philosophical Writings of Gottlob Frege**, eds. Peter Geach and Max Black, Oxford: Basil Blackwell, 1960, First published in 1884, p. 124.
[4] Bertrand Russell, **Our Knowledge of the External World**, New York: A Mentor Book, 1960, p. 73.
[5] John D. Baum, **Elements of Point Set Topology**, New York: Dover Publications, Inc., 1964, p. 20.
[6] Ian Stewart, **Concepts of Modern Mathematics**, New York: Dover Publications, Inc., 1995, p. 9.
[7] Baum p. 20.
[8] Stephen F. Barker, **Philosophy of Mathematics**, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1964, p. 77.
[9] Ruth Barcan Marcus, **The Journal of Philosophy**, Vol. 68, No. 7, April 8, 1971, p. 193.
[10] Ibid., p. 192.
[11] William James Earle, **Introduction to Philosophy**, New York: McGraw-Hill, Inc., 1992, pp. 68-69.
[12] Stewart, p. 71.
[13] Marcus, p. 191.
[14] Stewart, p. 69.
[15] Bertrand Russell, "Descriptions," ed. Robert R. Ammerman, **Classics of Analytic Philosophy**, New York: McGraw-Hill Book Company, 1965, pp. 15-24.
[16] An ordered set is said to be dense, if it contains at least two elements and no neighboring elements. A dense set is always infinite, because every finite set containing at least two elements has also neighboring elements.

E. Kamke, **Theory of Sets**, trans. Frederick Bagemihl, New York: Dover Publications, Inc., 1950, p. 70.
[17] William James' "stream of consciousness" manifests this, being,

> nothing jointed; it flows. . . . But now there appears . . . a kind of jointing and separateness among the parts, of which . . . . I refer to the breaks that are produced by sudden *contrasts in the quality* of the successive segments of the stream of thought.

William James, "The Stream of Thought," *Principles of Psychology*, in **Pragmatism: The Classic Writings**, ed. H. Standish Thayer, New York: Mentor, 1970, p. 142.
[18] Ernest Nagel and James R. Newman, **Gödel's Proof**, New York: New York University Press, 1986, p. 31.
[19] Carl B. Boyer, **A History of Mathematics**, 2nd ed., New York: John Wiley & Sons, Inc., 1991, p. 610.
[20] Nagel and Newman, pp. 15-16.
[21] Stewart, pp. 67-68.
[22] Nagel and Newman, pp. 74-75.
[23] Ibid., pp. 21-23.
[24] Stewart Shapiro, **Thinking about mathematics: The philosophy of mathematics**, Great Clarendon

Street, Oxford, UK: Oxford University Press, 2000, p. 265.

[25] . . . arrows . . . represent the 'rule' which tells us what $f(x)$ is. [Ian Stewart, *Concepts of Modern Mathematics* (New York: Dover Publications, Inc., 1995), 69.] . . . The standard notation expressing the fact that f is a function with domain $\boldsymbol{D}$ and target $\boldsymbol{T}$ is

$$f:\boldsymbol{D} \to \boldsymbol{T} \ldots .$$

If the range of f is the whole of $\boldsymbol{T}$, then f is said to be a function onto $\boldsymbol{T}$. Another common word for such a function is *surjection* (from the Latin: f throws $\boldsymbol{D}$ on to $\boldsymbol{T}$). . . . If every element of $\boldsymbol{T}$ lies on at most one arrow (perhaps on none) then f is an *injection*. [Stewart 70.] . . . If we have a function $f:\boldsymbol{D} \to \boldsymbol{T}$ which is both an injection and a surjection, then the arrows pair off elements of $\boldsymbol{D}$ and $\boldsymbol{T}$: an element of $\boldsymbol{D}$ at the tail end, an element of $\boldsymbol{T}$ at the head. [Stewart 70-71.] . . . [T]he situation is perfectly symmetrical; and if we turn all the arrows round we define another function

$$g: \boldsymbol{T} \to \boldsymbol{D}$$

in the opposite direction. . . . Functions which can be turned round in this way . . . . are known as *bijections*, or as *one-to-one correspondences*.

Stewart, p. 71.

[26] Bertrand Russell, **Our Knowledge of the External World**, p. 73.

Program design of STEAM education initiatives in urban communities

Dr. Joseph Bowman, Jr.
Center for Urban Youth and Technology
Department of Educational Theory and Practice
University at Albany
Albany, New York 12222 USA

Dr. Joseph Bowman, Jr.
Center for Urban Youth and Technology
University at Albany
Albany, New York 12222 USA

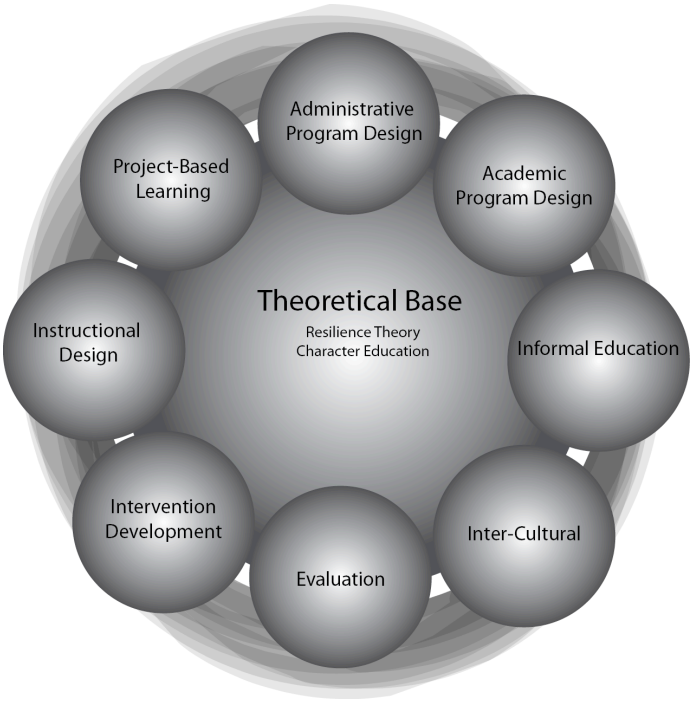
ABSTRACT

This presentation discusses the significance and relevance of program design research in Informal environments in urban communities. Design and project based research developed through the Center for Urban Youth and Technology (CUYT) model has produced several Science, Technology, Engineering, Arts, and Math (STEAM) projects in urban communities. Elements of our design model will be explored and defined. Connections between project, design, and intercultural research will be presented to define how the CUYT model has evolved. A case study intervention will be included to provide evidence and details of our model. An external project based research model will be provided for comparison and utilized to enhance further discussion.

Keywords: Project Based Research, Informal Education, Instructional Technology, Resilience, Character Education, Inter-Cultural, Intervention, and Evaluation.

1. INTRODUCTION

The CUYT model (Figure 1) is adaptable, flexible, and able to be implemented in any urban setting given knowledge of that community environment. It is a project based, hands-on, and non-threatening, student-centered learning environment. A driving theme is the importance of providing intercultural models that are based on the population of participants living in these urban settings.



CUYT Design Model (Figure 1.)
Design by Manulito Loman, M.L.S.

The model addresses “cradle to the grave” or “pipe line” notions, and provides continuity in STEAM education activities for urban youth, their parents, and their teachers. The theoretical base for the model centers around the “resilient“ nature of urban youth and the design of interventions that support and expand these attitudes and concepts.

Our presentation defines the model and discusses the basic themes for the design and development of program interventions that incorporate project-based design, informal education environments, and intercultural [1] models.

STEAM is identified as a critical factor in the development of innovation, economic opportunities, and social expansion of the United States of America. We are challenged by a "Quiet Crisis"[2]: we are not preparing our youth to become future 21st century leaders of the labor and university workforce. Public and private education must provide academic and scholarly pathways that support educational achievement for our youth. The challenge of the "Quiet Crisis" includes high drop out rates that are problematic. Each year, approximately 1.2 million students fail to graduate from high school, more than half are from minority groups. [3]. The average scores of 15-year-old students (PISA 2009) rank 25th out of 34 countries when compared with students elsewhere in the world. [4]. The average scores of 15-year-old students on a science literacy scale, Ranked 17th out of 34 OECD countries. [5]

The data suggests that educational program designers need to identify new, innovative, and creative strategies to address these issues and reach this "new majority" of learners.

The problem is how to reach the students of the 'Quiet Crisis" with cost effective program design models that support and supplement existing formal academic systems. Another option is to create alternative (informal) program models that provide authentic, hands-on activities to mirror or shadow educational and workforce experiences. Our model has selected an informal education, intercultural, and project based approach. The CUYT model can align its applications with the needs of the student population and with the needs of the academic and workforce environment where the model is deployed. STEAM activities, intercultural models, instructional design, and project-based research anchor the model. Flexibility allows the model to be used in formal and informal settings, but informal settings (summer, after school, and weekends) have provided greater creativity, larger collaborative opportunities, and immersion 'direct connection' STEAM experiences.

2. CUYT DESIGN MODEL ELEMENTS

The CUYT design model uses several elements of instructional design theory in the format, concept, and structure of this STEAM based model of instruction. Needs assessments, task analysis, learning theories, cultural awareness, and technology integration are aspects of this component area. The goal is to develop program intervention models that reach underserved, economically challenged youth (3rd grade – 12th grade, cross gender, ethnically mixed, all religions) who have an interest in learning about the STEAM fields. Many students are level 1 or 2 in middle school (under the New York State evaluation system) and special education in high school. Although written off, these students have great educational and academic potential if the academic environment can be changed and modified. Student centered, non-threatening, informal education environments and interventions need to be created to serve this population.

Elements of the CUYT design model (see Figure 1) are defined and include: Intervention Development, Program Design, Academic Program, Administrative Program, Intercultural models, Project based learning, and Informal and Public education around a Theoretical base.

3. INTERVENTION DEVELOPMENT

Two central themes need to be considered as this phase (intervention) evolves. (1) Knowledge of your audience, their concerns, academic background, attitude, and interest level. (2) What is the working or operating environment (school, university, CBO, informal or public education, private, and community) that we will operate in. The CUYT model has operated in public schools, community based organizations, churches, universities, and city community centers. We prefer university settings with significant resources (faculty, students, and facilities), but many of our most successful interventions were convened in community settings. We are embedded in urban settings and have the opportunity to interact with students, parents, and their environment. This action gains student and community trust, respect, and teaches us how to reach and serve this population.

4. ACADEMIC PROGRAM DESIGN

In this area we reverse engineer our design by asking what expectations, outcomes, assessments, and final projects would be evaluated for student success. A

series of course/workshop rubrics, activities, and presentations that identify student skills levels and content knowledge are created to support the academic program design. Discovery, hands-on immersion, STEAM exploration, and cultural awareness techniques are used to create problem solving and other higher order learning skill activities with the students. We raise the academic bar for these students and in active programs have observed that when challenged, they move toward and meet the challenge. Curriculum content is developed with subject area specialists, is examined and matches our content to the graduation specific standards at the national and state level.

This ensures that these interventions follow the same curriculum grade level standards that the students follow in their public school lives. Knowledge of student learning and achievement status provides us (curriculum designers, teachers, and university facility) with a guide to student's prior learning and skills. This allows our interventions to develop a mentorship and tutor resource for students and parents.

5. ADMINISTRATIVE PROGRAM DESIGN

Operational considerations are essential elements of a total program model and ensure effective handling of program activities. Elements include: salary, schedules, space-facilities, contracts, calendars, availability (staff, students, parents, faculty/instructors), transportation, food, securing funding, and grants writing. Program intervention sustainability is equally important to determine program resources, length of intervention, and quality of services.

Any program design model must have a strong leadership team, who are passionate about the work, willing to put in the required time for program success, and have the ability to work with a diverse range of staff, faculty, and students. Attention to detail is an important quality of the leader team. (It is the little things that can bring things to a halt and stop the show). Networking and the ability to create collaborative partnerships are important, as this impacts funding, establishes other program resources, and adds new content ideas to the interventions. (in this discussion interventions and activities are used in the same context to represent various aspects of our CUYT programs).

6. THEORETICAL BASE

The CUYT model has been rooted in the importance of cultural awareness and resilience theory [7]. Our view of resilience theory subscribes to the beliefs that all youth are resilient, creative, and ready to learn. By labeling or branding students, (at-risk, level 1 or 2, and/or special education) we place a stigma, bias, and attitude toward these students. A stigma that we project through our educational system and our society. It is these labels that our students buy into. When we stop the labeling or change the label type that we associate with our youth, their attitudes about who they are and their academic abilities will also change.

Resilience theory is an important concept that can be used to start the process. We view all students as gifted and/or talented and provide them with the resources and opportunities that allow them to succeed. We are not naïve and realize that these youth are at different levels on the social, emotional, and academic scales. The respect and passion that we demonstrate toward these youth; the non-threatening learning environment that the design model creates; and the unique content areas of study in STEAM through nanosciences challenge and stimulate students' desire to learn. This supports our integration of character education into the design model to address issues of self-respect, team building, honesty, loyalty, bullying, and motivation. Character education provides assistance to students that reside in communities where gang violence, crime, and drugs are prevalent.

7. INTER-CULTURAL MODEL

The CUYT model is designed for all youth and adult learners, but clearly focused on underserved, economically disadvantaged, and academically challenged urban populations (African-American, Caribbean-American, and Hispanic). As part of our CUYT design model and to address the multi-cultural needs of our program population, inter-cultural models are included in the design model. Cultural elements of historical contributions, STEAM role models, current tools/devices in STEAM, STEAM movies, shows, and theater

presentations, and demonstrations of economic and community development are integrated into the CUYT design model.

Specific program activities include: Culturally Situated Design Tools that provide web-based instruction on the cultural relationship between math and their culture; the "Black Book Project" sessions, where musicians interpret the images from the Hubble Telescope for youth; and the urban "Nano" theater, where students create skits and video productions about African and Hispanic American scientists.

Cultural design inclusion demonstrates how various cultures have supported the rich mosaic of STEAM discoveries and opportunities in the country. It provides evidence of our participation in science and math through human history.

8. PROJECT BASED LEARNINGS (PBL)

The days and times of the "sage on the stage" has given way to "coach teaching" classroom activities that are supported with simulations and interactive learning between multiple school sites. Technology of all forms is integrated in the classroom environment. Even the classroom can be transformed into an outdoor living lab or a mobile or remote site around the country or world. This is the world our students see and we are challenged to use real world experiences to assist in the academic and instructional development of their experiences. Larmer and Mergendoller (2010) presented seven elements of PBL which are supported by the CUYT model: a need to know, a driving question, student voice and choice, 21st century skills, inquiry and innovation, feedback and revision, and a publicly presented product. Student centered learning environments, team building, and collaborative teaching are included in the CUYT design model.

9. INFORMAL AND PUBLIC EDUCATION

Early interventions of the design model were implemented in school settings, adhering to class periods, block schedules, administrative red tape, class size, staff/instructor availability, and classroom/computer lab availability. We moved to an infused school day activity where one or two days and times were selected and program activities were provided. This was facilitated by the school administration, small school size, selection of students, small class size, and community and business participation.

After school interventions continued in elementary and middle school environments, but student external activities (Boy/Girl scouts, sports programs, and other social activities) compromised our attendance and completion rates. We scaled back the after school activities and increased the weekend and summer activities. Our informal education interventions have evolved into yearly weekend and summer (four to six week) programs. The CUYT model has created external partnerships with community based organizations, area businesses, school districts, state/local agencies, and universities/colleges. The model allows our program interventions to be flexible, current, and provide real world experiences for program participants.

10. INSTITUTE FOR NANOSCALE TECHNOLOGY AND YOUTH – HIGH SCHOOL CASE STUDY

This intervention focused on thirty high school and adult education students, who were considered "Special Education", on track for academic dismissal (drop out) from area schools, economically disadvantaged, and represented Afro-American and Hispanic populations in the capital district of New York State. Our goals were to introduce them to career opportunities in information technology, nanoscale sciences, and multi media design. Students met in a series of workshops sessions to explain program goals, session activities, benefits of this intervention, program expectations, student outcomes, and criteria for selection to the six week summer program (no summer school classes).

Our partners, College of Nanoscale Science and Engineering (CNSE), provided the Nanoscience training sessions (three weeks). The Center for Urban Youth and Technology (CUYT) provided the character education and multi media (e-publishing, video production, and robotics) sessions (one week). And the University Center for Academic and Workforce Development (UCAWD) provided the

Microsoft IT Academy Word training and certification (two weeks).

Youth development city resources provided employment salaries for students and they were required to make presentations about their program experiences to university faculty, district administration, teachers, parents, and fellow students. Students produced a program newsletter and a program video production. In these hands-on activities, students created articles, photos, power point presentations, and rap poetry for the newsletter. Scripts, program formats, production crew selection (camera person. audio, lighting, and video editor) had to be determined to complete the video production.

11. REFLECTIVE PROCESS

As part of our reflective process, other design models were identified and reviewed.
The CUYT model was compared with the *Research Methods for Community Change: A Project-Based Approach*, by Randy Stoecker.[6] Stoecker's project-based approach (diagnosis, prescription, implementation, and evaluation) (PBA) was similar to our model, population, and communities. PBA enhanced our research and evaluation methodology and we included surveys, writing samples, student presentations, focus groups, and interviews for the CUYT model and student achievement. We are creating a program evaluation report and an analysis of participant attitudes and achievement in the STEAM fields of study.

12. CLOSING

The CUYT design model is created to provide information and access to STEAM resources in urban communities. Many students. teachers, and parents in these communities are not exposed to the high tech bio-technical, alternative/renewable energy, nanoscience, e-transportation, robotic, radio frequency aircraft, high speed broadband/wireless, and information technology fields of today and the future. Our model is flexible and can be utilized across elementary, middle, high school, and adult students. We have focused our interventions on the STEAM fields to address the country's aging work force and under utilization of our underserved populations, and to increase the pool of innovative ideas into our society. The CUYT design model is bridging educational achievement with work force needs and economic development opportunities to demonstrate the effectiveness of this type of design process.

13. REFERENCES

[1] P. Young, **Instructional Design Frameworks and Intercultural Models.** IGI Global Hershey, PA, 2009, (p. XIV).

[2] S.A. Jackson, **The Quiet Crisis: Falling Short in Producing American Scientific and Technical Talent.** Building Engineering and Science Talent (BEST), 2005.

[3] Alliance for Excellent Education, **High School Dropouts in America: Fact Sheet.** 1201 Connecticut Avenue, NW, Suite 901, Washington DC 20036, 2009. www.all4ed.org

[4] H.L. Fleischman, P.J. Hopstock, M.P. Pelczar and B.E. Shelley, **Highlights From PISA 2009: Performance of U.S. 15- Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context,** (NCES 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, 2010, p. 18.

[5] H.L. Fleischman, P.J. Hopstock, M.P. Pelczar and B.E. Shelley, **Highlights From PISA 2009: Performance of U.S. 15- Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context,** (NCES

2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, 2010, p. 24.

[6] C. Rak & L. Patterson, "Promoting Resilience in At-Risk Children", **Journal of Counseling &Development:** JCD, 744, 1996, pp. 368-373.

[7] J. Larmer and J. Mergendoller, "7 Essentials for Project-Based Learning. Some "projects" border on busywork. Others involve meaningful inquiry that engages students' minds", **Educational Leadership,** Vol. 68, No. 1, 2010, p. 2.

[8] Randy Stoecker, **Research Methods for Community Change: A Project-Based Approach**, Sage Publications, 2005.

Integrated Modeling of Agricultural Production Systems: Achievements and Remaining Issues

François Guerrin
UR Recyclage et risque, CIRAD
Saint-Denis, 97408 La Réunion, France
UR Biométrie et Intelligence artificielle, INRA
Castanet-Tolosan, 31326, France

ABSTRACT

Improving the sustainability of agriculture has become crucial to deal with tomorrow's challenges such as supplying food to a continuously growing world population while mitigating its environmental impacts (e.g. climate changes). Recycling organic wastes to substitute chemical fertilizers for various organic ones (e.g. sewage sludge, household refuses, plant residues, livestock manures, agro-food industrial wastes) is one of the ways towards this end. Addressing this calls for the coordinated use of heterogeneous knowledge on both the biophysical (i.e. organic products, soils, crops) and managerial (i.e. farmers' practices) components of the whole production systems. Computer models, encompassing various pieces of that knowledge, are built to represent these systems as linked production and consumption units spread over a territory. These models are used for simulating management scenarios and assessing their performances against agronomical and environmental criteria. This paper describes our main achievements: (i) a methodology for modeling and analyzing material flows on a territory scale; (ii) a conceptual modeling framework of farming systems; (iii) a way of representing human activity in farming systems based on the 'situated action' theory. It points also out two remaining issues: (iv) assessing simulated management scenarios; (v) using models with stakeholders to support their management practices.

Keywords: Simulation modeling; Hybrid dynamical system; Activity representation; Situated action; Operations management; Agricultural production systems; Environmental assessment.

1. INTRODUCTION

The research discussed in this paper is focused on simulation modeling of agricultural production systems considered at two organization levels: single farms (individual management) and organized sets of farms (collective management). The aim of this research can be rephrased as designing simulation models to help design management policies of farming systems (and conversely). These models are conceived with farming systems agronomists to help evaluate farming systems management. They allow the dynamics of the various material flows (namely, biomass) operating within the production systems in interaction with the farming practices to be simulated. Two modeling approaches have been favored until now: hybrid dynamical systems, encompassing both continuous and discrete variables, and multi-agent systems.

Two research issues of unequal importance, the second being tackled since only recently, are dealt with:

- Finding representational structures (i.e. conceptual and formal frameworks) to make operational the available knowledge: designing the model is here the focus;
- Finding tools (i.e. computer models and the way to use them) to support agricultural stakeholders: designing management policies is here the focus.

In terms of models, the main achievements are the following:

- Material flow dynamic simulation models, based on the analysis of agricultural practices [1], to reason about various cases of livestock waste management: single farms (MAGMA model [5]); groups of farms (BIOMAS multi-agent system [4]); collective waste treatment plant supplied by multiple farms (APPROZUT model [6]); collective manure application plan considering the interaction between the individual (single farms) and collective (groups of farms) levels of management (COMET model [21]).
- Simulation of flow networks using timed automata and model-checking [13].
- Joint representation of farming practices and biophysical flows within dairy farms (GAMEDE model [23])
- Modeling framework of human activity at operations level with generic aim [9].

This paper provides details about the principal methodological findings:

- A methodology for modeling and analyzing material flows on a territory scale (Section 2);
- A conceptual modeling framework of farming systems (Section 3) illustrated on three models among those enumerated above (Section 4) ;
- A way of representing human activity in farming systems based on the 'situated action' theory (Section 5).

It also discusses two important issues that still remain incompletely resolved:

- How to assess simulated management scenarios? (Section 6);
- How to use our models with stakeholders to support their management? (Section 7).

The perspectives open to the different sides of this work in the coming years are pointed out.

2. MODELING AND ANALYZING MATERIAL FLOWS ON A TERRITORY SCALE: THE 'MAFATE' APPROACH

Beyond the development of the simulation models enumerated in Section 1, one of the main achievements is the formalization of the approach which actually constituted the driving thread of the research done in partnership with systems agronomists. This approach, termed 'Mafate' [11], encompasses several steps yielding the following outcomes:

1. Farm surveys, covering the diversity of management situations found in the considered territory;
2. Farm typology, defining the main farming types and characterizing both their structure and management policies;
3. Conceptual models, synthesizing the knowledge gained on farming practices from surveys;
4. Computer models, designed to simulate the interaction between the material flows and the farming practices at both 'individual' (intra-farms) and 'collective' (inter-farms) levels of organization;
5. Simulation outputs of management alternatives checked by experts (e.g. agronomists, technical staff, skilled farmers) according to agricultural and environmental criteria;
6. Model validation as virtual experiment tools in relation with agricultural stakeholders.

Steps 1 and 3 are deemed essential prior to constructing flow management models in order to account for actual farming practices, identify and explicitly describe actual management constraints and strategies. Step 2 is also very useful to take into account the diversity of situations found in the region considered. Performing simulations with the models (step 5), a long but interesting task, is mandatory to analyze dynamically (in contrast with widespread static methodologies as Life-cycle analysis) the functioning of the systems represented. Important work remain to be done, on the one hand, on multi-criteria assessment of simulated management strategies (step 5), on the other hand, on the use of models in management situations with agricultural stakeholders (step 6). These two issues, still incompletely resolved, are detailed below (Sections 5 and 6).

3. A CONCEPTUAL FRAMEWORK FOR MODELING AGRICULTURAL SYSTEMS

The ambition was to design a modeling framework with the following aims:

- Representing agricultural production systems on different temporal and spatial scales;
- Integrating the various pieces of knowledge available on these systems;
- Simulating the dynamics of interactions between management practices and material flows;
- Assessing the impact of these practices on the systems' viability and sustainability;
- Designing management strategies to improve the systems' performance against various criteria.

The material and work flow models that have been developed (cf. Section 1), the recent efforts of model generalization (extension of waste management to whole-farm operations in GAMEDE [23]; generic simulation of action [9]) and the design of a comprehensive approach ranging from the acquisition of knowledge to model building and simulation to support agricultural stakeholders (cf. Section 2) go in this direction.

These experiences allowed an understanding of the representation of agricultural production systems considered at different levels of organization on various temporal and spatial scales (farm, group of farms, agro-food supply chain) to emerge. According to this understanding, an 'Action-Flow-Stock' ontology has been devised [7].

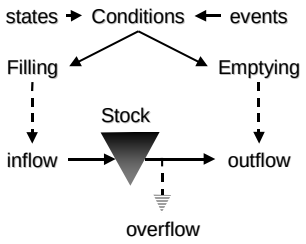


Figure 1. Action-Flow-Stock representation of a production unit (PU).

According to this ontology, agricultural production systems are represented as a set of stocks connected by flows of materials controlled by the farming activities (Fig. 1). Two types of flows are distinguished: "workable" flows, which take place only if there is human intervention, and "biophysical" flows, which take place even in the absence of human intervention. These flows interact through human activity, which aims to guide the biophysical flows, among which those leading to the "products" of the system, by the workable flows it generates. The management of the production system can then be seen as the control of a set of stocks by the activities of the operator (i.e. the farmer and farm workforce). These activities stem from the confrontation between encountered situations and strategies: implementing strategies helps create new situations; the experience gained by this implementation can, in turn, change strategies.

The relevance of this conceptual framework, derived by generalization of livestock effluents management models listed in Section 1 (i.e. MAGMA, BIOMAS, APPROZUT), has been verified, on the one hand, at the level of individual farm operation [23], on the other hand, at the level of collective management:

- Simulation of a hog slurry collective application plan in Brittany (Western France) using the COMET model [12][21];
- Draft modeling [8] and life-cycle analysis of the Reunion Island swine sector described as a supply chain.

The coupling of workflow management models with mechanistic models of biophysical processes may, however, be problematic when the data necessary to the setting of the latter are missing or when their generic feature is not guaranteed in the local situation investigated. To represent these

processes, we thus moved towards the synthesis of expert knowledge in the form of simple empirical rules or formulae validated locally. An example of such an empirical coupling is provided by the GAMEDE model [23].

4. EXAMPLES OF MODELS BASED ON THE ACTION-FLOW-STOCK ONTOLOGY

MAGMA: Livestock effluent management at farm level

The MAGMA model [5] addresses the case of livestock effluent management within a farm. Two types of units are involved in such a "distribution" (i.e., one-to-many) configuration (Fig. 2): livestock enterprises producing animal wastes and consumption units, such as crop plots or waste treatment plants, where effluents are spread or supplied.

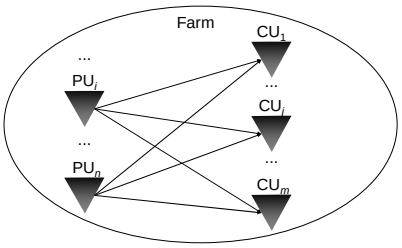


Figure 2. Distribution configuration in the MAGMA model to represent organic waste management within farms (PU: livestock enterprise; CU: consumption unit).

Simulating MAGMA allows management strategies of livestock effluents to be assessed with respect to several indicators: environmental (nitrogen losses due to stock overflowing, fallow land spreading, over-fertilization of crops); agronomical (nitrogen applied to crops); economical (working time, vehicle mileages…) and organizational (frequency and temporal distribution of spreading actions). MAGMA has been used to analyze waste management policies in livestock farms in Reunion Island, such as that described in [20].

APPROZUT: Supply of treatment plant by multiple farms

The APPROZUT model [6] deals with the case of simulating a two-stage production system where the first stage is a set of pig farms producing slurry scattered over a territory and the second is a unique collective treatment plant where slurry is brought in a many-to-one fashion (Fig. 3). Policy assessment is mainly done in terms of organization and logistics.

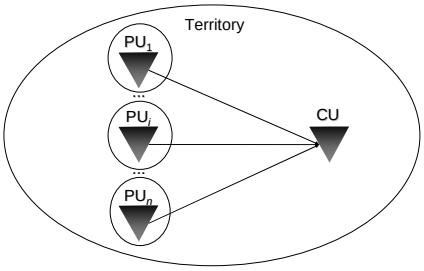


Figure 3. Supply configuration in the APPROZUT model (PU: livestock farm; CU: single waste treatment plant).

Approzut has been used to analyze a project of pig slurry treatment involving 51 pig farms located in a remote mountainous cirque in Reunion Island where available agricultural land was too scarce to spread raw slurry.

COMET: Mixed distribution and supply configuration

COMET [21] essentially results from coupling together the MAGMA and APPROZUT logistic models yielding the distribution/supply configuration displayed on Fig. 4. It also includes sub-models simulating biophysical processes used as environmental assessment criteria (e.g. the STAL model [19], which simulates ammonia emissions at spreading).

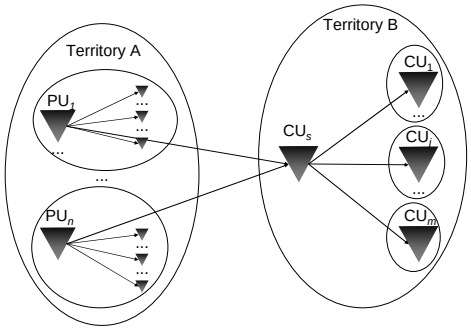


Figure 4. Mixed distribution/supply configuration in the COMET model (PU: livestock farm encompassing also crops; CUs: intermediate storage; CU: crop farm).

COMET has been used to jointly simulate individual manure spreading within single pig farms and the functioning of a collective spreading plan aimed at transferring manure surpluses from livestock farms to land loaned in remote crop farms in Brittany (Western France). The alternate use of dynamic simulation with COMET and static life-cycle analysis allowed the whole functioning of this case-study to be thoroughly assessed [12].

5. A CONCEPTUAL SHIFT: FROM PLANNED ACTION TO SITUATED ACTION

The confrontation of action representation in the flow management models based on the Action-Flow-Stock ontology [7] with the ontology of agricultural production systems devised by Martin-Clouaire and Rellier [17] led to question the paradigm of 'planned action' in favor of the theory of 'situated action' [22].

The management problems at the operational level are, indeed, typically formulated in terms of planning and decision. This is the very Western conception that actions necessarily result from deliberations made with representations (plans) to decide in response to previously established intentions. The study of many domains, however, shows that a very large part of human activity is non-deliberative or, even, reactive in nature; it takes place in interaction with the local situations in which each agent is involved [3][14]. Therefore, the theory of "situated action" alleges there is no need of representing explicitly the activity to be performed; plans, although they may be used to guide action, never determine it completely.

The action modeling framework already drafted [9] has the ambition to contribute to this situated action theory. The

first reason is the construction of models. If the goal is to represent detail of agricultural, large and complex production systems, basing any action on a comprehensive and coherent plan appears elusive, due to the complexity of planning itself. This challenge is also justified from a theoretical point of view, except to enter an infinite recursion loop: if any action is planned, then so is planning, and planning for planning also, and so on... The other reason is linked with the usefulness of the models. If the objective is to evaluate production systems, it is by representing as better as possible what is actually done, and not what should be done (i.e. tasks specified by the plan), that can allow the impacts of activity to be measured and, in turn, the mutual influence of the context, thus modified, on the activity itself to be appreciated. Taking an a priori defined plan of action as essential determining factor would be similar to taking a static referent in an inherently dynamic environment to generate a process which is, also, dynamic. In contrast, taking action as a focal point, the present approach is designed to meet Checkland's wishes: "...modelling purposeful human activity systems as sets of linked activities which together could exhibit the emergent property of purposefulness" [2].

It is, hence, the operational level that must the models represent in being primarily focused on action rather than on decision and planning. However, it is at the strategic level that these models must be used to assist researchers in experimenting the systems and, possibly, stakeholders in their decision processes, in keeping with Mc Cown's view [16]. In other words, if the model must represent the action of virtual agents at the operational level, its use must contribute to the decision-making of real actors at the strategic level. These are currently the main research objectives:

- Develop an ontology for representing systems of activities at the operational level by a minimum and consistent set of concepts;
- Formalize this framework to build simulation models of agricultural production systems;
- Analyze with these models these systems operation viewed as the interaction between biophysical processes and human activities;
- Infer practical lessons to help manage these systems.

In this perspective, the concepts relevant to describing the coordination between actors, the spatial location of activities, the physical structure of the work setting and the relationship between the concepts of agent and action shall be specified. This is part of an ongoing PhD thesis project supervised by the author.

6. THE ISSUE OF ASSESSING FARMING SYSTEM MANAGEMENT

Any management requires the assessment of the system's performance it relates to. The comparison of management policies, so far, was based only on a few indicators calculated by the flow models: agronomic (e.g. nitrogen applied relatively to crop needs), environmental (e.g. nitrogen excess, ammonia and methane emissions), economic (e.g. working time, distance traveled by vehicles) or organizational (e.g. temporal distribution of activity, robustness to disturbances). These indicators take into account only two dimensions: technical, measured in terms of efficiency, and environmental, measured in terms of risk, taking nitrogen as main criterion. The technical dimension assesses the system at the level where it is represented. If it qualifies its viability in the short term, it does only little in appreciation of its contribution to sustainable development in the long term. The environmental dimension concerns the system outputs on a scale that encompasses it immediately (i.e. the impact on its immediate environment). Environmental risk is addressed only as "hazard" (occurrence of a risk factor) and ignores the sensitivity and the particular nature of the receiving environment. In both cases, the assessment is performed with a normative view.

To address these problems, we must distinguish between two questions:

- How to evaluate the technical performance of production systems?
- How to evaluate their actual or potential environmental impacts?

In the first case, modeling biophysical flow is needed to simulate their interactions with the workable flows. This does not imply to represent all mechanisms in detail but, at least, to have a robust approximation of their evolution. To do this, the knowledge on the biophysical processes is synthesized by expressions for linking, as simply as possible, the causes and effects without going into the details of the underlying mechanisms (cf. Section 3). In the second case, comparing different management strategies is needed. The issue of sustainable development, which has become the essential assessment criterion, leads now to think the impacts of these systems in terms of risk (proven or alleged) on other time and space scales (often larger) than the ones on which they were previously considered. Hence, the interest in overall assessment approaches ("from cradle to grave"), such as life cycle analysis (LCA), which allows this comparison (although statically) through standardized indicators representative of different categories of impacts. An example of alternatively combining LCA with simulation modeling in a comprehensive approach to assess and help improve the design of a collective manure management plan by a group of farmers has been realized recently [12] [15].

These preliminary results are far, however, from exhausting the subject of environmental assessment which deserves to be rethought in the light of the objectives: what has to be assessed, for which purposes, with which actors? The goal of assessing the sustainability of farming systems striving to adapt to multiple change factors requires also defining the relevant space and time scales to be accounted for. The choice of the 'scale of representation' of a production system becomes, thus, a central issue for modeling, along with the methods of up- or downscaling the current models as soon as an extension or reduction of scope necessary to embrace larger or finer scales is sought. This questioning is a research perspective.

7. THE ISSUE OF USING MODELS FOR MANAGEMENT SUPPORT

The main question is: How to use simulation models to help stakeholders evaluate and design management strategies of production systems? This issue calls to other more specific questions related to:

- The ways of using the models: Which users? What situations? What modes of interaction?
- The engineering of simulation likely to facilitate users' learning: Which cases to simulate? What scenarios? Which protocol? How to capitalize the knowledge gained through simulations?

Dealing with these questions was first attempted in the period 2004-2007, unfortunately with too little achievements. If a first experience of participatory simulation had been made to assist in the choice of the treatment process for pig manure in the locality of Grand Ilet in Reunion [18], it was using a GIS and a spreadsheet model developed by fellow agronomists. The dynamic simulation models listed in Section 1, although quite used by these colleagues, have not yet been tested truly to design management strategies with "real" agricultural actors. When it could have been the case, actually, the projects aborted prematurely for unexpected reasons: in Grand Ilet (with APPROZUT), the action-research dynamics that had been initially launched by researchers was interrupted once the folder had been assigned to one of the institutional partners; in Brittany (with COMET), the collective manure application project was stopped due to the opposition of residents, not accounted for in the model...! The phase shift between the researchers' and the actors' time explains, in part, this state of affairs. However, deeper causes must also be sought in our inability to correctly grasp the social games of players in these organizational or political processes. In these contexts, beyond a purely technical rationality, one might ask if actual decision still requires the support of a model. It seems not.

Nevertheless, the work with the MAGMA [20], APPROZUT [10] and COMET [12] models allowed the way for a simulation approach to design management policies of production systems to be paved. The protocol was designed with an experimental logic: (i) construction of a base scenario corresponding to the current situation, (ii) assessment and analysis of the scenario through simulation, (iii) introduction of gradual changes for designing iteratively new scenarios. This dimming of the changes introduced in the simulation scenarios corresponds, from the point of view of operations management, to challenging firstly very short-term operational choices, then medium-term tactical decisions, and, finally, longer-term strategic decisions. The objective of this approach is not only to understand why farmers do what they do, but, above all, to understand their rooms for maneuver.

The production of documents allowing the user's approach to be represented and the knowledge gained by simulation to be capitalized is, for now, manually performed in a paper form. Using more sophisticated tools (e.g. mind maps, concept maps) to better organize this multimedia information (texts, graphics, data, etc.) should be considered in relation with the model users. If the simulation of actual cases of farms is interesting in view of advising individual farmers, reasoning on farm types can be useful for the purpose of supporting agricultural advisors or professional and public policy-makers to develop general scope alternatives at a micro-regional level. However, a too short experience in trying to elucidate the place of models in a decision process and finding the way to capitalize the knowledge gained from simulations has led, eventually, to consider cooperating with "real" researchers in management science, ergonomics or knowledge engineering to tackle these issues that are far from trivial.

8. SUMMARY

A way to improve the sustainability of agriculture is to design new management policies of agricultural production systems based on the integration of heterogeneous knowledge on their biophysical and human components. Simulation models, representing those systems as productive units spread over a territory, have been designed to assess these systems performance against agronomical and environmental criteria and, so doing, help design new management policies. Beyond the various simulation models realized to date, the main achievements were pointed out: a comprehensive approach and a conceptual framework for modeling and analyzing material flows on a territory scale; the challenge of the 'situated action' theory to represent human action in farming systems. Two incompletely resolved issues were also pinpointed: assessing the impacts of management policies at various scales and setting the practical ways to use simulation modeling with agricultural stakeholders. The research avenues that are thought of were also underlined: complete a generic modeling framework of human activity in agricultural systems, namely, by introducing the spatial dimension of action in addition to its temporal one; decide on the relevant temporal and spatial scales for assessing the sustainability of these systems and the related representational scale of the models used to simulate them; find the practical ways to use the models with agricultural stakeholders in decision making and capitalize the knowledge gained from practicing simulation.

9. REFERENCES

[1] C. Aubry, J.-M. Paillat, F. Guerrin, "A conceptual model of animal wastes management at the farm scale. The case of the Reunion Island", **Agricultural Systems**, Vol. 88, 2006, pp. 294-315.

[2] P. Checkland, **Soft systems methodology: A 30-year retrospective**, Chichester: Wiley, 1999.

[3] W. Clancey, "Simulating activities: relating motives, deliberation, and attentive coordination", **Journal of Cognitive Systems Research**, Vol. 3, 2002, pp. 471-499.

[4] R. Courdier, F. Guerrin, F.-H. Andriamasinoro, J.-M. Paillat, "Agent-based simulation of complex systems: Application to collective management of animal wastes", **Journal of Artificial Societies and Social Simulation**, //jasss.soc.surrey.ac.uk/5/3/4.html, 2002.

[5] F. Guerrin, "Magma: A model to help manage animal wastes at the farm level", **Computers and Electronics in Agriculture**, Vol. 33, No. 1, 2001, pp. 35-54.

[6] F. Guerrin, "Simulation of stock control policies in a two-stage production system. Application to pig slurry management involving multiple farms", **Computers and Electronics in Agriculture**, Vol. 45, No. 1-3, 2004, pp. 27-50.

[7] F. Guerrin, "Modelling agricultural production systems using an action-flow-stock ontology", **MAS 2008 International Workshop on Modelling and Applied Simulation**, Amantea, Italy, 2008a.

[8] F. Guerrin, "Modelling animal farming systems as supply chains", **18th Triennial Conference of the International Federation of Operational Research Societies (IFORS)**, Sandton, South Africa, 2008b.

[9] F. Guerrin, "Dynamic simulation of action at operations level", **Journal of Autonomous Agents and Multi-Agent Systems**, Vol. 18, No. 1, 2009, pp. 156-185.

[10] F. Guerrin, J.-M. Médoc, "A simulation approach to evaluate supply policies of a pig slurry treatment plant by multiple farms", **Joint Efita 5th Conference and 3rd World Congress on Computers in Agriculture and Natural Resources**, Vila Real, Portugal, 2005.

[11] F. Guerrin, J.-M. Médoc, J.-M. Paillat, "Mafate: modelling and analysing matter flows on a territory scale", **Cirad 2007**, 2008, pp. 48-49.

[12] F. Guerrin, J.-M. Paillat, "Combining individual and collective management of animal manure to reduce environmental impacts on a territory scale", **Modsim 2011 International Congress on Modelling and Simulation**, Perth, Australia, 2011.

[13] A. Hélias, F. Guerrin, J.-P. Steyer, "Using timed automata and model-checking to simulate material flow in agricultural production systems. Application to animal waste management", **Computers and Electronics in Agriculture**, Vol. 63, No. 2, 2008, pp.183-192.

[14] R. Johnston, M. Brennan, "Planning or organizing: the implications of theories of activity for management of operations", **Omega International Journal of Management Science**, Vol. 24, No. 4, 1996, pp. 367-384.

[15] S. Lopez-Ridaura, H.M.G. van der Werf, J.-M. Paillat, F. Guerrin, "Environmental analysis of agricultural systems: Coupling dynamic simulation models with Life cycle assessment", **8th International Conference on EcoBalance**, Tokyo, Japan, 2008.

[16] R. McCown, "Changing systems for supporting farmers' decisions: problems, paradigms and prospects", **Agricultural Systems**, Vol. 74, No. 1, 2002, pp. 179-220.

[17] R. Martin-Clouaire, J.-P. Rellier, "Modelling and simulating work practices in agriculture", **International Journal on Metadata, Semantics and Ontologies**, Vol. 4, No. 1-2, 2009, pp. 42-53.

[18] J.-M. Médoc, F. Guerrin, R. Courdier, J.-M. Paillat, "A multi-modelling approach to help agricultural stakeholders design animal wastes management strategies in the Reunion Island", **iEMSs 2004 International Environmental Modelling and Software Society Congress**, Osnabrück, Germany, 2004.

[19] T. Morvan, P. Leterme, "Vers une prévision opérationnelle des flux de N résultant de l'épandage de lisier : paramétrage d'un modèle dynamique de simulation des transformations de l'azote des lisiers (STAL)", **Ingénieries**, Vol. 26, 2001, pp. 17-26.

[20] J.-M. Paillat, F. Guerrin, J.-M. Médoc, C. Aubry, "Simulation de stratégies de gestion de matières organiques avec le modèle Magma. Application au cas d'une exploitation type", in F. Guerrin & J.-M. Paillat (eds), **Actes du séminaire de restitution de l'ATP 99/60** (cederom), Cirad, Montpellier, France, 2003.

[21] J.-M. Paillat, S. Lopez-Ridaura, F. Guerrin, H.M.G. van der Werf, T. Morvan, P. Leterme, "Simulation de la faisabilité d'un plan d'épandage de lisier de porc et conséquences sur les émissions gazeuses au stockage et à l'épandage", **Journées Recherche Porcine**, Vol. 41, 2009, pp. 271-276.

[22] L. Suchman, **Plans and situated actions: The problem of human-machine communication**, Cambridge University Press, 1987.

[23] J. Vayssières, F. Guerrin, J.-M. Paillat, P. Lecomte, "Gamede: A global activity model for evaluating the sustainability of dairy enterprises. Part I. Whole-farm dynamic model", **Agricultural Systems**, Vol. 101, 2009, pp. 128-138.

The Sustainable Engineering Design Model: Necessity or Luxury

By Anthony D. Johnson BSc(Hons) M.I.Mech.E, C.Eng,
Andrew G. Gibson BSc (Hons) DipM MIEx and
Dr. S.M.Barrans BSc(Hons) F.I.Mech.E, C.Eng

Abstract

Sustainability in the field of the design of the built environment has been successfully applied for thousands of years, where materials have been reused and recycled. More recently there has been a great emphasis on sustainability in the field of geographic sciences.

Engineering design is a vast subject covering an enormous range of disciplines, but sustainability issues have rarely been applied to engineering design.

This paper outlines the normally accepted general design model and proposes a model for sustainability as applied to mechanical engineering design. Issues such as sustainable sourcing of materials, ecological design approach, sustainable use of new equipment and sustainable decommissioning using the 4r approach are all explored.

Taguchi proposed that the quality of engineering products could be defined at the design stage rather than at the manufacturing stage. The same is true of the application of sustainability where engineering designers should instigate sustainable engineering in new designs. Furthermore, correctly applied sustainable design techniques will reduce costs and improve the Triple Bottom Line.

The model proposes that mechanical engineering designers apply sustainable design techniques (4R's) to new equipment design. The 4R approach is outlined as follows:

Reduce: Designing products that reduce consumption by utilising resources efficiently, both inherently and in terms of energy utilisation is the single easiest way to reduce business costs as well as one of the best means by which to improve ecological credibility.

Reuse: The manufacturing of goods is hugely resource intensive. Designing products that have extended life spans, are upgradable or refurbishable allows for the optimum use of resources.

Sustainability Past and Present

It seems that sustainability can be considered to be a "hot air" subject. There is much written but results seem to be somewhat lacking. There are, however, some individuals as well as organisations that have embraced sustainability for many years and are very successful. There is much work to do to educate and empower the majority of people and organisations into adopting practical mechanical engineering sustainability.

Architects and builders have long since built sustainable structures. Even early man built dwellings that were self sustaining. There are many modern examples of sustainability in the built environment. Perhaps some of the better examples can be found in the recycling of building materials. Plates 1 and 2 below show the reuse of building materials applied to the Citadel Walls in Ankara, Turkey. This can perhaps be described as an over enthusiastic reuse of building materials.

Pates 1 and 2: Examples of the Reuse of building Materials in the Citadel Walls, Ankara, Turkey

The Geophysical environment has also been active in the application of sustainability projects. Beach groynes are an excellent example of sustainability of coastline. Plate 3 shows beach groynes in place in Bournemouth, UK. These wooden structures are built like fingers out in to the sea perpendicular to the shore, thus preventing long shore drift and preserving the shore line.

Plate 3: Beach Groynes, Bournmouth UK

There are some excellent examples of sustainability in Mechanical Engineering, however it could be argued that not enough is being done since most of the sustainability focus is applied to recycling.

Classic Engineering Design and Manufacture Model

The classic engineering design model has a natural extension of manufacturing, component use and eventual disposal. This model has been used for thousands of years and though some elements of sustainability have been practiced, the mind set and pressure on many modern engineers is to design the job so that it can be manufactured to a low cost. See Figure 1.

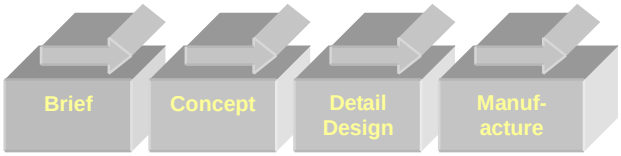


Figure 1: Classic Design and Manufacture Model

Sustainability in Mechanical Engineering has been practiced in some forms for years but it could be argued in a limited sense. Recycling of steel is well practiced as is the recycling of plastics. Steel recycling in the US, figure 2, has reached some dizzying heights reaching recycling rates 103% of output in 2009 [2], though averages are nearer 80%.

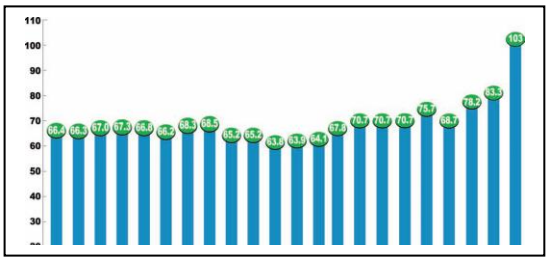


Figure 2: Overall Steel Recycling Rate (Steel Recycling Institute)

Vehicle manufacturers normally aim for 90% recycling of vehicles at the end of their life. This includes plastics and other components.. In the US steel recycling from vehicles reached rates of 121% of output in 2009, with averages close to 100% [2].

Vehicle manufacturers have also been attempting to design and manufacture the sustainable use of vehicles by optimising new designs to reduce mass, thereby reducing fuel consumption.

Power plants have also been the target of Design Engineers' creativity in the development of leaner internal combustion engines and more specifically the development of hydrogen engines and electrically driven vehicles.

In terms of sustainable mechanical engineering the question has to be asked, "Is this enough?" It is a fact that engineers are still stripping resources from the earth at an alarming rate. Worldwide steel production is around 127.5 million Tonnes [5] with only around 70million tonnes being recycled. Steel however is the world's most recycled commodity. If Design and Manufacture Engineers are to have a conscience the answer must be "We can do more!"

Many global companies have sustainable policies in place but each has its own approach and its own agenda. There is a great need for a cohesive and coherent approach so that all designers can work towards similar goals and have a significant effect.

Taguchi, a manufacturing and statistical engineer whose most significant work was publicised in the 1950's and 1960's expounded theories relating to the quality of manufactured goods. He noticed that the quality of manufactured goods was usually left to the manufacturing craftsmen to achieve. Taguchi suggested that quality could be designed-in before components were manufactured. He also recognised that to achieve designed-in quality the mind set of the designers had to be changed.

Similarly the Smallpiece Trust, set up in the 1960s aimed to change the mind-set of designers to embrace simplicity of design and efficiency of materials usage.

As designers were eventually made more aware of the "new" design method, manufactured goods became more available at a lower cost and higher quality. Engineering designers had created better components by introducing standards, tolerances, machine tools, which lead to high production methods.

The Taguchi analogy can be applied to Sustainable Engineering Design (SED). Sustainability has to begin with the designer who creates the products. It is he who is the key and who must envisage and design components using sustainable techniques, equipment and methods.

It must be acknowledged that although some designers have sustainability in the forefront of their design practice the majority of designers may only pay lip service to SED. In order to achieve true SED the design mind set has to be modified.

The Engineering Design Process can no longer be related purely to cost benefits. The Engineering Design process must now accommodate cost and sustainability.

It is proposed to offer a new design and manufacture model that combines the original model with sustainable issues.

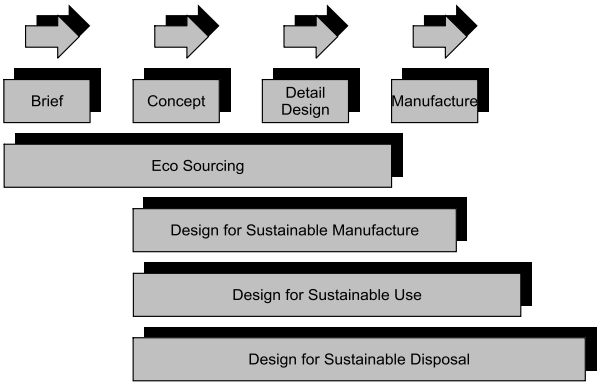


Figure 3: Sustainable Engineering Design Model

It was only a few years ago that reduction in pollution was the watchword for environmentalists. This was a crude yardstick for what has now become sustainability. Within the engineering design environment the emphasis on pollution has now evolved into a fairly detailed approach under the new label of Sustainability. This now encompasses the whole process from sourcing to disposal.

Sustainable Sourcing (Eco Sourcing)

Transportation: It is inevitable that raw materials will always be hewn from the ground and then transported to the processing point using fossil fuels to provide the energy to propel the transporters. This practice is often over very long distances.

The current common practice is for western organisations to source products in the Western Pacific Rim; China, Japan, Korea. This is largely done on the basis of reduced cost. It should be remembered that the environmental impact of producing these goods is roughly similar in the Pacific Rim as it would be in the west. The real impact on the planet's resources is in the burning of fossil fuels used in the energy generated for transportation. Responsible sourcing would mean that designers would specify local suppliers, thus reducing the environmental impact of transporting goods long distances. An added benefit is that local industries would thrive.

Some commodities have to be transported since they are available only in another part of the world. In such cases the question would be "how can this transport be arranged in a sustainable way?" Sailing ships could be employed or perhaps modern sailing versions that used the natural elements for propulsion. This is no pipe dream. Examples such as the MV Beluga (see fig 4 below) have shown that the wind can be harnessed to provide part of the necessary propulsive power for large modern freighters. For example, Gerd Wessels of Wessels Reederei says "There is enormous free wind-energy potential on the high seas. With *Skysails* [7] we can reduce energy by 50% on a good day, giving at least 15% annual fuel savings."

Flettner rotors [8], rotating sails, have also been fitted to freighters with some success.

Figure 4: Skysails MV "Beluga"

Courtesy of Skysails

Similarly, Eco Marine Power [9] are already designing solar powered craft from small ferries to freighters. Figure 5 shows Eco Marine Power solar ferry "Medaka". Critics may scoff at using sail or solar power for freight but what would we use if there were no more fossil fuels?

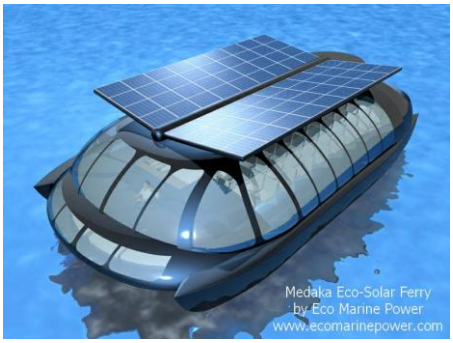


Figure 5: Solar Power Ferry "Medaka"

Courtesy of Eco Marine Power

Techniques: Techniques could be changed to accommodate processes that gave a sustainable benefit over current techniques. An emerging technology is that of rapid prototyping. This has grown alongside the development of 3D computer models and has usually been associated with the 3D printing of actual sized plastic models.

Techniques are now being developed which create 3D components using laser fused metal powders. This technique effectively reduces time to manufacture and reduces transport costs and environmental impact to almost zero since the component can be formed with no waste at the assembly plant.

Managed Source: All raw materials should be labelled with a sustainable source value (SSV). The main feature of this would be to inform the designer of the environmental cost of the raw material. This may seem a tall order but the system already exists in the form of managed exotic timber, which carries a certificate of authenticity of sustainable sourcing. With such a system in place designers could select a material according to its sustainable impact.

Material flow systems – open and closed loop: This concept, introduced in the 1990s is now being embraced by, amongst others, the EU. Joke Schauvlike, [11] President of the EU Environmental Council is of the opinion that "We must deal with our materials, and with our energy, more efficiently. At the end of their life we must be able to reuse materials as new raw materials. This is called completing the cycle." This approach was discussed in economic and energy terms by Clift and Allwood under the title "Rethinking the Economy" [12].

Taking the matter further, we can see that the present linear materials flow systems model, figure 6 is not sustainable over the longer term as manufacturers take no account of the issues of raw material extraction and transport discussed above, nor of the end-of-life issues once the product is no longer usable or obsolete.

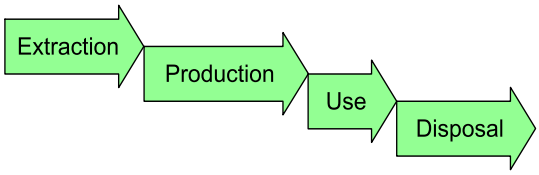


Figure 6: Linear Flow Systems Model

In the closed-loop flow system model, figure 7, materials and components would be recovered and reused reducing material inputs and outputs as close as possible to zero. This produces a *hierarchy* of sustainable end-of life techniques.

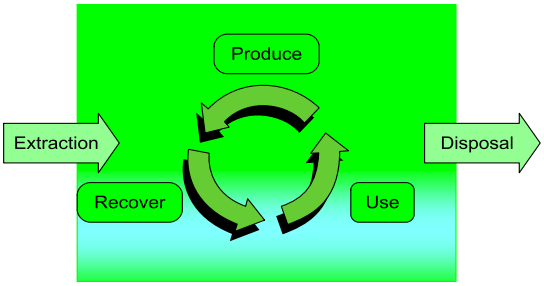


Figure 7: Closed Loop Flow Systems Model

Recycle, refurbish, re-use: Re-used and refurbished products and materials use even less energy in restoration. Designing machinery and equipment such that it can be repaired and refurbished became less common as companies sought to save cost by reducing the labour content and by moving production to cheaper labour sources. In line with the thinking of Stahel of the Product Life Institute in Geneva, It is our contention that legislation, lobby pressure and tax-based initiatives will drive a resurgence in equipment designed for ease of refurbishment and re-use, and that forward thinking producers will use a positive marketing message similar to campaigns such as Fair Trade will begin to place a premium on sustainably designed products. Stahel explores how moving from disposable products to service delivery could lead to restructuring of a post-industrial economy. Energy use would partly be substituted by labour, mainly skilled labour, as re-engineering substitutes for primary material demand. Activities which are labour- rather than capital-intensive are less subject to the economies of scale which characterise the chemical and material industries. Thus Stahel's concept of the performance economy also embraces more localisation of economic activity under the maxim "Do not repair what is not broken, do not remanufacture something that can be repaired, do not recycle a product that can be remanufactured"

Reduce: This is perhaps the oldest of the sustainable design techniques. By optimising design, the use of materials and hence the energy associated with transportation etc. can be

reduced. This has been seen particularly in packaging improvements, where a revised approach can improve the material usage efficiency without compromising on product life, safety or security

Designers Duty; It is the duty of the designer to source materials from sustainable sources or at least from sources which have a reduced impact on the planets resources. This emphasis would reduce the Sustainable Source Value (SSV)

Design for Sustainable Manufacture

The designer or design team is the entity who selects the manufacturing process. He can chose processes, techniques and materials. The formation of components has for centuries mostly relied upon the removal of material to create a shape. This process results in a great deal of waste both in material and energy required to remove the material. Casting components defines the shape with much less waste, but even this process requires a great deal of energy to produce molten material and then machine to final size. Sometimes the energy expensive process cannot be avoided but the designer should fix his gaze on reducing energy and material waste. Focus should also be directed towards the selection of materials that can be processed easily and select processes that are not energy hungry.

Sustainable manufacture may not always be focussed at the component. Factories can improve their energy usage. The use of natural light, intelligent building management systems, recycled waste disposal, LED lighting, rainwater harvesting, use of bio-waste for the generation of biogas are all ways in which manufacturing plants can better their Sustainable manufacture Value (SMV) These achievements were implemented by the Brandix Group in Sri Lanka [10] who converted their thirty year old factory to meet “green” factory standards. Brandix achieved 80% carbon emissions reduction and 46% energy reduction.

Packaging reduction and the use of recycled materials is also a major method of improving the SMV. Mattel, the toy company has for some time been instrumental in reducing the source fibre for its packaging and has reduced the amount of packaging thus not only saving on cost but improving the SMV.

Design for Sustainable Use

For certain classes of machinery and equipment, this is arguably the element in a product’s life which has the most impact on sustainability.

In the field of construction equipment and road transport, the energy consumed by a machine in use during its lifetime far outweighs the energy consumed in its production. JCB, for example have optimised the design of their machines over the years to use lean-burn diesels, minimising engine size and emissions by using flywheels to reduce peak demand. This is also the approach adopted by Caterpillar Industrial power Systems under Gwynne Henricks. At the CEA conference 2011, Ms Hendricks clearly indicated that design for sustainable usage and extended product life cycle was the key challenge facing the industry as it developed new products.

Similarly, Emerging technologies have allowed radical improvements in electric car charging, reducing the use of low-efficiency internal combustion engines at least for short journeys.

Designers have to take responsibility for the impact of their equipment on the environment. The complete non-use of fossil fuel power may not be practically achieved but it may be significantly reduced. This can be done in several ways:

Design Optimisation: One optimised aspect of vehicle design is that of reduced weight. Structures are optimised for strength; buildings are optimised to accommodate earthquake oscillations. Optimisation can also be applied to sustainable use in selecting appropriate power systems and methodology in use that improves the Sustainable Use Value (SUV).

Incorporate equipment that gives back: Emerging and young technologies such as solar power and wind power can easily be incorporated into many products. New build houses for instance could incorporate solar panels (PV panels) on the roof. Vehicles could also be fitted with PV panels and make use of the air they disturb in travelling by incorporating micro wind generators.

Reduce energy usage: There are many options here. A few are listed: Design equipment which is lighter in weight, apply smaller power units, specify leaner power units, use energy from renewable power sources, insulate against heat loss.

Use of natural energy: Power is the driver for any usage process. It makes sense to use naturally generated power and low energy solutions where possible. It is the designer’s role to select the lowest energy option and to design that option in to the new products thereby improving the Sustainable Usage Value (SUV). One of the many ways this could be done is by applying electrical drive units such as those in electrically powered vehicles and other transport vessels. Hydrogen engines, whist still in their infancy, have a zero impact on the environment.

Energy Storage: No matter how clever the application it is inevitable that there will always be “take” from the environment. Devices have to be built therefore which have the capacity or to generate energy for those processes that demand it. Some of these devices which actually generate energy are dealt with elsewhere, however, energy storage must be considered. Large chemical batteries are useful and efficient in use, though their manufacture and eventual disposal can take a heavy toll on natural resources.

An alternative to chemical storage is the use of of Kinetic Energy Storage Devices. These devices are essentially flywheels which rotate at high speed storing kinetic energy which can then be converted back through generators into usable electricity. Flywheel batteries can be very high tech systems requiring a significant amount of manufacturing resource. Other systems can be low tech, made to normal engineering principles, which demand fewer resources from the environment for manufacture A version of the low-tech device is currently being developed at the University of Huddersfield in conjunction with ESP Ltd.

Whichever method is adopted the resources used in its manufacture are given back during the use of the device. These storage devices are able to accommodate energy when there is low demand and introduce it back to the grid when there is high demand. This evens out the electricity demand and reduces the necessity to generate electricity on demand, thus reducing resources required for generation.

In a different application a large bank of flywheel batteries could store the output from a several power stations when demand is low and return it to the grid during high demand periods, the requirement for power stations would be reduced.

The emerging technology of electrically powered vehicles requires not only infrastructure but also a quick means of recharging. A bank of flywheels in strategic locations would provide that means, perhaps domestically or whilst parked at the store.

Energy storage devices possess a very low Sustainable Use Value (SUV) simply because the resources needed to create them is more than compensated during the life of the device.

Design for Sustainable Disposal

The designer is the creator of the product and has the power to create a sustainably friendly disposal technique. The designers mind set is always to reduce cost and should now refocus slightly into sustainable disposal. There are several ways that a product at the end of its life may be utilised or disposed of in a sustainable way:

Recycling and re-use: Thus far the material sourcing that has been considered has been from an original source, however this need not be the case since materials can be gleaned from several other sources. Perhaps the most obvious source is from recycled materials.

Some materials, such as building materials have been a recycled source for thousands of years. In more recent years steel has been successfully recycled and is now the world's most recycled material. There has also been a surge in the variety and diversity of recycled materials. These include: shoes and clothes, electrical appliances, glass, non-ferrous metals, vehicle tyres. It is estimated [6] that up to 90% of discarded items and products can be recycled or reused. Materials gleaned from recycling processes are less costly and use less energy than the original source material.

Refurbish / Repair: Die cast components and products were the norm in the 1950's. Items were held together with screws and could be dismantled and repaired. During the early 1960's the advent of plastics became the popular use for toys, kitchen implements, garden tools, household devices and many other products. These were normally "snap-together" and were almost impossible to dismantle without breaking the product and hence difficult to repair. It was the beginning of the "throw-away society." The mind set of throw away and buy another has to some extent started to turn towards refurbish and re-use.

Refurbishment means that products are not thrown away but restored so that the product's life can be extended. The current recession has focussed companies into refurbishing components rather than buying new. The civil engineering industry in the UK has been hit hard by the recent recession which means fewer building projects and fewer purchases of new items of plant and ancillary equipment.

A West Yorkshire manufacturer of brick and block crane attachments has found a lucrative market in refurbishing equipment and supplying spares as the new-equipment market has evaporated. Figure 8 shows a typical brick /block crane attachment. This is a product which may be welded if it breaks and components replaced when worn. It can be restored to a working product with much less input and with a much smaller impact on resources than a new product. This is an excellent example of extended life giving a very low Sustainable Disposal Value (SDV)

Figure 8: Brick/Block Clamp

Courtesy of HE&A Ltd

An excellent example of repair is that of motorcycles used in India and Pakistan. In these countries the favoured individual transport is the 70cc or 100cc motorcycle shown in figure 9.

Figure 9: 70cc Motorcycle

Here the designers have taken the initiative and designed a vehicle with a low resource impact value. These motorcycles are designed to have simple parts, be low cost, easy to repair and have a relatively low impact on resources when manufactured and also in use. They can be refurbished as long as parts are available. This is an excellent example of low SUV.

Give-back

No matter how products are re-used, refurbished or recycled or how clever the usage impact is reduced the plain fact is that the usage of resources is being merely being slowed. There will always have to be some form of "take" from the earth's resources.

Give-Back is a technique where designers actually build devices which give back to the environment or perhaps design give back components in to current products.

Solar power panels on car roofs, micro wind generators built into vehicles are just two ideas that could be explored. Most vehicles are left outside for much of their life. PV panels set into the roof and built-in micro wind generators could produce energy which could then be stored. Trading this stored energy in to a central repository could give discount to the liquid fuel or the recharging of the vehicle. In another application solar panels could be incorporated on the roof of new buildings.

Imagine the energy generation possibilities that solar panels would make if introduced on the millions of homes in the UK. The power generation would be enormous and reduce the reliance on power stations.

The energy thus generated needs to be stored. As part of collaboration between the University of Huddersfield, UK and ESP Ltd a battery flywheel system is being developed shown in figure 10. These Kinetic Energy Storage Devices have the capacity to store large amounts of energy in a spinning flywheel. At peak times these devices can be tapped to provide the electricity grid with power. Several thousand of these in a single facility could provide enough storage capacity to eliminate a power station. This is an excellent example of low Sustainable Give Back Value (SGBV)

Figure 10: Idealised Flywheel Battery 20KWh Storage

An excellent example of Sustainable Give Back Value (SGBV) is the World Trade Centre Building in Bahrain, figure 11. This building incorporates wind generators. The building rises 240m. The shape of the towers is designed to funnel the wind on to the wind turbines generating 675KW in total which is up to 15% of the total power consumption of the building.

Figure 11: Bahrain World Trade Centre with Wind Turbines

General Overview Tool Requirement.

Carbon dioxide is always painted as the ogre of greenhouse gases, but it is not the largest constituent. The proportions of the main greenhouse gases are as follows [4]:

Water vapour	36-70%
Carbon Dioxide	9-26%
Methane	4-9%
Ozone	3-7%

It can be seen that moisture is the biggest greenhouse gas yet the common assumption is that it is harmless. As a liquid, water is life giving but as moisture in the atmosphere it prevents sunlight from reaching the earth and also acts as a blanket, keeping heat within the envelope of clouds.

Though carbon is certainly a large proportion of the whole of greenhouse gases perhaps the question should be asked “why is carbon dioxide used as an environmental yardstick?” The focus on carbon dioxide cannot be considered as wrong since it is a useful tool but it does not really cover the whole picture

When reviewing the sustainable design options it is clear that the conversion of materials into products uses energy. One of the by-products of the energy usage is carbon dioxide however **energy usage** is a much clearer indicator of resource usage and is a much better indicator of the environmental impact.

Energy is used in one form or another to extract raw materials, convert them in to products, drive them during their useful life and dispose of them at the end of their life.

It is therefore proposed that each element of the sourcing-conversion-use-disposal process is given a Sustainable Value. These elements are mentioned above and can now be assembled in to a tool by which designers and environmentalists can judge the sustainable impact of a product. The definitions are listed below:

SSV: Sustainable Source Value

SMV: Sustainable Manufacturing Value

SUV: Sustainable Use Value

SDV: Sustainable Disposal Value

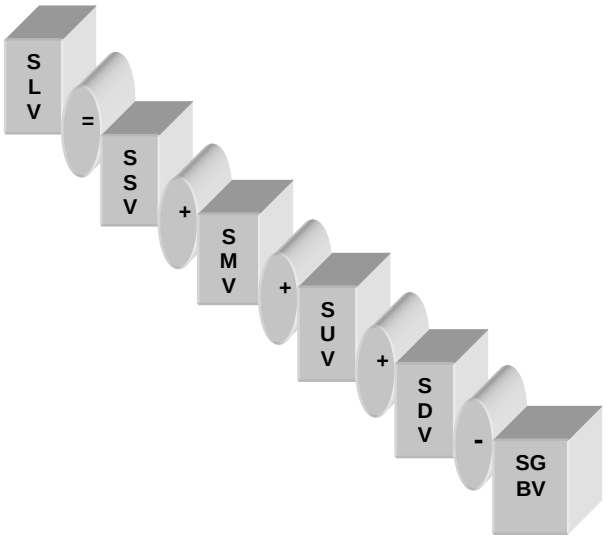
SGBV: Sustainable Give-Back Value

These indicator values should be kept as low as possible since the lower the value, the lower the impact on the earth’s resources.

These values combined below give the overall sustainable impact tool or “Sustainable Life Value” or SLV.

SLV is derived from the addition of SSV, SMV, SUV and SDV and is a measure of the resource impact during the life of a product.

SGBV is a measure of how much resource is returned and can therefore be deducted from the resource impact (sustainable Life Value: SLV) as the model shows in figure 12.



Conclusions

In this world-wide consumer society it is inevitable that products will be produced used and disposed of, but every product reduces the earth's resources. Perhaps this process cannot be prevented, but it can be slowed substantially.

The task falls to our Engineering Designers and Engineering Design teams to instigate a change in attitude and approach towards Sustainable Engineering Design. As in the Taguchi model for quality engineering it can only be the designer who can instigate the shift in attitude and the change in design practice. Design is the key to Sustainable Engineering

The model outlined above reduces transport distances reduces manufacturing resources reuses, recycles and repairs goods. Adopting these elements will lead to a localised economy model reducing financial costs and providing local, specialist employment.

It can also be seen that an element of "Give-Back" will help to reduce the impact of a product by reducing the Sustainable Life Value (SLV) and that the implementation of give-back technologies such as flywheel batteries can greatly offset the impact products may have on earth's resources.

The use of carbon foot printing is useful but is viewed as a commercial tool only and does not embrace the whole life cycle of the product from sourcing to disposal. It is proposed that a much more useful and accurate measurement is energy usage since this is involved at every stage of the products life.

The introduction of a measurement tool, Sustainable Life Value (SLV) is a great step forward in the measurement of resource impact. There is some work yet to accomplish to ensure this measurement tool becomes viable.

Sustainable Engineering Design: Necessity or Luxury

The earth's resources are dwindling at a heavy rate and over the last 30 years in particular there has been much debate and suggestion as to how the use of resources could be reduced. A few lone practitioners began to act, but still the extravagance continued. Recent years has seen increased climate change and a greater thrust towards sustainability. Recently The Institution of Mechanical Engineers has urged its members and partner institutions to adopt a sustainable approach. The initiative has begun to focus the minds of Mechanical Engineers and initiated action. Many Global Companies are developing their own strategy but there is a great need for a cohesive and coherent approach so that all designers can work towards similar goals and have a significant effect. It is the proposal of this paper to propose a complete model from sustainable sourcing to sustainable disposal and further suggest a much needed measurement method, that of Sustainable Life Value.

Bibliography

[1] American Iron and Steel Association: www.steel.org

[2] Steel Recycling Institute: www.recycle-steel.org

[3] The Street: www.thestreet.com Mattel: 03/10/11

[4] U.S. Environmental Protection Agency, Greenhouse gas emissions,

http://www.epa.gov/climatechange/emissions/index.html

[5] ISSB Monthly World Steel Production Review: www.steel onthenet.com 6/10/11

[6] City of London MBC: www.cityoflondon.gov.uk

[7] Skysails GmbH www.skysails.info

[8] German Aerospace research Council (DLR) www.dlr.de

[9] Eco Marine Power: http://www.ecomarinepower.com/en/solar-ferry-medaka

[10] Towards a Sustainable Industrial System: University of Cambridge and Cranfield University: ISBN: 978-1-902546-80-3

[11] Ellen MacArthur Foundation http://www.ellenmacarthurfoundation.org/

[12] Rethinking the Economy Clift & Allwood, published in The Chemical Engineer Mar20

Modeling of Agile Avionics Software Development Processes through the Application of an Executable Process Framework

Patrick KINGSBURY, André WINDISCH
Cassidian Air Systems
Rechliner Strasse
85077 Manching, Germany

and

Wolfram HARDT
Chemnitz University of Technology
Department of Computer Science
09107 Chemnitz, Germany

ABSTRACT

In software intensive avionics projects the problem of missing adherence to the complex process landscape has been known for decades. This problem is significantly aggravated when the combination of business needs, such as improving productivity and responsiveness to technical changes are required in addition.
The modeling of the Agile Avionics Software Development Processes through the Application of an Executable Process Framework shows first useful results in improving the situation of missing process adherence and is increasing transparency of process changes.

Keywords: Process Modeling, Complex Avionics System Software, Executable Process Framework, Agile Avionics Software development, BPMN2.0

1. INTRODUCTION

In today's avionics software development the survival and growth of business requires effective means to align organizational business objectives with software project management and software processes.

The continuous technological advancement of computer technology over the past decades is accompanied by a similar growth of the complexity of avionics systems which in turn caused an exponential increase of the complexity of aircraft software [9] as indicated by figure 1. For decades this increasing software complexity has been standing in strong contrast to the problem of insufficient or missing software process adherence in the complex avionics software engineering process landscape. The results are observed in many civil and military aviation programs leading to severe cost and schedule overruns. It's not that the software doesn't work; it's the traceability of the software [7], i.e. the proof that it has been developed according to the standards.

Furthermore, this problem is significantly aggravated in a competitive environment where improved productivity, faster time to market and better quality are required. Traditionally the approach to Avionics software development follows the waterfall lifecycle model as depicted by figure 2 which provides less development speed compared to the Agile lifecycle model shown in figure 3.

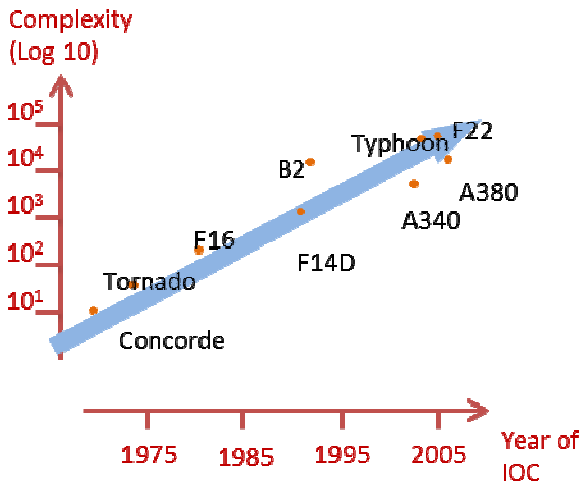


Figure 1: Increasing Complexity of Aircraft Avionics

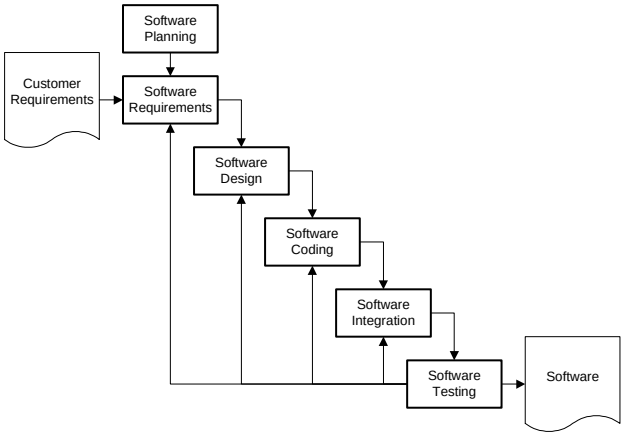


Figure 2: Waterfall Lifecycle Model

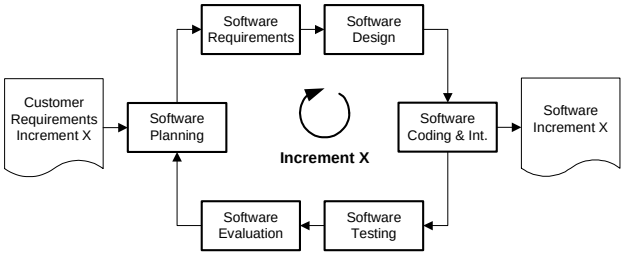


Figure 3: Agile Lifecycle Model

This paper describes the new idea to model the complete avionics software process landscape incorporating both development and certification standards with the Business Process Modeling Notation 2.0 (BPMN). The first result of this research project currently undertaken at Cassidian is the development of a BPMN 2.0 based process framework for the specification and deployment of complex agile avionics software engineering processes.

Besides its formal static semantics the BPMN standard also specifies execution semantics for the implementation of business processes in corporate IT infrastructures. In the context of our research project this feature has been used to deploy and execute the complex avionics software development process landscape transparently via web-browser in the complex software development environment.

Through the application of this process framework the introduction of the Agile Lifecycle Model for the avionics software development became feasible in the context of traceability for certification.

With this solution two new ideas are presented to the area of software engineering and process modeling: (a) to use the Business Process Modeling Notation (BPMN 2.0) for the formal specification of all software project management and software engineering processes and (b) to use a process engine to deploy and execute agile avionics software development processes.

The remainder of this paper is structured as follows: chapter 2 will provide a brief overview on related work before the executable process framework is discussed in chapter 3. Subsequently, in chapter 4 the formal process models for agile avionics software engineering and their application in the executable process framework are presented. Chapter 5 concludes with an outlook on future activities.

2. RELATED WORK

Today the area of process modeling and execution is mainly restricted to the business process level. In particular the new BPMN 2.0 standard has gained wide acceptance in industry. Several application areas have been reported, such as internal process management in large health care institutions [4], customer management [4] and process migration in telecommunications [5], customer support management in aerospace [6], and many others.

This acceptance in industry is based on the need for a common, cross-domain process standard which not only supports the modeling, but also the static verification of complex process landscapes and their deployment on enterprise IT-infrastructures. BPMN 2.0 fulfils all these requirements: its static semantics are formally defined by a UML-Metamodel and its execution semantics in terms of WEB services.

Although the new BPMN 2.0 standard explicitly lists engineering processes as a possible area of application, no references to Avionics software development processes could be found. One of the reasons could be, that traditionally the application of software process standards in industry is defined by a set of authorized planning documents (e.g. software development plan, software verification plan, etc.) which specify the individual processes, their inputs / outputs, and the process stages to be performed. Even though these processes are usually depicted in some graphical form, no formal process modeling is applied.

In the context of the EUREKA-ITEA AGILE Projects A. Wils et al [15] investigated the applicability of agile

methods to the embedded software domain. At a first glance the combination of agile development with certification of Avionics Software seems to be a contradiction, but it is feasible. However, no particular agile process solution was presented. No further publications on agile software development for avionics systems have been found.

3. THE EXECUTABLE AVIONICS SOFTWARE ENGINEERING PROCESS FRAMEWORK (EASE-P)

Framework Requirements

The lack of formal process modeling in the agile avionics software engineering domain manifests itself in inconsistent software process planning documents and insufficiencies in software engineering process adherence. The consequences are severe project delays, cost overruns, and quality problems - the most recent one being reported in [7]. From our experience the reasons for project failure are manifold, however, the top 3 addressed by this paper are:

(1) Inconsistencies in the software engineering process landscape
(2) Lack of adherence to software engineering processes or methods
(3) Insufficient project metrication, solely based on Earned Value Management (EVM)

The source of problem (1) is the complexity of the process landscape required for agile avionics software engineering. Figure 4 shows the landscape of the relevant processes of which most not only run in parallel but are also of a highly iterative nature. However, the resulting complex process interaction pattern are typically neither modeled nor verified. Instead, different process areas are defined by software planning documents which provide an informal picture of the processes, textually detail their activities, and describe their input / output relation with other processes.

The non-adherence to software engineering processes (2) cannot just cause major project delays but also endanger Avionics software certification. Although detailed textual descriptions of all processes and process stages exist, the overall complexity of the process landscape obstructs the situational awareness of the individual software engineers. This problem is intensified by the fact that engineers are typically assigned to one process stage only, e.g. software requirements analysis, and typically have very different educational backgrounds and skills.

The third reason for project failure addressed by this paper is the one-dimensional project measurement and control process implemented in most organizations. The standard approach is the utilization of Earned Value Management (EVM), i.e. project progress is measured in terms of man hours spent vs. values earned in terms of project milestones achieved. However, a project milestone does not denote quantitative and qualitative product information. To get a clear picture of productivity rates and product quality the EVM system has to be complemented with a product metrication process.

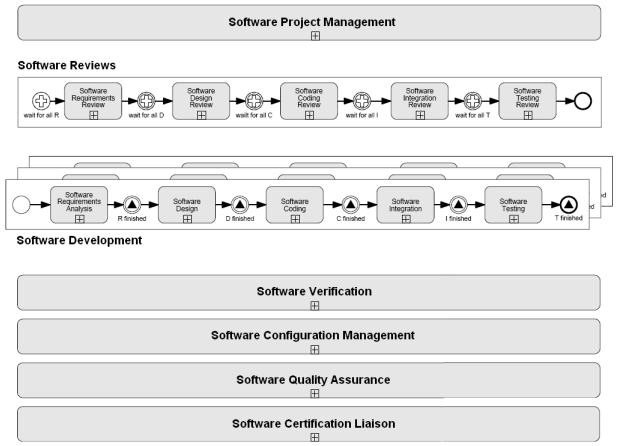


Figure 4: The agile avionics software process landscape

To eliminate the aforementioned deficiencies in Avionics software engineering a new approach to process modeling and execution is required. This new approach should

- Utilize formal process specifications which lend themselves to the application of formal verification techniques in order to eliminate process inconsistencies
- Provide explicit and graphical process guidance to increase the individual situational awareness and to reduce the impact of personal educational backgrounds and skill sets
- Support the implementation and integration of product metrication's to complement the traditional EVM based project control process.

Framework Concept

This paper proposes the new idea to adopt the business process modeling and execution approach to the domain of Avionics software project management and software engineering. All processes on the business, project, and

engineering levels are then specified in the unified formal notation BPMN 2.0. Based on the static semantics of this notation formal verification approaches, e.g. model checking, can be applied which enable the detection and elimination of inconsistencies in process interaction and process data exchange by simulation.

Besides this strong advantage the utilization of BPMN 2.0 offers additional benefits for Avionics software project management and Avionics software engineering. This paper proposes the new idea to deploy and execute the formally specified processes on a process execution engine such as jPBM [8]. However, this engine not only executes the individual project and engineering processes but also ensures overall process orchestration. Based on the graphical syntax of BPMN 2.0 both execution and orchestration of these processes can be graphically represented in a tool to provide an explicit visual guidance for the software engineers and to reduce the direct impact of personal educational background and skills. Moreover, for training purposes the process execution can be simulated for training projects. This allows for a seamless integration of training and engineering activities.

For avionics software project management the presented approach can be extended by integrating product metrication activities. This requires the specification and deployment of product metrication activities in the scope of the project measurement and control process and the implementation and integration of metrication procedures on the process execution engine. The invocation of these metrication procedures is then triggered whenever the corresponding activities of the project measurement and control process are executed. The gathered quantitative and qualitative product data – i.e. number of base lined requirements, implemented LoC, completed test procedures – provide a far more detailed project status than that solely based on EVM.

Framework Implementation

The conceptual ideas presented in the previous section were validated in the aerospace industry by means of an implementation prototype. The resulting EASE-P process framework combines existing tools, such as configuration management systems and task databases, with a new process execution engine and process visualization tools.

The overall tool architecture of the EASE-P process framework is depicted by figure 5. It utilizes a web-based client / server architecture where the jPBM process execution engine is integrated on the web server. The processes can be visualized and controlled via standard web browsers interactively. However, the type of interactions allowed is restricted for the different users depending on their role and responsibility in the project. The implementation of this approach is based on system access rights which also govern access to the framework tools. This ensures for instance that project metrication can be executed only from an account with project management rights. However, the same account is not permitted to introduce software configuration baselines or to check-in source code.

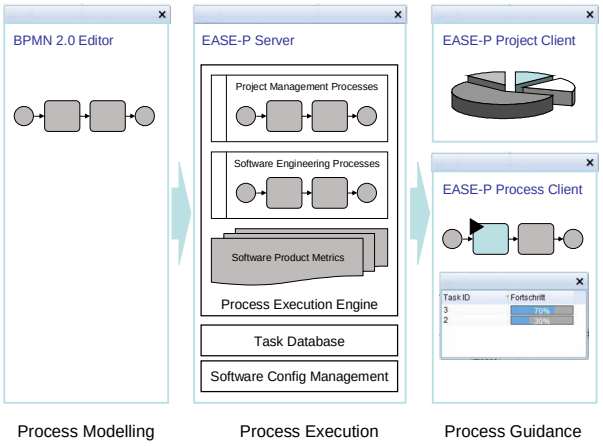


Figure 5: The executable avionics software engineering process (EASE-P) framework

The integration of the process execution engine and the existing tools is based on web-services. This approach is conforms to that used on business level and hence allows the future integration of the EASE-P framework into the enterprise IT-infrastructure.

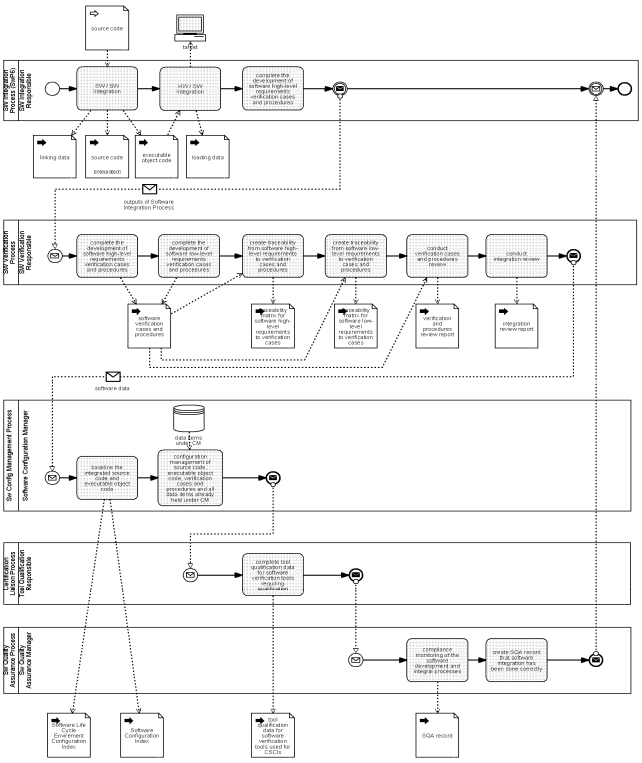


Figure 6: BPMN specification of software integration process [14]

4. AGILE AVIONICS SOFTEWARE ENGINEERING PROCESS FRAMEWORK

The implementation prototype of the EASE-P process framework was used to model and deploy the agile avionics software development processes as depicted by figure 4. Thereby, the top-level BPMN specification closely follows the planning document structure defined by the avionics software certification standard DO-178B [3]. In this model the agile software development process formally captures all activities usually described by the Software Development Plan (SDP). The same applies to software verification, quality assurance, configuration management, and certification liaison which are traditionally defined by the Software Verification Plan (SVP), the Software Quality Assurance Plan (SQAP), the Software Configuration Management Plan (SCMP), and the Plan for Software Aspects of Certification (PSAC).

The only exception to this rule is the separate specification of the software review activities which normally constitute one specific part of the software verification activities. However, this modification was necessary due to the fact that we allow multiple software development process instances to be executed in parallel in order to achieve a agile software development. Figure 7 shows the holistic view on the agile avionics software process realization.

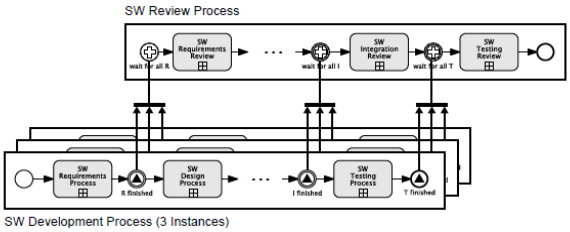


Figure 7: Holistic view on agile avionics software process [14]

In this context the explicit software review process is used to synchronize the parallel development phases before the formal software review is conducted, which is mandatory to fulfill the certification requirements for Agile avionics software development.

The presented modeling approach extensively uses BPMN process composition to roll-up the implementation details. As an example consider the software integration process shown as a single process box by figure 4. The formal specification of the detailed activities of this process is depicted by figure 6. In this BPMN specification the parallel execution of software processes is modeled by means of BPMN pools each of which encapsulates the process-specific sequence of activities. The information and data flow between these processes are modeled in terms of BPMN events which trigger and synchronize the internal activities of the concurrently executing processes.

The process models have been deployed on the EASE-P process framework which provides graphical process guidance to software project managers and software engineers.

5. CONCLUSION

This paper presented the two new ideas to implement agile software development processes for avionics software engineering: (a) to use the Business Process Modeling Notation (BPMN 2.0) for the formal specification of all software project management and software engineering processes and (b) to use a process engine to deploy and execute agile avionics software development processes.

Besides the strong advantage that BPMN 2.0 provides both a formal process specification semantics and an execution semantics the EASE-P Framework offers the following additional advantages:

1. Visualization of all software processes and their complex interaction to both software developers and project managers

2. Process guidance for all software developers through step-wise process execution to ensure subsequent avionics software certification

3. Situational awareness at each state for project managers through integration of metrication to ensure schedule adherence, productivity level and objective metrication based on development artifacts for Software Project Management which complements the traditional EVM approach

The presented approach shows that all necessary agile processes for the development of certifiable embedded Avionics software can be specified in BPMN 2.0 and integrated into EASE-P process framework. These processes and interactions have been based on the relevant standards for software development ISO/IEC 12207 und DO-178B.

REFERENCES

[1] Allweyer, Thomas. **BPMN 2.0 – Introduction to the Standard for Business Process Modelling.** Books on Demand GmbH, Norderstedt, Germany, 2009. ISBN 978-3-8391-4985-0

[2] International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC). **ISO/IEC 12207 – Information Technology – Software Life Cycle Processes.** 1995

[3] Radio Technical Commission for Aeronautics (RTCA). **DO-178B – Software Considerations in Airborne Systems and Equipment Certification.** 1992

[4] Signavio. **Introduction of a common process platform at AOK.** (In German). http://www.signavio.com/de/referenzen/kunden.html, 2010.

[5] Casewise Ltd. **How PBM and BPMS tools are supporting migration capabilities for Alcatel-Lucent.** http://www.casewise.com/NR/rdonlyres/45EF7DA6-55B5-4013-A2FF-F9ECBDEAE68F/0/alcatellucent.pdf, 2010.

[6] Interfacing Technologies Corporation. **Case Studies – Bombardier Aerospace.** http://interfacing.com/Literature/Business-process-case-studies, 2010.

[7] Reuters - **Brake software latest threat to Boeing 787 –** 15.07.08

[8] Red Hat. **Jboss jBPM Process Execution Engine**. http://www.jboss.org/, 2010.

[9] Robert A. Dietrick, Major, USAF: **Impact of Weapon System Complexity on Systems Acquisitions** - Air Command and Staff College Air University 2006

[10] From Wikipedia, the free encyclopedia - **Education**

[11] Silvia T. Acuna Universidad Autonoma de Madrid Spain and Natalia Juristo Universidad Politecnica de Madrid Spain - **Software Process Modeling**

[12] From Wikipedia, the free encyclopedia - **Productivity paradox**

[13] Berglas, Anthony (2008) - **Why it is Important that Software Projects Fail**

[14] F. Triebsch, Thesis - **Konzeption und Umsetzung eines Frameworks für die Modellierung, Ausführung und Metrikation von Agilen Prozessen in der Avioniksoftware-entwicklung**

[15] Wils, Andrew et al - **Agility in the avionics software world. K.U.Leuven DistriNet, 2006** http://www.agileitea.org/public/papers/agileavionics.pdf

[16] Beck, Kent et al - **Manifesto for Agile Software Development.** http://www.agilemanifesto.org**, (23.07.2010)**

[17] Sommerville, Ian – **Software Engineering 6th Edition**, Addison Wesley

Designing Contents for a Serious Game for Learning Computer Programming with Different Target Users

Danu Pranantha, Cai Luo, Francesco Bellotti, Alessandro de Gloria
Elios Lab, University of Genoa, DIBE Via Opera Pia 11A, Genoa, Italy 16145
Ph./Fax.: +39-010-3532795
{danu, cai.luo, franz, adg}@elios.unige.it

ABSTRACT

In recent years, all over the world, students are less interested in science, especially in computer science where the number of students is shrinking. Also, 50% or more students who initially choose computer science soon abandon the study. Hence, it is essential to attract students to learn computer science, in particular computer programming, via non conventional teaching approach. A promising medium is serious game for learning computer programming, given its fun gameplay characteristics. However, several games intended for computer programming are either hard to be extended, lack of dynamicity, or are not challenging which lead to boredom and lack of motivation in learning. On the other hand, computer science unplugged learning for primary-aged children via mathematical based activities using, for instance, cards and boards, is considered to be very compelling and improving conceptualization. Moreover, problem solving skill acquisition in form of puzzle based learning for university students are immediately pertinent to their problem solving skills development due to its attracting and intellectually challenging nature. Given these facts, this paper presents game contents targeted for different prior knowledge in programming which will be used in serious game for learning programming.

Keyword: Game Contents, Serious Games, Computer Aided Learning, Learning Computer Programming

1. INTRODUCTION

The advancement of information technology has brought many changes on how we work and live. Learning as central part of human activity is also affected by this innovation. Digital learning has been widely adopted for various objectives e.g. learning basic knowledge in early childhood development [1], learning mathematics in elementary school [2], and learning various subjects in universities [3, 4]. Prensky [5] stated that mostly people see learning as a painful activity that someone has to go through to acquire necessary knowledge and skills. Prensky [5] compared it with playing games. The clear distinction is that playing is a fun and engaging activity which gives motives to people to play it voluntarily. Therefore, Prensky strongly argued that conventional learning is outdated and the modern way of learning is through a real gameplay so that student will learn while having fun. More moderate views from Blunt[3] and Bellotti et al [6, 7], suggested the use of serious games as tools to support learning and to motivate students. According to Zyda [8], a serious game is "a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives". Therefore, any game built to differ from pure entertainment can be considered as a serious game. Serious games have been extensively studied in recent years for their ability to enhance general development such as logic, memory, problem solving [9, 10], induce motivation in learning and improve academic performance [3, 11]. Serious games are adopted in academic curriculum mostly in the area of management and economics [3]. Hence, there are still many benefits of their application which could be further investigated. For instance, the benefit in collaboration development [1, 11, 12], motivational factors [2, 13, 14, 15], contents selection and authoring [1, 6, 7], genre [14, 15], strategy and competition, behavior [10], personalized learning, etc.

2. BACKGROUND

In recent years, all over the world, students are less interested in science, especially in computer science where the number of students is shrinking. 50% or more students who initially choose computer science soon abandon the study [16]. Computer science curriculum 2008 of Association for Computing Machinery (ACM) mentioned that fluency in a programming language is prerequisite to the study of most of computer science [17]. Being the core of study in computer science, it is essential to introduce computer programming in an attractive style while keeping students interests on the subject after entering university. According to Bayliss [18], three approaches exist to motivate students to learn programming: (1) building novice programming environment, (2) introducing programming contest, (3) game.

Alice2 [19] is one of example of novice programming environment to teach programming using drag and drop environment. This allows users to learn logic and

programming structures without being involved in the syntaxes. Overmars [20] suggested GameMaker to develop simple games through drag and drop environment to assist students in understanding program structure. The second approach, programming contest, adopted in [21, 22] were able to motivate students to learn programming due to its competitive nature where dropout decreased from 72% to 45% and passing rate increased to double. The last approach makes the students play game to learn. An example is Colobot, a video game about colonizing a planet using robots that players have to program in specific object oriented (OO) language. However, the real engagement with rich experiences and fun in learning can be provided through playing games due to its gameplay nature [5]. As a matter of fact, exploitation of games as part of educational curriculum has been investigated in literatures.

Some works have already been done in introducing games to computer programming either as assignments [20, 23, 24, 25] or actually playing real games such as Role Playing Game (RPG), Real Time Strategy game, and 2D adventure game [13, 16, 26]. Muratet et al [16] developed a real time strategy (RTS) game about how to acquire resources and locate opponents using some given functions. Mohammed et al [26] developed an adventure 2D game that utilized local culture and provided interaction with the users via a cellular phone like graphical user interface (GUI). The GUI enables users to write code snippet per line until the required functionality achieved. Chang et al [13] built a mini and incomplete MORPG game with tasks in form of simple quests such as answering true or false, multiple choices, filling the blanks, and simple coding.

Game as assignments was done by Chen et al [24] by developing a restricted graphical like game of airplanes. Overmars et al [20] discussed the use of GameMaker as a development tool for teaching object oriented (OO) concepts, whereas [23] use 3D computer game like tool to teach computer animation using C-Sheep and Open GL as renderer. Sung et al [25] developed GTA (Game-Theme Programming Assignment) to teach introductory programming using Microsoft XNA. All of these works wanted to attract students to learn programming.

The work of Muratet et al [16] is good for training logics and competition. However, it only provides limited number of functions which can be used. Moreover, it has merely five missions which are inextensible. There is no mechanism to introduce new levels or other type of missions or tasks. Other work by Mohammed et al [26] provides a good interaction by providing cellular phone like programming GUI. However, there is only coding quest that is supported. New type of tasks, additional tasks and stories require some amount of efforts to be implemented. Moreover, students have to learn the game mechanics before playing the game which diverge students from learning the actual skill i.e. programming. MORPG game developed by Chang et al [13] suffers from boredom since it lacks of dynamicity within the game i.e. attributes of players such as health and strength, and the virtual world are static. There is no reward in gaining more health or strength by solving given tasks.

Sung et al [25] provides good examples of game as assignments. This work introduced programming logic into some visualized problems which is good for beginner and intermediate level students to boost their interests to programming. At first, the GTA GUI induced experimentation, yet as students become more proficient, the advantage of visualized problem diminished. The fact that more experienced students prefer assignment without elaborate setups was confirmed by Guzdial [27]. Hence, it lacks of tasks variety with respect to different targeted skills.

In short, up to now there is no serious game designed and developed to learn computer programming, in particular contents with rich variety of tasks intended for different targeted skills while maintaining the motivational level high.

3. RESEARCH QUESTIONS

Creating well suited contents for different targeted skills is essential in learning. This can be observed from an extension program named computer science unplugged learning, intended for primary-aged children designed by Bell et al [28]. This program introduces knowledge on computer science such as data, information, and some basic algorithms and it has been perceived well by students. The activities in the program highly utilize mathematics and analogy represented by for instance cards and boards are considered to be very compelling and improving conceptualization for their age. Other observation is puzzle based learning used to enhance problem solving skill to university students [9]. The result was immediately observable in their problem solving skills development due to its attracting and intellectually challenging nature.

This paper aims at investigating ideas on contents for different type of users based on their knowledge and skills in programming for serious game for learning programming, as well as nurturing problem solving skills, to be effective, and to be motivational. Contents will be developed based on topics covered in computer science curriculum 2008 of ACM for computer programming [17]. This paper is a preliminary stage of developing a novel serious game which will lead to future research in

answering the possibility of computer game to support teaching, and learning programming in particular. The early steps will try to determine the following questions.

- Who are the target user group of the game?
- What prior knowledge and skills of each target has?
- What are the examples of contents for each targeted skill?

4. THEORETICAL BASIS

As aforementioned, learning is less motivating compared to playing game since playing game is a fun and engaging activity. Generally fun comes from activities that we enjoy to do and the more we do, the better we get. Prensky [5] listed a collection of fun activities defined by Garneu: beauty, immersion, intellectual problem solving, competition, social interaction, love, creation, power, discovery, advancement and completion, and application of ability. The key of engaging characteristic of games is the gameplay factor such as game's rules, players' choices, the difficulty of the road to success, the game balance which keeps the player in the "Flow Zone", and a long term goal with clear short term goals. Depending on the game genres, the engaging activities can be as it follows.

- Puzzle game: the physical and mental challenge within the puzzles
- First Person Shooter (FPS) game: the opponents' speed and abilities.
- Strategy game: the available options and tactics to be employed.

Another example of engaging activity is mind game by asserting uncertainty of information in a class since students need to sort out the correct information from the false.

The difficulty of road to success and ability to keep the game balance according to the flow zone is the important aspect in creating contents for game. On pedagogical point of view, incongruity theory by Lankveld et al [29] stated learning may takes place if there is low positive difference between environment complexity against the internal mental model of the context, see Figure 1. Moreover, Vygotsky, a constructivism theorist, stated that in order learning to be optimal, the knowledge and skills should be kept in the Zone of Proximal Development (ZPD) [30]. This means learners should be kept in a narrow zone where the knowledge and skills is neither too difficult nor too easy to master under a proper guidance. These theories emphasized on the conformity of learning material with the prior knowledge of the learners. Another theory is Sweller's delivery strategy called Cognitive Load Theory (CLT) [31]. Sweller divided the memory of the brain into two i.e. short term and long

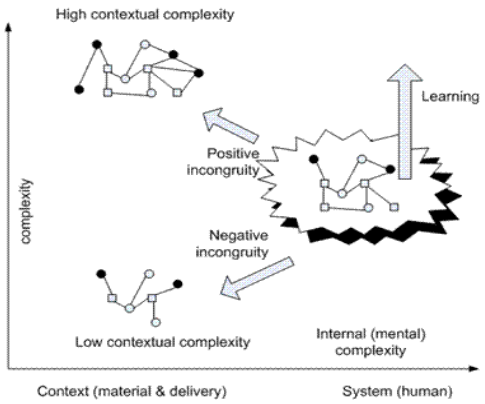


Figure 1.: Incongruity, Lankveld et atl [28]

term. Short term memory is essential to process perceived information prior to be stored in long term memory as knowledge. Yet, the short term memory is limited and overloading it inhibits learning. Therefore, information should be delivered in gradual manner which will give learners time to structure their knowledge and develop understanding. Providing partially worked example called Faded Working Example (FWE) for the learner to be completed is one way in CLT.

In general, the idea is that the game/contents should leverage the player's personal connects (previous knowledge and experience), in order to better support knowledge acquisition.

5. REQUIREMENTS ON DESIGNING CONTENTS

Prior to providing contents with various difficulty levels to better suit with the skill of the users, the target user groups should be defined. Afterward, in each target group, various tasks with multi-level challenges can be introduced. This will allow the game to adjust with the progress of the users. There are four dimensions of multi-level challenges can be employed [2] as follows.

- Task difficulty: prior knowledge needed to solve problems
- Task complexity: the number of sub-tasks which compose the main task
- Resources: what users have to complete the tasks e.g. time constraints
- Opponents: the number of opponents and their abilities

Subsequently, a delivery scenario can be developed in order to correctly cast the suitable tasks for learners. This is called adaptive learning i.e. a feature will be made by taking account the user profiles in term of their corresponding user groups and their performance which is highly suggested by Frazer et al [15]. However, this feature will not be discussed further. Instead, some examples of tasks with different targeted skills with their multi-level challenges will be presented.

Target User Groups

Prior to developing the contents, targets should be defined. Computer science unplugged [28] intended as extension program for primary aged children to introduce basic computer science are considered to be very compelling. Therefore, in order to attract prospective students to computer science, this type of approach is interesting to be explored as part of the contents where pre-university students are mostly the target. In addition, students who already enrolled in computer science major should be retained and motivated. GTA [25] are good examples for beginner students to keep them interested with the computer science. As students become more proficient, more challenging should be exploited [27]. Programming contests [21] or programming from scratch using Faded Working Examples (FWE) described in Section 4 may provide these challenges with topic of higher level problem solving. Therefore, there are three targets that will be on focus as shown by Table 1.

Table 1: *Targeted users with their corresponding skills*

Targeted Users	Knowledge and Skill Level	Tasks Type
Pre-university students	High school mathematics and algebra No previous knowledge	• mathematical based activities using cards and boards
First year university students	No previous knowledge	• support curricular activities • low to medium level problem solving and logic
Second year and upper university students	Basic programming	• support curricular activities • higher level of problem solving and logic

Computer Programming Curriculum

Given defined targets in Table 1, the tasks for the second and the third targets should be based on computer science curriculum. According to joint ACM and IEEE curriculum, fundamental programming stresses on fundamental programming concepts, basic data structures, and algorithmic processes to reach fluency in programming language (PL) [17]. Table 2 shows the topics covered by programming fundamentals.

Table 2: *Curriculum on programming*

Courses	Materials	Topics
ACM/IEEE: Programming fundamentals	▪ Fundamental constructs ▪ Algorithmic problem solving ▪ Data structures	▪ Basic syntax and semantic, variables, types, simple I/O, control structures ▪ Problem solving strategies, role of algorithm
	▪ Recursion ▪ Event driven programming ▪ OO programming ▪ Foundations in information security ▪ Secure programming	▪ Representation of numeric data, arrays, strings, pointers and references, linked structures ▪ Concept of recursion, mathematical functions, divide and conquer strategies ▪ Event handling methods, event propagation, exception handling ▪ OO design, encapsulation, classes and sub classes, inheritance, polymorphism ▪ Role of computer and network security, security standards, worms, viruses, risk assessment ▪ Overflowing checks, secure run-time stacks
UNIGE: fundamentals of computer	▪ Introductory concepts ▪ Software design ▪ PL C++ ▪ Data structures ▪ Algorithms ▪ OO programming	▪ Computer architecture, hardware, software, operating system, compiler, applications, binary encoding of information ▪ problem analysis; concept of algorithm, software design methodologies ▪ structure of a program in C + +, basic data types, variables, control structures, pointers, dynamic memory allocation, functions and recursion, I/O, debugging ▪ queues, stacks, lists, trees and their operations: access, insertion, deletion, sorting ▪ search and sorting algorithms, fusion algorithms, recursive algorithms ▪ classes, objects, messages and methods, design of simple classes, constructors and destructors, overloading, of functions and operators, inheritance, polymorphism

Joint ACM and IEEE curriculum is a suggestion of the materials needed in order students to have equal competences. Therefore, it does not strictly have to employ all the materials in the lesson. Some materials e.g.

secure programming can be excluded if the competence output does not require that specific areas [17]. A comparison for fundamental of computer course in University of Genoa (delivered in two semesters) can be observed in Table 2. Both curriculums have similar materials to develop fluency in programming which mainly consist of fundamental concepts, basic algorithms, data structure and OO programming.

Topics for Each Target User Group

Given the target users, and teaching materials in Table 1 and Table 2, we have preliminary constructed the topics for each target user in Table 3.

Table 3: *Topics for each targeted user groups*

Target Users	Topics
Pre-university students	Binary numbers, data representation, text and image representation and compression, simple algorithms
First year university students	All materials for pre-university students, PL, data structures, algorithmic problem solving , recursion
Second year and upper university students	All materials for 1st year students, OO programming, algorithms complexity

Bloom taxonomy classified the learning objectives into three domains: cognitive, affective, and psychomotor [33]. Within each domain, learning at higher level is dependent on having attained prerequisite knowledge at lower level. This taxonomy justified the topics selection presented in Table 3. The three domains of learning objectives will be further investigated to determine the level of taxonomy of each task given the domains. Nonetheless, an illustration of tasks to support cognitive domain is as follows.

- Pre-university students: to give ideas on computer science. Hence, it will evolve around knowledge and comprehension (level two in the cognitive domain [33]).
- First year university students: previous level added with application, problem solving, and logical thinking (level four in the cognitive domain [33]).
- Second year and upper university students: previous level added with synthesis and evaluation (level six in the cognitive domain [33]).

The tasks can be in form of true/false, multiple choice, pair matching, put in the right place, and find the error. However, more variety of tasks will be considered in the future. In each targeted skill, multi-level challenges can be implemented by defining tasks difficulty level, tasks complexity, and resources restriction to complete the tasks. An important thing which should be noted is the Bloom taxonomy merely drives the topics delivered, not how the game will progress.

6. CONTENTS ILLUSTRATION

Based on the topics with respect to different targets above, examples are proposed as preliminary idea on the contents of the game for computer programming. Basically the game will be designed to have a low learning curve for mastering it and motivate the users to play using Csikszentmihalyi's flow theory [34] instead of using conventional Bloom taxonomy driven learning [33].

Pre-University Students

The objective of this segment is to introduce computer science as an attractive field to learn. One of possible ways is through mathematical based activities [28] which are syntax free as follows.

1. Activity 1: Binary number

- Given a group of jars with the following volume in liters (L): 1, 2, 4, 8, 16 (Figure 2). How to obtain 5L from those jars if '?' can only be represented by 0 or 1 which means you don't use the jar or you use the jar, respectively.
- Similar as before, how you can obtain 9L and 30L.

Figure 2.: Group of jars

This activity gives an idea to students about binary representation and its use in computers in representing data and information. For instance, the answers of the question are *00101*, *01001*, *11110* for 5L, 9L, and 30L, respectively. It allows students to understand the amount of bits that required in representing certain number in the computers. Increasing complexity can be done by, for instance, adding jars that represent additional bits used. Also, this type of activities can be expanded to introduce the concepts of ASCII codes.

2. Activity 2: Understanding logic and algorithm

Five cards are drawn from the stack which results the numbers in Figure 3a. Then, the cards are shuffled and put in face down position (Figure 3b). The task is to sort the cards in Figure 3b in increasing order. However, only two cards are comparable at a time. What is the minimum number of comparisons until the cards successfully sorted in order?

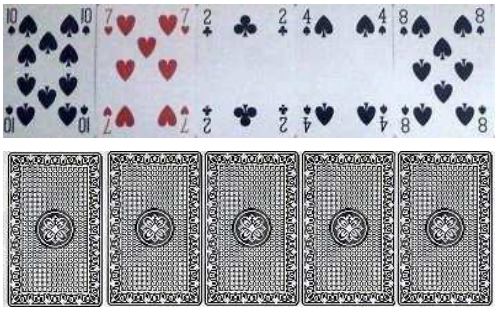


Figure 3.: a) cards face up, b) cards face down

This gives an idea to students about sorting algorithm. The activity requires interactive graphical user interface (GUI) to flip card and compare two cards at a given time instance. This activity can be sliced into minor steps or expanded to various searching algorithms e.g. binary searching, selection sort, and hash table for different level of challenges.

First Year University Students

The objective of this segment is to retain the interest of the new students in computer science, in particular programming course. Puzzles and visualized tasks may help them understand the materials as follows.

1. Task 1: Data structure

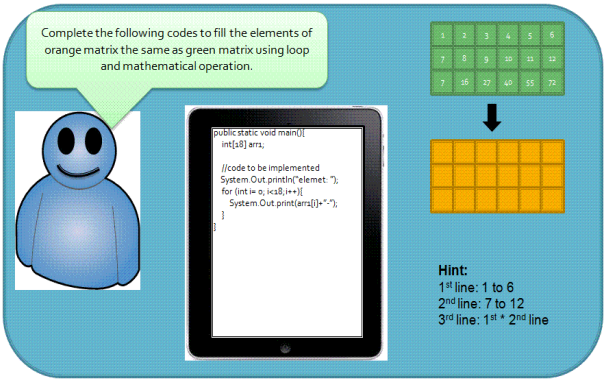


Figure 4.: Task 1: Data Structure

Figure 4 visualizes an empty array, a filled array, and a piece of codes that represent the empty array used within one pass loop structure. The task is to fill the elements of the empty array as the elements of the filled array by using the loop.

The purpose is to give an idea to students about value assignment in array and problem solving using simple mathematical operations. The extension of this task can be different type of mathematical operation, sorting, or two arrays addition or subtraction. Adding the size of the array is one of examples in increasing complexity. Also, worked example can be removed which make the problem more complex.

2. Task 2: Algorithm analysis

You are given the following code in Java.

```
public static void main(){
    k=0;
    for(int i=0;i<10;i++){
        k++;
    }
    System.Out.println(k);
    System.Out.println(i++);
}
```

Which one is the correct output?

A. 10 and 11
B. 9 and 11
C. 10 and 12

The purpose is to allow students to perform syntax and semantic analysis. The level of difficulty can be increased by adding nested loop. Adding possible answers or variables to be observed are possible to increase complexity. Resource restriction can be in the amount of time to answer the question.

Second and Upper Year University Students

This segment differs from previous segments in term of the complexity, since target users are expected to have acquired basic programming. This target skill needs more challenging problems to retain the amusement such as complex puzzles used in programming contests and OO programming as follows.

1. Problem 1: Mathematics

Find the solution for equation: $4x^2 + 12x - 20 = 0$ by using all coefficients as input.

This problem assesses the mathematical capability of the students as well as their analytical skill to observe all possible cases in the problem. This problem can be sliced into several steps using Faded Working Example (FWE) to assist students. However, the more advanced the students, the less assistance will be provided. Competition can be induced by recording the best time used by students beside the correctness of the program.

2. Problem 2: OO Programming

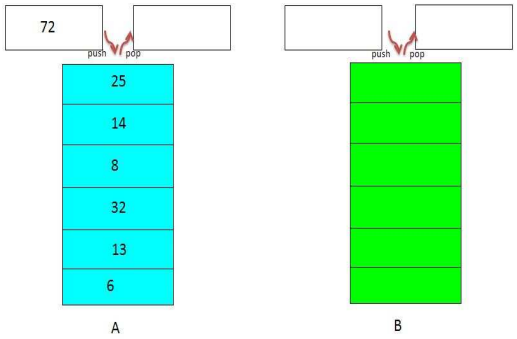


Figure 5.: Problem 2: OO, data structure, and sorting

Given two instantiated stack objects A and B with their corresponding operation i.e. *push* and *pop* (Figure 5), sort the numbers in stack A using stack B in decreasing order.

The purpose is to allow students to understand a type of data structure i.e. stack and utilize it for processing information. This problem can be sliced into several steps using Faded Working Examples (FWE) depending on the students' level.

For both problems above, FWE can be in form of how the problems are presented. For instance, students with lower skill level have to arrange a given set of unordered pieces of codes, whereas students with higher skill level have to complete a given program by filling several lines of missing codes. Moreover, visualization such as elements' movement in the stacks (Problem 2) may be provided as hints for different skill level.

7. CONCLUSION AND FUTURE WORKS

This paper served as the first stone for developing serious games to support learning in computer programming. It defined different target user groups of computer programming games with respect to the users' prior knowledge and learning objective. Also, preliminary examples of tasks as part of the contents for educational game were presented. Utilizing familiar subjects such as mathematical based activities are proposed for pre-university students to introduce computer science and attract them to study in computer science. Puzzles and visualized tasks are applied for first year to retain their interests in studying computer programming, whereas complex puzzles akin to problems in programming contests are employed for second and upper year university students.

Future work is planned to develop a richer variety of tasks for the contents which supports Faded Working Example (FWE), and to design the tasks and user model, the gameplay with rich interaction, the assessment, and personalized delivery for the game. Available technologies will be observed to realize the game and tested to students in computer programming course. Indicators represent students' skills and motivation in learning will be constructed and later on measured to confirm the expected impact on students' perception in learning computer programming with the objective of the game. Afterward, automated learning can be developed which will propose personalized learning for each different user. Some sensors and actuators may be incorporated in order to acquire users' needs and to provide feedbacks within the game. Also, the results of playing game will be evaluated to draw research finding.

8. ACKNOWLEDGEMENT

This work is supported by the European Commission and Education, Audiovisual and Culture Executive Agency (EACEA) in form of doctoral research grant namely Erasmus Mundus Interactive and Cognitive Environment (EM-ICE) PhD. Therefore, we would like to express our gratitude for providing us such grants. Moreover, we would like to thank the non-blind reviewers of this work: Dr. Wei Chen and Prof. Matthias Rauterberg (TU Eindhoven, the Netherlands), Prof. Carlo Regazzoni (University of Genoa, Italy), and Dr. Mahendrawati Erawan (ITS Surabaya, Indonesia).

9. REFERENCES

[1] Rauterberg, M. "Positive effects of VR technology on human behavior", 2004. **Proceedings of the 14th International Conference on Artificial Reality and Telexistence** (pp. 85-88). KAIST and VRSJ.

[2] Cheng H.N.H, Deng Y.C., Chang S.B., Chan T.W. "EduBingo: Design of multi-level challenges of a digital classroom game", 2007. **1st IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning**.

[3] Blunt, R. "Does Game-Based Learning Work? Results from Three Recent Studies", 2007. **The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)**, NTSA, Orlando, Florida, USA, p.945-954.

[4] Riedel, J., Pawar, K. "A report on the experiences gained from evaluating the cosiga NPD simulation game", 2009. **16th International Product Development Management Conference**, Twente University, The Netherlands, EIASM, Brussels.

[5] Prensky, M. **The Motivation of Gameplay: The Real Twenty-first Century Learning Revolution**, 2002. On the Horizon, 10(1), 5-11.

[6] Bellotti F., Berta R., De Gloria A., Primavera L. "Enhancing the educational value of videogames", 2009. **Magazine for Computers in Entertainment (CIE) – Special Issue: Media Arts and Games (Part II)**, Vol. 7, Issue 2.

[7] Bellotti F., Berta R., De Gloria A., Primavera L. "A task annotation for SandBox Serious Games", 2009.

Proceedings of the 5th IEEE international conference on Computational Intelligence and Games.

[8] Zyda, M. "From visual simulation to virtual reality games", 2005. **Computer**, Vol. 38, no. 9 (pp. 25-32).

[9] Falkner N., Sooriamurthi R., Michalewicz Z. "Puzzle-based learning for engineering and computer science", 2010. **IEEE Computer** (pp. 20-28), Vol. 43, No. 4.

[10] Rauterberg, M. "Positive effects on entertainment technology on human behavior", 2004. **Proceedings of the IFIP international working conference system concepts** (pp. 51-58). Kluwer Academic Press.

[11] Zielke M.A., Evans M.J., Dufour, F., Christopher T.V., Donahue J.K., Johnson P., Jennings E.B., Friedman B.S., Ounekeo P.L., Flores R. "Serious Games for immersive cultural training: creating a living world", 2009. **IEEE Transaction on Computer Graphics and Applications**, April 2009 (pp. 49-60), Vol. 29, Issue 2.

[12] Kurniawan, S.H. "Intergenerational Learning through World of Warcraft", 2008. **Proceeding on 2nd IEEE Conference on Digital Game and Intelligent Toy Enhanced Learning** (pp. 98-102).

[13] Chang M., Kinshuk "Web-based multiplayer online role playing game (MORPG) for assessing students' Java Programming Knowledge and Skills", 2010. **IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning** (pp. 103-107).

[14] Frazer A., Argles D., Wills G. "The Same, But Different: The Educational Affordance on Different Gaming Genres", 2008. **8th IEEE International Conference on Advanced Learning Technologies** 2008 (pp. 891-893).

[15] Frazer A., Argles D., Wills G. "Is Less Actually More? The Usefulness of Educational Mini-games", 2007. **7th IEEE International Conference on Advanced Learning Technologies** 2007.

[16] Muratet M., Torguet P., Jessel J.P., Viallet F. "Towards a Serious Game to Help Students Learn Computer Programming", 2009. **International Journal of Computer Games Technology**, Vol. 2009, Hindawi Publishing Coorporation.

[17] Association for Computing Machinery (ACM). "Computer Science Curriculum 2008: An Interim Revision of CS 2001", 2008. **ACM and IEEE Computer Society**.

[18] Bayliss J.D. "Using Games in Introductory Courses: Tips from Trenches", 2009. **ACM Special Interest Group on Computer Science Education** (pp. 337-341), Chattanooga, Tennessee, USA.

[19] Kelleher C., Cosgrove D., Culyba D., Forlines C., Pratt J., Pauch R. "Alice 2: programming without syntax errors, 2001". **In proceeding of 15th Annual Symposium on the User Interface Software and Technology**, Paris, France.

[20] Overmars M. "Learning object oriented design by creating games", 2005. **IEEE Magazine** December 2004/January 2005.

[21] Garcia-Mateos G., Fernandez-Aleman J.L. "Make Learning Fun with Programming Contests", 2009. **Computer Science Transaction on Edutainment II**, LNCS 5560, pp. 246-257, Springer-Verlag.

[22] Trotman A., Handley C. "Programming Contest Strategy", 2008. **Journal on Computer and Education**, v50 n3 p821-837 Apr 2008, Elsevier.

[23] Anderson E.F., McLoughlin L. "Critters in the classroom: A 3D computer-game-like tool for teaching programming to computer animation students", 2007. **SIGGRAPH International Conference on Computer Graphics and Interactive Techniques** 2007.

[24] Chen W.K., Cheng Y.C. "Teaching Object oriented programming laboratory with Computer Game Programming", 2007. **IEEE Transactions on Education**, Vol. 50, No. 3.

[25] Sung K., Hillyard C., Angotti R.L., Panitz M.W., Goldstein D.S., and Nordlinger J. "Game-themed programming assignment modules: a pathway for gradual integration of gaming context into existing introductory programming courses", 2010. **IEEE Transactions on Education** July 2010.

[26] Mohammed P., Mohan P. "Combining digital games with culture: a novel approach towards boosting student interest and skill development in computer science programming", 2010. **2nd IEEE Conference on Mobile, Hybrid, and On-line Learning**.

[27] Guzdial M. "Contextual computing education". Invited presentation, **Microsoft Research Faculty Summit**, Jul. 2008 [Online]. Available from http://home.cc.gatech.edu/guzdial/169 [Accessed in 20 March 2011]

[28] Bell T., Witten I.H., Fellows M., **Computer Science Unplugged**: "an enrichment and extension programme for primary-aged children". Available from http://www.tielt.org/publications/IJCAI-WS-Ponsen.pdf. [Accessed in 11 March 2011].

[29] Lankveld van G., Spronck P., Rauterberg M. "Difficulty scaling through incongruity", 2008. **Proceedings of the 4th Artificial Intelligence and Interactive Digital Entertainment Conference** (pp. 228-229). AAAI Press.

[30] Vygotsky, L. S. "Mind in Society: The Development of Higher Psychological Processes" (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.), 1978. Cambridge, MA: Harvard University Press

[31] Sweller, J. "Cognitive load during problem solving: Effects on learning". **Cognitive Science**, 1988. 12, 257-285.

[32] Faculty of Engineering, University of Genova. "Curriculum for Computer Engineering: Fundamental of Computer Course", 2011 [Online]. Available from http://www.informatica.ingegneria.unige.it/php_files/insegnamenti.php [Accessed in 10 May 2011]

[33] Bloom B. S., Engelhart M. D., Furst E. J., Hill W. H., and Krathwohl D. R. "Taxonomy of educational objectives: the classification of educational goals", 1956. **Handbook I: Cognitive Domain** New York, Longmans, Green.

[34] Csikszentmihalyi, M. "Flow: The psychology of optimal experience", 1990. New York: Harper Perennial.

[35] Chen W., Bambang Oetomo S., Feijs L. M. G., Andriessen P., Kimman F., Geraets M., and Thielen M. "Rhythm of Life Aid (ROLA) – An Integrated Sensor System for Supporting Medical Staff during Cardiopulmonary Resuscitation (CPR) of Newborn Infants", 2010. **IEEE Transactions on Information Technology inBioMedicine**, vol. 14, no. 6, pp. 1468-1474, Nov. 2010.

Designing Virtual Worlds for Inquiry: Can It Be Done?

Heather DODDS
Liberal Arts, Western Governors University
Salt Lake City, UT 84107

ABSTRACT

Simulations in the form of games are prevalent in the social and learning environments. One of the most complex areas to learning in science is inquiry, the act of acting like a scientist while asking questions and exploring the unknown. Science simulations afford learners the ability to manipulate time, space, and learning conditions; simulations can take learners to impossible places. Discovery-based learning techniques like simulations can be cognitively overwhelming. There is criticism that these approaches are not effective forms of learning. Prudent use of simulations for the topic of science inquiry is advocated. Early evidence indicates that this is an area of growth in instructional design as simulations can provide for the open structure required in a scientific inquiry.

Keywords: Virtual worlds, Science inquiry, Virtual environments, simulations, instructional design

INTRODUCTION

Learners are surrounded by visually immersive, interactive, and stimulating culture outside of education. Games, smart phones, and entertainment compete for learner attention. Kzero, a virtual worlds statistics consulting company, reported that there are a total of 1.399 billion registered accounts within virtual worlds as of July 2011 [1]. Thus growing numbers of people now have at least one digital presentation, an avatar, in a virtual world. Zynga, the Farmville game creator, shared statistics that there are 30 millions virtual farms in the United States whereas there are only two million actual farms [2]. Thus, virtual worlds are becoming reality for our learners.

This paper proposes a discussion of one of the most recent and advanced instructional design approaches, the use of immersive simulations. Inquiry is the act of investigating and asking questions. Inquiry holds a special role within science education as it is both content (learners need to know the scientific method) and it is pedagogical approach (one learns to be a scientist through scientific behaviors such as questioning). Tools that instructional designers use are never one size fits all. At the same time, what is the role of virtual environments within inquiry learning? Can inquiry-based learning methods be built into these stimulating experiences? What are the myths of inquiry-based learning in online environments? What is on the future horizon for teaching and learning science inquiry online? Educators, designers, and researchers of the 21st century must address these concerns if education is to engage the learner to become a true citizen of this new learning age.

SCIENCE SIMULATIONS

Simulations and science share a natural relationship. Science educators have to explain and illustrate concepts that are often beyond the realm of the eye. For this reason, simulations are a natural fit for this discipline. Simulations can take the learner in to situations that would be too dangerous such as inside a nuclear reactor, too expensive such as repetitions of DNA testing, or physically impossible such as journeying into the center of a volcano. Additionally, computer-based simulations can let the learner manipulate variables that would normally be unavailable in an experiment [3]. Blake and Scanlon [4] explain that time or other simulation variables can be altered by the learner so that the learner can explore the desired inquiry of 'what if' questions. Note that even though variables can be altered within a simulation, there are times when simulations should restrict learner control in order to maintain learning focus.

Blake and Scanlon [4] further define simulations as "programs that contain a representation of an authentic system or phenomenon" and "allow students to change some of the parameters in the program and observe what happens as a result" (p 491). There are many types of educational simulations. According to Alessi and Trollop (as cited in Akpan), types of simulations include physical, procedural, situational, and process [5]. Thus, Bills [6] proposed that simulations are meant to simulate a real –world situation within a computer-based environment.

Simulations, however, represent much more than moving diagrams. One of the most powerful ways simulations can be used in science is to test prior learner conceptions. It has been shown by Rea-Ramirez, Nunez-Oviedo, and Clement [7] that learners often enter science courses with persistent and incorrect concepts about how science works. These prior conceptions are based on life experiences and as such, they are very difficult to overcome. Learners will often accept a 'book concept' during a course and answer exam questions correctly but then continue to retain an incorrect concept for their lifetime. A designer needs to ascertain those misconceptions and carefully explore them with the learners in order to dismantle misconceptions and form new accurate science concepts. It follows that if learners formed incorrect concepts through life experiences, a good way to correct those concepts is with other immersive experiences. Simulations fill that gap.

RECOMMENDATIONS FOR USE

Simulations are good for situations that are otherwise: (a) dangerous such as viewing nuclear reactions, (b) expensive such as military flight training, or (c) impossible such missions to the Moon. The benefits of simulations go beyond these as Akpan

explained, "Even if real-life exploration is feasible, such experimentation can be supplemented by simulations that offer students the opportunity to explore a wider range of variables more rapidly. Such simulated experiences potentially can be used to confront alternative conceptions, produce disequilibration and with appropriate scaffolded instruction, lead students to a new accommodation" (Introduction, para. 4) [5]. Simulations can also shorten teacher preparation time and allow more control over experimental variables [4]. For these reasons, simulations can represent a prudent selection for some science instructional circumstances.

Blake and Scanlon [4] recommended, "To be scientifically useful, simulations should be based on real events and data" (p.499). This would align with Squire's assertion that simulations must be tied to real-world events in order for their instructional effectiveness to transfer beyond the simulation to the actual learner performance [8]. Merrill's First Principles of instructional design also support this point [9].

Science simulations are more than just viewing animated pictures. Blake and Scanlon [4] recommend that "Use of multiple representations, graphs and an opportunity to observe any graphs forming while the experiment is running (in real time) is also a very useful feature for simulations" (p. 500). Graph reading skills are critical in science. This tip may not be possible in all simulations, but there is value in linking math and science in this dynamic way.

Game creators, whose concerns are very similar to simulation creators, are careful to include a compelling story. Blake and Scanlon [4] point out that "For all simulations, facilities to tailor activity to student ability levels and a narrative for students to follow ought to be provided either online in the simulation or by the accompanying notes" (p. 500). With computer-based simulations, the program can be created so that it automatically adjusts for each learner while carrying a similar narrative through the simulation. Additionally, the learner can keep notes with an accompanying virtual journal or notebook.

Beyond use in science instruction, simulations can break access barriers for learners. Klemm and Tuthill [10] emphasized the use of simulations and virtual field trips not only for science education but to meet the needs of students with disabilities that may not be able to visit locations in real life [11]. Although this may be true, simulations are not appropriate at all times for all learners in science.

CRITICISM OF SIMULATIONS

It should be noted that implicit forms of instruction have been accused of being ineffective [11]. Interestingly, the solution for effective use of simulations harkens back to the wise use of personalized instruction from the 1950s [12]. Blake and Scanlon [4] point out that "Many researchers emphasize the importance of a good instructional plan when using simulations. de Jong et al. (1994) argue that the reason for finding no conclusive evidence for effectiveness and efficiency of simulations, despite their popularity in instruction, is the lack of support for learners in some simulations, that is, if learners encounter difficulties they may not overcome these on their own" (p.499). Also, because simulations are challenging forms of instruction, designers should be careful to align the assessment with the method of learning.

Simulations, however, are not instructional 'eye candy' and this form of instruction moves beyond simple animations of content. Learning must be specified and built into the experience. Bills (2010) stated "the system of instruction needs to not only include mastery of concepts and procedures, but also achievement of metaskills and the transformation to tacit knowledge" (p. 396). Gibbons, McConkie, Seao and Wiley [13] warn that "An unaugmented model has limited instructional value, can create instructional inefficiencies, and can lead the learner into misinterpretations and misconceptions" (pp. 171-172). Thus the use of simulations within instruction must always contain some kind of instructional goal, even if that goal is not made explicitly known to the learner. This contrasts with Squire's [8] assertion that games "create an emotionally compelling context for the player" (p. 445) and that "there are still overarching narratives at work" (p. 447). Simulations do not necessarily contain narratives as games contain. Simulations can be shorter and the learner often has an implied role upon entering the interface. However, both simulations and games often do not strive for Squire's [8] "perfect representation of reality" (p. 446). It is this aspect of a 'version' of reality that has an effect upon learners.

Mayer's [14] multimedia studies warned that there are combinations of media that can be overwhelming to the learner. This is in accord with van Merriënboer, and Ayres' assertions about Cognitive Load Theory [15]. This theory posits that incoming information briefly enters the learner's visual or auditory senses. Then working memory within the brain has a limited capacity to process, make sense, and sort information into irrelevant, which is then ignored, or into important experiences which are transferred to long-term memory. Long-term memory is considered to be limitless for human purposes and this is where cognitive load theory and experiential learning theory intersect. By creating meaningful learning experiences, learners can internalize their learning for later transfer outside of the educational experience. Said another way, learners can remember the experience long after the lesson is done. Simulations strive to create multiple versions of meaningful learning experiences.

Cognitive load theory predicts that the beginning phases of any simulation should be simplistic and not reflective of reality so that the learner does not need to struggle to figure out how to sort the incoming information. This reduces extraneous information and allows the learner to focus on the germane information. Once the learner moves beyond novice stage, more detail and higher fidelity and authenticity can be incorporated into the simulation. Bills [6]paraphrases Dr. Allison Rosette's description of authenticity as the "parallels between the learning experience and the learner's life and real-world application" (p. 398).

CONCLUSION

There have been several examples of successful science simulations. Ketelhut, Nelson, Clarke, and Dede used the virtual world, Riverworld© as an example of a disease outbreak and then middle school science students needed to use scientific inquiry skills to explore the cause of the disease [16]. Various simulations need to be explored for their value in giving learners true inquiry experiences. When the future horizon of simulations is only limited by imagination, the learner should engage in a journey of asking unlimited questions and finding scientific answers.

REFERENCES

[1] Kzero. **VW registered accounts for Q1 2011 reach 1.185 bn.** 2011. Retrieved from http://www.kzero.co.uk/blog/?p=4580

[2] Zynga. **Numbers.** 2011. Retrieved from http://www.zynga.com/about/numbers.php

[3] F. M. Coleman, F. M. Software simulation enhances science experiments. **T H E Journal,** *25,* 1997, 56-58.

[4] C.C. Blake, & E. E.Scanlon, E. E. Reconsidering simulations in science education at a distance: features of effective use. **Journal of Computer Assisted Learning,** 23(6), 2007, 491-502. doi:10.1111/j.1365-2729.2007.00239.x

[5] J.P. Akpan. Issues associated with inserting computer simulations into biology instruction: A review of the literature. **Electronic Journal of Science Education,** *5*(3) 2001. Retrieved from http://ejse.southwestern.edu/article/view/7656

[6] C.G. Bills, C.G. High engagement strategies in simulation and gaming. In K. H. Silber and W. R. Foshay (Eds.), **Handbook of improving performance in the workplace:** Vol. 1. Instructional design and training delivery, 2010 (pp. 396-434). San Francisco, CA: Pfeiffer.

[7] M.A. Rea-Ramirez, M.C. Nunez-Oviedo, & J. Clement. Role of discrepant questioning leading to model element modification. **Journal of Science Teacher Education***, 20*(2), 2009, 95-111.

[8] K. D. Squire. Video game-based learning: An emerging paradigm for instruction. In K. H. Silber and W. R. Foshay (Eds**.), Handbook of improving performance in the workplace: Vol. 1. Instructional design and training delivery**. 2010, (pp. 435-467). San Francisco, CA: Pfeiffer.

[9] M. D. Merrill, M. D. First principles of instruction. **Educational Technology Research and Development,** *50*(3), 2002, 43-59.

[10] E. B. Klemm & G. Tuthill, G. Virtual field trips: Best practices. **International Journal of Instructional Media,** *30*(2), 2003, 177-194.

[11] R. E. Clark, K. Yates, S. Early, & K. Moulton, K. An analysis of the failure of electronic media and discovery-based learning. Evidence for the performance benefits of guided training methods. In K. H. Silber and W. R. Foshay (Eds.), **Handbook of improving performance in the workplace:** Vol. 1. Instructional design and training delivery, 2010 (pp. 263-297). San Franisco, CA: Pfeiffer.

[12] M. Molenda, M. Origins and evolution of instructional systems design. In Silber, K. H., & Foshay, W. R. (2009). **Handbook of improving performance in the workplace:** Vol. 1. Instructional design and training delivery, 2010, (pp. 53-92). San Francisco: Pfeiffer.

[13] A. Gibbons, M. Mcconkie, K. Seo, D. Wiley, D. Simulation approach to instruction. In Reigeluth, C. M., & Carr-Chellman, A. A. (2009). **Instructional design theories and models, volume III: Building a common knowledge base.** 2009, (pp. 167-193). New York: Routledge Education Taylor & Francis Group.

[14] R. E. Mayer, R. E. **The Cambridge handbook of multimedia learning.** 2009 New York: Cambridge University Press.

[15] J. J. G. van Merriënbor, &P. Ayres, P. Research on cognitive load theory and its design implications for e-learning. **Educational Technology, Research and Development** *53*(3) 2005, pp. 5-13.

[16] D. J. Ketelhut, B. C. Nelson, j. Clarke, & C. Dede, A multi-user virtual environment for building and assessing higher order inquiry skills in science. **British Journal of Educational Technology,** *41*(1), , 2010, 56-68. doi:10.1111/j.1467-8535.2009.01036.x

Production of Fuel Grade Ethanol: Optimization – Based Design, Operation and Control.

Manuel A. RAMOS

Departamento de Ingeniería Química, Universidad de Los Andes

Bogotá D.C., Colombia.

Pablo GARCÍA – HERREROS

Departamento de Ingeniería Química, Universidad de Los Andes

Bogotá D.C., Colombia.

Jorge M. GÓMEZ

Departamento de Ingeniería Química, Universidad de Los Andes

Bogotá D.C., Colombia.

And

Jean M. RENEAUME

Laboratoire de Thermique, Energétique, et Procédés (LaTEP), Université de Pau

Pau Cedex, France.

ABSTRACT

The extractive distillation of ethanol using glycerol as entrainer is studied, where a Numerical – Optimization based design, operation and controls were performed. The research includes several steps, where different kinds of numerical optimization are involved. Through the Optimization - based design, it is possible to obtain both optimal operating conditions and optimal design. The solutions of all the optimization problems are achieved using state of the art computational tools and solvers. The results of each stage establishes the process that maximizes an economic criterion for the industrial production of bioethanol satisfying each problem constraints

Keywords: Fuel Grade Ethanol, Numerical Optimization, Non Linear Programming, Extractive Distillation, Mixed – Integer Non Linear Programming, Biofuels.

1. INTRODUCTION

Biofuels are nowadays viable alternatives to replace fossil fuels worldwide. Biofuels research has been carried out in the past few years thanks to recent interest in renewable energy sources and the benefits of low contamination associated with their use. In several countries, biofuels are produced using traditional chemical processes: for example, in countries like Brazil, Colombia and Thailand, biodiesel is produced from palm oil and alcohol, producing great quantities of glycerol as a byproduct of the process. This overproduction of glycerol has been a source of research for engineers, since an important question emerges: what to do with it?

On the other hand, fuel grade ethanol production implies the need to remove water from it. One way to achieve this is through extractive distillation, an operation that is very energy intensive and that contributes in a considerable way to the total energy requirements of the whole process. In this context, extractive distillation needs an entrainer (a high boiling point compound) in order to achieve the desired ethanol purity, and here is where glycerol comes into play. Using the excess glycerol produced in the biodiesel process it is also possible to produce fuel grade ethanol, as it is shown in figure 1.

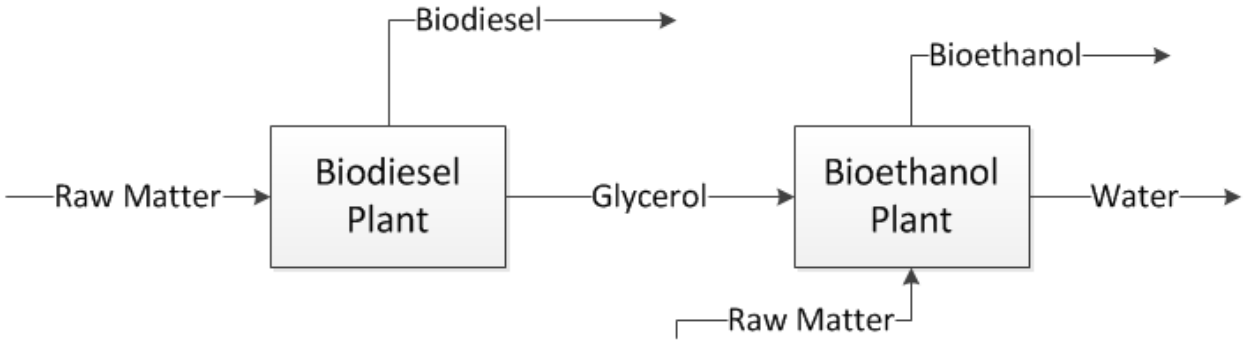


Figure 1. Simplified Flowsheet of the integration between Biodiesel and Bioethanol production.

This is why the present work is focused in the extractive distillation unit, where a Numerical Optimization – based design, operation and control were performed. It is very important to remark that optimization plays a very important role in the production of biofuels, since it is necessary to produce them in such a manner that makes their prices competitive against fossil fuels [1]. In this context, our research includes several steps, going from the most general case to others more complex. Nevertheless, every step in the research process is essential to achieve the next one.

2. EXTRACTIVE DISTILLATION PROCESS

Extractive distillation is the partial vaporization process that occurs in the presence of a miscible entrainer that alters the relative volatilities of the components present in the mixture to be separated [1]. Adding a new compound to the mixture shifts the thermodynamic equilibrium, modifying the feasible compositions that can be achieved by ordinary distillation. The extractive distillation system is made up by two distillation columns: the first is the extractive distillation column and the second is the entrainer regeneration unit. The flowsheet diagram is shown in figure 2.

3. PRELIMINARY WORK

The stages (past, present and future) included in our research process are addressed next: In the first place, the steady – state (no time - dependence) preliminary design and simulation of the extractive distillation column via shortcut methods was conducted using available commercial software (Aspen Plus ©), in order to obtain the design parameters that adjusts to certain product constraints (purity and quantity of raw matter to use). With the results of this first simulation, the steady – state simulation of the extractive distillation column via rigorous methods was implemented, in order to approve the previously obtained design. After this validation, the rigorous model was implemented in a programming environment, such as Matlab ® or using a programming language such as FORTRAN in order to simulate the column in an equation – oriented formulation.

The rigorous equation – oriented model implemented in this stage and in the subsequent ones is a thermodynamic equilibrium –based model, where the *MESH* equations describe the physical phenomena governing the operation of the extractive distillation column. The *MESH* equations

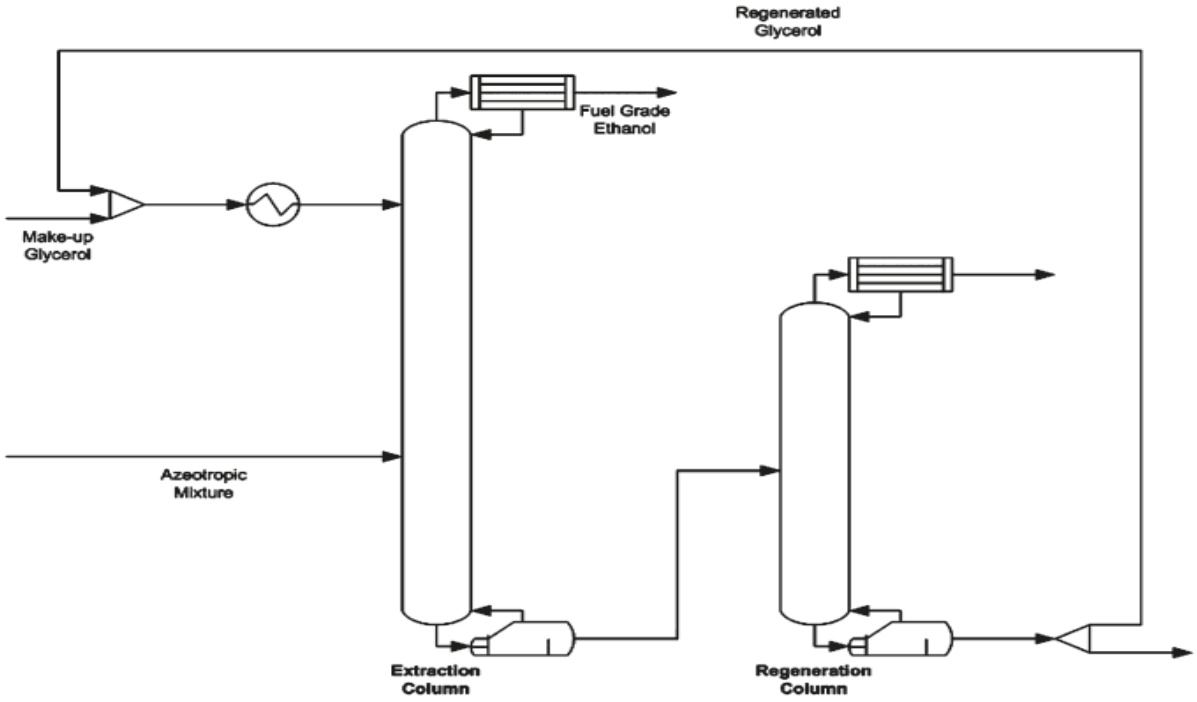


Figure 2. Flowsheet of the extractive distillation process [1].

are widely used in the industry and in research due to their apparent simplicity, its mathematical elegancy and its predictive capacities [2]. The equations comprised in this model are the following [2]:

-Mass balance: Total and partial for each stage.

-Equilibrium relations: for each component and each stage.

-Mole fraction summations: one for each stage.

-Energy balance: one for each stage.

Here is where the first optimization stage appears, which consists on finding the optimal values of all the column variables in order to minimize annual operating costs. This type of optimization problem is classified as an *NLP* problem (*Non Linear Programming)*, where there are only algebraic variables, and the model is nonlinear. Up to this point, only preliminary simulations, optimizations and designs were carried out.

4. OPTIMIZATION – BASED DESIGN

In the next stage an optimization – based design was implemented. With the rigorous model tested, we proceed to find (numerically) the optimal column configuration (in terms of design and construction) to minimize the annual operating and capital cost, using an *MINLP (Mixed Integer Non Linear Programming)* approximation. Results of this stage are very important *per se*, since its use will allow the extractive distillation column to be built and operated in a real processing facility.

The rigorous design of the extractive distillation system implies establishing the following: areas of the heat exchangers, column diameters, column heights and feed stage locations. It is important to note that these parameters are strictly related to the number of stages of the column. This is why the design variables of the extractive distillation system are the stages in each one of the five column sections: three for the extractive column (rectifying, extractive and stripping) and two for the regeneration column (rectifying and stripping).

The optimization problem of this stage consists of an economic objective function (which consists of discrete and continuous variables) subject to the model constraints (the *MESH* equations of both columns) and the operational constraints:

$$\begin{aligned} &\max \ Z = f(x, y) \\ &s.t. \quad h(x) = 0 \\ &\qquad g(x) \le 0 \\ &x \in X, y \in Y \end{aligned} \tag{1}$$

In Eq. (1) x are the continuous variables, y are the discrete variables, h(x) are the model constraints and g(x) are the operational constraints.

The operational constraints are product requirements in order to meet certain standards for its commercialization:

-Minimum molar purity of ethanol produced: 99.5 %.

-Maximum operating temperature allowed in the process (decomposition temperature of glycerol): 555K.

Objective Function.

The Objective Function is made up by the following elements:

-Market value of products.

-Raw materials cost.

-Operating Costs: value of the utilities required in the columns operation.

-Infrastructure cost: cost of the columns, additional equipment and installation.

Solution strategy.

The solution of the optimization problem is achieved through a two – level strategy. The discrete variables are considered in a master problem that uses a stochastic algorithm in order to evaluate different configurations of the system. The continuous variables are considered in an NLP subproblem that uses a deterministic algorithm in order to find the optimal operating conditions of the system. The implementation of a stochastic algorithm increases the probability of obtaining the global optimum [3]. The model was programmed in Matlab® and was solved on an Intel Core 2 Duo CPU with a 3.07 GHz frequency and 3.21 GB of RAM. The results of this stage are shown in figure 3.

5. OPTIMAL CONTROL

However, to achieve steady – state operation in such equipment is a very difficult task. In addition, process variables (e.g. feed flow rate, temperature) can vary as a function of time. Here is where automatic control arises: in response to these variations, the column control system regulates control variables to the point of desired operation of the state variables.

In terms of optimization, an optimal control strategy can be designed to obtain the optimal profiles for a specific time period taking into account the state variables in the extractive distillation column [4]. In this stage, the optimal profile of the control variables (e.g. glycerol feed, energy requirements) that minimizes the operating cost in a determined period of time was obtained, assuming uncertainty (sinusoidal wave) in the feed conditions [5].

It is important to highlight that in this stage it was necessary to change the programming environment to a more specific one, like *GAMS (General Algebraic Modelling System).* This change was made due to the scale of the model and optimization generated due to the introduction of the time variable (algebro – differential model). This type of problem is classified as a *DNLP* problem (*Dynamic Non Linear Programming),* because it has differential and algebraic variables, such as the ones the model has. In this stage, the operational constraints stay the same as the previous stage.

The general formulation of a *DNLP* problem is as follows:

$$\min_{\substack{xd \\ xa \\ u \\ d}} \; J(x_d(t), x_a(t), u(t), d, \theta(t)) \tag{2}$$

Subject to:

$$\begin{aligned} & f(\dot{x}_d(t), x_d(t), x_a(t), u(t), d) = 0 \quad \forall t \in \left[t_o, t_f\right] \\ & q(\dot{x}_d(t), x_d(t), x_a(t), u(t), d) \le 0 \quad \forall t \in \left[t_o, t_f\right] \\ & u \in R^u, \quad d \in D_c \\ & x_d \in X_d \subseteq R^{x_d} \qquad x_a \in X_a \subseteq R^{x_d} \end{aligned} \tag{3}$$

In the last equations, f is the model of the column and the equations of the control model, and q are the operative constraints of the process. $\dot{x}_d$ are the differential terms of the state variables, x_d are the differential state variables, x_a are the algebraic state variables, u is the vector of control variables, d are the design variables and θ are the uncertainty parameters. In this case, no design variables are taken into account.

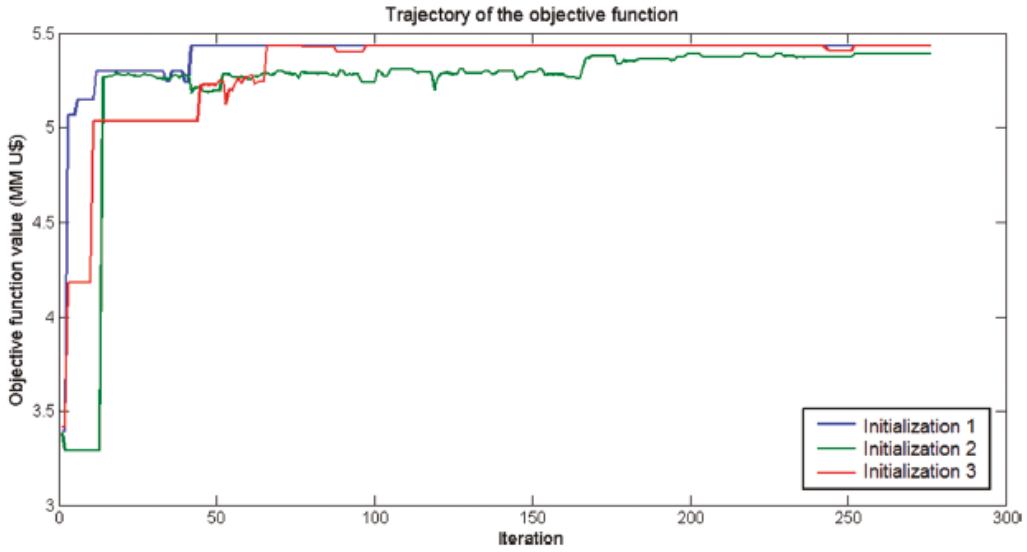


Figure 3. Trajectory of the objective function in the MINLP stage [1].

Objective Function.

The objective function of this stage is different from the optimal design in steady state, because it is not an algebraic objective function but an integral objective function over the time variable. It is made up by the following elements:

-Market value of products.

-Raw materials cost.

-Operating costs.

-In addition, this objective function has three elements in order to minimize the difference between the steady state operating point and the dynamic operating point. This is made since it is needed to make the transition between two steady states as smooth as possible [6].

Solution strategy.

The algebro – differential model was discretized using finite differences, transforming the DNLP problem into an NLP problem. Because of this, the problem became a sparse large – scale problem. To achieve the solution, an interior – point, large scale algorithm available in GAMS was used: *IPOPT*. This state - of – the - art solver was proven to solve efficiently dense large – scale problems, as well as sparse large – scale problems [7]. The optimization was solved in an Intel i5 CPU with 4 cores, 3.2GHz frequency and 4 GB of RAM. A typical solution of an optimal control problem is shown in figure 4.

6. CURRENT AND FUTURE WORK

The research stages described up to this point are already developed. In the next section the present and future work is described:

It is the intention of the project to combine the optimal design and control in order to accomplish simultaneously these two tasks. The benefits associated with this procedure are that the column can be designed taking into account its controllability and therefore improving its design and operating/capital costs.

If this optimization is accomplished successfully, the column can be far more cost – efficient that the one designed in the *MINLP* stage. This optimization problem is classified as *MIDO* (*Mixed Integer Dynamic Optimization)* in which there are integer, differential and algebraic variables.

Another research that is being conducted intents to compare results of the above extractive distillation column with another case in which external heat sources or sinks are added to the column structure, probably under the *MINLP* formulation.

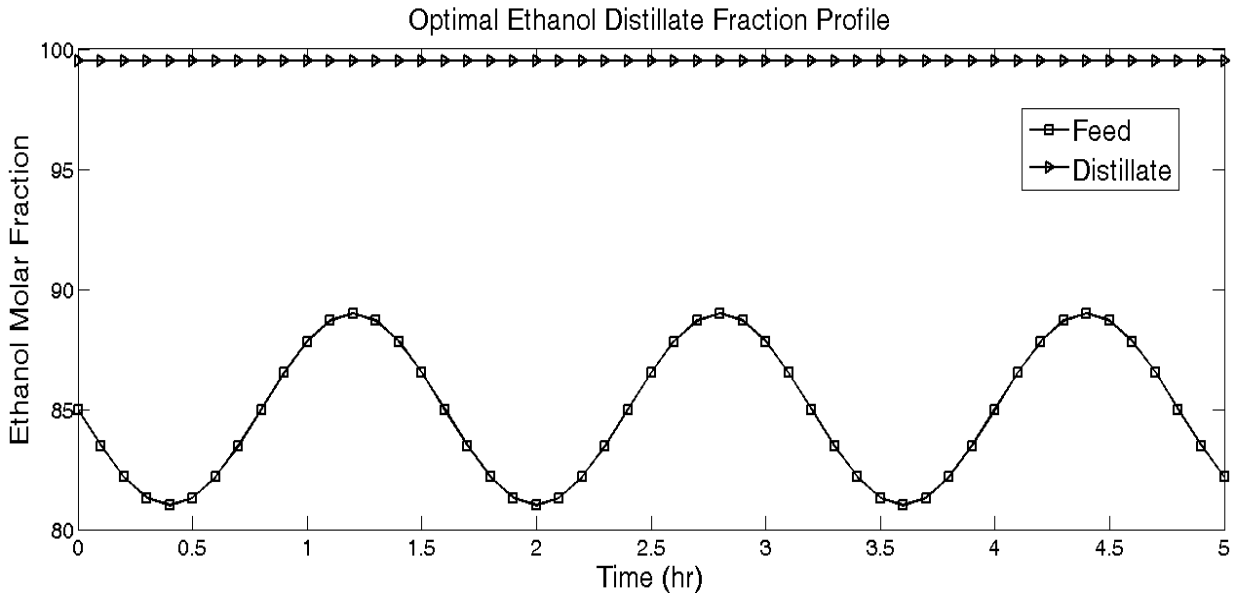


Figure 4. Optimal control results profile.

7. CONCLUSIONS

The optimization – based research and design proposed considers the elements of design and control with the operation of the extractive distillation for the production of fuel grade ethanol using glycerol as solvent. The approaches reviewed here allow analyzing the feasibility of the process in order to find the most suitable conditions to produce bioethanol. This is achieved through systematic research.

Optimization nowadays, making use of advanced technology and state – of – the – art tools, algorithms and solvers, has become an essential tool in modeling, design and deploy. With optimization techniques one can turn economically infeasible processes into feasible ones. That is why optimization is so important to the industrial sector.

The results obtained in every stage propose a process that offers a very good projection for the industrial production of fuel grade ethanol using glycerol as solvent. As said in the introduction, it is important to produce biofuels in such a way that makes its prices competitive in the market.

8. REFERENCES

[1] P. García-Herreros, J. M. Gómez, I. D. Gil, and G. Rodríguez, "Optimization of the Design and Operation of an Extractive Distillation System for the Production of Fuel Grade Ethanol Using Glycerol as Entrainer," *Industrial & Engineering Chemistry Research*, vol. 50, pp. 3977-3985, 2011.

[2] K. R. Taylor R., "Real - World Modeling of Distillation," *Chemical Engineering Progress*, vol. 98, pp. 28-39, 2003.

[3] I. G. José A. Caballero, "Una Revisión del Estado del Arte en Optimización.," *Revista Iberoamericana de Automática e Informática Industrial*, vol. 4, no. 1, pp. 5-23, 2007.

[4] L. Biegler, "An overview of simultaneous strategies for dynamic optimization," *Chemical Engineering and Processing: Process Intensification*, vol. 46, pp. 1043-1053, 2007.

[5] M. Millán, "Intégration du design et de la commande optimale: Application à la distillation réactive," Université de Pau et des Pays de l' Adour and Université de Los Andes, France and Venezuela, 2005.

[6] R. L.-N. de la Fuente and A. Flores-Tlacuahuac, "Integrated Design and Control Using a Simultaneous Mixed-Integer Dynamic Optimization Approach," *Ind. Eng. Chem. Res.*, vol. 48, no. 4, pp. 1933-1943, 2009.

[7] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, pp. 25-57, Apr. 2005.

ARCHITECTURE AS INFORMATION PROCESSING

Manfred Wolff-Plottegg
Institute of Architecture and Design, Department of Building Theory and Design,
Vienna University of Technology;
Vienna, 1040, Austria

ABSTRACT

Architecture seen as information processing, taking place on an open communication platform with a plenitude of new planning approaches, is becoming operative, methodical and systemic.
The Institute of Architecture and Design, Department for Building Theory and Design, at the Vienna University of Technology is focusing on the redefinition of functions and planning methods, simulation and 3D-modeling are basic tools.
Traditional planning conceptions: pragmatic target definition, auratic objects, beautiful buildings, optimisation, specialisation, autocratic and determinist processes are no longer state of the art. It is therefore obvious that the latest developments in science and technology should be exploited for the relatively slow architecture medium.

Keywords
Architecture, Generator, Hybrid, Information, Process, Second Life, Simulation, 3D Modelling;

1. INTRODUCTION

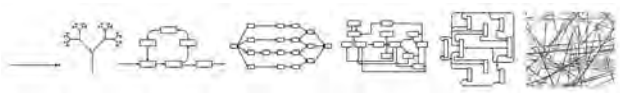
The Institute of Architecture and Design, Department of Building Theory and Design analyses current developments in architecture by means of computer-aided design, putting a main focus on functions. By applying higher-level methods using web-based information and communication technology during the planning process, on the one hand new architectures are produced in the design stage, whilst on the other hand an innovative and more comprehensive approach to computer-based methods in architectural planning is developed. The basic principles of architecture algorithms on the technological planning level are thus extended to include the end user scripting. Architecture, computer and www are regarded as being equal information processing media. They are information editors.
The theory related to this management of planning was published in "ARCHITEKTUR-ALGORITHMEN" [1], "HYBRIDARCHITEKTUR & HYPERFUNKTIONEN" [2], "ARCHITECTURE … SCRIPTING" [3] . The installation HYPER HYBRID GENERATOR was developed by Manfred Wolff-Plottegg and Jochen Hoog (Institute of Architecture and Design, Department of Building Theory and Design) together with the programmers Lukas Ofner and Johannes Sperlhofer (Faculty of Computer Science) and was presented for the first time at the Biennale in Seville 2008. 3D modelling, self-regenerative random architecture is constantly generated on the Second Life platform. The potential of Second Life as a communication platform simulates beyond normal visualisation and surfing.

2. THE EXTENDED PLANNING ENVIRONMENT

Planning is the name of the game: planning procedures, planning methods, basic planning principles, planning management, planning of the nth order, the planning of the planning and thus project monitoring. Traditional planning strives to serve the assignment of requirements > fulfilment; a predetermined spatial programme is followed by the fulfilment of functions. Every linear assignment goes with an encapsulated function!…the kitchen is for cooking…the bed in the bedroom is for sleeping…the flat in the housing estate …the vacuum-cleaner for cleaning …The specified target definition has to comply with an optimised implementation ruling out any deviations. This planning behaviour is determinist; today's planning > tomorrow's concreting > to be utilised as long as possible (sustainability). That architecture of rigid elements (of forms, proportions and suitability) and function separation corresponds to the above paradigm.
Reality of everyday use, constant change in requirements and shifts due to economic frame conditions, make insisting on former decisions obsolete. What is needed, is an architecture of mutation and the potential of a process-controlling architecture. An architecture of processes no longer defines deterministic objects, but simulates developments, chains of functions, permanent changes.
The perspective in architectural project management, especially the Theory of Function, has been newly contextualised, changed and opened up. After the paradigm shift from determinism to an open system, architecture is not so much involved with itself (function fulfilment, material, and form), but rather operates with planning processes and interprets architecture as being process-oriented. The product / object element / building is relativized and the proceeding itself (production methods, inclusion of the producers, the users and the extended fields of reference) is redesigned in a process-like way. Instead of optimising spaces for a specific function, a building is organised as a capacity building. Capacity buildings react to the reality of a constantly changing use. This would mean that the main focus

would lie on hybrid utilisation in its architectural function, instead of on rigid buildings.
Today, even urban development is forced to deal with process control. Architecture is no longer a building or an object; urbanism is no longer a function, but an organ. We can promote the aspect of architecture as a medium (up to now very slowly) and as a process in a specific way, thus creating architecture of acceleration. We shall then realise that this kind of purported architecture has little to do with personal preferences (in harmony with belly and brain), but rather with "external" procedures and more global system variables: product > tool > tool-machine > program > program control > system control.… After cause and effect > participation and interaction > basic principles, paradigms > subsystems, system changes, system planning > open systems.

3. FROM LINEAR PLANNING TO NETWORKING



1.) Systematic diagrams of different planning processes

Non-sequential reading – after linearity, tree structures, closed loops, feedback loops, fractals, chaos, self-generating systems are now available, beyond process control, architecture of avatars, the vision of autocatalytic components:

from determinist target definitions to open systems
from buildings to variable utilisations
from variation of form to variation of function
from autocratic planning to interconnected autocatalytics
from predefined economically planned regimes to democratic participation and further to an interactive multiplayer interface
from passive users to active planners
from typologies to procedures
from objects to processes (from pictures to films)
from planning of a finished status to planning of development potential
from buildings to process architecture and further to process control
from separation of functions to networking and further to hybrids
from specialisation to variability / flexibility to buildings at the disposal
from mono function to multifunction and further to hyper function

All these developments are based on the new system theories of the 20th Century – cybernetics, fractals, chaos, fuzzy logics, game of life, constructivism, complexity, surplus, autopoiesis, etc. – which accompany the paradigm shift. On the operative level, the computer serves as an instrument, especially in planning architecture by means of new display technology, CAD and animations for simulation and parameterisation, by exemplary random generation. LAN, WAN and www networks grant access to new communication platforms.

4. ARCHITECTURE, SCIENCE, TECHNOLOGY, MEDIA

Drawing techniques have always exerted an influence on architecture; the central perspective determined the dominance of the front main façade in the Renaissance Period, and later on, following the two-point-perspective, the side façade became more important. Today, the splines and the nurbs determine free form geometry with bubble and blob architecture in its wake. Scientific results too, have always had an influence on architecture: Johannes Kepler discovered the laws of elliptic planetary motion and shortly afterwards, the first elliptic domes were built ….
After the eclectic canon of styles and the reduction to the narrow vocabulary of form of the Classical Modern style, computer-generated form-finding changed architecture in a revolutionary way; form follows function has become obsolete. Over the past 25 years, we have proved that we can generate all kinds of forms. All forms are being built: high rises like cucumbers or ice lollies – independent of their function. In order to integrate uses & functions into planning more closely, a further feature of CPU (central processing unit of the computer), information processing and process control, can be activated. After generating forms for architecture, the utilisation procedures are then to be controlled. To steer architecture away from formal design, it is necessary to regard planning as information processing; architecture is an information editor. By applying new technologies and operative process control, architecture returns to its origins in science and technology.

5. ARCHITECTURE AS INFORMATION PROCESSING

Electronic telematic communications override local correlations. People are constantly using their mobile phones or surfing on the web, regardless of personal presence and/or local circumstances. The presence of different spheres of reality is represented by binary streams of signs (bit strings) acting as a vocabulary for communication – interconnecting neurons as well as computers. Information and visions and ideas about the real world and the world in our brain, as they are actually encoded by biological organisms, and the world of digital data processing are interconnected by information processing as a common denominator. Generating images, spaces or architecture must no longer be done manually, mentally or anthropocentrically / expressionistically, because this can be carried out on a data processing / data manipulation level. The exclusive handling of classic elements like columns, walls or ceilings for the purpose of controlling utilisation is transferred into the software sphere.

2.) Web of Life, ZKM Karlsruhe 2001

The www in particular, as a non-target-oriented, collective generator of information, changes the disposability and structure of information; the changed potential controls communication and work procedures. So this appears to have a stronger influence than the Theory of Interfaces. It is not the input (individual will), but the procedure that influences the output. Architecture has always been a global medium, its function being – comparable to electronic media – an operative system to control functions.

Hybrid architecture is to be understood as a self-organised process-orientated planning system for self-organised process-controlling architecture. This planning method is thus www-adequate and corresponds to today's information processing (access and processing) … it works at high speed and is non-selective … and is an extrinsic procedure for the planner, but as far as the system is concerned, it is intrinsic … it has versatile logics … has contingency as a consequence of the hybrid combinatorics … is quasi a compiler language … the operating system of hybrid architecture.

6. COMMUNICATION PLATFORMS

The internet provides manifold possibilities of communication: www (hypertext documents), e-mail (asynchronous, private), chat (synchronous, conference), forums (multidirectional, public), blogs (monodirectional, public), VoIP (voice) and video conferences (image).

The developments of web 2.0 like for example "social networking" (facebook, StudiVZ, Xing or Twitter) and personalised internet (delicious, netvibes, google applications, flickr and youtube) offer openly designed interfaces (api's – Application Programming Interfaces). Here the democratic basic principle of the web is evident.

A major characteristic of modern communication is the openness of its communication platforms. Information flow is no longer mono or bi-directional and does not distinguish between author and user. Everyone can be a player or a planner, a quasi avatar as part of the system. Nobody is passive. Data exchange is multiple (networking); everyone can upload and download – multiplayer media. The autocratic privilege is followed by interaction, participation and democratisation of communication; we are in the "youniverse". Multi-user virtual environments (MUVE) or massive multiplayer online role playing games (MMORPG) sum up those communication technologies and position them as three-dimensioned entities in a virtual world.

Computer games play a major role as a driving force for technological developments (hard and soft). They provide three-dimensional worlds. The 3D-visualisation in cyberspace is one of the reasons for its success as a mass medium (media hype). The initially envisioned virtual playground is now fully acknowledged and used as a space where events and social contact take place. The virtual world of telematic communication is as real as Ludo.

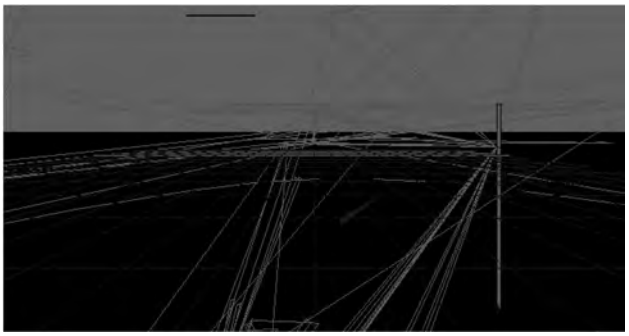
Second Life unites the world of www and games. In this virtual world, everyone can additionally and actively build and change three-dimensionally, can work on spatial design individually or on a collaborative basis. Basic features of the "real world" can be activated and properties (behaviour) can be added via scripting. Second Life is a virtual three-dimensional online platform that enables thousands of users (e.g. 55.275 users on 29/1/2009) to come together at one virtual place. Currently, the software offers the most stable and accepted multi-user virtual environment, free of charge. The reasons for this are its good usability (user-friendly), its balance of graphical display and transmission speed (realism versus spectrum). Last, but not least, it is possible to visit the same 3D-world together with other users, even with different identities.

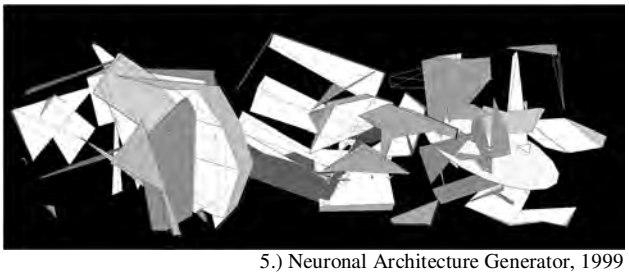
The architect no longer has the sole autocratic planning privilege; everyone can intervene in the real world of Second Life.

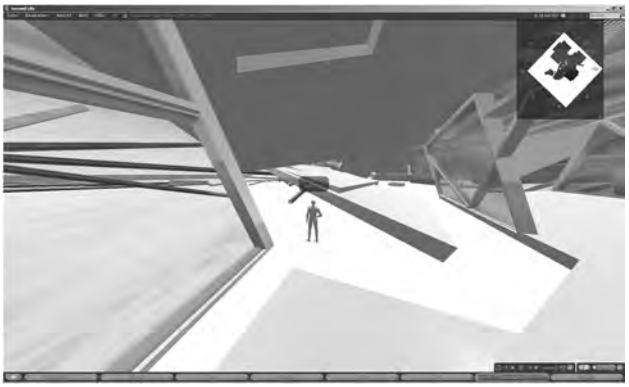
7. ARCHITECTURE GENERATORS

The Hyper Hybrid Generator is the latest product of a series of architecture generators. Its development is based on the shift of conceptual design away from the personal stimulus (brain, handwriting, will to create) to algorithms. Random controlled processes, data processing and data manipulations produce architecture… tirelessly, constantly producing new configurations, a plenitude of various designs, exchangeable and freely accessible. Free interpretability is placed on an equal level with free usability (at the disposal).

3.) Analog Architecture Generator, 1987

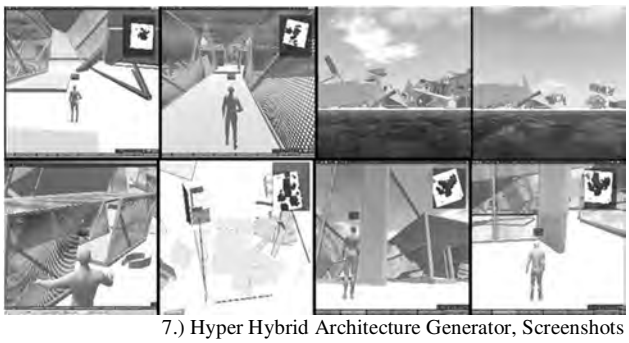

4.) Digital Architecture Generator, 1990


5.) Neuronal Architecture Generator, 1999


6.) Hyper Hybrid Architecture Generator, 2008

Hyper Hybrid Generator,
3D-modeling and Simulation

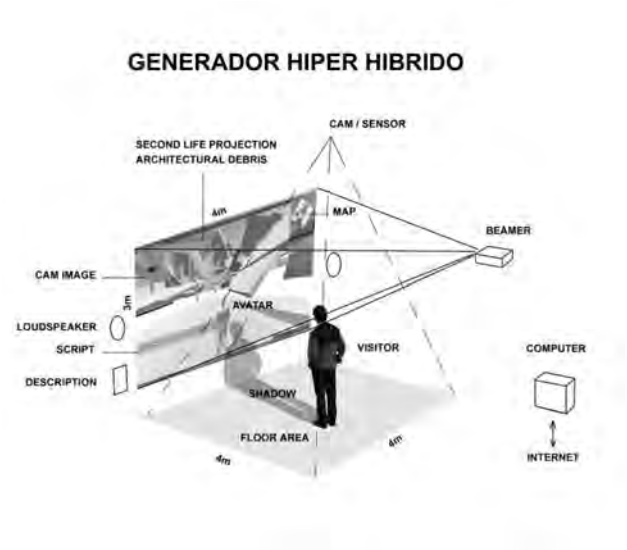
The basic modules of the Second Life platform control the planning process; mutual intercommunication and interaction of all planning participants and components (objects and properties). Construction takes place as a consequence of communication (interaction). Generation occurs in the background on the basis of scripting. It is not the computer but architecture itself which is the interface.


7.) Hyper Hybrid Architecture Generator, Screenshots

On the Second Life Island "Generador hiper hibrido", a script permanently generates components (building debris) which, controlled by a physics engine, fall onto the island. The components quasi automatically create architecture agglomerations, connect to form manifold combinations and constructions, join up to create spatial configurations and form different internal spaces and series of free spaces. After certain time periods (controlled by the position of the avatar in relation to the sensor cupola), individual elements disintegrate and vanish spontaneously and later on fall onto the island again as fragmented components generating new configurations. This happens persistently and can be visited and viewed by every Second Life user per avatar. The diversity of the components plus the three-dimensioned virtual environment permanently offers new spatial experiences.

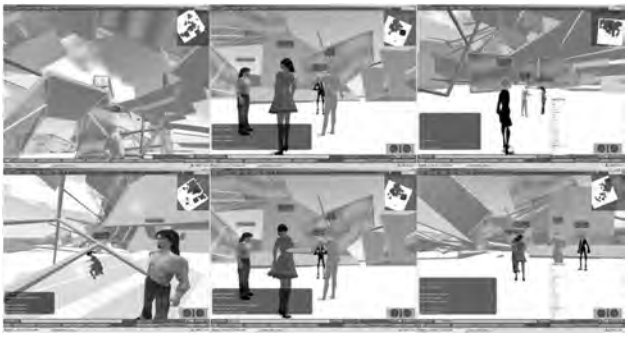
Visitors of BIACS3 can see the island on the screen from the perspective of a different avatar. This avatar also moves through a script, algorithmically controlled, quasi automatically passing through the continuously changing and self-renewing architecture of the island. The avatar's perspective (camera) is positioned such that the avatar is invisible (first person perspective).

As soon as a visitor enters the installation zone (representing the Second Life Island), he/she is tracked by a video camera and his/her movements are captured by means of a video tracking process. The visitor now controls (instead of the algorithmic movements) the avatar analogously. Thus the avatar enters the projection field becoming visible on the island (third person perspective). In the same way the visitor controls his own movements, he also controls the avatar as well as the movement of his own shadow. Additionally, superimposed images of the video camera as well as the background script can be viewed.



8.) Hyper Hybrid Architecture Generator, Setup

Therefore, the installation shows a total of five different representations / appearances / aggregate states of any visitor simultaneously within one configuration in Second Life.
The thus created architecture can be experienced anywhere by the internet. It is a continuous script with instructions for the computer (instead of for a construction company). The genius loci is the computer itself which can be experienced virtually by an alter ego. Visitors change into a cursor and become part of the installation, either as a visitor of the exhibition in Seville or as an avatar in Second Life. Real and virtual manifestations begin to overlap one another.



9.) Hyper Hybrid Architecture Generator, Screenshots

Technical description
It uses the programming language of SL and creates a new kind of morphing by simulating common architectural design processes in which fragments of memories / wishes / dream images would usually initiate new architecture.
The script of the Hyper_Hybrid_Generator generates architectural hybrids. Hybrid results are achieved by random selection and coincidental grouping of different elements (input). The input may consist of architectural debris falling onto the island:

```
for( i = (integer)llFrand(giMax - giMin) + giMin; i >= 0; i-- ) {
objName = llGetInventoryName(INVENTORY_OBJECT, (integer)llFrand(i));
llRezObject(objName, vThisPos, ZERO_VECTOR, <llFrand(360.0), llFrand(360.0), llFrand(360.0), llFrand(1.0)>, giSyncID);
}
```

From a large number of architectural projects individual debris are generated at random by means of a script. They are then placed on the island as a unit under the “parent script” and are subject to a random rotation.

```
state_entry() {
gintHandle = llListen(gintPrimChan, gstrCommander, NULL_KEY, "");
llSetStatus(STATUS_PHYSICS,1);
}
```

After its generation, the “child object” awaits instructions from its “parent script”, like for example from the physics engine integrated in SL, to which it reacts (it falls down).

```
integer iHops = llAbs(llCeil(llVecDist(llGetPos(), gvDestPos) / 10.0));
for( x = 0; x < iHops; x++ ){
lParams += [ PRIM_POSITION, gvDestPos ];
}
llSetPrimitiveParams(lParams);
```

The child-element computes the collision data together with the simulated island or other objects and quantifies the path of fall.

```
if (gvDestPos == llGetPos()) llSetStatus(STATUS_PHYSICS, 0);
```

8. CONCLUSION

The Hyper_Hybrid_Generator is a web-based permanent running script embedded in the multi-user virtual environment of Second Life (SL). In order to avoid the possibility of being moved by other influences, when the element stops moving, it loses its ability to react to the physics engine and awaits its deletion.

9. REFERENCES

[1] M. Wolff-Plottegg, **ARCHITEKTUR-ALGORITHMEN**, Wien: Passagen Verlag, 1996.

[2] M. Wolff-Plottegg, **HYBRID ARCHITEKTUR & HYPER FUNKTIONEN**, Wien: Passagen Verlag, 2007.

[3] M. Wolff-Plottegg, Hrsg., **Architecture ... Scripting**, Wien: Sonderzahl, 2011.

E-didactic Strategies with Peer Feedback Processes for Online Learning

Lisbeth AMHAG
Faculty of Education and Society, Malmö University
Malmö, 205 06, Sweden

ABSTRACT

This article focuses on strategies for how online course outlines can be designed to improve the use of collaborative peer feedback in distance education and how different dialogic patterns can be identified. Two separate studies were conducted to investigate students' use of own and others' texts meaning content in their peer feedback as a tool for learning and how the content can be analysed. Data were collected from two student groups; one from 40 student teachers' peer feedback and discussions of four assignments (N=759) from two 15 credit web-based courses; and one from 30 student teachers' argumentations and discussions of one assignment (N=253) from one 15 credit web-based course. An analytical framework, based on Bakhtin's theories of dialogues in study one, and combined with Toulmin's argument pattern (TAP) in study two, are employed to assess the quality of the meaning of peer feedback and argumentations. A close investigation of the dialogical patterns shows the extent to which students distinguish, identify and describe the meaning content in their peer feedback that emerge in collaboration with other students in an online setting as an important aspect. The dialogue patterns that developed are illustrated in selected excerpts.

Keywords: *Computer Supported Collaborative Learning; Computer-mediated Communication; Distributed Learning; Interactive Learning; Discourse analysis.*

1. INTRODUCTION

The background of this research is to meet the increased request on distance education with high quality and performance level in professional degree programs. Another setting is to to improve the use of peer feedback in distance education, which can promote students' learning and development, as well critical ability. Peer feedback uses in this study as information provided by students with aspects of each one's understandings as well alternative strategies and solutions based on literature. University assessments such as reports, articles and project presentations are more complex work. Students need to have emphasis on the learning processes in writing, inquiring and problem solving. A practical benefit of implementing peer feedback is that the feedback becomes available during the learning process and in much larger quantities, than the teacher could ever provide alone.

A clear trend is that distance education in whole or in part is organized with support of online learning environments, is steadily increasing and is currently the higher education sector that is growing fastest (ICDE, 2009). The development of distance education has thus resulted in a *new way of teaching* and to learn *in* and *with*. The importance of developing critical reasoning and self-reflective learning has been highlighted in several studies within the field of distance learning and education (e.g. Vonderwell, 2003; Finegold & Cooke, 2006; Wegerif, 2006; Swann, 2010). While many models are available for content analysis of asynchronous discussion groups and the design of online activities to promote e-learning (De Wever et al., 2006; Schrire, 2006; Strijbos et al., 2006; Weinberger & Fischer, 2006; Sun et al., 2008), there are considerably fewer models that analytically investigate the meaning and quality of peer feedback.

2. PREVIOUS RESEARCH

A general overview of the state of research in the last decade of online learning shows that the research design in most studies in the area primarily involved experimental, descriptive and iterative studies (Suthers, 2006). Either have researchers examined the technical opportunities, how individual learning can be described and explained and compared how learning is developed in campus courses and in online courses. A frequent pedagogical problem in web-based education, discussed by Stahl and Hesse (2008), and Garrison and Arbaugh (2007), is that students and teachers mainly focus on the individual learning process. Self-regulated learning through using web-based tools and wireless technology module systems on their own is not nor enough. Another educational problem, described by Stahl and Hesse (2008), is that students and teachers tend to focus on procedural learning and ignore the conceptual learning intended by the curriculum designers. These courses tend to result in relatively superficial or unreflective re-productions among both individual students and student groups. The dialogues investigated in these studies soon assume the character of transmitting 'information'. They become a simple confirmation of what others already have written, and therefore the participants do not succeed in developing deeper knowledge construction.

When looking for studies on peer learning, Dochy et al. (1999) and Topping (2005) emphasize that by assessing the work of fellow students, students also learn to evaluate their own work. Producing and receiving peer feedback have a considerable profit in order to account for the time and effort that is required to engage in the learning process of peer feedback. This view is also supported by Shekary and Tahririan (2006), who state that peer assessment in language-related episodes (LRE) resembles any other form of collaborative learning. LRE are mini-dialogues, in which students ask or talk about language, or explicitly or implicitly questions of their own language use or that of others. The result of the study suggested that it offers students the potential to develop new knowledge and understanding. Most benefit to students was the nature of acceptance, not its mere presence. Another studies (e.g. Dysthe, 2002; Amhag & Jakobsson, 2009; Amhag, 2010; 2011) illustrate the potential and voices in online peer feedback as the range of meaning-mediating possibilities, as an active tool with self-reflective and interdependent arguments and thoughts, where each student can contribute with his or her own expertise and receive new information and experiences from others. Compared with Saunders (1989) combination of two factors: 1) what students do together with the tasks assigned to them as collaborators, and 2) the roles and responsibilities the students assume as collaborators and the interactive structure underlying the activity, is peer assessment often more limited than other forms of collaborative learning in the sense that it generally offers a lower degree of interactivity. He calls this process as "co-responding" and affects students' possibilities for interactive meaning making and collaborative knowledge construction.

In order to shed more light on the meaning and quality of collaborative peer feedback online, this study aims to investigate in two studies; one study with response activity and one with argumentative activity, how online course outlines can be organized to improve students' use of their own and others experiences, texts and productions to develop critical thinking, as well as peer feedback ability, individual as collective, as a

tool for learning. Additionally, the aim is to develop patterns of qualitative peer feedback, who are allowing to distinguish, identify and describe the meaning content that emerge in collaboration with others in an online learning setting, both directly and retrospectively as an important aspect. Response ability is here related to a concrete answer to a specific text in order to become a more conscious writer. Argumentative ability is related to the process of assembling and reassembling different components of the students' own and others´ words and meanings. There is also a need for the students to understand the "ground rules" of peer feedback and to respond, argue and discuss with one another in a reasonable way. According to Scheuer et al. (2010), students not only need to "learn to argue" or "learn to respond", they also need to learn good responding and argumentation practices, through aspects of each one's understandings as well alternative strategies and solutions about specific topics. In other words, collaborative peer feedback for online learning in the sense with practicing of responding and argumentation skills that supports critical thinking, as well as other important aspects in collaboration with others. The research question in this study is:

- How can the quality of peer feedback be analyzed and practiced online in which students' in collaboration with others can use own and others' texts meaning content as a tool for learning?

3. THE STUDY

Method and data collection

Each of the two present studies follows one student group. The first study monitored 40 student teachers (of which 22 were women and 18 were men), who were studying teacher education at distance as part of the credits they needed in order to become qualified teachers. During the six courses, the students continued working as teachers in upper-secondary schools in Sweden. The majority had already worked between one and five years, while around one fifth had worked for more than five years. Data were collected from the student teachers' peer feedback and discussions which was given as part of the first two consecutive 15 credit web-based courses called *Teacher Assignment* and *Learning and Development* with two assignments in each with peer feedback activity (N=759; 350 in course 1; 409 in course 2).

The second study monitored 30 student teachers (of which 19 were women and 11 were men) at a Swedish School of Education. Data were collected from the student teachers' peer feedback and argumentations of one assignment (N=253) from the first 15 credit web-based course *Teacher Assignment*. In the first course assignment, about school development in their subject, the students had trained providing peer feedback in their groups. The study focuses the second course assignment, where the students worked both individually and collaboratively with 31 cases of teacher leadership (one official case and one from each student).

In both studies the students were divided into groups, with five to seven individuals in each. Each group included both men and women. The students first submitted their own particular contribution to the assignments. Afterwards, they had to give peer feedback and discuss in study one and in study two argue in their peer feedback and discuss the contributions of the other members of their group, over a period of a week. The purpose of the assignments was to start a discussion and an argumentation concerning different solutions to the underlying problems in the content of the assignments and relate to own experiences and literature.

Analysis of peer feedback in study 1

The analysis and interpretation of the students' meaning content in the online peer feedback in study 1, the following quotation by the Russian linguist Mikhail Bakhtin´s theoretical framework of dialogues (1981; 1986, 2004a; 1986, 2004b) was used and implemented: "as a neutral word of a language, belonging to nobody, as an *other's* word, which belongs to another person and is filled with echoes of the other's utterance; and, finally, as *my* word, for, since I am dealing with it in a particular situation, with a particular speech plan, it is already imbued with my expression" (1986, 2004b, p. 88). The first aspect is the *neutral* word that reflects the world of others, in the sense of more general meanings. This word is not built on specific words from literature or personal experiences. The second aspect is *others´* word, which is filled with echoes of others' voices, based on others' experiences and reasoning from others´ texts, including references and paraphrases of other people's words from literature. Others´ words have been created in another context. They are negotiated, and confirm a certain meaning relating to the argument at hand, but they do not originate in the person him/herself, and are not necessarily related to the person's own experience. Finally, the third aspect is *my* word, because the speaker or writer has experience of a particular situation, and connects a certain line of reasoning with internal reflections and feelings. A summary of Bakhtin´s multiple voices is outlined in Figure 1.

Multiple voices	Patterns of meaning in peer feedback
1. Neutral word of a language	• reproduces other people's world view • aims at any general meanings and thinking • is not built on words from literature or personal experiences
2a. Others' word	• reproducing reproductions of previous voices • contains echoes of other voices, dialogic overtones • explicit voices can be heard presenting voices • the voices do not originate in the person himself
2b. Others' word from literature [my addition]	• reproducing reproductions of other authors' voices • drawing on other subject experience and reasoning from other texts • references to and paraphrases of other people's words from literature, expressing these in their own words • creating, negotiating and confirming the meaning
3. My word	• carries internal reflections and feelings • contains their own and others' voices, arguments, justifications, contradictions, experience etc. as appropriated to the speaker's own words • constructs and reconstructs a mutual meaning or a part of it • creating, negotiating and confirming the meaning

Figure 1: Summary of multiple voices and patterns of meaning in peer feedback in study 1.

The analysis phase involved taking into account that every utterance, spoken or written, always is formed by a voice, and expressed from a particular viewpoint or perspective (Bakhtin, 1980, p. 293). Voice shall here be understood as person's utterance, including meaning of own and others' words from different contexts, and expressed from a particular viewpoint or perspective. Bakhtin (1981, p. 427) talks about a `discourse´ [Rus. slovo] in the dialogue, and points to social and ideological differences within a single language. In Bakhtin´s account, the notion of utterance is inherently linked with that of voice. It is "the speaking personality, the speaking consciousness. A voice always has a will or desire behind it, its own timbre and overtones" (1981, p. 434). In other words, the utterances contain

dialogic overtones, which can, for example, be composed of assertions regarding the world, ontological conclusions, or hypotheses regarding a phenomenon. Bakhtin emphasises that language has multiple functions, and every utterance, with its attitudes and values, places humans in a cultural and historical tradition.

Analysis of peer feedback in study2

In the analysis and interpretation of the students' meaning content in the online peer feedback in study 2 was Bakhtin's theoretical dialogic framework combined with Toulmin's argument pattern. Toulmin (1958, pp. 98, 101, 103) describes how writers and readers can deal with texts, and how they can use the resources of texts to determine what they mean – or rather, some possible meanings – and how it can be achieved with an argument model containing six elements. Three elements are mandatory, while the remaining three are more voluntary or optional, since they occur often, but not always. The basic argument model consists of three mandatory elements: C (*claim*), D (*data*) and W (*warrant*). The extended argument model includes three more optional elements; Q (*qualifier*), R (*rebuttal*) and B (*backing*). The task is to show students how to present their ideas in an understandable and coherent manner, based on these data and the claims of the original opinion. A summary is given in Figure 2.

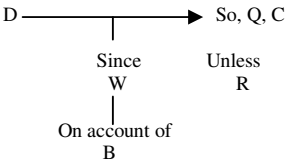


Figure 2. Summary of Toulmin's argument pattern (Toulmin, 1958, p. 104)

The first mandatory element, *claim* (C), is a superior standpoint, with a relationship to any determination or assertions about what exists, or the justification of the norms or values that people hold or desire for acceptance of the claim. The second mandatory element, *data* (D), is the information which the claim is based on, and may consist of previous research, personal experience, common sense, or statements used as evidence to support the claim. The third mandatory element, *warrant* (W), is explicit or implicit argument that explains the relationship between data and claim, for example, with words such as *because* or *since*. The first optional element, *qualifier* (Q), is related to the claim, and indicates the degree of strength in the claim of using peculiar comments, for example, with words such as *probably*, *maybe*, *therefore* or *so*. The second optional element, *rebuttal* (R), is connected to the *qualifier* (Q), providing statements or facts that either contradict the claim, data or rebuttal, or qualify an argument, with words such as *but* and *unless*. The third optional element, *backing* (B), can be connected directly to the warrant (W), with often implicit motives underlying claims, expressed with words such as *because of* or *on account of*. According to Toulmin, all terms of the basic argument model (C, D & W) are required to describe or analyse the argument. A revised version of Toulmin's argument pattern with the mandatory and optional elements, inspired by developments of the specific features in the TAP made by Kneupper (1978) and Simon et al. (2006), is given in Figure 3.

The first phase of analysis was focus placed on specific features: the extent to which students had made use of Toulmin's mandatory elements; data, claims and warrants, the optional elements; qualifiers, rebuttals and backings (which in English are often presented by characteristic words, such as *because, so* or *but*), and how the different elements in the same argument are related to each other. However, this phase of the analysis does not show how the elements relate, explicitly or implicitly, to other arguments in a chain of utterances. The dialogical interaction with other claims, data, warrants, etc. cannot be distinguished, as such, in the first phase of analysis, or the creation of meaning, when two or more voices or discourses encounter each other, as Bakhtin emphasizes. The second phase of analysis involved discovering and identifying another set of relevant aspects, using an approach based on Bakhtin's theories of *double-voiced discourse* (1984, p. 185), which inevitably occurs under conditions of dialogic interactions. On the one hand, Bakhtin broadens the concept of language, by pointing to the fact that dialogic interaction and a dialogic relation are inherent to all communication. On the other hand, Toulmin's practical argument pattern makes the structure visible that connects various data, claims and support for the arguments to each other. Using a combination of these perspectives thus makes the analysis of written asynchronous responses and arguments more explicit, reliable and valid.

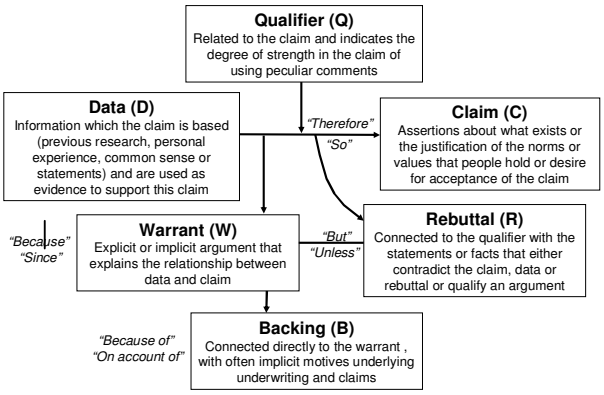


Figure 3. Revised version of Toulmin's argument pattern (TAP) in study 2.

The first phase of analysis was focus placed on specific features: the extent to which students had made use of Toulmin's mandatory elements; data, claims and warrants, the optional elements; qualifiers, rebuttals and backings (which in English are often presented by characteristic words, such as *because, so* or *but*), and how the different elements in the same argument are related to each other. However, this phase of the analysis does not show how the elements relate, explicitly or implicitly, to other arguments in a chain of utterances. The dialogical interaction with other claims, data, warrants, etc. cannot be distinguished, as such, in the first phase of analysis, or the creation of meaning, when two or more voices or discourses encounter each other, as Bakhtin emphasizes. The second phase of analysis involved discovering and identifying another set of relevant aspects, using an approach based on Bakhtin's theories of *double-voiced discourse* (1984, p. 185), which inevitably occurs under conditions of dialogic interactions.

4. RESULTS

Results in study 1

The two studies led to two main sets of results. First in study 1, that the students' task-related meaning content and multiple voices in the responses gradually change character, as the personal, social dialogic interaction in course 1 becomes more objective and task-related during course 2. When the voices are "half someone else's", they can also become "one's own", when the students appropriate the words of others, and invest them with their own intentions and capabilities. When students can communicate their knowledge in more insightful ways than before, they become aware of what is understandable or incomplete. This process *generates new meaning* between

writer/author and reader/addressee. In the following excerpt from course 2, *Learning and Development*, we can observe that the students' utterances contain examples of all of Bakhtin´s multiple voices. Three student teachers are engaged in a discussion about the impact the pupils' social situation may have on their learning and development. The students have studied literature on socio-cultural learning and development processes. In the assignment, they are requested to describe a concrete teaching situation that is linked to the literature, and to reflect on the course of events in the situation. In this assignment, the multiple voices arise primarily from the students' own examples, and from how they are able to use the literature to analyse the described teaching situation.

1. Harry	[…] Maria is a girl who has chosen the Vehicle Programme because she imagined a future as a driver. She is keen and forward and really wants to learn to drive a truck. Maria has, as I see it, two characteristics that have not located her in a barrel on the Vehicle Programme. Firstly, she is female. Prejudices are many from both classmates, other pupils and, unfortunately, also teachers. Truck driving is not for "womenfolk". This has certainly meant that Maria has been viewed as less knowledgeable right from the first years at upper secondary school. Teachers who have prejudices against certain pupils, regardless of the type of prejudice, it must be difficult, if not impossible to practice the kind of dialogic teaching that Dysthe (1996) describes. (One example she highlights is Ann in the class of Baywater who really understood the importance of authentic issues in the classroom).
2. Carl	Hi Harry! I think you grabbed the issues that Maria had in a very exemplary manner. You took not only time to show everything from scratch, you might also build up the confidence of Maria so that she passed the driving test and could proceed in training with the others.
3. Eva	Hi Harry and everyone else! I agree with Carl in a lot, your, Harry, exemplary manner gave Maria confidence back. Just by being taken seriously and therefore being respected, a pupil shows respect back. And it is good to be respected, isn't that what all teachers want most often? The previous teacher driving attitude is somewhat to my surprise something I also have encountered among other teachers. It is assumed that one's own way of teaching is the right thing ("It has been operating for 100 years before!") and that some of the pupils are "uneducated". For obviously these pupils can not learn what other pupils can. The specific learning style works for some pupils and not for everyone, they don't want to think about it, and blame everything on heavy workload, lack of time or all upper secondary schools forms. (Just school forms, I have noticed is a popular target to blame ...) We are probably ourselves here in the group a bit envious of your situation with the luxury to teach individually, with one student at a time. But on the other hand, you said that you have been teaching all day and have no time for planning, so maybe we should not whine... Fun to read!
4. Harry	Hi Eva! Yes, maybe you are right that it is a luxurious situation with one to two students at a time, but I can assure you that I am quite out of the box after a working day. To move around with an 18-year-old in a carriage that is 22 meters long and weighs 35-40 tons requires my undivided attention and concentration throughout. It is like driving myself while coaching. But it is a great advantage with only 1-2 students at a time. I come very close to the pupil, and can devote myself to one pupil at a time. It is an advantage.

The initial argument raised by Harry: "[...] Truck driving is not for `womenfolk´[...]" may be an example of *others'* words reproduced convincingly by Harry. The utterance is likely to have been expressed by another person, and thus contains echoes of other voices, something Bakhtin (1986, 2004b) describes as *dialogical overtones*. It can also be interpreted as the manifestation of *written polyphony*, because the argument using an other's word has the same value or authority as Harry's utterance above (Møller Andersen, 2002; 2007). As a conclusion to the initial argument, Harry writes: "[...]. This has certainly meant that Maria has been viewed as less knowledgeable right from the first years of upper secondary school [...]". This claim can be interpreted as both Harry's *own* words, based on his own reflections and the words of *others´,* based on the arguments of others from the school, but which Harry appropriates to become his *own.* Harry continues in the course assignment with more *neutral* words when he writes: "[...] Teachers who have prejudices against certain students, regardless of the type of prejudice.... [...]". This statement is neutral in the sense that it contains notions which are generally approved by teachers and colleagues. The view expressed can therefore be interpreted as not over-built with Harry's own words. He continues his argument by writing: "[...] it must be difficult, if not impossible to practice the kind of dialogic teaching Dysthe (1996) reports [...]". This claim can be considered as referring to evidence of *others'* words from the literature. It thereby indicates that Harry has insight into certain characteristics of a dialogic classroom, and that he relates his ideas on what classrooms look like – or should look like – on words of *others´* from the literature. In this case, the reference is to the literature of Dysthe, describing the classroom of multivoicedness. She gives examples of how the teacher Ann in the class of Baywater has authentic and open questions, which means that the pupils can think and freely articulate what they understand, regardless of whether their suggestions are simply temporary opinions, or if the responses are inadequate. Harry's knowledge of what a dialogic classroom is can be seen as an example that learning is not created from a single word or from the language system alone, but in the relationship and interaction between his *own* words and *others'* words from the literature. We are thus in the presence of a form of *intertextuality* in Harry's contribution, with different subject experiences and reasoning from other texts. In the responses, Carl (2) confirms with *neutral* words that he has read Harry's answer to the course assignment, but also uses to some extent his *own* words, when he writes "[...] you might also build up the confidence of Maria [...]". But Carl does not broaden and develop further Harry's arguments about Maria's situation, by using his own words, from experiences or from the literature. Eva (3) reflects more of the mutual respect between pupils and teachers. This may be considered as an example of her *own* words to express her own problems of workload and lack of time, as well as echoes of other's teachers' voices concerning the importance of being respected as a teacher. The tension or potential difference between Harry and Eva with respect to the "luxury" of a situation where the teacher only has to teach one pupil at a time, can be seen as an example that they are both shareholders and co-authors of a common narration (Rommetveit, 2003), and that they become aware of each other's words.

Results in study 2

The second set of result in study 2 shows the importance of dialogic interaction with both responding and argumentation activity. The students had before trained providing feedback in their groups. In the following excerpt, the argument patterns in written, asynchronous arguments will be distinguished, identified and described, as well as the dialogical relations between written contributions. The students' names are fictitious. The excerpt, in Figure 4, is from a discussion between Chris and Katrina.

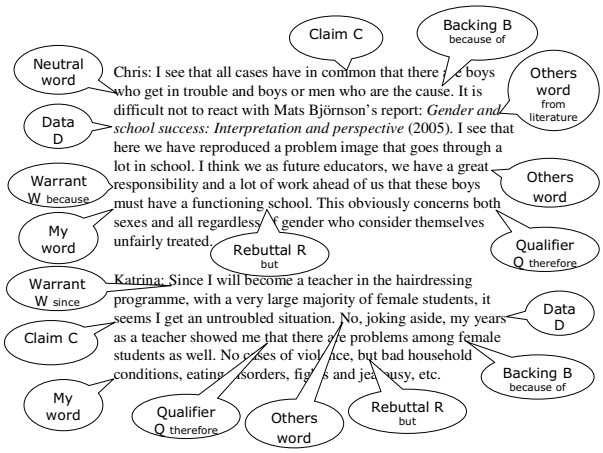


Figure 4: The specific elements and words with different voices in the meaning content of the argument.

The discussion illustrates the significance of comparing opposing arguments between classmates' cases of teacher leadership, when Chris starts a discussion about what he considers has been developed in their collective contributions. Chris' standpoint that "there are boys who get in trouble and boys or men who are the cause", and Katrina's counterarguments that "there are problems among female students as well", are the *claims* in this excerpt. Both claims point to problem areas that exist at school today. The *data* of gender and school success in Chris' statement is supported in the literature, while Katrina's argument is based in personal experiences from her years as an unqualified teacher. The *warrant* in Chris' statement is here also explicit, *because* it explains the relationship between teachers' responsibility to have a functioning learning environment for both boys and girls and by Katrina *since* she will become a teacher with a large majority of female students. Chris writes: "I see that here we have reproduced a problem image that goes through a lot in school". This statement is the *backing* in the argument, *because* the meaning or motivation of the statement can be understood as: What I write is supported by the literature, therefore, I write it in my post. The corresponding meaning found in Katrina's statement can be understood as: What I write is supported by my personal experience. If we look at the *qualifier* of the two claims, it is confirmed in Chris' contribution by all, regardless of gender, who consider themselves unfairly treated, while in Katrina's contribution, the statement applies to female students suffering from bad household conditions, eating disorders, fights and jealousy, etc. Chris' statement that teachers have a lot of work in order to achieve a functioning school is the *rebuttal* in his argument. If we look at the continued discussions between Chris and Katrina, the creation of meaning here also depends on the discourse, with *neutral* word, *others´* word and *my* words, as well as the context in which these voices are expressed. According to Bakhtin (1981, p. 293), the word in language is half someone else's, and becomes "one's own" when the speakers or writers populate it with their own intention and appropriate the words as their own. In the excerpt above, Chris uses words from the literature, while Katrina puts words on her own experiences and ontological conclusions. The utterances thus contain *dialogic overtones,* since they are filled with echoes of other people's words, arguments, evidence and reasoning from other texts (Bakhtin, 1986, 2004b). The mandatory elements, *claim* (on gender-related problems), *data* (from the literature) and *warrants* (concerning the teachers´ responsibility), can here be related to the corresponding *backing*, degree of strength in the *qualifier,* and connecting *rebuttal*. The relation between *neutral* words (with general meanings) can be evaluated with respect to *others´* word (from literature) and experiences. Some are appropriated to become *my* words. The students become shareholders and co-authors in a joint meaning, in which knowledge and understanding develops. In short, the excerpt illustrates the fact that these mutual negotiations emerge in dialogues between students, and their *meaning potentials* arise as the range of *meaning-mediating* possibilities (Rommetveit, 2003). Such negotiations are illustrated in this argumentation about gender and functioning school for boys in trouble, and girls with bad household conditions, eating disorders, or fights and jealousy.

5. DISCUSSION AND ONLINE IMPLICATION

In present two studies, based on Bakhtin's theories of dialogues (1981; 1986, 2004b; 1986, 2004a) in study 1, combined with Toulmin's argument pattern (1958) in study 2, appears a new quality dimension in which the specific words with voices and elements and voices in the online peer feedback – as well as the dialogical relations between them – makes more explicit and more visible. It may be concluded from results emerging in these studies, that using assignments drawing on authentic assignments and cases with collaborative peer feedback, is indeed a way to make the words more genuine and living (Bakhtin, 1984). The peer feedback strategies with group activities over a specific period, where dialogue exchange and collaboration are in focus, opens for the manifestation of *written polyphony*, because the students' independent voices in their peer feedback have the same value or authority as authors in books (Møller Andersen, 2002; 2007). A more complex peer feedback character develops when the content is confronted with others' utterances, consisting of comparing different statements and justifying opposing words and voices. There may be direct and explicit opinions in the contributions and assertions about what exists, or statements that contradict, confirm, complement or develop further. Common sense or implicit or unspoken motives may also be expressed.

Students also learn to evaluate their own work when they are producing and receiving peer feedback. In this particular form of discourse, the students' peer feedbacks consist partly of their own words and voices, and partly of others'. Each peer feedback is an intersection of words, where at least one aspect of others´ words can be read, and each utterance can be considered as an answer to preceding utterances, that is, it has *addressivity* (Bakhtin, 1986, 2004b). This addressivity is made more possible in collaboration with other students and can be compared with Hattie and Timperley´s (2007) three major questions of effective feedback: Where am I going? (What are the goals?), How am I going? (What progress is being made toward the goal?), and Where to next? (What activities need to be undertaken to make better progress?). The questions correspond to the design of *feed up, feed back* and *feed forward* and they are partly dependent on to reduce the gap across the level of task performance, the level of process of understanding how to do a task, the metacognitive process level, and/or the self level. The combination of what students do together with the tasks assigned to them as collaborators, and the roles and responsibilities the students assume as collaborators and the interactive structure underlying the activity offer the potential to develop and expand the space of learning and understanding (Saunders, 1989; van del Pol et al., 2008).

The implications and results that the studies highlights are that it is in collaborative peer feedback understanding of different meaningful meanings is clarified and develops. A strategy to promote collaborative peer feedback with critical review and meta-reflection can be to let a) the students after the peer feedback processes compile their own posts and self-assesses them with further reflection, theoretically and practically. Another option can be to let b) the students compile others' peer feedback and analyze them further, theoretically and practically. A further strategy to promote collaborative and

dialogic exchange may be to let c) peer feedback and critical review of and between students be a part of the examination. These processes creates the conditions for students to find structure and patterns of how peer learning and reasoning can be shaped, negotiated and confirmed "between I and other", in an online context. The dialogue patterns that developed during the two studies provide examples of how the meaning content in collaborative peer feedback can be distinguished, identified and characterised. The analysis offers students, student groups and teachers further insights into how they can use Bakhtin's theories of double-voiced discourse and Toulmin's argument model, and thereby gain greater awareness of how "arguing to learn" and "responding to learn" can be promoted, evaluated and developed in online education at distance.

References

Amhag, Lisbeth. (2010). *BETWEEN I AND OTHER. Web-based student dialogues with arguments and responses for learning*. Malmö University: Malmö Studies in Educational Sciences: Doctoral Dissertation Series 2010:57.

Amhag, Lisbeth. (2011). Students´ Argument Patterns in Asynchronous Dialogues for Learning. In *Research Highlights in Technology and Teacher Education*: SITE Research Book 2011, Ed/ITLib Digital Library, http://www.editlib.org/.

Amhag, Lisbeth, & Jakobsson, Anders. (2009). Collaborative Learning as a Collective Competence when Students Use the Potential of Meaning in Asynchronous Dialogues. *Computers & Education, 52*(3), 656-667.

Bakhtin, Mikhail M. (1981). Discourse in the novel (Caryl Emerson & Michael Holquist, Trans.). In Michael Holquist (Ed.), *The Dialogic Imagination: Four Essays by M. M. Bakhtin* (pp. 259-422). Austin: University of Texas Press.

Bakhtin, Mikhail M. (1984). Discourse in Dostoevsky. In *Problems of Dostoevsky´s Poetics* (pp. 181-272). Minneapolis: University of Minnesota Press.

Bakhtin, Mikhail M. (1986, 2004a). From Notes Made in 1970-71 (Vern W McGee, Trans.). In Caryl Emerson & Michael Holquist (Eds.), *Speech Genres & Other Late Essays* (Vol. 9, pp. 132-158). Austin: University of Texas Press.

Bakhtin, Mikhail M. (1986, 2004b). The Problem of Speech Genres (Vern W McGee, Trans.). In Caryl Emerson & Michael Holquist (Eds.), *Speech Genres & Other Late Essays* (Vol. 9, pp. 60-102). Austin: University of Texas Press.

De Wever, B., Schellens, T. , Valcke, M. , & Van Keer, H. . (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education, 46*(1), 6-28.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education: a review. *Studies in Higher Education, 24*(3), 331-350.

Dysthe, Olga. (1996). *Det flerstämmiga klassrummet - att skriva och samtala för att lära. [The classroom of multivoicedness - write and talk for learning]*. Lund: Studentlitteratur.

Dysthe, Olga. (2002). The Learning Potential of a Web-mediated Discussion in a University Course. *Studies in Higher Education, 27*(3).

Finegold, Adam R.D., & Cooke, Louise. (2006). Exploring the attitudes, experiences and dynamics of interaction in online groups. *The Internet and Higher Education, 9*, 201-215.

Garrison, D. Randy, & Arbaugh, J.B. (2007). Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education, 10*, 157-172.

Hattie, John, & Timperley, Helen. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81-112.

ICDE, International Council for Open and Distance Education. (2009). *Global Trends in Higher Education, Adult and Distance Learning*. Retrieved 2011-03-13. from http://www.icde.org/filestore/Resources/Reports/FINALICDEENVIRNOMENTALSCAN05.02.pdf.

Kneupper, Charles W. (1978). Teaching Argument: An Introduction to the Toulmin Model. *College Composition and Communication, 29*(3), 237-241.

Møller Andersen, Nina. (2002). *I en verden af fremmende ord. Bachtin som sprogbrugsteoretiker*. Denmark: Akademisk Forlag.

Møller Andersen, Nina. (2007). Bachtin og det polyfone. In Rita Therkelsen, Nina Møller Andersen & Henning Nølke (Eds.), *Sproglig polyfoni. Tekster om Bachtin og ScaPoLine*. Aarhus: Aarhus Universitetsforlag.

Rommetveit, Ragnar. (2003). On the Role of "a Psychology of the Second Person" in Studies of Meaning, Language, and Mind. *Mind, Culture, and Activity: An International Journal, 10*(3), 203-218.

Saunders, William. (1989). Collaborative writing tasks and peer interaction. *International Journal of Educational Research, 13*, 101-112.

Scheuer, Oliver, Loll, Frank , Pinkwart, Niels , & McLaren, Bruce M. (2010). Computer-supported argumentation: A review of the state of the art. *Computer-Supported Collaborative Learning* (5), 43-102.

Schrire, Sarah. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education, 46*(1), 49-70.

Simon, Shirley, Erduran, Sibel, & Osborne, Jonathan. (2006). Learning to Teach Argumentation: Research and development in the science classroom. *International Journal of Science Education, 28*(2-3), 235-260.

Stahl, Gerry, & Hesse, Friedrich (2008). The many levels of CSCL. *International Journal of Computer-Supported Collaborative Learning, 3*(1).

Strijbos, Jan-Willem, Martens, Rob L. , Prins, Frans J. , & Jochems, Wim M.G. (2006). Content analysis: What are they talking about? *Computers & Education, 46*(1), 29-48.

Sun, Pei-Chen, Tsai, Ray J., Finger, Glenn, Chen, Yueh-Yang, & Yeh, Dowming. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education, 50*(4), 1183-1202.

Suthers, Daniel D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*(3).

Swann, Jennie (2010). A dialogic approach to online facilitation. *Australasian Journal of Educational Technology, 26*(1), 50-62.

Topping, Keith. (2005). Trends in Peer Learning. *Educational Psychology, 25*(6), 631-645.

Toulmin, Stephen E. (1958). *The uses of argument*. Cambridge, England: Cambridge University Press.

van der Pol, J., van den Berg, B.A.M., Admiraal, W.F. , & Simons, P.R.J. . (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education 51 (2008) 51*(4), 1804-1817.

Wegerif, Rupert. (2006). A dialogic understanding of the relationship between CSCL and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning, 1*(1).

Weinberger, Armin, & Fischer, Frank. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71-95.

Vonderwell, Selma. (2003). An examination of asynchronous communication experiences and perspectives of students in an online course: A case study. *The Internet and Higher Education, 6*, 77-90.

Serious Gaming to Improve the Safety of Central Venous Catheter Placement

Daniel Katz
Department of Anesthesiology, Mount Sinai Medical Center
New York, NY 10029
and
Samuel DeMaria
Department of Anesthesiology, Mount Sinai Medical Center
New York, NY 10029

ABSTRACT

Approximately 5 million central venous catheters (CVCs) are placed by physicians annually in the United States, with a complication rate of 15%.[1] Guidelines and recommendations are continually being established and updated regarding CVC placement.[2] While much has been done regarding training the technical skills of CVC placement using part-task trainers (i.e., mannequins), successfully finding and cannulating a central vein is but one part of the process. In fact, many steps designed to prevent untoward complications involve non-technical skills which are perhaps more important in training practitioners to safely place CVCs.

First in aviation and now in healthcare, practitioners are being trained in realistic and highly interactive simulated environments so they can learn not just technical skills , but the key management and non-technical steps which make their task safer.[3] One modality being used to improve performance is video gaming simulation, or "serious gaming." Gaming as a learning tool is being increasingly utilized in health care fields and can lead to better skill-based outcomes.[4] As such, we have developed a game based around the placement of CVCs that will be used as a new teaching modality in a pilot program for instructing residents in safe CVC placement.

Keywords:
Serious gaming, video game simulation, central venous catheter, transfer of learning, game-based learning

INTRODUCTION

Approximately 5 million central venous catheters (CVCs) are placed by physicians annually in the United States. [1] Regrettably, as with any medical procedure, complications occur. Several studies have approximated the complication rate from these procedures to range from 5%-26%.[5,6]Common complications include infection, pneumothorax, arterial puncture, thrombosis and embolism with rates that are often inversely correlated with clinical experience.[7,8] The subsequent costs of catheter-related complications are high, with a single catheter-related infection, for example, costing from $4000 - $56000.[9] Additionally, certain complications from improper placement of catheters are already affecting reimbursement rates of medical centers nationwide, placing additional value on proper placement.

Guidelines and recommendations are continually being established and updated regarding CVC placement in an attempt to minimize these complications, including the use of principles such as aseptic technique and antibiotic-coated catheters.[2] While much has been done regarding training the technical skills of CVC placement using part-task trainers (i.e., mannequins), successfully finding and cannulating a central vein is but one part of the process. In fact, many key steps designed to prevent common untoward effects such as barrier precautions involve non-technical skills, which are perhaps more important in training practitioners to safely place CVCs. Traditionally, these additional steps are learned by practitioners through an apprenticeship type method which can lead to non-standardized practices that may be a detriment to patient safety or lead to confusion as to what best practices are for a particular procedure. Additionally, given the rotation based approach of medical training, it is often the case that trainees will go through brief periods of intense training followed by long periods without placing a CVC which can lead to further skill deterioration. Simulation and gaming may be a way to standardize these practices, improve patient outcomes, and prevent technical skill decay.

Initially in aviation and now in healthcare, practitioners are being trained in realistic and highly interactive simulated environments so they can learn not only psychomotor skills (e.g., adjusting throttle on a plane or intubating a patient), but the key management and non-technical steps which make their task safer.[3] Such simulators have already been proven as effective teaching tools in a variety of healthcare environments including laparoscopy[10,11], bronchoscopy[12], and even team training exercises in areas such as ACLS.[13,14] Additionally, it has been shown that skill retention when using simulators is often superior to standard practices.[15,16] and that the use of simulators reduces the learning curve of many standardized procedures.[14,15] Likewise, it has been shown that not only can simulators improve outcomes, but they can improve efficiency of performing procedures as well.[3]

One specific modality being used to improve performance with simulators is screen-based video gaming simulation, or "serious gaming." Serious gaming as a learning tool is being increasingly utilized in health care fields and can lead to better skill-based outcomes.[4] The theoretical benefits of gaming environments include the ability of the participant to familiarize themselves with an otherwise unfamiliar environment or situation. Additionally the participants can review their progress and have the ability to make errors and learn from them without negative consequences. They can also proceed at their own pace, allowing for participants with different skill levels to learn at a speed that is comfortable for them, without added time pressure. Gaming as a training tool for physicians has not been widely available as it is relatively novel. Game development can be very time consuming and expensive. Fidelity is also a concern, as many of the video game developers

have little medical and clinical experience. However, more opportunities are becoming available.

Currently, laparoscopy gaming for surgeons is the best established medical gaming application. A positive correlation has been shown between increased skill in the gaming simulator and increased skill on actual patients.[17] Additionally, Aggarwal et al showed effectiveness of their game simulator by using the standard set by the airline industry; the transfer-effectiveness ratio (TER).[3,11] Broadly speaking, the TER is means of expressing a ratio of time spent learning a skill on simulator versus normal training. Specifically, to obtain the TER, one must take the difference of the number of trials or time taken to perform the task between the control group and the simulator-trained group divided by total training time received by the simulator group(see Figure 1).

$$TER = \frac{Yo-Yx}{X}$$

Yo=Median time required by control group

Yx= Median time required by gaming group

X=Amount of time spent on simulator

Figure 1

This ratio is an approximation of cost/time effectiveness of the addition of the simulator to the standard program.[18,19] Given that serious gaming has been shown as an effective teaching tool in a variety of areas, a similar game for CVC placement might improve practitioners' ability to safely place these devices and improve patient safety.

As such, the aim of this project is to create an interactive screen-based simulation of internal jugular venous cannulation that will incorporate all aspects of the procedure including setup, sterile preparation, technique of catheter placement, and catheter maintenance. Once the game is created we aim to investigate the usefulness of our serious gaming program in enhancing the ability of the participants to place CVCs. We will do so by assessing the rate of compliance with the previously described procedural steps as measured in the simulated and actual operating room environments.[20]

METHODS

Our study will be divided into three phases: Game Development, Game Launch, Game Validation and Revision

Game Development:
The simulation group at The Mount Sinai Human Emulation, Education, and Evaluation Lab for Patient Safety and Professional Study (HELPS) Center collaborated with the Human Symbiosis Lab group at Arizona State University (ASU) for the gaming project. Staff at ASU who are expert developers of serious medical games, in consultation with the HELPS Center designed and developed the game.

Our game was developed using the Unreal Software Platform on which the ASU Team has successfully implemented and constructed games for the Nintendo Wii for surgical training.[21] The freeware developers version was used for ease of access and game construction. Graphics, templates, and sounds found in the default kit were used. After creation of the virtual world we began to design the platform for CVC placement. Our game design incorporated current best-practices for CVC placement (as outlined by the American Society of Anesthesiologists and the Institute for Healthcare Improvement Central Line Bundle)[3],[22] and current protocols used at The Mount Sinai Medical Center (MSMC) Department of Anesthesiology.[23] The central line checklist described by Dong et al[20] was the basis for the internal computerized grading scheme (See Figure 2).

CVC Proficiency Scale Checklist:

- Preprocedure ID verification
- Informed consent communication
- Trendelenberg position
- Operator maximal barrier precautions
- Hand hygeine
- Chlorhexidine skin antisepsis
- Sterile gloving and gowning
- Patient maximal barrier precautions
- Ultrasound sterile technique
- IJ compressibility by ultrasound
- Procedural pause
- Successful independent IJ Venipuncture
- Transduction/Manometry to verify venous access
- Correct securing of the catheter
- Successful independent SC venipuncture

Adopted from Dong et al

Figure 2

Users are first asked to visit a website that allows for game download and registration. They are instructed to pick a username and password for their anonymous account. After downloading the game and importing their user ID participants

have access to two game modes. The first mode is a practice mode whereby their patient, CVC kit, and environment are preprogrammed to teach the participant the proper steps. At each point the participant is given prompts to the proper order of the steps and is not allowed to click on objects in the environment that are not in the correct sequence. Additionally, an information panel is displayed to guide the participant to both the step currently on the previous step completed. There is a timer in the top right corner to let the user know how long the current attempt is taking. In this game mode there is no penalty for time taken, nor any visual or audio prompts for taking more than the normal allotted time. Once completed the user is taken to a scoring screen that shows them each step completed and their score for that step. In the instruction mode they are given a perfect score since they are taken through step by step. Once the instruction mode is completed users are able to access the gaming mode. In this mode no visual or audio prompts are given to guide the user. Additionally the clock timer will give visual feedback to the user for taking too much time. Users are free to place the CVC in whichever manner they choose, since now the environment is completely unlocked to them. The internal scoring system checks which steps they perform and the order in which they perform them and awards points accordingly. For example, should the user not wash his/her hands prior to donning his gown and gloves he is given zero points for the wash hands and scrub steps, even if the user goes back later to scrub since sterile technique has already been broken. Upon completion of the task the user is directed to a scoring screen. This screen again displays the correct order of steps and shows the user which steps where done correctly and which steps where either missed or performed at the incorrect time. The score is then uploaded onto an internet server that will log the score and display it on our leader board anonymously. This way, participants can compare their scores amongst each other to foster friendly competition, without being individually targeted. We anticipate that by allowing game-like incentives within the system, we will have high retention of the users and subsequently high skill gain

Game Launch:

After over a year of game development we have launched the initial version of our game. Both medical students and anesthesiology residents have access to our game. Additionally we have begun to receive feedback on game design and effectiveness of teaching. It has thus far been very well received, with many residents and students reporting that it has helped them learn and maintain their CVC placement skills.

Game Validation and Revision:

After a brief launch phase we have begun our game validation and revision phase. We have currently enrolled twenty four anesthesiology residents from the department of anesthesiology at Mount Sinai Medical Center to validate our game. From the group of twenty four residents, two groups have been formed. They are currently being randomized either to have full access to the CVC game (gamer group) or to continue their usual practice after standard departmental training in the surgical intensive care unit using actual patients (non-gamer group). Study participants will be classified into sub groups based on years in clinical practice as well as experience and comfort with CVC placement to control for varying experience with CVC placement.

Prior to game access all participants will come to the Mount Sinai HELPS Center to perform a standardized central line placement on a mannequin (Blue Phantom, Redmond, WA). They will be timed and evaluated based on the Dong et al grading scheme that was used for the design of the gaming scoring system.[20] After baseline data collection, subjects given access to the game will have a "warm up" period to familiarize themselves with the gaming process. This will involve a group session which demonstrates the game and educates participants as to its use. Participants will then be allowed to use the game as often as they would like, with mandatory use of the game at least once per week. Use of the game will be tracked via a web-based platform which records user logins and game completion. This is the same website that hosts the leader board for participants to compare scores.

After three weeks of gaming, qualitative and quantitative analyses of the participants' abilities in CVC placement will be examined. We will bring the participants into the HELPS Center simulation lab and have them attempt CVC placement on a mannequin (Blue Phantom, Redmond, WA). These data will be ultimately be part of our primary outcome data, with raw time and an overall global assessment of performance score given by the expert raters as well. Participants will also be asked to complete a survey about how they perceived their own placement of the central venous catheter including; ease of procedure, comfort with all the steps of the procedure, adherence to safety and infection control protocols and overall performance. Those who were in the gaming group will additionally be asked if they felt the game improved their comfort and ability in placing CVCs. Additionally, to calculate the TER for the gaming group versus the control groups, the equation TER= (Yo-Yx)/X will be used, where Yo is the median time required by the control group to place a central line and Yx is the same measurement for the gaming group after using the game for X amount of time(see Figure 1).

Should we experience positive results we intend to further develop our game. We hope to develop multiple levels of central line placement to be performed once the basic level has been mastered. Once a basic line placement has been performed the user will unlock other difficulty levels wherein their knowledge of best practices are tested. These distinct difficulty levels will include minor obstacles to line placement that will allow the user to adapt to the situation while still maintaining proper techniques. Points will be awarded for adhering to standard practices and will be combined with the difficulty level to make a total score. Participants will be penalized points for skipping steps or not adhering to standard practices. Additionally, the virtual patient will now experience complications that will change the outcome of the procedure based on specific steps missed and overall score. For example, should a user not use sterile technique the patient will suffer an infection. This information will be included in the final report the participant receives after game completion. . Our online interface will continue to allow for participants to play the game as often as they would like, from any location.

CONCLUSIONS

We aim for our serious gaming project to impact different areas. First, we hope that the implementation of the game at the Mount Sinai Medical Center will improve the clinical practice of CVC placement in our department. If it is found to be an effective tool and we hope to expand the game to the medical center itself, where hopefully we can reduce the complication rate of CVC placement in actual patients throughout our institution. This might not only result in substantial financial savings for the institution, but could also save lives. Lastly, and more broadly, we hope to show that the implementation of a web-based, serious medical game which reinforces best practices for CVC placement may be an efficient, inexpensive, and widely dispersible way of reducing CVC-associated complications across multiple institutions.

[1] Gould M, Mcgee D. Preventing Complications of Central Venous Catheterization. NEJM 348;12 2003 1123-1131.
[2] www.asahq.org/clinical/CentralVenousAccessGuidelinesDraft06142010.pdf *(accessed August 28, 2010).*
[3] Toff NJ. Human Factors in Anaesthesia: Lessons From Aviation. British Journal of Anaesthesia. 105(1) 21-5 2010.
[4] Conkey C et. al. Relationships Between Gaming Attributes and Learning Outcomes. Simulation and Gaming. V40N2 217-266 2009.
[5] Merrer J, De Jonghe B, Golliot F, et al: French Catheter Study Group in Intensive Care. Complications of femoral and subclavian venous catheterization in critically ill patients: a randomized control trial. JAMA. 2001;286(6):700-707
[6] Raad I, Darouiche R, Dupuis J, et al; The Texas Medical Center Catheter Study Group. Central venous catheters coated with minocycline and rifmapin for the prevention of catheter-related colonization and bloodstream infections. A randomized, double blind trial. Ann Intern Med. 1997;127(4):267-274
[7] Fares LG II, Block PH, Feldman SD. Improved house staff results with subclavian cannulation. Am Surg. 1986;52(2):108-111
[8] Sznajder JI, Zveibil FR, Bitterman H, Weiner P, Bursztein S. Central vein catheterization. Failure and complication rates by three percutaneous approaches. Arch Intern Med. 1986;146(2):259-261
[9] Heard S et al. Prevention of Central Venous Catheter Bloodstream Infections. Journal of Intensive Care Medicine. 25(3) 131-138 2010.
[10] Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. Ann Surg. 2004;240(3):518-525
[11] Aggarwal R, Ward J, Balasundaram I, et al Proving the Effectiveness of Virtual Reality Simulation for Training in Laparoscopic Surgery. Ann Surg. 2007:246(5):771-779
[12] Blum MG, Powers TW, Sundaresan S. Bronchoscopy simlulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. Ann Thorac Surg. 2004;78(1):287-291
[13] Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' non-techincal skills (ANTS): evaluation of a behavioural marker system. Br J Anaesth. 2003;90(5):580-588
[14] Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie WC. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. Chest. 2008;133(1):56-61
[15] Stefanidis D, Korndorffer J, Sierra R, et al. Skill retention following proficiency-based laparoscopic simulator training. Surgery. 2005;138(2):165-170
[16] Andreatta P, Chen Y, Marsh M, Cho K. Simulation based training improves applied clinical placement of ultrasound-guided PICCs. Supp Care Cancer. 2010.
[17] Ewy et. al. Simulation Technology for Health Care Professional Skills Training and Assessment. JAMA 282:861-866 1999
[18] Rantaned EM, Talleur DA. Incremental transfer and cost effectiveness of ground based flight trainers in a university aviation program. Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting. 2005;764-768
[19] Taylor HL, Talleur DA. Emanuel TW Jr, et al. Transfer of training effectiveness of a flight training device (FTD). Proceedings of the 13th International Symposium on Aviation Psychology. 2005;1-4
[20] Dong Y, Suri HS, Cook DA, Kashani KB, Mullon JJ, Enders FT, Rubin O, Ziv A, Dunn WF. Simulation-based objective assessment discerns clinical proficiency in central line placement: a construct validation. Chest. 2010;137(5):1050-6.
[21] Bokhari, R., Bollmann, J., Kahol, K., Smith, M., & Ferrara, J. Design, Development, and Validation of a Take-Home Simulator for Fundamental Laparoscopic Skills: Using Nintendo Wii for Surgical Training. American Surgeon, Accepted for Publication in 2010.
[22] http://www.ihi.org/IHI/Topics/CriticalCare/IntensiveCare/Changes/ImplementtheCentralLineBundle.htm
[23] http://www.youtube.com/watch?v=coEpM7IBzsM

The Management and Engineering Model for Sustainable Development in an organization

Jan BAGINSKI
Faculty of Production Engineering, Warsaw University of Technology
Warsaw, 02-524, Poland

and

Aldona KLUCZEK
Faculty of Production Engineering, Warsaw University of Technology
Warsaw, 02-524, Poland

ABSTRACT

The paper presents a management and engineering model for sustainable development taking into consideration organizational levels of organizations and management systems. Developed the management and engineering model based on nine modules have the feature of the adjustment to changes in the macro and competitive environment, and tracking all the changes. The model combines the concept of innovation with a set of activities and resources necessary for its implementation, enabling the organization to achieve those objectives, and contribute to sustainable growth.

Keywords: Sustainable development, engineering, management, commercialization, systems

1. INTRODUCTION

The main contribution to ecological and environmental development was made during the United Nations' *Conference on Environment and Development* in 1992, which took place in Rio de Janeiro [1]. Result of the conference was developed documents such as *Agenda 21* [1] and Rio Declaration on Environment and Development. Recommendations of Agenda 21 should be implemented by nations, local authorities, as well as business units. According to this document, every unit (acting globally, nationally or locally) should concentrate on few, general areas of focus. First one concerns environment protection and reasonable management of natural resources. It combines reduction of wastes and pollution, as well as protection of endangered species. Second area of focus concerns economic growth and fair split of profits; for example grants and donations for development of less developed entitles. The last main area considers social development and it mainly includes provision of common access to welfare and education.
Examination of the concept of sustainable development at the business unit level means reducing material consumption, energy intensity of production, raising productivity of environmental resources and reducing contaminants while achieving economic, and social goals [2]. This implies, therefore, the efficient use of natural resources and environmental protection and management systems. A tool that will enable the sustainable development management system. To fully implementation of sustainable development concept at the business level it is necessary the most important international standards, which are presented as follows (Figure 1 The management and engineering model for sustainable development).

[1] Agenda 21 is a plan of actions which should be taken globally, nationally and locally in order to achieve sustainable development in economic, social and ecological dimensions.

2. THE PRINCIPLES OF THE MODEL

The goal of developed model is to provide the framework in which each organization operates. This allows to ensure ongoing access to people, capital and natural resources. This in turn helps organizations to deliver better return for shareholders, manage risk effectively, reduce environmental impacts, cut operating costs and provide more business development opportunities. Activities toward sustainable development lead to a development of product/technology which relies on the rational application of raw materials, water and energy, at all stages of the product life cycle at simultaneous reducing the impact on the environment. Such approach should be supported by strategies and international standards or a set of different tools for eco-design and manufacturing. Thus, the model describes how to manage organization today in understanding of an interaction between its elements of systems, and the ability to harmonize and seeking an appropriate balance between the different dimensions of activity: economic, social, environmental [3].
In this paper, it was assumed that the constructed model of Management and Engineering for Sustainable Development will be universal and can be applied to companies operating in both manufacturing and service. The model will be aimed at optimizing and streamlining business processes focused on the implementation of green innovation. Extension of innovation issues, however, requires a broader form of expression analysis of the needs of business and legal regulations on environmental protection and implementation of new systems and process management methods have led to conducting research in this area and to define the model.

The basic premise of the model is designed to prepare organizations to operate in the new organizational model that provides effective and aimed at sustainable development of the business processes. Model should be implemented to ensure that the processes provided by the organizations meet the expected quality requirements of customers.

It was proposed to make the business model consisted of the following modules, shown schematically in Figure 1 in relation to the individual identified elements included in its scope (the module should implement the functions defined in the system for each module separately). The whole creates a kind of adhesive for the business model to take account of the processes, methods, management systems that enable the smooth functioning of the company. Nevertheless, the following characterizes each of the modules in order to more fully explain their methods of operation in the overall model.

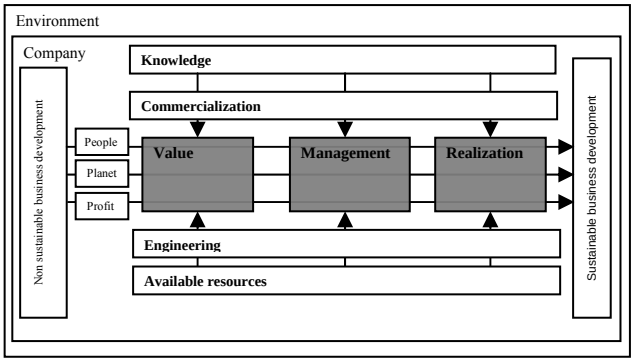


Figure 1 The management and engineering model for sustainable development)

The presented model consists of nine modules, each dependent (resulting from the integration of activities) between structural elements of the company:

1. Non – sustainable development of the company
2. Available knowledge
3. Management
4. Engineering
5. Available resources
6. Value
7. Realization
8. Commercialization
9. Sustainable development of the company

Non-sustainable development of the company: Each company operates in a turbulent environment. Changing ways of doing business, management systems, the values that guided the organization and the expectations of stakeholders. The result is a structurally unbalanced development, due to the unlimited exploitation of resources [4].
Currently, the public expects more companies' involvement in building a better quality of life and growth of enterprises in such a way that "it does not affect a significant irreversible environmental and human life, would not lead to degradation of the biosphere, which menaces laws of nature, economics and culture"[5]. Companies struggle with their place in the market and the need to measure itself against the best in the conditions of constantly changing environment the need to impose on the undertakings of its development.

Available knowledge: Knowledge is primarily, although not exclusively, scientific knowledge, it mainly deals with the epistemology and philosophy of science. In most organizations have become parties collect and process knowledge. Knowledge is identified in the database record of the different forms of knowledge. It is associated with factors such as culture, ethics, values, intuition, working conditions, management style, which creates conditions for the assimilation of new information in the form of all sorts of documents, standards and procedures [6]. Organizations that have the resources they are "able to prepare a proper strategy to their competence", defining thus the transition to a module which is *"available resources"* defined as the intellectual capital of organizations and intangible resources. Thus, knowledge has become a cohesive link information, human resources, IT systems, available technology in the process of intra-organizational activities.

Management: Knowledge management or information storage and processing of all data used in computerized systems. IT systems are one of many factors determining the effectiveness of using knowledge. They support all processes in the organization. The development of organizational knowledge is based on information processes[2] occurring in the enterprise and ICT. The use of information systems involves the need to make significant investments in tools and technologies. For information systems are based on even standard management systems, which include quality management systems, environmental systems, management systems, occupational health and safety. Presented systems can be integrated into one integrated management system. The basis for building an integrated system is ISO 9001, ISO 14001 and PN-N 18001, because these standards have a similar approach to the management and relatively consistent requirements. Application of the above information technology affects the acceleration processes in the organization [7]. Remain a pillar of the IT systems for customer relationship management CRM (Customer Relationship Management), which integrated with other applications such as: supply chain management (SCM) create an environment conducive to the sustainable management of the organization.
The most universal management concept in the organization is a comprehensive Quality Management. TQM should be understood as a holistic approach to management of the organization including management functions such as planning organizing, directing staff focused on the rational use of company resources while maintaining proper relations with the environment.

Engineering: In order to effectively use the principles of TQM in organizations, a number of techniques, methods, tools and management systems were developed. These tools can be divided depending on the programs of action into projective activities (field) in the range of:

- qualitative (QFD, FMEA, SPC, "SixSigma"
- environmental (clean production, recycling, CO2)
- occupational health and safety (PY, 5S, Keizen)
- social (codes, best practices).

Available resources: "Company resources are called all of its assets, capabilities, skills, organizational processes, attributes the company, information, knowledge, etc. controlled by the enterprise, enabling him to conceptualize and apply strategies for enhancing the efficiency and effectiveness*."* [8;9]. Companies, whose motto is to work towards sustainable development (who want to move towards sustainable development) should determine the available resources of knowledge and those resources that are impossible to achieve as a result of this strategy. In this way it becomes possible to define the strategy competence.

[2] A. Domanski defines as "a distinct temporal and spatial information processing system, which is a set intentionally interrelated elements, which are: data sources, methods of their collection and processing, information flow channels, material and human resources and the destination information", Introduction to Informatics, Author: Niedzielska E., Publisher PWE, Warszawa 1993.

”New working methods, modern machinery and equipment, new legislation (...), new technologies and new motivation strategies (...)” [10] are the resources that may become a source of competitive advantage, while being a source of sustained and sustainable development of enterprises, which enable company to operate in a competitive environment.

The sustainable development: The idea of sustainable development is the assumption of continuous economic and social progress harmonized with natural environment. It is a challenge to balance the activities in the areas of environmental, economic and social, while respecting the goods of nature “in order to ensure the possibility of satisfying the basic needs of individual communities or citizens of both the present generation and future generations” [11]. The contribution to ecological and environmental development was described in *Standards & guidelines: preliminary report* [12], in which there were included opportunities to maintain relationship between natural processes and human activity with improved practices that in turn reflect and sustain the contributions of ecological system services. This development leads to increased business competitiveness through sustainable investments carried out in accordance with the principle of so-called triple bottom line[3]. This concept is based on the description of three factors: the financial result (profit) combined with social responsibility (people) and concern for the ecological dimensions of activity (planets) that should form the basis for measuring and evaluating the functions of organizations in sustainable development. Companies that implement the objectives of sustainable development recognize the primacy of ecological requirements of economic activities, “all undertaken actions take into account the needs of future generation**”** [13;14].

The concept of sustainable business development is often associated with the concept of CSR (Corporate Social Responsibility). The idea of CSR is based on activity-based initiatives for sustainable development, respecting the economy, ecology and ethics. In a narrow sense, a commitment to business in favor of ethical conduct and contributing to economic development while demonstrating respect for people, communities and care for the environment. In supporting the idea of CSR can help the Global Compact principles, which are a kind invitation to drive in all spheres of activity, the ten principles of human rights, labor standards, environmental protection and anti-corruption. Compliance with these rules leads to making positive changes in the sphere of business operations.

By implementing a strategy for sustainable development organization based on social responsibility becomes possible to develop innovative solutions (eco-products). The innovation process should be involved all staff members who should be encouraged to develop sustainable solutions in a continuous and systematic. This ensures that all departments understand the organization recognize the impact of organization on the environment, economy and society.

Realization: Designing technologies in the field of sustainable development (environmental or ecological called), depending on their business, organizations should take into account the following criteria:

- sustainable production,
- “model of sustainable community”,
- protection and restoration of ecosystems and natural areas,
- efficiency in energy use,
- renewable energy,
- use of alternative energy sources in order to obtain solar and wind energy, geothermal source,
- environmentally friendly transport,
- waste, air pollution, water pollution, noise.

In the aspect of sustainable design / construction products interact with other concepts and forms of production organization, providing changes in the process and system of governance, mainly such as:

- Lean Management – slimming management so the idea is to simplify all processes and flows, in order to avoid errors and waste. It is visualized in the phase of construction or a change in production techniques and organization of work. The most popular Lean tools include: 5S, Kanban, SMED, TPM and standardization work (some of the tools described in the module, “Engineering”);
- Total Quality Management - a comprehensive quality management;
- Computer-Integrated Manufacturing (CIM), based on information technologies, which include Computer Aided Design (CAD), Computer Aided Engineering (CAE) and Computer Aided Quality (CAQ) as well as Flexible Manufacturing Systems. Through CIM is meant to assist the functions of product development, development of production planning and control of the production process, as well as the process to ensure quality [15].

Commercialization: This module is one of the factors, and also a module of the business model which has an impact on the process of enterprise sustainability management. Commercialization is to convert innovative ideas into ready to enter the market products. The introduction of technology to the market requires a lot of research and development and is also associated with huge costs and risks, which are accompanied from the moment a concept of technology. Commercialization process begins with a thorough diagnosis of the advantages of new technology (check the basic elements or components of the product in a laboratory environment and real, and then build a prototype product, and the final phase of the demonstration version of the product in use) and to assess the potential effectiveness and feasibility of the technology. The effectiveness of the commercialization of technology depends on market value, on the applicability of the technology in the economy. First of all, new technology has to find buyers. In the case of technology, environmental technology find buyers when it is competitive compared with other technologies for environmentally friendly, energy efficient, safe, and above all efficient. This last factor raises the most controversy in obtaining the necessary permits required for the release of technology on the market [16]. Choice of implementation strategies for product/technology on the market (including the sale of property rights, licensing, joint ventures, spin-off/out) depends on the organization and is linked to compliance with all legal requirements, environmental and financial.

Sustainable development of the company: Integration of activities grouped in the various modules leads to the concept of sustainable development companies. It appears that expectations for sustainable investment clients (environmental technology) to reduce the consumption of energy and raw materials reduce

[3] „Triple bottom line”, a concept for reporting business activities, introduced by John Ellington and described in his book titled: *Cannibals With Forks : The Triple Bottom Line of 21 Century Business,* New Society Publishers, Stony Creek, CT, 1998.

waste and pollution, "forcing" from peeling the right business strategies. European organizations also require the preparation of reports on the activities of the company, its influence on the development of the environment and showing that the company is able to strike a balance between the different aspects of your business.

3. CONCLUSIONS

More and more is spoken about the concept of an economy based on sustainable development in the context of production and service. Category of sustainable development of organizations associated with a range of systems, methods, tools and regulations to protect the environment. Hence, it has been proposed to build a model of the management and engineering for sustainable development model, using the strengths in existing and new areas of business activities. Proposed model described in the paper:

- combines the concept of innovation with a set of activities and resources necessary for its implementation, enabling the organization to achieve those objectives, and contribute to sustainable growth;
- takes into account elements of the business organization and management systems and essential to the knowledge of management science, in modern terms, using methods and tools and technologies;
- may be applied in contemporary managed enterprises.

It will allow for an adequate response to the frequent changes in the market and the changing needs of customers. This is possible only when companies begin to compete against themselves for the availability of manufacturing technology, capacity, and above all, quality customer service and offered them services.

4. REFERENCES

[1] A. Wycislik, B. Gajdzik, 2008, *Jakość, środowisko i bezpieczeństwo pracy w zarządzaniu przedsiębiorstwem* (Quality, environment and industrial safety in business management) Silesian Technical University, Gliwice, op. cit.; p.83 (http://en.wikipedia.org/wiki/Earth_Summit, (accessed on August 17th 2011).

[2] Mazur-Wierzbicka E., 2005, "Koncepcja zrównoważonego rozwoju jako podstawa gospodarowania środowiskiem przyrodniczym (Concept of sustainable development as the basis for the management of the natural environment)", [In:] Funkcjonowanie gospodarki polskiej w warunkach integracji i globalizacji (The functioning of the Polish economy in terms of integration and globalization), Department of Microeconomics of University of Szczecin; p.33-44; also available at http://mikro.univ.szczecin.pl/bp/pdf/18/2.pdf (accessed on August 17th 2011).

[3] Witek-Crabb A., 2005, Zrównoważony rozwój przedsiębiorstw – więcej niż ekorozwój. *Zrównoważony rozwój przedsiębiorstw a relacje z interesariuszami (*Sustainable development companies - more than sustainable development. *Sustainable development and business relations with stakeholders*) (red.) H. Brdulak, Oficyna Wydawnicza SGH, Warszawa, s. 561-568.

[4] Klos L., 2005, "Ekorozwój jako podstawa aplikacyjna założeń polityki ekologicznej (Sustainable developmet as a basis for environmental policy)", [In:] Teoretyczne aspekty gospodarowania (Theoretical aspects of management), Department of Microeconomics of University of Szczecin, p. 211-218; also available at http://mikro.univ.szczecin.pl/bp/index.php?a=f17g22 (accessed on August 17th 2011).

[5] Kozlowski S., 1996, Czy transformacja polskiej gospodarki zmierza w kierunku rozwoju zrównoważonego? (Does the transformation of the Polish economy tend towards sustainable development?) [In]: Mechanizmy i uwarunkowania ekorozwoju (Mechanisms and Determinants of Ecodevelopment), Bialystok University of Technology, Bialystok, p.19

[6] Davenport T.H., Prusak L, 1998, How Organizations Manage What They Know, Harvard Business Review.

[7] Lech P., 2003, Zintegrowane systemy zarządzania EEP/ERP II. Wykorzystanie w biznesie, wdrażanie"(Integrated management systems ERP/ERPII. Use in business, implementation), Difin, Warszawa.

[8] Daft R.,1987, Organization Theory and Design, New York, West, p.143.

[9] Word Commision on Environment and Development (WCED), 1987, Our common future, Oxford Univeristy Press, Oxford.

[10] Piasecki B. (eds.), 1999, Ekonomika i zarządzanie małą firmą (Small Business Economics and Management), PWN, Warszawa-Łódź, s.226.

[11] Ustawa z dnia 27 kwietnia 2001 r. Prawo ochrony środowiska: Dz. U. z 2001 r. Nr 62, poz. 627 (Act of 27 April 2001 Environmental Protection Law: Law Acts, 2001, No. 62, item. 627).

[12] Acknowledgments Product Development Committee, Standards & guidelines: preliminary report, Sustainable Site Initiative, 2007, p.5-17.

[13] Kozlowski S., 1994, Droga do ekorozwoju (The Way Towards Ecodevelopment), PWN.

[14] Pakulska P., Poniatowska-Jaksch M., Rozwój zrównoważony – „szeroka i wąska" interpretacja (Sustainable development – „broad and narrow interpretation", [In:] [online], available at www.sgh.waw.pl (accessed on August 17th 2011).

[15] Durlik L., 2004, Inżyniera zarządzania. Strategia i projektowanie systemów produkcyjnych (Engineering Management: strategies and designing productions systems), Publisher „Placet", Gdansk, p.149.

[16] Kluczek A., 2009, Komercjalizacja technologii dla zrównoważonego rozwoju (Commercialization of technologies for sustainable development), Przegląd Techniczny, 24/2009.

.

Harnessing the Chaos: Understanding Barriers to Inter-organizational Communication and Collaboration within the Grid Network

Angela C. Dalton
Gariann M. Gelston
Lucas C. Tate
Pacific Northwest National Laboratory
Richland, WA 99352

Extended Abstract for Interdisciplinary Communication

Keywords: inter-organizational communication, collaboration, decision making, emergency response, network effectiveness, power grid.

1. INTRODUCTION

With escalating demand and intensifying market competition, the implementation of advanced technologies that make more data available at a much faster pace to grid operators can help reduce the potential for contingencies and system events but not eliminate them entirely due to a high level of external uncertainty and the potential for human errors in the workflow. Thus, an unrelenting challenge for power grid entities remains: in time of emergency, how can grid operators resolve system event effectively and rapidly?

2. BACKGROUND AND SIGNIFICANCE

In recent years, much research energy has been invested in advancing computational modeling of grid failures [1] and grid operation simulation [2, 3] in the hope of making operations more reliable, and forecasts more accurate and dynamic. Other research aims at utilizing new technologies such PMU to enable near real-time system monitoring and control [4, 5]. Innovative communication technologies and visualization techniques have also been introduced for faster and more secure data transmission and for providing greater decision support capabilities [6, 7]. On the human decision making front, research on naturalistic decision making and situation awareness has gained much currency, securing its significance as a key programmatic component for operator training [8].

Yet, the power grid is not simply a network of interconnected physical infrastructures; it is a network of organizations with not only shared stakes, but also competing goals and divergent client requirements. In a system event, the human and organization network needs to act quickly and effectively to maintain the integrity of the physical network. Thus, to have an accurate understanding of how the network as a whole responds to and mitigates emergencies, an understanding of the intricacies in interpersonal and inter-organizational interaction and coordination is prerequisite. While the research field has certainly benefited from the abundance of individual operator level analysis, there is, however, surprisingly little research being undertaken to bring insights from the organizational theory and behavior fields to bear on improving inter-organizational collaboration in grid operations in general, and in system event remediation in particular.

3. RESEARCH QUESTIONS

A. Objectives

To help address this research void, in this paper we propose to 1) identify organizational theories and models relevant to power grid operations and emergency response from domains such as communication theories, decision making performance research, and organizational network effectiveness research; 2) apply the organizational perspective to grid operations through analogical reasoning; 3) generate new insight about organizational and human barriers to inter-organizational collaboration in grid operations; and 4) provide recommendations to align the current collaboration practices with the resilience requirements of the future grid.

At the outset, we refer to inter-organizational collaboration in grid operations as a relationship

signaled by cooperation and coordination between and among multiple organizations from dispersed geographic locations. Such a relationship is enacted, modified, and negotiated through continuous communication, and is influenced by the forces of the market, institutional rules, legal control, and network topology [9].

In the electric power grid, inter-organizational communication and coordination is often carried out between transmission operators, generator operators, balancing authorities, and reliability coordinators [10]. This paper will focus on the transmission operator entity and examine how it communicates and coordinates with other functional players in the grid system in event mitigation. The primary goal of this paper is to identify the critical concepts in group communication theories, decision making performance, and organizational network effectiveness research, and apply these concepts to analyzing the relationship between communication effectiveness and decision making performance by the power grid organizational network in the context of system emergencies.

B. Understanding the grid from an organizational perspective

In network effectiveness research, the effectiveness evaluation framework proposed by Provan and Milward [11] might provide useful insight for examining the effectiveness of the grid. This framework assesses organizational network effectiveness at three levels: network participant, network itself, and the community served by the network. Although the set of organizations in their study were publicly-funded health and human services organizations, applying this framework to studying the effectiveness of power grid organizational network will prove helpful for advancing the frontier of this research domain and contributing to a better understanding of how well the grid functions as a network in emergency response and mitigation. A number of grid network characteristics warrant elaboration. First, dissimilar to network membership uniformity in their study, the power grid network consists of diverse members, including public organizations, not-for-profit organizations, and for-profit entities. The resulting network has heterogeneous membership composition, and more intense divergence in organizational imperatives, stakeholder expectations, accountability requirements, and client needs. Second, similar to state-funding agencies in the health and human services network in Provan and Milward's study, the North American Electric Reliability Corporation (NERC) also functions as a quasi-NAO (network administrative organization) that coordinates its members' action. However, given its limited control over network resources (its primary function is setting standards and overseeing compliance), NERC may have a more relaxed grip on the network compared to a NAO in a stereotypical NAO-centered network. With these differences in mind, we can still superimpose the evaluative framework on power grid organizations in order to examine their network effectiveness. At the organizational level, four evaluative indicators seem relevant: client, organizational legitimacy, acquisition of resources, and cost factors [11]. At the network level, a key effective indicator is the quality of service to the intended customers. For grid entities, the quality of service can be interpreted as reliability (adequacy will not be addressed in this paper although it is a key service indicator as well), which can be operationalized as the frequency of system events. The fewer the events in a given time frame, the more reliable is an organizational network. A particularly important indicator of a highly effective network is one in which information and resources are shared freely and seamlessly from one organizational member to another. For power grid organizations, due to their physical interdependence, there is an undisputable incentive to cooperate within the network. However, as somewhat autonomous institutions, full cooperation is dampened by reluctance in communication and information sharing, possibly out of concerns for market competition, information security, and organizational survival. Thus, grid organizations' operational imperative and institutional constraints may create tension and, in turn, barriers to more expansive inter-organizational communication and coordination which make a network effective.

Insight from group-based decision making performance and communication may also enhance our understanding of the barriers to inter-organizational communication and coordination. Extant research suggests that the quality of decision making is influenced by communication modalities, norms, procedures, quality, and behavioral characteristics [12].

C. Hypotheses and methodology

Following this stream of inquiry as well as the network effective evaluation framework, we propose to test hypotheses with regard to network effectiveness and decision making performance in grid event mitigation. Network effectiveness and decision making performance is measured as the

event count of system events reported to NERC from power grid organizations throughout North America. Lower rates of event occurrence indicate a more efficient network and better decision making performance. We operationalize inter-organizational communication based on observable and quantifiable communication behavior, norms, and information sources, and measure it in terms of communicative relationships (communication with whom), frequency (how often), scope (what is communicated), and modalities of communication (face to face, audio only, and web-based methods). We hypothesize that these aspects of inter-organizational communication will have an important influence on effectiveness of the grid network in its mitigation effort. Other factors such as an organization's position (central vs. peripheral) in the communication network, organization size, age, business volume, and information availability will also influence these organizations' decision making performance and network effectiveness. Additionally, other contextual variables such as the weather, size of client pool and whether or not an entity performs multiple functional roles in the network will also be considered. Individual-level variables such as operators' education, skills, and training, though important, will be excluded from this analysis due to the higher-level unit of analysis that underpins this research.

Structured survey will serve as the primary research method. The projected sample size will be 50 -100. The target population will consist of individuals from transmission operator organizations, balancing authorities, and reliability coordinators. Subject matter experts will be recruited to participate in an online structured survey with both open- and close-ended questions. Statistical analysis will be performed on the survey results.

4. POTENTIAL RESEARCH CONTRIBUTIONS

Through this research, several important theoretical and empirical contributions can be made to studies of organizational behavior as well as to power grid research. First, quantifying concepts such as decision making performance and network effectiveness will hopefully bring more clarity to these important but abstract subjects, which has received sustained criticism in organizational research for lacking operational definitiveness. The power grid network provides organizational researchers with a well-defined structural arrangement, clear functional division of labor, and concrete timeframes and organizational contexts for investigating communication and collaborative decision making. These features make it relatively easy to quantify conceptual framework and constructs that are otherwise abstract or fuzzy in other organizational contexts. Furthermore, this study will quantitatively test the relationship between inter-organizational communication, organizational decision making performance, and network effectiveness, and provide some empirical evidence from the field of power grid organizations to inform theory development in organizational research, which has, to date, remained an under-explored topic.

By the same token, this paper has the potential to contribute to the research on human factors and power grid operations by 1) identifying barriers to efficient and effective inter-organizational communication in event mitigation and system restoration; 2) helping grid entities recognize areas for improvement and investment to meet the growing demand on grid operators; and 3) enabling the conceptualization of innovative communication technologies that will bring communication tools up to speed with system monitoring and control tools that have begun to make strides in current operations, making the network as a whole more reliable and efficient for the future.

5. ACKNOWLEDGEMENT

This research is supported by the Future Power Grid Initiative, a Laboratory Directed Research and Development Project at the Pacific Northwest National Laboratory. We thank Henry Huang, Jeff Dagle, Paul Whitney, William Pike, and Garill Coles for their contributions.

6. REFERENCES

[1] R. Kinney, P. Crucitti, R. Albert and V. Latora, "Modeling Cascading Failures in the North American Power Grid", **European Physical Journal B - Condensed Matter and Complex Systems**, vol. 46, no. 1, 2005, pp. 101-107,.

[2] S. R. Nassif, "Power Grid Analysis Benchmarks", in **Proceeding ASP-DAC '08 Proceedings of the 2008 Asia and South Pacific Design Automation Conference, IEEE** Computer Society Press, Los Alamitos, CA, 2008.

[3] D. Chassin, and C. Posse, "Evaluating North American Electric Grid Reliability Using the Barabasi-Albert Network Model", **Physica: A Statistical Mechanics and Its Applications**, vol. 355, 2005, pp. 667-677.

[4] A. Bose, K. Tomsovic, D.E. Bakken, and V. Venkatasubramanian, "Designing the Next Generation of Real-Time Control, Communication, and Computations for Large Power Systems", Invited Paper, **Proceedings of the IEEE**, vol. 93, No. 5, 2005, pp. 965-979.

[5] D. Novosel, V. Madani; B. Bhargava, V. Khoi Vu, and J. Cole, "Dawn of the Grid Synchronization", **Power and Energy Magazine, IEEE**, vol. 6 no.1, 2008, pp. 49-60.

[6] T. J. Overbye, J. D. Weber, and K. J. Patten, "Analysis and Visualization of Market Power in Electric Power Systems", **Decision Support Systems**, vol. 30, no. 3, 2001, pp. 229-241.

[7] C. H. Hauser, D. Bakken, and A. Bose, "A Failure to Communicate: Next Generation Communication Requirements, Technologies, and Architecture for the Electric Power Grid", **Power and Energy Magazine, IEEE**, vol. 3, 2006, pp. 47-55.

[8] M. Endsley, B. Bolte, and D. Jones, **Designing for Situational Awareness: A User Centered Approach**, London: Tylor and Francis, 2003.

[9] T.B. Lawrence, C.Hardy, N. Phillips, "Institutional Effects of Interorganizational Collaboration: The Emergence of Proto-Institutions", **Academy of Management Journal**, Vol. 45, No. 1, 2002, pp. 281-290.

[10] North American Electric Reliability Corporation, **Reliability Functional Model: Function Definitions and Functional Entities**, 2009, pp. 1-55. Available: http://www.nerc.com/files/Functional_Model_V5_Final_2009Dec1.pdf

[11] K.G. Provan, and H.B. Milward, "Do networks really work? A framework for evaluating public-sector organizational networks," **Public Administration Review**, vol.61 no. 4, 2001, pp. 414-423.

[12] R. Hirokawa,and M. S. Poole (ed.), **Communication and group decision making**, Thousand Oaks, CA: Sage Publications, 1996.

The Utilization of High-Frequency Gravitational Waves for Global Communications

Robert M L BAKER, Jr. and Bonnie S. BAKER
Transportation Sciences Corporation and GravWave® LLC
8123 Tuscany Avenue, Playa del Rey, CA 90293

ABSTRACT

For over 1000 years electromagnetic radiation has been utilized for long-distance communication. heliographs, telegraphs, telephones and radio have all served our previous communication needs. Nevertheless, electromagnetic radiation has one major difficulty: it is easily absorbed. In this paper we consider a totally different radiation, a radiation that is not easily absorbed: gravitational radiation. Such radiation, like gravity itself, is not absorbed by earth, water or any material substance. In particular we discuss herein means to generate and detect high-frequency gravitational waves or HFGWs, and how they can be utilized for communication. There are two barriers to their practical utilization: they are extremely difficult to generate (a large power required to generate very weak GWs) and it is extremely difficult to detect weak GWs. We intend to demonstrate theoretically in this paper their phase-coherent generation utilizing an array of in-phase microelectro-mechanical systems or MEMS resonator elements in which the HFGW flux is proportional to the square of the number of elements. This process solves the transmitter difficulty. Three HFGW detectors have previously been built; but their sensitivity is insufficient for meaningful HFGW reception; greater sensitivity is necessary. A new Li-Baker HFGW detector, discussed herein, is based upon a different measurement technique than the other detectors and is predicted to achieve a sensitivity to satisfy HFGW communication needs.

Keywords: Gravitational waves, communications, Li-Baker detector, microelectromechanical systems, high-frequency gravitational waves.

1. INTRODUCTION

Since the dawn of civilization electromagnetic radiation has been utilized for long-distance communication: heliographs, telegraphs, telephones and radio have all served our previous communication needs. Nevertheless, electromagnetic radiation has one major drawback: it is easily absorbed. In this paper we consider a totally different radiation, a radiation that is not easily absorbed: gravitational radiation. Such radiation, like gravity itself, is not absorbed by earth, water or any material substance. In particular we discuss herein a means to generate and detect high-frequency gravitational waves or HFGWs and how they can be utilized for communication. HFGWs are defined as GWs having frequencies in excess of 100 kHz (Douglas and Braginsky [1]) and long-wavelength GW detectors such as LIGO, Virgo and GEO600 cannot sense HFGWs [2]. Global communications by means of HFGWs would be the ultimate wireless system. HFGW communication would greatly reduce communications costs since it would not require the following:

- No satellite transponders
- No microwave relays
- No underwater cables
- No coaxial cables

Since the Nobel Prize winning observations of Hulse and Taylor in the 1970s no one has doubted the existence of gravitational waves. There are two barriers to their practical utilization: they are extremely difficult to generate (a large power required to generate very weak GWs) and it is extremely difficult to detect weak GWs. In the past several decades hundreds of peer-reviewed journal articles have addressed these issues, for example Beckwith [3] and Grishchuk [4]. We intend to demonstrate theoretically in this paper that their generation utilizing superradiance (Scully and Svidzinsky. [5]), which involves a linear double-helix array of in-phase micro-electromechanical systems (MEMS) resonator elements, in which the HFGW flux is proportional to the square of the number of elements, solves the HFGW generation or transmitter difficulty. The use of a new, but well documented in peer-reviewed literature, effect discovered by Fangyu Li (Chongqing University, China, [6]) solves the detection difficulty. This Li-effect is the basis for the very sensitive Li-Baker HFGW detector, designed by Robert Baker and developed jointly by United States and Chinese HFGW research teams. As documented in peer-reviewed literature [7, 8, 9, 10] such a detector has sensitivity more than sufficient to receive the transmitted HFGW signal at a significant distance from the transmitter. Dehnen in Germany showed in an article [11] that HFGWs could be generated in the laboratory, using General Relativity, through the use of crystal oscillators. His work is the basis for an efficient HFGW generator or transmitter. The critical element in Dehnen's HFGW generator or transmitter had been the large size and power requirements of his crystal oscillators. This difficulty is removed through the use of modern MEMS technology. There have been other challenges to HFGW communications based upon the mistaken belief that GW generators or transmitters can only be designed using spinning rods or the effect described by Gertsenshtein in 1962 [12] and analyzed by Eardley in 2008 in the JASON report [13]. Both of these methods for generating GWs are unsatisfactory and produce negligible GW power.

2. HFGW GENERATORS (Transmitters)

There exist several sources for HFGWs or means for their generation. The first generation means is the same for gravitational waves (GWs) of all frequencies and is based upon the quadrupole equation first derived by Einstein in 1918.[14] A formulation of the quadrupole that is easily related to the orbital motion of binary stars or black holes, rotating rods, laboratory HFGW generation, etc. is based upon the jerk or shake of mass (time rate of change of acceleration), such as the change in centrifugal force vector with time; for example as masses move around each other on a circular orbit. Figure 1 describes that situation. Recognize, however, that change in force Δf need NOT be a gravitational force (see Einstein, 1918 [14]; Infeld quoted by Weber 1964 [15] p. 97; Grishchuk [16]). Electromagnetic forces are more than 10^{35} larger than gravitational forces and should be employed in laboratory GW generation. As Weber ([15] p. 97) points out: "The non-gravitational forces play a decisive role in methods for detection and generation of gravitational waves ..." The quadrupole equation is also termed "quadrupole formalism" and holds in weak gravitational fields (but well over 100 g's), for speeds of the generator "components" less than the speed of light and for the distance between two masses r less than the GW

wavelength. Certainly there would be GW generated for r greater than the GW wavelength, but the quadrupole "formalism" or equation might not apply exactly. For very small time change Δt the GW wavelength, $\lambda_{GW} = c\,\Delta t$ (where $c \sim 3\times10^8$ m s^{-1}, the speed of light) is very small and the GW frequency ν_{GW} is high. The concept is to produce two equal and opposite jerks or Δf 's at two masses, such as MEMS, a distance $2r$ apart. This situation is completely analogous to binary stars on orbit as shown in Figs. 1 and 2**.**

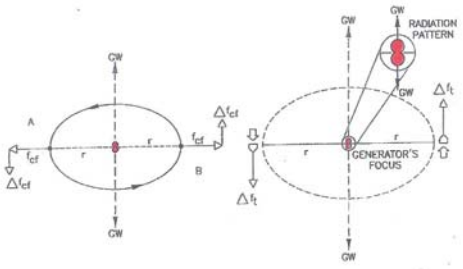


Figure 1. Change in Centrifugal Force of Orbiting Masses, Δf_{cf}, Creates Radiation.

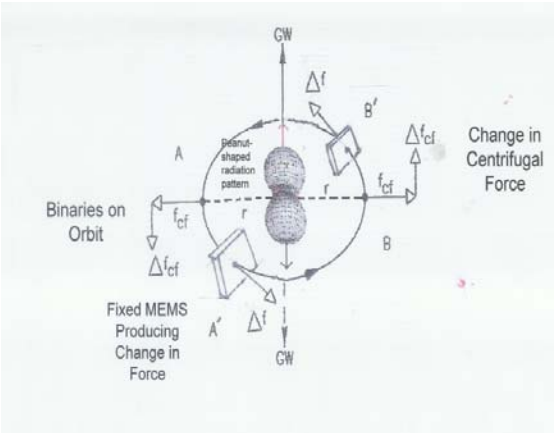


Figure 2. Radiation Pattern Calculated by Landau and Lifshitz [17] Section 110 Page 356.

Next we consider an array of GW sources. Consider a stack of orbit planes, each one involving a pair of masses circling each other on opposite sides of a circular orbit as in Fig. 3. Let the planes be stacked one light hour apart (that is, $60\times60\times3\times10^8 = 1.08\times10^{12}$ meters apart) and each orbit exactly on top of another (coaxial circles). According to Landau and Lifshitz [17] on each plane a GW will be generated that radiates from the center of each circular orbit. The details of that generation process are that as the masses orbit a radiation pattern is generated. In simplified terms (from the equations shown on page 356 of Landau and Lifshitz [17]) an elliptically shaped polarized arc of radiation is formed on each side of the orbit plane (mirror images). As the two masses orbit each other 180^0 the arcs sweep out figures of revolution. Together these figures of revolution become shaped like a peanut as shown in Fig. 2.

The general concept of the present HFGW generator is to utilize an array of force-producing elements arranged in pairs in a cylindrical formation such as a double helix as in Fig. 4. This is analogous to the binary-star arrays of Fig. 3 in which an imaginary cylinder could be formed or constructed from the collection of orbits. As a wavefront of energizing radiation proceeds along the cylindrical axis of symmetry of such a double-helix array, shown in Fig. 4 the force-producing element pairs (such as pairs of film-bulk acoustic resonators or FBARs) are energized simultaneously and jerk, that is they exhibit a third time

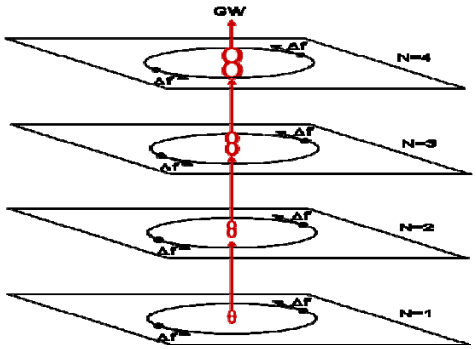


Figure 3. GW Flux Growth Analogous to Stack of *N* Orbital Planes.

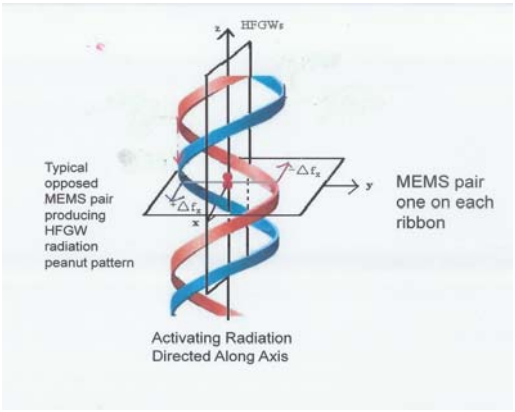


Figure 4. Double-Helix HFGW Generator Array (Patent Pending)

derivative of motion and the flux (W m^{-2}) thereby increased. Utilizing General Relativity, Dehnen and Romero-Borja [11] computed a superradiance build up of "… needle-like radiation …" HFGWs beam emanating from a closely packed but very long linear array of crystal oscillators. Their oscillators were essentially two vibrating masses a distance b apart whereas a pair of vibrating FBAR masses is a distance $2r$ apart as shown in Fig. 5, but operates in an analogous fashion as piezoelectric crystals.

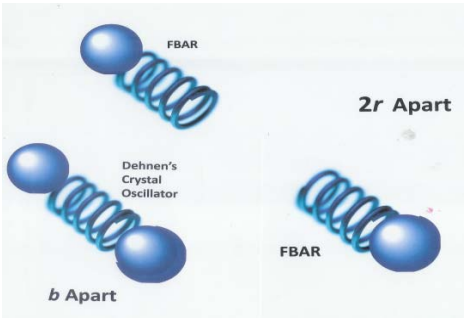


Figure 5. Comparison of Dehnen and Romero-Borja [11] crystal oscillator and FBAR-pair system.

Superradiance also occurs when emitting sources such as atoms "…are close together compared to the wavelength of the radiation …" (Scully and Svidzinsky [5] p.1510). Note that it is not necessary to have the FBAR elements perfectly aligned (that is, the FBARs *exactly* across from each other) since it is only necessary that the energizing wave front (from Magnetrons in the case of the MEMS or FBARs as in Baker, Woods and Li [18]) reaches a couple of nearly opposite FBARs at the same time so that a coherent radiation source or focus is produced

between the two FBARs. The energizing transmitters, such as Magnetrons, can be placed along the helixes' array axes between separate segments of the array or, more efficiently, at the base of the double helixes so that a superradiance force change, Δf, produced by energizing one off-the-shelf FBAR is 2 N microwave beam is projected up the axis of the helixes. The according to Woods and Baker [19], so that the power is given by the equation derived in Baker [20]:

$$P = 1.76\times10^{-52}\,(2r\,\Delta f/\Delta t)^2 \quad \text{W.} \qquad (1)$$

Let the activating radiation for the FBARs be conventional Magnetrons as employed in one-thousand watt microwave ovens. The frequency would be ν_{EM} = 2.5 GHz (thus Δt = 4×10^{-10} s and λ_{EM} = 12 cm**).** The HFGW frequency is twice that of the activating EM radiation or ν_{GW5} = 5 GHz**.** For Eq (1) the calculation of the combined Δf of all the pulsating MEMS or FBARs requires more consideration. We will set the length of a double-helix array cylinder as 20 m, but recognize that it can be separated into segments along the same axis with energizing transmitters, e.g., Magnetrons installed on the cylinder axis between the segments. The transmitters could also be phase coherent and arranged in a line along the double-helix axis at its base. If, for example, there were 1000 one-kilowatt Magnetrons feeding in on one hundred 12-cm, (λ_{EM}, wide levels) and each of their beams covered a 10-cm radius circle, then the energizing radiation flux would be 3.2×10^4 W m^{-2}. According to superradiance there would result a needle-like microwave radiation directed along the axis of the double helixes amounting to 32 gigawatts per square meter. In order to create a perfectly planner wave front, with no irregularities, the cylindrically symmetric MEMS array could be contained in a wave guide or possibly a very wide coaxial "cable," surrounded by a robust one megawatt heat sink. To increase instantaneous power to the array, bursts of gigawatt power, for example, every millisecond could be employed that would maintain a megawatt average power input.

The walls of the cylindrical array are taken to be 30-cm thick. Thus the volume of the twenty meter long array is $\pi(r_1^2 - r_2^2)$ ×20 m^3, where r_1 is the outside radius = 0.35 m and r_2 is the inside radius = 0.05 m. Thus the volume is 7.5 m^3. The FBAR is a mechanical (acoustic) resonator consisting of a vibrating membrane (typically about 100×100μm^2 in plan form, and about 1μm thickness), fabricated using well-established integrated circuit (IC) micro fabrication technology. A typical off-the-shelf FBAR as shown schematically in Fig. 6, usually has overall dimensions 500 μm by 500 μm by approximately 100 μm thick. For our purposes, in which a high number density is important, we will trim the FBARs to a minimum size. In order to account for fabrication margins we will take the dimensions as 110 μm by 110 μm by 20 μm for an FBAR volume of 2.42×10^{-13}m^3.

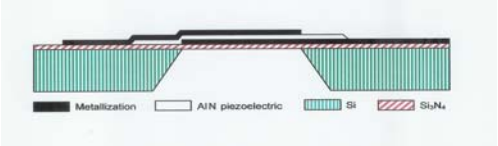


Figure 6. Basic FBAR Construction (cross-section side view, not to scale).

Thus the total number of FBARs in the double-helix cylindrical array is 2.85×10^{12} and the number of pairs is half of that. There would be N = 1.425×10^{12} FBAR pairs in the double-helix cylindrical array. Since each FBAR exhibits a jerking force of 2 N the combined Δf of all the jerking FBAR pairs is 2.85×10^{12} N. if the jerking pairs (or "orbits") moved in concert**.** From Eq. (1) the total power produced by the double-helix array is P = $1.76\times10^{=52}(0.2\times2.85\times10^{12}/4\times10^{-10})^2 = 3.57\times10^{-10}$ W. But due to the N levels, each one of which represents an individual GW focus, there exists a "Superradiance" condition in which the HFGW beam becomes very narrow as shown schematically in Fig. B of [5]. Thus the HFGW flux, in W m^{-2}, becomes much larger at the cap of the radiation pattern. According to the analyses of Baker and Black [21] the area of the half-power cap is given by

$$A_{cap} = A_{1/2(N=1)} / N \quad \text{m}^2 \qquad (2)$$

A more conservative approach would be that there are N individual GW power sources each with a Δf = 2 N. Thus from Eq. (1), with $2r_{rms} = 2\sqrt{[(r_1^2 + r_2^2)/2]} = 0.5$ m, the total power produced by the double-helix array is $P = 1.55\times10^{13}$ $\times1.76\times10^{-52}(0.5\times2/4\times10^{-10})^2 = 1.69\times10^{-20}$ W. But due to the N levels, each one of which represents an individual GW focus, there exists a "Superradiance" condition in which the HFGW beam becomes very narrow as shown schematically in Fig. B of Scully and Svidzinsky [5]. Thus the HFGW flux, in W m^{-2}, becomes much larger at the cap of the radiation pattern. According to the analyses of Baker and Black [21] the area of the half-power cap is proportional to 1/N and the GW flux is:

$$S(1) = (P/4)(1.71/N) = (1.69\times10^{-20}/4)/(1.71/1.55\times10^{13}) = 3.8\times10^{8}\ \text{W m}^{-2} \quad . \qquad (3)$$

From Baker, et al. [22], Eq. (6A) of the Appendix, the amplitude of the dimensionless strain in the fabric of spacetime is

$$A = 1.28\times10^{-18}\sqrt{S}/\nu_{GW} \quad \text{m/m} \quad . \qquad (4)$$

So that at a one-meter distance $A = 5\times10^{-32}$ m/m. If the FBARs in all of the helix levels are not activated as individual pairs, then the situation changes. For example, let all of the FBARs in a 6-cm wide level (½ λ_{EM}) be energized in concert. The number of levels would be reduced to N = 20 m/0.06 m = 333. But, because the FBAR-pairs in each level act together, Δf = (2 N)(1.55×10^{13} / 333). Thus the changes in Eq. (1) cancel out and there is no change in HFGW flux.

The HFGW beam is very narrow. From Eq. (4b) of [21] for N = 1.55×10^{13} it would be $\sin^{-1}$ (0.737)/ $\sqrt{1.55\times10^{13}}$ = 1.87×10^{-7} radians. For N = 333 the angle is 0.0022 radians. This is still narrow, but the double helix configuration certainly reduces the width of the HFGW beam. Additionally multiple HFGW carrier frequencies can be used with modulation schemes e.g., pulse carrier phase shift key, so the signal is very difficult to intercept, and is therefore useful as a low-probability-of-intercept (LPI) signal, even with widespread adoption of the HFGW technology.

From Woods, et al. [23] the current estimated sensitivity of the Chinese Li-Baker HFGW Detector is A = 1.0×10^{-30} m/m to 1.0×10^{-32} m/m with a signal to noise ratio of over 1500 (Woods, et al [23] p. 511) or if we were at a 1.3x10^7 m (diameter of Earth) distance, then S = 1.33×10^{-20} Wm^{-2}. and the amplitude A of the HFGW is given by A = 3.8×10^{-39} m/m. Although the best theoretical sensitivity of the Li-Baker HFGW detector is on the order of 10^{-32} m/m, its sensitivity can be increased dramatically (Li and Baker, [9] by introducing superconductor resonance chambers into the interaction volume (which also improves the Standard Quantum Limit; Stephenson [24]) and two others between the interaction volume and the two microwave receivers. Together they provide an increase in

sensitivity of five orders of magnitude and result in a theoretical sensitivity of the Li-Baker detector to HFGWs having amplitudes of 10^{-37} m/m. There also could be a HFGW superconductor lens, as described by Woods [25], which could concentrate very high frequency gravitational waves at the detector or receiver. Thus with Chinese Li-Baker HFGW detector program successful and the Wood's lens practical, **the Li-Baker detector will exhibit sufficient sensitivity to receive the generated HFGW signal globally.**

3. HFGW DETECTORS (Receivers)

Operational HFGW Receivers

In the past few years HFGW detectors, as exhibited in Figs. 7**,** 8 and 9 have been fabricated at *Birmingham University*, England, *INFN Genoa,* Italy and in Japan. These types of detectors may be promising for the detection of the HFGWs in the GHz band (MHz band for the Japanese) in the future, but currently, their sensitivities are orders of magnitude less than what is required. Such a detection capability is to be expected, however, utilizing the Li-Baker detector. Based upon the theory of Li, Tang and Zhao [6] termed the Li-effect, the detector was proposed by Baker during the period 1999-2000, a patent for it was filed in P. R. China in 2001, subsequently granted in 2007 [26.] (http://www.gravwave.com/docs/Chinese%20Detector%20Patent%2020081027.pdf).

Preliminary details were published later by Baker, Stephenson and Li [22]. This detector was conceived to be sensitive to relic HFGWs (or high-frequency relic gravitational waves, termed HFRGWs) having amplitudes as small as 10^{-32} to 10^{-30}, but using resonance chambers to 10^{-37} or possibly smaller [9].

The Birmingham HFGW detector measures changes in the polarization state of a microwave beam (indicating the presence of a GW) moving in a waveguide about one meter across as shown in Fig. 7. (Please see Cruise [27]; and Cruise and Ingley [28].) It is expected to be sensitive to HFGWs having spacetime strains whose amplitudes are $A \sim 2 \times 10^{-13}$.

Figure 7 *Birmingham University* **HFGW Detector**

The *INFN Genoa* HFGW resonant antenna consists of two coupled, superconducting, spherical, harmonic oscillators a few centimeters in diameter. Please see Fig. 8. The oscillators are designed to have (when uncoupled) almost equal resonant frequencies. In theory the system is expected to have a sensitivity to HFGWs with size of about $A \sim 2\times10^{-17}$ with an expectation to reach a sensitivity of $\sim 2\times10^{-20}$. (Bernard, Gemme, Parodi, and Picasso [29]); Chincarini and Gemme [30]). As of this date, however, there is no further development of the *INFN Genoa* HFGW detector.

Figure 8 . *INFN Genoa* **HFGW Detector**

The Kawamura 100 MHz HFGW detector has been built by the *Astronomical Observatory of Japan.* It consists of two synchronous interferometers exhibiting an arms length of 75 cm. Please see Fig. 9. Its sensitivity is now about $A \approx 10^{-16}$ (Nishizawa et al., [31]). According to Cruise [32]) of *Birmingham University* its frequency is limited to 100 MHz and at higher frequencies its sensitivity diminishes.

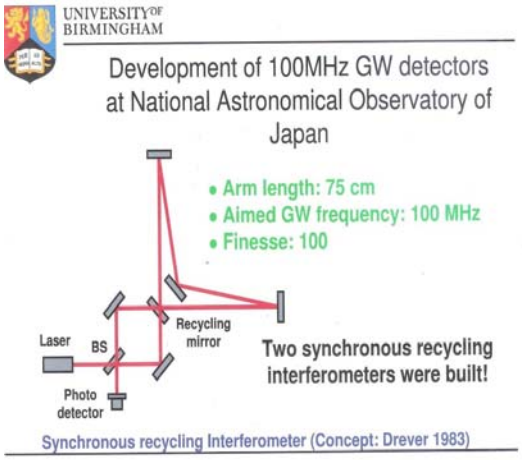


Figure 9. The National Astronomical Observatory of Japan 100MHz Detector. Cruise [31].

Concept (Li-Effect)

The Li-Effect or Li-Theory was first published in 1992 [6]. Subsequently the "Li Effect" was validated by secveral journal articles, independently peer reviewed by scientists well versed in General Relativity, [7, 8, 9, 10] including capstone paper, Li, et al [32]). The reader is encouraged to review the key results and formulas found in Li et al., [10] and the detailed discussion of the coupling among HFGWs, a magnetic field and a microwave beam found in Li et al. [10]. The Li-Effect is *very different* from the classical *(inverse)* Gertsenshtein- Effect. With the Li-Effect, a gravitational wave transfers energy to a separately generated electromagnetic (EM) wave in the presence of a static magnetic field. That EM wave has the same frequency as the GW and moves in the same direction. This is the "*synchro-resonance condition***,**" in which the EM and GW waves are synchronized and is **unlike the Gertsenshtein-Effect**.[12] The result of the intersection of the parallel and superimposed EM and GW beams, according to the Li-Effect, is *new EM photons moving off in a direction perpendicular to the*

beams and the magnetic field directions. These photons signal the presence of HFGWs and are termed a "perturbative photon flux" or PPF. Thus, these new photons occupy a separate region of space (see Fig. 10) that can be made essentially noise-free and the synchro-resonance EM beam itself (in this case a Gaussian beam) is not sensed there, so it does not interfere with detection of the photons. The existence of the transverse movement of new EM photons is a **fundamental physical requirement**; otherwise the EM fields will not satisfy the Helmholtz equation, the electrodynamics equation in curved spacetime, the non-divergence condition in free space, the boundary and will violate the laws of energy and total radiation power flux conservation. In this connection it should be recognized that *unlike the Gertsenshtein effect,* the Li-effect produces a *first-order* perturbative photon flux (PPF), proportional to the amplitude of the gravitational wave . A not A^2. In the case of the Gertsenshtein-Effect such photons are a second-order effect and according to Eq. (7) of Li, et al. [33] the number of EM photons are "…proportional to the amplitude squared of the relic HFGWs, A^2," … and that it would be necessary to accumulate such EM photons for at least 1.4×10^{16} seconds in order to achieve relic HFGW detection (Li et al., [33]) utilizing the Gertsenshtein-Effect. In the case of the Li theory the number of EM photons is proportional to the amplitude of the relic HFGWs, $A \approx 10^{-30}$, not the square, so that it would be necessary to accumulate such EM photons for less than about 1000 seconds in order to achieve relic HFGW detection (Li et al., [10]). The JASON report (Eardley, [13]) confuses the two effects and erroneously suggests that the Li-Baker HFGW Detector utilizes the inverse Gertsenshtein effect. It does not and does have a theoretical sensitivity that is about $A/A^2 = 10^{30}$ greater than that incorrectly assumed in the JASON report for the detection of relic HFGWs.

The Li-Baker HFGW detector operates as follows:

1. The perturbative photon flux (PPF), which signals the detection of a passing gravitational wave (GW), is generated when the two waves (EM and GW) have the same frequency, direction and suitable phase. This situation is termed "synchro-resonance." These PPF detection photons are generated (in the presence of a magnetic field) as the EM wave propagates along its *z*-axis path, which is also the path of the GWs, as shown in Figs. 10 and 11.

2. The magnetic field **B** is in the y-direction. According to the Li effect, the PPF detection photon flux (also called the "Poynting Vector") moves out along the x-axis in both directions.

3. The signal (the PPF) and the noise, or background photon flux (BPF) from the Gaussian beam have very different physical behaviors. The BPF (background noise photons) are from the synchro-resonant EM Gaussian beam and move in the z-direction, whereas the PPF (signal photons) move out in the x-direction along the x-axis and only occur when the magnet is on.

4. The PPF signal can be intercepted by microwave-absorbent shielded microwave receivers located on the *x*-axis (isolated from the synchro-resonance Gaussian EM field, which is along the *z*-axis).

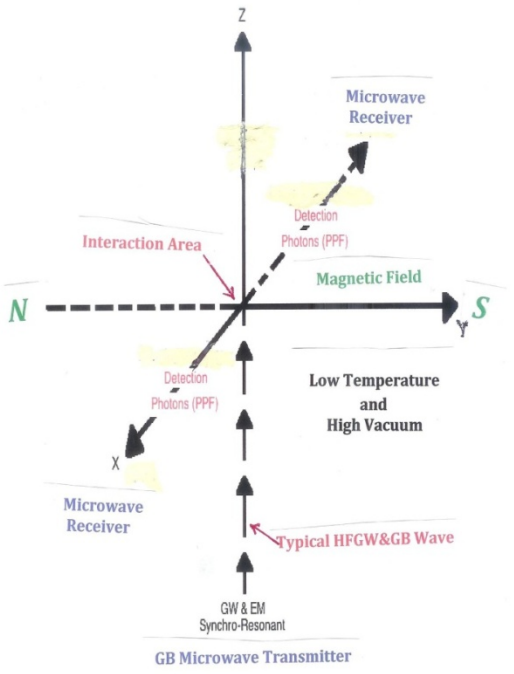


Figure 10.. Detection Photons Sent to Locations that are Less Affected by Noise.

5. The absorption is by means of off-the-shelf -40 db microwave pyramid reflectors/absorbers and by layers of metamaterial or MM absorbers (Landy, et al. 2008, Woods et al. [23] and Patent Pending) that also seal off out gassing. As discussed in detail by Woods, et al. [23] absorption of about –220 dB or an absorption coefficient of 10^{-22} for the two double MM layers, can be achieved. As noted by Landy, et al. [34] since "…impedance matching is possible, and with multiple layers, a *perfect* [absorbance] can be achieved." In addition, isolation is further improved by cooling the microwave receiver apparatus to reduce thermal noise background to a negligible amount as has been accomplished in single-photon receivers (Buller, [35]). In order to achieve a larger field of view and account for any curvature in the magnetic field, an array of microwave receivers having multiple horns (the two receivers having, for example, 12 cm by 12 cm horns (four such horns some two HFGW wavelengths or $2\lambda_{GW}$ on a side) could be installed at x = ± 100 cm (arrayed in planes parallel to the y-z plane). As noted in the following Table, all sources of noise in the Li-Baker HFGW detector such as *diffraction* from the intense Gaussian beam (Woods [36]), *dark-background shot noise, signal shot noise, Johnson noise, preamplifier noise, quantization noise, mechanical thermal noise, phase* or *frequency noise,* can be reduced to negligible amounts in a properly designed Li-Baker detector.

Noise Contributor	Brief Description of Noise source	Mitigation/Elimination Means	Nominal Computed Value photons s^{-1}, NEP W
Dark-background shot noise	GB noise especially diffraction	Wall geometry and absorbing wall materials	$4.2\times10^{-23}s^{-1}$,2.8 10^{-46}W
Signal shot noise	Noise in the signal itself	Part of useful data and not to be eliminated	--
Johnson noise	Thermal agitation in a power amplifier resistance	Refrigeration to low temperature	$5\times10^{-3}s^{-1}$,3×10^{-28} W
Preamplifier kTC noise	Stray capacitance and load resistance	Reducing bandwidth, load resistance and/or stray capacitance.	$1\times10^{-6}s^{-1}$, 8×10^{-30} W
Quantization noise	Analog to Digital Converter	Increasing the number of bits used	$2\times10^{-3}s^{-1}$, 1×10^{-28} W
Mechanical thermal noise	Brownian motion of sensor components.	Refrigeration to low temperature	$3\times10^{-4}s^{-1}$,2×10^{-27} W
Phase or Frequency noise	Fluctuations in the frequency of the microwave source for the GB.	Cavity-lock loop or a phase-compensating feedback loop	$5\times10^{-15}s^{-1}$,3×10^{-38} W

Summary Table of Li-Baker detector noise based upon experimental data concerning its components (Woods et al. [33]).

The total noise equivalent power or NEP is 1.02×10^{-26} W (noise flux is 1.54×10^{-3} photons per second). If need be the receivers could be further cooled and shielded from diffraction noise by baffles and optimum detector geometry as shown in Woods et al. [23]. Given a signal that exhibits the nominal value given in the Summary Table above of Woods et al. [23] of 99.2 s^{-1} photons, one quarter of which is focused on each of the microwave receivers, which is 24.8 s^{-1} photons or 1.6×10^{-22} W, the signal-to-noise ratio for each receiver is better than 1500:1 [23].

4. CONCLUSIONS

The utilization of modern MEMS technology and a double-helix array of them would allow for the construction of a HFGW generator or transmitter involving superradiance that exhibits sufficient strength to transmit HFGW signals globally. This is possible even though the conversion rate of EM power to GW power is exceedingly small and, like EM radiation, the GW signal power falls off as the inverse square of the distance. It is shown herein that a properly designed double-helix array of MEMS (or FBARS) can generate sufficient power to reach a receiver on the opposite side of the globe. Three HFGW detectors or HFGW receivers have previously been fabricated and others theoretically proposed, but analyses of their sensitivity suggest that for meaningful HFGW reception, greater sensitivity is necessary. The theoretical sensitivity of the Li-Baker HFGW detector discussed herein, that is based upon a different measurement technique than the other detectors, is predicted to satisfy HFGW communication needs The detector can be built from off-the-shelf, readily available components and, when coupled with the double-helix MEMS or FBAR array transmitter, could provide for transglobal HFGW communications.

Acknowledgments

This research was supported by *Transportation Sciences Corporation* in the United States and by *GravWave® LLC* internationally. The assistance of Professor Giorgio Fontana, *Trento University,* Italy, Professor Fangyu Li, *Chongqing University,* China and Gary V. Stephenson of *LinQuest Corporation,* Los Angeles, USA is gratefully acknowledged'

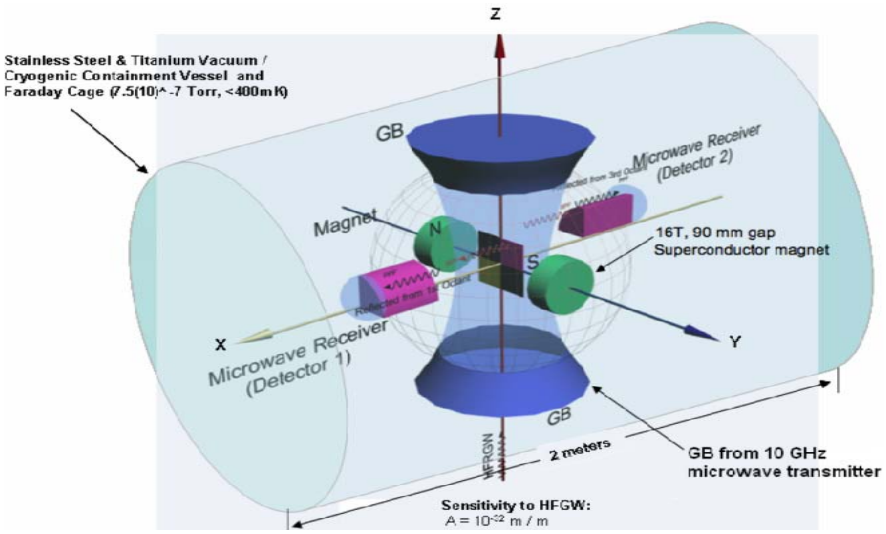


Figure 11. Schematic of Li-Baker HFGW Detector (Peoples Republic of China Patent Number 0510055882.2 [25])

. REFERENCES

[1] Douglass, D.H. and Braginsky, B. (1979), "Gravitational-radiation experiments," in "General relativity: an Einstein centenary survey" Ed. Hawking S.W. and Israel W. (CUP, UK), 90-137.

[2] Shawhan, P. S. (2004), "Gravitational Waves and the Effort to Detect them," *American Scientist* **92**, 356.

[3] Beckwith, A. W. (2009), "Relic High Frequency Gravitational waves from the Big Bang, and How to Detect them," *American Institute of Physics Conference Proceedings*, Melville, NY **1103, p.** 571... . arXiv:0809.1454v1 [physics.gen-ph] (Paper 031).

[4] Grishchuk, L. P. (2003), *Gravitational-Wave Conference,* The MITRE Corporation, May 6-9..

[5] Scully, M. O. and Svidzinsky, A. A. (2009), "The Super of Superradiance," *Science* **325**, pp.1510-1511.

[6] Li, F. Y., Tang M. and Zhao P. (1992), "Interaction Between Narrow Wave Beam-Type High Frequency Gravitational Radiation and Electromagnetic Fields," *Acta Physica Sinic*a **41,** pp. 1919-1928.

[7] Li,, F. Y., Meng-Xi Tang, Jun Luo, and Yi-Chuan Li (2000) "Electrodynamical response of a high-energy photon flux to a gravitational wave," *Physical Review D* **62**, July 21, pp. 044018-1 to 044018 -9.

[8] Li, F. Y.,, Meng-Xi Tang, and Dong-Ping Shi, (2003), "Electromagnetic response of a Gaussian beam to high-frequency relic gravitational waves in quintessential inflationary models," *Physical Review B* **67**, pp. 104006-1 to -17.

[9] Li, F. Y. and Baker, R. M L, Jr. (2007), "Detection of High-Frequency Gravitational Waves by Superconductors," *International Journal of Modern Physics* **21**, Nos. 18-19, pp. 3274-3278.

[10] Li F. Y., Baker R. M L, Jr., Fang Z., Stephenson G.V. and Chen, Z. (2008), "Perturbative Photon Fluxes Generated by High-Frequency Gravitational Waves and Their Physical Effects," *European Phys. J.* ***C*** **22**, Nos. 18-19, 30 July; peer reviews and manuscript available at: http://www.gravwave.com/docs/Li-Baker%206-22-08.pdf .

[11] Dehnen, H. and Fernando Romero-Borja (2003), "Generation of GHz – THz High-Frequency Gravitational Waves in the laboratory," paper HFGW-03-102, *Gravitational-Wave Conference,* The MITRE Corporation, May 6-9. Peer reviews and manuscript available at: http://www.gravwave.com/docs/Analysis%20of%20Lab%20HFGWs.pdf.

[12] Gertsenshtein, M. E, (1962), "Wave resonance of light and gravitational waves,"*Soviet Physics JETP*, Volume 14, Number 1, pp. 84-85.

[13] Eardley, et al. (2008) "High Frequency Gravitational Waves," JSR-08-506, October, the JASON defense science advisory panel and prepared for the Office of the Director of National Intelligence.

[14] Einstein, Albert, (1918) Über Gravitationswellen. In: Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, Berlin (1918), 154–167.

[15] Weber, J. (1964), "Gravitational Waves" in *Gravitation and Relativity*, Chapter 5, pp. 90-105, W. A. Benjamin, Inc., New York.

[16] Grishchuk, L. P. and Sazhin M. V. (1974), "Emission of gravitational waves by an electromagnetic cavity." *Soviet Physics JETP* **38**, Number 2, pp. 215-221.

[17] Landau, L. D. and Lifshitz, E. M. (1975), *The Classical Theory of Fields*, Fourth Revised English Edition, Pergamon Press, pp. 348, 349, 355-357.

[18] Baker, R. M L, Jr., Woods, R. C. and Fangyu Li (2006), "Piezoelectric-Crystal-Resonator High-Frequency Gravitational Wave Generation and Synchro-Resonance Detection," American Institute of Physics Conference Proceedings, Melville NY **813**, pp. 1280-1289. Manuscript available at: http://www.drrobertbaker.com/docs/AIP;%20HFGW%20Piezoelectric%20Generator.pdf

[19] Woods,, R. C. and Baker, Jr., R. M L (2005), "Gravitational Wave Generation and Detection Using Acoustic Resonators and Coupled Resonance Chambers," in the proceedings of *Space Technology and Applications International Forum (STAIF-2005)*, edited by M.S. El-Genk, American Institute of Physics Conference Proceedings, Melville, NY **746**, 1298.

[20] Baker, R. M L, Jr. (2006) "Novel formulation of the quadrupole equation for potential stellar gravitational-wave power estimation" *Astronomische Nachrichten* **327**, No. 7, pp. 710-713. Peer reviews and manuscript available at: http://www.gravwave.com/docs/Astronomische%20Nachrichten%202006.pdf

[21] Baker, R. M L, Jr. and Black, C. S.(2009), "Radiation Pattern for a Multiple-Element HFGW Generator," American Institute of Physics Conference Proceedings, Melville, NY **1103.** pp. 582-590. Manuscript available at: http://www.drrobertbaker.com/docs/Analyses%20of%20HFGW%20Generators%20and%20Radiation%20Pattern.pdf

[22] Baker, R. M L, Jr., Stephenson, G. V. and Li, F. Y. (2008), "Analyses of the Frequency and Intensity of Laboratory Generated HFGWs," American Institute of Physics Conference Proceedings, Melville, NY **969**, pp. 1045-1054. Peer reviews and manuscript available at: at: http://www.gravwave.com/docs/Analysis%20of%20Lab%20HFGWs.pdf

[23] Woods, R. C., Baker, Jr. R. M L , Li, F. Y. Stephenson, G. V. Davis, E. W. and Beckwith, A. W. (2011), "A new theoretical technique for the measurement of high-frequency relic gravitational waves," *Journ Mod. Phys.* **2**, No. 6, pp. 498-518; manuscript available at: . http://www.gravwave.com/docs/J.%20of%20Mod.%20Phys%202011.pdf and abstract available at http://vixra.org/abs/1010.0062**.**

[24] Stephenson, G. V. (2009) "The standard quantum limit for the Li-Baker HFGW detector**,"** in the *Proceedings of the Space, Propulsion and Energy Sciences International Forum (SPESIF***),** 24-27 February, Edited by Glen Robertson. (Paper 023), American Institute of Physics Conference Proceedings, Melville, NY **1103** , 542-547**; manuscript available at:** http://www.gravwave.com/docs/HFGW%20Detector%20Sensitivity%20Limit.pdf

[25] Woods R. C . (2007), "Modified Design of Novel variable focus lens for VHFGW," in the proceedings of *Space Technology and Applications International Forum (STAIF-2007)*, edited by M.S. El-Genk, American Institute of Physics Conference Proceedings, Melville, NY 880:1011-1018.

[26] Baker, Jr. R. M L (2001)) Peoples Republic of China Patent Number 01814223.0, "Gravitational Wave Detector," issued September 19, 2007; patent claims available at: **http://www.gravwave.com/docs/Chinese%20Detector%20Patent%2020081027.pdf**

[27] Cruise, A. M. (2000), "An electromagnetic detector for very-high-frequency gravitational waves," *Class. Quantum Gravity* **17**, pp. 2525-2530.

[28] Cruise, A. M. and Ingley, R. M. J. (2005), "A correlation detector for very high frequency gravitational waves," *Class. Quantum Grav.* **22**, 5479-5481.

[29] Gemme, B. F.., Parodi, R. and Picasso, E. (2001), "A detector of small harmonic displacements based on two coupled microwave cavities," *Review of Scientific Instruments* **72,** Number 5, May, pp. 2428-2437.

[30] Chincarini, A and Gemme, G. (2003), "Micro-wave based High-Frequency Gravitational Wave detector," HFGW-03-103, *Gravitational-Wave Conference,* The MITRE Corporation, May 6-9.

[31] Nishizawa, A. et al. (2008), "Laser-interferometric detectors for gravitational wave backgrounds at 100 MHz: Detector design and sensitivity," *Phys. Rev. D* **77,** Issue 2**,** 022002.

[32] Cruise, A. M. (2008), "Very High Frequency Gravitational Waves," Gravitational Wave Advanced Detector Workshop (GWADW), Elba Conference, 17 May, slide presentation 132.

[33] Li, F.Y., Yang, N., Fang, Z., Baker, Jr., R. M L, Stephenson, G. V. and Wen, H.,(2009), "Signal photon flux and background noise in a coupling electromagnetic detecting system for high-frequency gravitational waves," *Phys. Rev. D.* **80**, 060413-1-14; manuscript available at: http://www.gravwave.com/docs/Li,%20et%20al.%20July%202009,%20HFGW%20Detector%20Phys.%20Rev.%20D.pdf

[34] Landy. N. I., Sajuyigbe, S,. Mock, J. J., Smith, D. R. and Padilla (2008), "Perfect Metamaterial Absorber," *Physical Review Letters* **100**, pp. 207402-1-4, May 23.

[35] Buller, G. S. *et al*. (2010) ,"Single-Photon Generation and Detection," *Measurement Science and Technology* **21**, No. 1, pp. 1-28.

[36] Woods, R. C. (2011), "Estimate of diffraction from Gaussian Beam in Li-Baker HFGW detector," *Proceedings of the Space, Propulsion and Energy Sciences International Forum (SPESIF 2011***),** University of Maryland, Edited by Glen Robertson. (Paper 001); manuscript available at: http://www.gravwave.com/docs/Woods%202010.pdf

Analysis of dynamics fields systems accelerated by rotation. Dynamics of non-inertial systems.

Gabriel Barceló Rico-Avello
Doctor I. I.
Advanced Dynamics S.A., Spain, gabarce@iies.es

ABSTRACT

Starting from certain dynamic presumptions and based on a new interpretation of the behaviour of bodies which dispose of intrinsic angular momentum, when exposed to successive torques, new dynamics hypotheses have been developed, ending up with the conclusion that a new mathematical model in rotational fields dynamics can be set up. This would allow us to justify certain behaviours, until then not understood. Through this model different results are obtained, for certain assumptions, basing ourselves exclusively on a new interpretation of the composition or superposition of the motions originated by the acting torques.

We believe that the achieved results allow us to obtain a new perspective in dynamics, unknown up to date, making it possible to turn given trajectories which, until now, have been considered as chaotic, into deterministic terms. We have come to the conclusion that there still exists an unstructured scientific area in the present general assumptions and, specifically, in the area of rigid bodies exposed to simultaneous non-coaxial rotations.

For this purpose, it is necessary to analyze the velocity and acceleration fields that are generated in the body which disposes of intrinsic angular momentum, and assess new criteria in these speeds coupling. In this context, reactions and inertial fields take place, which cannot be justified by means of the classical mechanics.

It is the aim of this Paper to inform of the surprising results obtained, and to attract the interest towards the investigation of this new area of knowledge in rotational non inertial dynamics, and of its multiple and remarkable scientific and technological applications. [i]

I – Initial speculations and conjectures

It is possible to find new fields of research in new rotational dynamics of non-inertial systems. The foundations of rotational dynamics might be relevant to unsolved significant problems in physics.

Systems in the universe are in motion, in constant dynamic equilibrium. In the real universe, the general dynamic behaviour of rigid solids is characterized by its dynamic equilibrium. Through time, orbitation coexists with the intrinsic rotation. This aporia, and also the professor Miguel A. Catalán[ii] conjectures were our initial speculation.

The importance of our mathematical model is obvious. In this model not only the forces are leading players, but also the momentums of those forces which, while staying constant, will generate orbiting and constantly recurrent movements, generating a system in dynamic balance, and not being in unlimited expansion. This new dynamics theory will give us a better understanding of how universe and matter behave.

We would suggest a detailed and deep analysis of these dynamics hypotheses and propose continuing experimental testing necessary for confirmation.

II- Investigation project

We have been involved in an investigation project of non-inertial systems, to know the behaviour of rigid bodies exposed to simultaneous non-coaxial rotations. As a result of this investigation we have proposed new hypotheses in order to explain the dynamic behaviour of these bodies, insisting on the need of extending our studies on field theory.

We define the inertial reactions that are manifested in the matter, when it is subjected to accelerations, as *Dynamic Interactions* (ID). These are manifested in nature at any scale of magnitude. Any physical system and boundary conditions can be represented by a Lie group. The phase space of this dynamic is described in a quaternionic Kähler variety of 8-dimensional symplectic geometry. This would have two types of field's forces simultaneously: one corresponding to the actual applied forces and the other corresponding to inertial forces due to the *dynamic interactions* (ID) generated.

According to the *Relativity Principle of Galileo*, physical laws are identical in any system of *inertial reference*. The Classical Mechanics, and even the majority of modern physical theories, have been formulated for inertial reference frames, their validity having been proven in said systems.

Nevertheless, beyond these limits, in our opinion, there may be other assumptions in nature for which we nowadays still do not know exactly the laws to explain their behaviour, because the models of analysis we use are wrong. An example of this is the analysis of rigid solid bodies equipped with rotational movement. It is necessary in these cases to take into account possible inertial reactions.

The composition or superposition of the motions was understood also by Galileo to explain the *path* of a canyon ball. The question is to understand the superposition of fields originated by acting torques. We shall use the expression "coupling" in relation with the composition or superposition of the velocity field that are generated in these cases.

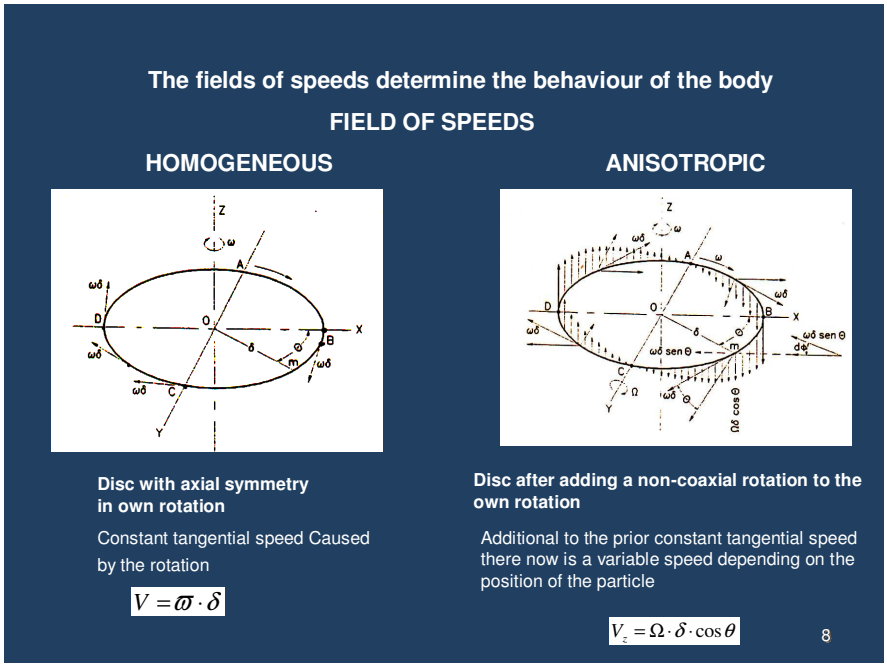


Figure I. The fields of speeds determine the behaviour of the body.

III- Rational deduction

This new non-inertial and non-Newtonian rotational dynamics of accelerated rigid solid bodies can be inferred in different ways: relativistic deduction, through equations of the generated fields, or through a rational deduction. We will concentrate on the last supposition.

In the case of a flat disc rotating around its symmetry axis, a field of speeds due to the rotation of the disc can be identified. In each particle of the disc, the generated tangential speed, will be identified in line with the equation:

$$\vec{v} = \delta\vec{\omega} \qquad (1)$$

With

δ: Distance from the particle to the rotation axis.

$\vec{\omega}$ *:* Rotation speed of the disc.

δ is also the circumference radio or the geometric place that contains the particles that are equidistant to the rotation axis and whose dynamic state we are analysing.

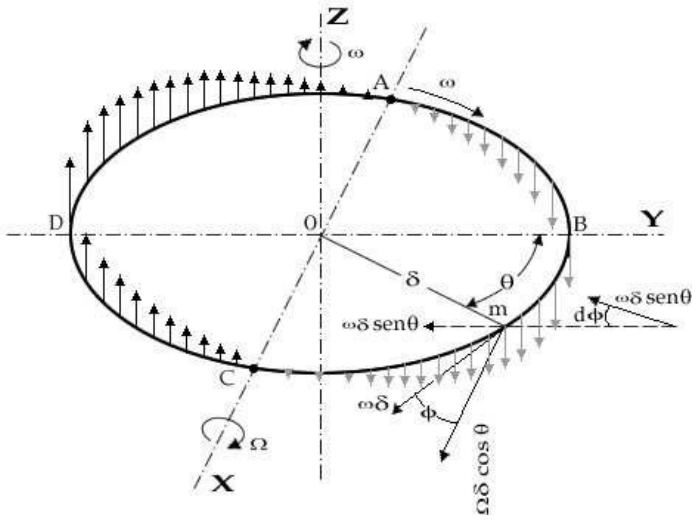


Figure II. In a body with $\vec{\omega}$ speed rotation on its principal axis, when it is subjected to a new no- coaxial rotation $\vec{\Omega}$, a non-homogeneous field of speeds is generated.

Therefore, all the particles situated in that circumference will have the same module of tangential speed but with a different orientation. As such, we obtain a homogeneous and balanced field of speeds as the result of the turn of the disc on its symmetry axis, figure I.

In the assumptions of simultaneous non-coaxial rotations, the rigid body experiences non-homogeneous speed fields, figure II. These fields generate anisotropic acceleration fields. These acceleration fields can be interpreted as fields of inertial forces, created in space through the effect of simultaneous non-coaxial rotations, figure III.

This rational deduction can be enlarged via an analysis inside the Fields Theory and its equations.

IV- Initial Paradox

We now express a paradox that permits us to introduce the concept of rotational inertia. When a body, with rotating intrinsic movement, is not submitted to external forces (or to its moments), according to the equations of Newton-Euler's mechanics:

$$d/dt\,(I\boldsymbol{\omega}) = 0 \qquad (2)$$

Result: $\boldsymbol{\omega} = Const.,$

Where I is the moment of inertia of the body, and $\boldsymbol{\omega}$ is its angular speed. The angular speed will be kept constant eternally due to its inertia.

Any rotation is an accelerated motion, since the linear velocity of every particle of the rigid body, though it remains constant in module, it will be constantly changing position. But being ω constant, we find the contradiction of having the example of a rotating movement accelerated by inertia, without any external force. This allows us to suppose the existence of one **Rotational Inertia**, fundamentally different from the Translational Inertia.

The Rotational Inertia would correspond to the inertia of the body when it has a movement of rotation; it will tend to maintain this rotation, despite the cessation of forces acting on it.

From the concept of rotational inertia, it is easy to infer a dynamic model based on constant rotation. A system maintains a constant angular speed when two points of the system remain in time, on the same dynamic state. In this case the system is in a state of **constant rotation** on a fixed axis.

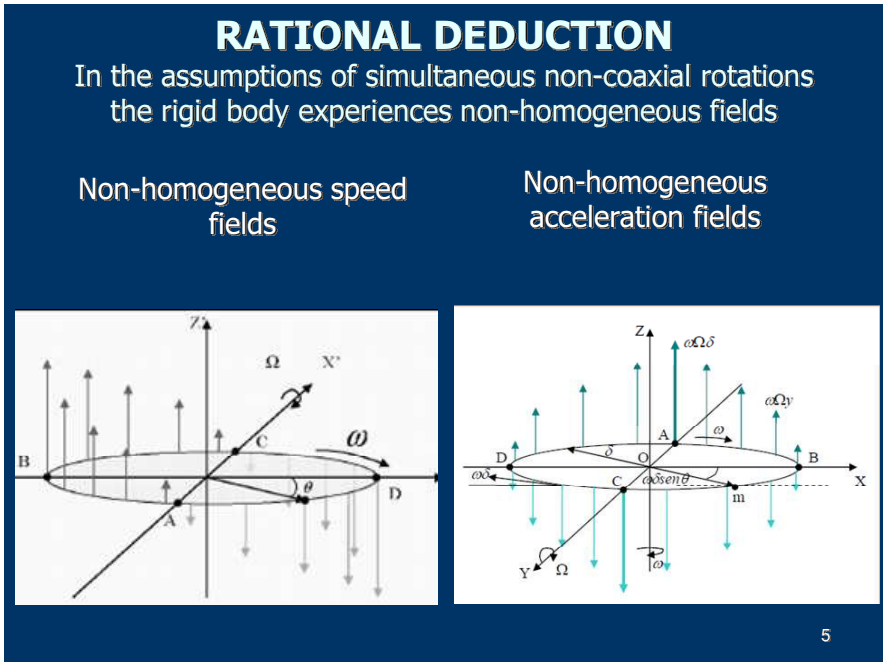


Figure III. Rational deduction

V- Axioms

The rotational dynamics based on the assumption of inertial reactions, is based on the principles of conservation of certain measurable quantities: **the motion quantity, the total mass and total energy**, and in this concepts: **Dynamics interactions, speed coupling, rotational inertia** and **constant rotation.**

From these principles of conservation of measurable magnitudes, and after the observation of the inertial reactions that occur in nature, we can deduce certain specific axioms:

1. **The rotation of space determines the generation of fields**

From a relativistic point of view, an intrinsic rotation can be interpreted as a fixed moving element plus a turn of the space of events which contains it. In a given body, two simultaneous non-coaxial intrinsic rotations can be generated around different axes. Two simultaneous intrinsic non-coaxial rotations of a given body, can be interpreted as two rotations of the space of events around different axes. The rotation of space determines the generation of anisotropic speeds and accelerations fields.

2. **Result of the action of non-coaxial moments**

When a solid is subjected to non-coaxial successive moments, non-homogeneous distributions of speeds and accelerations are generated. This can be identified as inertial fields.

3. **Inertial fields cause dynamic interactions**

The anisotropic speeds fields generated, interact with other fields of the rigid body, changing its dynamic state. For instance, the non-homogeneous tangential velocity field that is generated is coupled to the field of translation speeds.

4. **The action of successive non-coaxial torques on a rigid body cannot be determined by algebraic addition or calculated by the resultant force or torque**

This axiom reminds us of the impossibility of the use of vector algebra to solve these phenomena.

We understand that the inertial behaviour of the mass of the rigid solid body, when exposed to these movements has not been studied thoroughly.

VI- Motion Equation

Based on the **Principle of Conservation of the Motion Quantity,** and on the above mentioned axioms, we can obtain the motion equation.

We suppose that an infinitely flat disc with radio δ is submitted to a momentum *M,* whose axis coincides with the axis Z of the disc's symmetry. If this momentum is instantaneous, it generates a constant

rotation speed $\vec{\omega}$. A second momentum ***M´***, non-coaxial with the former will generate a new dynamic state which has been defined in Classic Mechanics as the gyroscopic effect, and which is attributed to a supposed gyroscopic momentum. This explanation of Classic Mechanics does not respond to the rational structures of the rest of the theory and represents a singularity in its conceptual development, as the axis of the new generated rotation does not coincides with the axis of the torque that generates such rotation.

We can interpret that the gyroscopic momentum does not exist physically, as it is simply the observable effect of a field of inertial forces generated by the simultaneous, non-coaxial, rotation of space (First axiom). We will check this starting point further below. In any case, based on the principle of **Conservation of the Motion Quantity,** the gyroscopic momentum ***D*** will be equivalent to the one acting from the outside ***M´*** and be the one that generates the second rotation non-coaxial with the first, and therefore:

$$M' = D \tag{3}$$

Supposing that the momentum ***M´*** will stay constant in time, it will keep its dynamic action on the body. Nevertheless, the referred gyroscopic momentum has been quantified through multiple methods of the classical mechanics with the following formula:

$$D = I\,\Omega\,\omega \tag{4}$$

If we observe the behaviour of this disc with own rotation $\vec{\omega}$ and submitted to a new non-coaxial torque ***M´*** (second axiom), we observe that it initiates a new rotation $\vec{\Omega}$ around an axis perpendicular to the new torque ***M´,*** and not around its own axis. Therefore we can infer that the field of inertial forces generated in the rotating space by a new non-coaxial momentum ***M´***, upon a moving body with a rotatory movement $\vec{\omega}$ and an inertial momentum *I* upon that rotation axis, and thus with an angular momentum $\vec{L}$, will oblige the moving body to acquire a precession speed $\vec{\Omega}$ defined by the scalar quotient:

$$\Omega = M'/(I\,\omega) = M'/L \tag{5}$$

The precession speed $\vec{\Omega}$ can be observed simultaneously with the initial $\vec{\omega}$, which remains constant within the body. Instead of the discriminating Poinsot[iii] hypothesis, which supposes that the angular momentums were coupled between each other and separate from the linear dynamic momentums, in the case of translation movement of the body, we propose the dynamic hypothesis which states that (Third axiom), **the field of translation speeds couples to the anisotropic field of inertial speeds generated by the second non-coaxial momentum**, forcing that the center of masses of the mobile modifies its path, without an external force having being applied in this direction.

As such, we obtain an **orbiting movement $\vec{\Omega}$, which is simultaneous with the constant intrinsic rotation of the moving body $\vec{\omega}$**. This new orbiting movement, generated by a non-coaxial momentum, is defined by **the rotation of the speed vector**, the latter $\vec{\omega}$ staying constant in module.

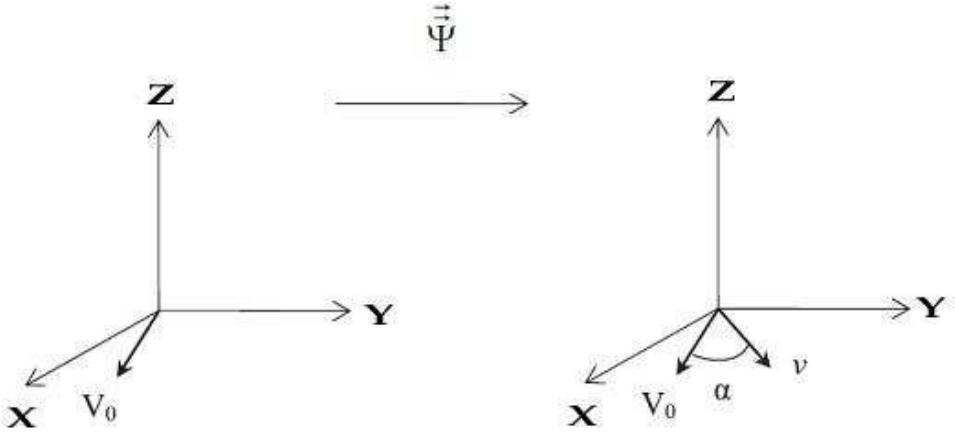


Figure IV. The rotational operator $\vec{\Psi}$ transforms, through one rotation α, the speed vector $\vec{V}_0$ into the speed vector v, both always situated on an identical plane, in this example in the XY plane. In reality in the plane that contains the acting momentum M'

In the assumption of figure IV, the new external momentum *M´*, which is supposed to be located on the *X* axis, will generate an inertial rotation around the *Z* axis, so that if the initial translation vector $\vec{V}_0$ was located on the *XY* plane, the resulting speed will stay on that plane after the rotation. Any rotation in space can be identified by a matrix. Therefore, in our supposition, the spatial rotation matrix $\vec{\vec{\Psi}}$ will be as follows:

$$\begin{pmatrix} \cos\alpha & -\text{sen}\,\alpha & 0 \\ \text{sen}\,\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

And will, in our assumption, generate a turn of the referred translation vector $\vec{V}_0$ in that *XY* plane.

Given that no external force modifying the quantity of translation movement has acted on the solid, its kinetic linear momentum will have to stay constant and therefore also its translation speed. However, if we accept that the homogeneous field of translation speeds couples to the field of tangential speeds due to the torque *M´* (Third axiom); we can determine what will be the new dynamic state of the body.

In this supposition, we then obtain as **motion equation** that the translation speed of the body's masses centre has not varied in magnitude and will therefore be equal to the initial in module, but submitted to the spatial rotation mentioned above:

$$\vec{v} = \vec{\vec{\Psi}}\ \vec{V}_0 . \quad (7)$$

The non-discriminating coupling proposed in our hypothesis is therefore identified by a spatial rotation of the speed vector and therefore:

$$\vec{v} = \begin{pmatrix} \cos\alpha & -\text{sen}\,\alpha & 0 \\ \text{sen}\,\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \vec{V}_0 . \quad (8)$$

As such we obtain:

$$\begin{pmatrix} \vec{v}_x \\ \vec{v}_y \\ \vec{v}_z \end{pmatrix} = \begin{pmatrix} \cos\alpha & -sen\,\alpha & 0 \\ sen\,\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} V_0^1 \\ V_0^2 \\ V_0^3 \end{pmatrix} = \begin{pmatrix} V_0^1\cos\alpha - V_0^2 sen\,\alpha \\ V_0^1 sen\,\alpha + V_0^2\cos\alpha \\ V_0^3 \end{pmatrix} \quad (9)$$

However, we can suppose that the new path will be the arch of a circle and the components of speed will thus be:

$$\begin{aligned} v_x &= -\delta \cdot \Omega \cdot sen\ (\Omega t) \\ v_y &= -\delta \cdot \Omega \cdot \cos\ (\Omega t) \\ v_z &= 0 \end{aligned} \quad (10)$$

We have:

$$\begin{pmatrix} -\delta.\Omega.\,sen\,(\Omega t) \\ -\delta.\Omega\cos\,(\Omega t) \\ 0 \end{pmatrix} = \begin{pmatrix} V_0^1\cos\alpha - V_0^2 sen\,\alpha \\ V_0^1 sen\,\alpha + V_0^2\cos\alpha \\ V_0^3 \end{pmatrix} \quad (11)$$

From which we obtain:

$$\begin{aligned} V_0^{\,1} &= 0 \\ V_0^{\,2} &= -\delta.\Omega \\ V_0^{\,3} &= 0 \end{aligned} \quad (12)$$

However, as: $\alpha = \Omega\, t$ (13)

Therefore the rotational operator $\vec{\vec{\Psi}}$ is in this supposition:

$$\vec{\vec{\Psi}} = \begin{pmatrix} \cos(\Omega t) & -sen\,(\Omega t) & 0 \\ sen\,(\Omega t) & \cos(\Omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (14)$$

And

$$\vec{V}_0 = \begin{pmatrix} 0 \\ \vec{V}_0 \\ 0 \end{pmatrix} \quad (15)$$

And, according to equation (5):

$\alpha = M´t/(I\,\omega)$ (16)

The motion equation finally can be written:

$$\vec{v} = \vec{\vec{\Psi}}\ \vec{V}_0 = \begin{pmatrix} \cos\ M´t/I\omega & -sen\,M´t/I\omega & 0 \\ sen\,M´t/I\omega & \cos\,M´t/I\omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \vec{V}_0 . \quad (17)$$

The rotational operator $\vec{\vec{\Psi}}$ transforms, through one rotation, the initial speed vector $\vec{V}_0$ into the speed vector v, both always situated on an identical plane.

It is observed how the rotational operator $\vec{\vec{\Psi}}$ is a function of the equations of sinus or cosinus of Ω t, existing a clear relation between the angular speed $\vec{\Omega}$ of orbit and the acting torque/couple M' and the initial angular speed $\vec{\omega}$. As such, we possess a simple mathematical relation between the angular speed $\vec{\omega}$ of the body and its translation speed $\vec{v}$.

In general, the mobile's path will be defined by intrinsic coordinates by the successive speeds of the body $\vec{v}$, determined by the matrix product of the rotational operator $\vec{\vec{\Psi}}$ on the initial speed vector $\vec{V}_0$. The referred equation (7) results, as a general equation of the movement for the bodies with angular momentum, when they are submitted to successive non-coaxial torques. For this equation, the rotational

operator $\vec{\vec{\Psi}}$ is the matrix transforming the initial speed into the one that corresponds to each successive dynamic state by means of a rotation.

Starting from the equation $\vec{v} = \vec{\vec{\Psi}} \ \vec{V}_0$ we can determine the surface in which all possible paths of a solid body with intrinsic rotation $\vec{\omega}$ would be present when varying simultaneously as parameters this speed or the time. As well the surfaces of paths for different speeds V_o when those are not a constant value, but also a function of another variable.

In short, in this simplified mathematical model, it would be possible that, in space, the mobile bodies submitted to successive non-coaxial torques, as a result of inertial dynamic interactions, would initiate an orbital movement so that, while maintaining the initial angular momentum and the second torque constant, its masses' centre would follow a closed orbital path without any need for real central forces.

So we can associate **dynamic effects to speed and a clear mathematical correlation between rotation and translation**. This mathematical connection allows us to identify a physical relation between **transfers of kinetic rotational energy to kinetic translation energy, and vice-versa.**

On the analysis of the dynamic behaviour of the bodies submitted to acceleration by rotation a change of mentality is necessary. In these cases, you cannot apply the same axioms and premises as those used in inertial systems (Fourth axiom).

VII – Experimental tests and physical-mathematical simulation model

The starting hypotheses as well as the mathematical model were confirmed by a series of experimental tests and also by a physical-mathematical simulation model of this behaviour.

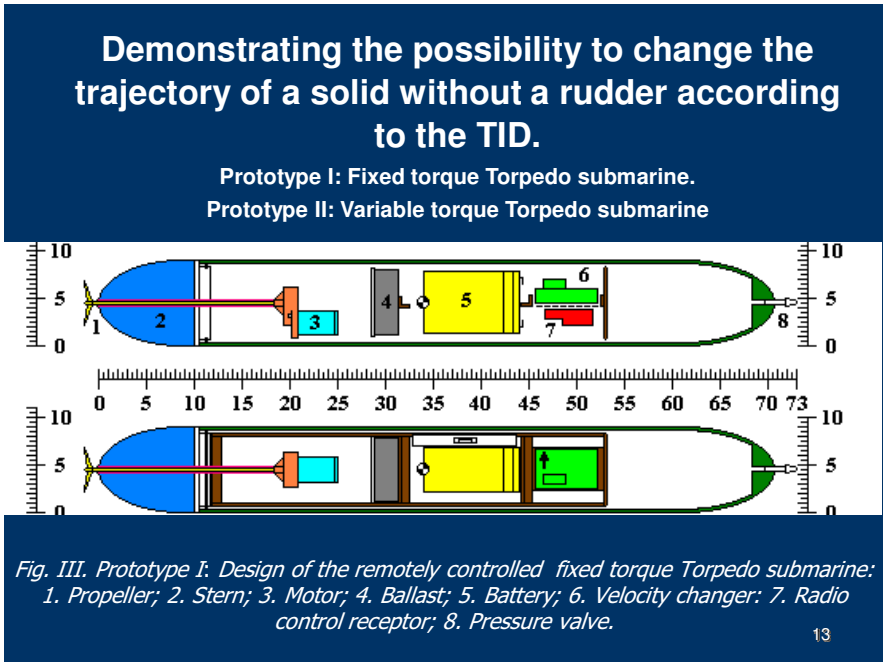


Figure V. Design of the remotely controlled submarine: 1. Propeller; 2. Stern; 3. Motor; 4. Ballast; 5. Battery; 6. Velocity changer: 7. Radio control receptor; 8. Pressure valve.

The challenge was to find a mobile with simultaneous angular momentum and linear speed. Because of the difficult availability of a device with these characteristics in space, it seemed sensible to continue the experiments with bodies floating in water. In this hypothesis, a cylinder or "Torpedo submarine" could be designed rotating around its longitudinal axis while at the same time driven by a propeller on its stern, provided as well with a gravitational torque perpendicular to the rotation axis. We make two different prototypes:

I Fixed torque Torpedo submarine, with constant unbalance.

II Variable torque Torpedo submarine, with pump and two deposits for water.

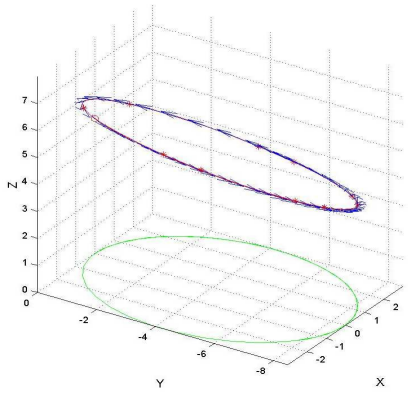


Figure VI. Path of the mass center of a mobile with intrinsic rotation and simultaneously submitted to an external momentum non-coaxial with its intrinsic angular momentum, obtained via computer simulation, in the supposed case that both, the applied moment and the translational linear speed of the mobile are constant. Simulating conditions: Tangential speed 5 m/s.

In accordance with the proposed dynamic hypotheses, a simulation of the behaviour of this solid in space, with intrinsic rotation and simultaneously submitted to an external momentum non-coaxial with its intrinsic angular momentum was realized, obtaining open or closed traces, equivalent to the trajectories of real bodies in space.

The equation of movement deducted and applied to a simulation model, determines a path that coincides with the one experimentally observed (Figure VI).

Also, quite a number of examples can be thought of for checking these dynamic hypotheses (see ***Un Mundo en Rotación***[iv]), which would allow us to interpret many, in our opinion, still unexplained assumptions in nature, using the interactions which result from rotating the space of events, as for example, the behaviour of so many rotating solid elements like the boomerang, the hoop or the wheel.

This new non-inertial rotational dynamics is developed in laws and corollaries, allowing a number of new, unknown scientific and technological applications. These *Rotational Dynamic Laws* are based on the inertial impossibility of matter to change their dynamic state in certain cases and propose the concept of ***rotational inertia*** as an *invariant of mass*. These laws are understood as a negation of nature to selective and discriminating couples established by Poinsot, and allow developing an alternative and specific *Theory of Dynamic Interactions (TID)* for bodies with angular momentum.

VIII - Conclusions

The present text is only a brief referential summary of the works carried out during the last twenty years in order to propose a *Rotational Dynamics of Interactions* applicable to bodies submitted to multiple successive non-coaxial torques. The initial hypotheses are based on new criteria about **speed coupling and rotational inertia**, and have been confirmed by experimental tests and by a mathematical model allowing the simulation of the real behaviour of bodies submitted to these excitations. A clear correlation between the initial speculations, the starting hypotheses, the mathematical simulation model, the deduced behaviour laws, the realized experimental tests and the mathematical model corresponding to the movement equations resultant of the proposed dynamic laws, have been obtained.

This research can be extended with the Field Theory and a relativistic deep analysis, and may allow the physical knowledge of new space systems and brings potential applications for the future, along with numerous relevant technology developments.

The *Theory of Dynamic Interactions* generalizes the concept of gyroscopic momentum, and of other inertial phenomena, incorporating them into the unified structure of a new *non inertial rotational dynamics of Interactions*.

According to the defended *Theory of Dynamic Interactions*, we can conceive a universe in a constant dynamic balance, in which a force momentum, with a zero resultant, will generate, as long as it works, a movement of constant orbiting, within a closed path. The importance of this mathematical model is obvious: not only the forces are leading players, but also the momentums of forces which, while staying constant, will generate orbiting and constantly recurrent movements, generating a system in dynamic balance, and not in unlimited expansion. This *Theory of Dynamic Interactions* is reasoned out and described in the book: ***Un Mundo en Rotación*** (2008), and its antecedents and fundamentals were presented in the book of the same author: ***El Vuelo del Bumerán*** (2005).

We want to suggest that interest should arise in physics in the exploration of non-inertial accelerated systems, and also to express a call for the need to develop scientific investigation projects for their evaluation and analysis, as well as technological projects based on these hypotheses. In our opinion, these hypotheses suggest new keys to understanding

the dynamics of our environment and the harmony of the universe. A universe composed not only of forces, but also of their momentums; and when these act constantly upon rigid rotating bodies, with an also constant translation speed, the result is a closed orbiting movement, thus a system which is moving, but within a dynamic equilibrium.

The application of these dynamic hypotheses to astrophysics, astronautics and to other fields of physics and technology possibly allows new and stimulating advances in investigation.

The result of this project is the conception of an innovative dynamic theory, which specifically applies to rigid rotating physical systems and which has numerous and significant scientific and technological applications,

Anyone interested in cooperating with this independent investigation project is invited to request for additional information to:

gestor@advanceddynamics.net

Or to look up at:

www.advanceddynamics.net.

---

[i] *On the Equivalence Principle,* presented by de author to the 61st International Astronautical Congress, Prague, CZ. Copyright ©2010 by Advanced Dynamics S.A. Published by the American Institute of Aeronautics and Astronautics, Inc. with permission.

[ii] Professor Miguel A. Catalán Sañudo. Spectroscopist. (Zaragoza 1894-Madrid 1957). Refer in *Un Mundo en rotación.* (Gabriel Barceló) Ed. Marcombo 2008, page 56.

[iii] Poinsot, L. *Théorie nouvelle de la rotation des corps*, 1834, refer by Gilbert: *Problème de la rotation d'un corps solide autour d'un point solide*, Annales de la société scientifique de Bruxelles, 1878, page 258 and by G. Barceló: *El vuelo del Bumerán*. Ed. Marcombo 2006, page 121.

Markovian Agents: A New Quantitative Analytical Framework for Large-Scale Distributed Interacting Systems

Andrea Bobbio,
Dipartimento di Informatica, Università del Piemonte Orientale, 15121 Alessandria, Italy
and
Dario Bruneo,
Dipartimento di Matematica, Università di Messina, Messina, Italy
and
Davide Cerotti,
Dipartimento di Informatica, Università di Torino, 10149 Torino, Italy
and
Marco Gribaudo,
Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milano, Italy

bobbio@mfn.unipmn.it, dbruneo@unime.it, davide.cerotti@di.unito.it, gribaudo@elet.polimi.it

ABSTRACT

A Markovian Agent Model (MAM) is a stochastic model that provides a flexible, powerful and scalable way for analyzing complex systems of distributed interacting objects. The constituting bricks of a MAM are the Markovian Agents (MA) represented by a finite state continuous time Markov chain (CTMC) whose infinitesimal generator is composed by a fixed component (the local behaviour) and an induced component that depends on the interaction with the other MAs. An additional innovative aspect is that the single MA keeps track of its position so that the overall MAM model is spatial dependent. MAMs are expressed with analytical formulas suited for numerical solution. Extensive applications in different domains have shown the effectiveness of the approach. In the present paper, we propose an example that illustrates how the MAM technique can cope with extremely large state spaces.

Keywords Markovian Agents, Distributed Interacting Systems, Performance Evaluation, Wireless Sensor Networks, Swarm Intelligence.

1. INTRODUCTION

Markovian Agents (MAs) are stochastic entities introduced with the aim of providing a flexible, powerful and scalable technique for modeling complex systems of distributed interacting objects, for which feasible analytical and numerical solution algorithms can be implemented. Each object has its own local behaviour that can be modified by the mutual interdependencies with the other objects. MAs are scattered over a geographical area and retain their spatial position so that the local behaviour may be related to the geographical position and the mutual interdependencies may depend on the relative distances and the transmittance characteristics of the interposed medium.

MAs are modeled by a discrete-state continuous-time finite CTMC whose infinitesimal generator may be influenced by the interaction with other MAs. The interaction among agents is represented by a *message passing model* combined with a *perception function*. When resident in a state or during a transition, an MA is allowed to send messages that are perceived by the other MAs, according to a spatial dependent *perception function*, modifying their behaviour. Messages may model real physical messages (as in wireless sensor networks) or simply the mutual influences of a MA over the other ones. Further, MAs may belong to a single class or to different classes with different local behaviours and interaction capabilities, and messages may belong to different types where each type induces a different effect on the interaction mechanism. The perception function describes how a message of a given type emitted by a MA of a given class in a given position in the space is perceived by MA of a given class in a different position. By consequence of the interaction mechanism, the entries of the infinitesimal generator of each MA are determined by the superposition of local terms and interaction induced terms.

In the previous literature, the modelling and analysis of large scale stochastic systems composed by interacting objects has been mainly faced by resorting to the superposition of interacting Markov chains, to asymptotic models or to fluid models. In the first case, the available techniques require the generation of the global state space, defined as the Cartesian product of the state spaces of the individual interacting objects. The explosion of the state space can be mitigated by exploiting symmetry properties, often included in the system definition, and producing the global transition rate matrix by means of tensor algebra operators applied to the local matrices [11]. Representative attempts in this direction define the interacting objects directly as Markov chains [8], [1], [10], or as finite state automata [22], [23] or as Petri nets [9], [21]. However, the compositional approaches, based on finite state objects, do

not account for interactions related to the relative position of the local objects. Recently, asymptotic models based on mean field theory have been proposed in the domain of performance evaluation [4], [2], [3], but they are not able to include spatial dependencies. Finally, fluid models [20], [17], [18] are able to capture the global behaviour of the system but loosing the capability of detailing the local behaviour.

In our approach, the local objects are finite state MAs and their interaction is represented by a fluid model. In this way, we do not need to explore the product state space, but we account for the effect of the global dependencies on the individual infinitesimal generators and we solve stochastic equations defined on individual sub-models. Interactive Markovian Agents have been introduced in [15], [16] for single class MAs and extended to Multi-class Multi-message Markovian Agent Model in successive works [14], [5], [13], [7]. In [12], mobility properties for the MAs have been introduced. Several application examples have been described and analyzed across the above papers and validation through simulation [7] or measuring real physical systems [6] has been provided.

The aim of this paper is to present the new modeling framework based on Markovian Agent Models (limited for simplicity to single-class MAs) and to show how the analytical model can be defined and implemented. A very large scale example has been selected to conclude the paper with the aim of illustrating the capabilities of the approach in dealing with extremely large state spaces.

2. MOTIVATION

Often complex systems are composed by many interacting objects that have their own local behaviour that can be modified by the mutual influences exercised with the other objects. As a representative example of this class of systems, Figure 1 represents a Wireless Sensor Network where the nodes of the network are sensors that behave according to their own specification but interact with the other sensors through the exchange of messages (represented by the links in the figure). In real networks, some nodes may fail or may be switched off, thus modifying not only the topology of the network but also the mechanism of the interaction and the mutual dependencies. Furthermore, the interaction may be dependent on the position of the objects and on their mutual distances.

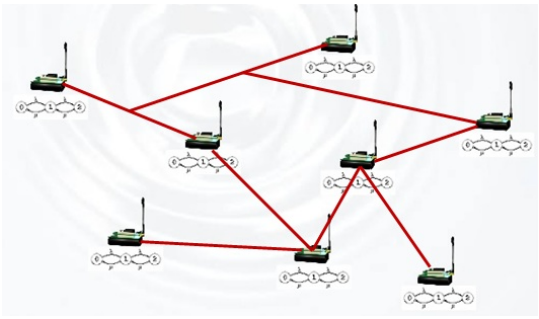


Fig. 1. Example of a complex system of interacting objects

Distributed systems with similar features, problems and difficulties may be found in many different technological areas and in many practical situations, like grids of computers, smart grids of power stations, and, in general, any system that can be represented in the form of a network where the nodes have an autonomous behaviour but, at the same time, are interdependent on the other nodes. The Markovian Agent Model was specifically studied to cope with the following needs:

i Provide analytical models that can be solved by numerical techniques, thus avoiding the need of long and expensive simulation runs;
ii Provide a flexible and scalable modeling framework for distributed systems of interacting objects;
iii Provide a framework in which local properties can be coupled with global properties;
iv Local and global properties and interactions may depend on the position of the objects in the space (space-sensitive models);
v The solution algorithm self-adapts to variations in the system topology and in the interaction mechanisms.

The constituent elements of a MAM are the MAs. MAs are represented by a finite state continuous time Markov chain (CTMC) whose infinitesimal generator is composed by two parts: a fixed component (the local behaviour) and an interaction induced component that depends on the interdependencies with the other MAs. MAs are scattered over a geographical area. A position dependent density function takes into account the density of MAs in each location and in each state of the CTMC characterizing the MA. The local properties of a MA may depend on its position and the mutual interdependencies may depend on the relative distances and the transmittance characteristics of the interposed media. The interaction among MAs is represented by a *message passing* model combined with a *perception function*. Messages may represent either real physical messages (as in wireless sensor networks) or, in general, the mutual influences exercised by an MA over the other MAs. The perception function rules the propagation of messages by taking into account the MA position in the space, the routing policy for the messages and the transmittance of the medium.

3. MARKOVIAN AGENT SPECIFICATION

The structure of a single MA is represented in Figure 2. States $i, j, \ldots, k$ are the states of the CTMC representing the MA. The transitions among the states are of two possible types that are drawn differently:

- Solid lines (like the transition from i to j or the self-loop from i) represent the fixed component of the infinitesimal generator and represent the local or autonomous behaviour of the object that is independent on the interaction with the other MAs (like, for instance, the time to failure distribution, or the reaction to an external stimulus). Note that we include in the representation also self-loop transitions that require a particular notation since are not visible in the infinitesimal generator of the CTMC [24].

- Dashed lines (like the transition from i to k or the transition into i) represent the transitions induced by the interaction with the other MAs. The way in which the rates of the induced transitions are computed is explained in the following section.

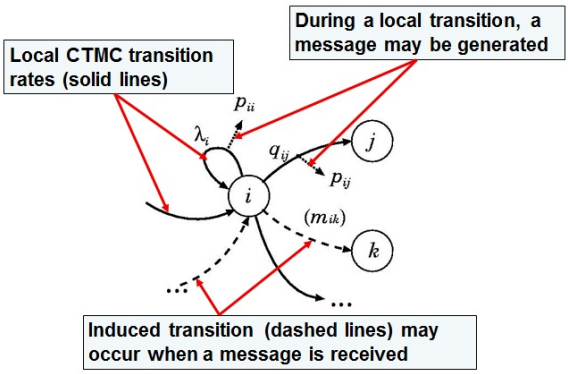


Fig. 2. Schematic structure of a Markovian Agent

During a local transition (or a self-loop) an MA can emit a message with an assigned probability, as represented by the dotted arrows in Figure 2 emerging from the solid transitions with a label denoting the corresponding message generation probability. Messages generated by an MA may be perceived by other MAs, according to a suitable perception function, and the interaction mechanism between emitted messages and perceived messages generates the induced transitions (dashed lines).

A MAM is a collection of interacting MAs defined over a geographical space $\mathcal{V}$. Given a position $\mathbf{v}$ inside $\mathcal{V}$, we define $\rho(\mathbf{v})$ the density of MAs in $\mathbf{v}$. According to the definition of the density $\rho(\mathbf{v})$, we can classify a MAM with the following taxonomy:

- A MAM is *static* if $\rho(\mathbf{v})$ does not depend on time, and *dynamic* if it does depend on time;
- A MAM is *discrete* if the geographical area on which the MAs are deployed is discretized and $\rho(\mathbf{v})$ is a discrete function of the space or it is *continuous* if $\rho(\mathbf{v})$ is a continuous function of the space.

Mathematical formulation

For the sake of simplicity we provide here the formal definition of a MAM with a single class of MAs and a single type of messages [16]. The extension to multi-class multi-message MAMs has been described in [7].

An MA is formally specified by the following tuple, where $\mathbf{v}$ is the position of the MA in the space $\mathcal{V}$:

$$MA(\mathbf{v}) = \{\mathbf{Q}(\mathbf{v}), \Lambda(\mathbf{v}), \mathbf{P}(\mathbf{v}), \mathbf{A}(\mathbf{v}), \mathbf{p_0}(\mathbf{v})\} \qquad (1)$$

where:

$\mathbf{Q}(\mathbf{v})$ is the local component of the infinitesimal generator;
$\Lambda(\mathbf{v})$ is the array of the self-jump transition rates;
$\mathbf{P}(\mathbf{v})$ is the probability of message generation;
$\mathbf{A}(\mathbf{v})$ is the probability of message acceptance;
$\mathbf{p_0}(\mathbf{v})$ is the initial probability vector.

Note that in the single-class MAM even if the structure of the CTMC associated to each MA is the same for all the objects, the values of the parameters may depend on position $\mathbf{v}$ and, therefore, may vary from MA to MA.

From the above definitions, we can compute the total rate $\beta_j(\mathbf{v})$ at which messages are produced by an MA in state j, in position $\mathbf{v}$:

$$\beta_j(\mathbf{v}) = \lambda_j(\mathbf{v})\, p_{jj}(\mathbf{v}) + \sum_{k \neq j} q_{jk}(\mathbf{v})\, p_{jk}(\mathbf{v}) \qquad (2)$$

where the first term in the r.h.s is the contribution of the messages emitted during a self-loop from j and the second term is the contribution of messages emitted during a transition from j to any $k\,(\neq j)$.

The interdependencies among MAs are ruled by a perception function $u(\mathbf{v}, i, \mathbf{v}', j')$ that defines how messages generated from an MA in state j' at position $\mathbf{v}'$, are received by an MA in state i at position $\mathbf{v}$. The perception function is a structural part of the model and it contributes to quantify the interdependencies among MAs. The perception function defines how messages issued by an MA in a given spatial location and in a given state propagate in the space and how they are perceived by other MAs in different locations. The transition rates of the induced transitions are primarily determined by the structure of the perception function.
A pictorial and intuitive representation of how the perception function $u(\mathbf{v}, i, \mathbf{v}', j')$ acts, is given in Figure 3. The MA in the bottom right portion of the figure in position $\mathbf{v}'$ broadcasts a message of type m that propagates in the geographical area until reaches the MA in the top left portion of the Figure in position $\mathbf{v}$. Upon acceptance of the message according to the acceptance matrix $\mathbf{A}(\mathbf{v})$, a new induced transition (represented by a dashed line) is generated in the model.

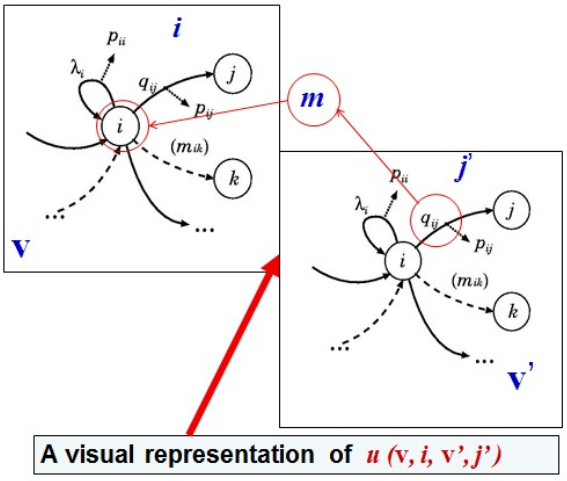


Fig. 3. Message passing mechanism ruled by a perception function

With the above definitions we are now in the position to compute the components of the infinitesimal generator of an MA that depend on the interaction with the other MAs and that constitute the original and innovative part of the approach.

We define $\gamma_{ii}(t, \mathbf{v})$ the total rate at which messages coming from the whole volume $\mathcal{V}$ are perceived by an MA in state i in location $\mathbf{v}$.

$$\gamma_{ii}(t, \mathbf{v}) = \int_{\mathcal{V}} \sum_{j=1}^{n} u(\mathbf{v}, i, \mathbf{v}', j)\ \beta_j(\mathbf{v}')\ \rho_j(t, \mathbf{v}') d\mathbf{v}' \qquad (3)$$

$\gamma_{ii}(t, \mathbf{v})$ in Equation (3) is computed by taking into account the total rate of messages emitted by all the MAs in state j and in a given position $\mathbf{v}'$ (the term $\beta_j(\mathbf{v}')$) times the density of MAs in $\mathbf{v}'$ (the term $\rho_j(t, \mathbf{v}')$) times the perception function (the term $u(\mathbf{v}, i, \mathbf{v}', j)$) summed over all the possible states j of each MA and integrated over the whole space $\mathcal{V}$. From an MA in position $\mathbf{v}$ and in state i an induced transition to state k (drawn in dashed line) is generated with rate $\gamma_{ii}(t, \mathbf{v})\, a_{ik}(\mathbf{v})$ where $a_{ik}(\mathbf{v})$ is the appropriate entry of the acceptance matrix $\mathbf{A}(\mathbf{v})$.

We then group the rates $\gamma_{ii}(t, \mathbf{v})$ in the diagonal matrix $\Gamma(t, \mathbf{v})$ (see Equation (4));

$$\mathbf{\Gamma}(t, \mathbf{v}) = [\, \gamma_{ii}(t, \mathbf{v}) \,] \qquad (4)$$

combining the local component $\mathbf{Q}(\mathbf{v})$ with the induced component $\Gamma(t, \mathbf{v})$, and taking into account the message acceptance matrix $A(\mathbf{v})$, we finally obtain the complete form of the infinitesimal generator $\mathbf{K}(t, \mathbf{v})$ of the MA as defined in Equation (5).

$$\mathbf{K}(t, \mathbf{v}) = \mathbf{Q}(\mathbf{v}) + \sum \mathbf{\Gamma}(t, \mathbf{v})\, [\mathbf{A}(\mathbf{v}) - \mathbf{I}], \qquad (5)$$

Once the complete generator for any MA is computed from (5) (and this is the most intensive computational part of the solution algorithm [7]) we can solve for the density $\boldsymbol{\rho}(t, \mathbf{v})$ of MAs in $\mathcal{V}$ the standard Chapman-Kolmogorov (second Equation in 6) under initial condition (first Equation in 6).

$$\begin{cases} \boldsymbol{\rho}(0, \mathbf{v}) &= \rho(\mathbf{v})\boldsymbol{\pi}_0(\mathbf{v}) \\ \dfrac{d\boldsymbol{\rho}(t, \mathbf{v})}{dt} &= \boldsymbol{\rho}(t, \mathbf{v})\mathbf{K}(t, \mathbf{v}). \end{cases} \qquad (6)$$

Note that each equation in (6) has the dimension of the CTMC of a single MA. In this way we have decomposed a problem defined over the product state space of all the MAs into several subproblems, one for each MA, having decoupled the interaction by means of Equation (5). Solution of each Equation in (6) is obtained by resorting to standard numerical techniques for differential equations, and provides the basic time-dependent measures to evaluate more complex performance indices associated to the system

4. APPLICATIONS

To show the flexibility of the MAM framework, several applications in different domains have been recently discussed also in cooperation with different research groups that were attracted by the capabilities offered by the model. The original and motivating application was related to Wireless Sensor Networks [15], [16]; then we showed applications related to the propagation of seismic waves in inhomogeneous media [14] and, in collaboration with colleagues of the Universidad Politecnica de Valencia (Spain), propagation of fire in inhomogeneous terrains and wind fields [13]. Finally, in collaboration with the University of Messina (Italy) we applied MAMs to swarm intelligent protocols [5], [6], [7].

The illustrative final example, presented in the following section, refers to a different field and regards an application in which a MAM provides a solution to a very-large-scale multi-body optimization problem, whose solution may be very time consuming or even unfeasible also by a simulative approach.

Large-scale multi-body optimization

The problem can be formulated in the following terms. The geographical space is in the form of a square grid of $n \times n = n^2$ cells, and we put one MA per cell. In the geographical space a number of randomly assigned cells are special aggregation points called *sinks*. Each cell of the grid must find the shortest (optimal) path to one of the sinks. The sinks may vary in time, in position and in quantity (some sinks may disappear, or new sinks may appear).

The defined optimization problem has been solved by resorting to swarm intelligent concepts [19], [7], based on the exchange of *pheromone* messages. In analogy with the biological process of ant colonies, each node sends a message containing its pheromone level and updates its value based on the level of its neighboring nodes, creating a pheromone gradient toward the sinks. The search for the shortest path is driven by the pheromone gradient, and in particular the shortest path is obtained by following the steepest pheromone gradient toward a sink. Any change on the network condition is reflected by an update of the pheromone intensity distribution.

The pheromone intensity is discretized in P levels from 0 to $P-1$, and the gradient construction is triggered by the sinks that emit messages with the highest pheromone level $P-1$ with rate λ. The MA representing the sink is reported in Figure 4a). In all the experiments we have assumed $P = 20$, i.e. we have discretized the pheromone intensity in $P = 20$ levels.

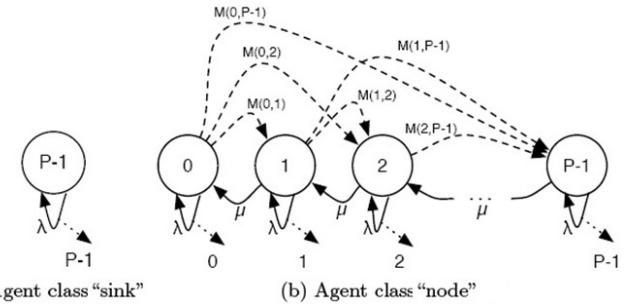


Fig. 4. Structure of swarm intelligent MAs

An MA modeling a cell node has P states representing the discretized pheromone levels (Figure 4b). When an MA node in state i receives a message encoding a pheromone level m, generates an induced transition (dashed lines) and jumps to a

state j that represents the pheromone level that is the mean between its current level i and the one encoded in the received message m. More formally, the dashed transitions in Figure 4b are labelled with label $M(i,j)$ defined as [7]:

$$M(i,j) = \{m \in [0 \cdots P-1] : round((m+i)/2) = j\} \quad \forall i,j \in [0 \cdots P-1] : j > i. \tag{7}$$

In other words, an MA in state i jumps to the state j that represents the pheromone level equal to the mean between the current level i and the level m encoded in the perceived message.

Then the MA emits its pheromone message with the actual level intensity with rate λ (mechanism represented by the self-loops). Furthermore, the pheromone evaporates with rate μ (local solid transitions) allowing the system to forget old information. Hence, the actual pheromone level in any cell at any time is a stochastic balance between the emission and evaporation rates.

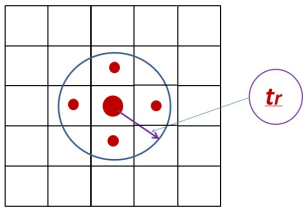


Fig. 5. The effect of the transmission range t_r

The pheromone propagation algorithm and the formation of the pheromone gradient depend on the pheromone emission rate λ, the pheromone evaporation rate μ and the transmission range t_r. The transmission range t_r takes into account that the energy of the messages emitted by the MAs is limited so that a message emitted in a given position can be perceived only by the MAs located inside a circle of radius t_r as depicted in Figure 5. In all the experiments the transmission range covers only the 4 first neighboring cells (Figure 5). The effect of a limited transmission range is reflected in the structure of the perception function whose definition is given by:

$$u(\mathbf{v}, i, \mathbf{v}', j') = \begin{cases} 0 & \text{dist}(\mathbf{v}, \mathbf{v}') > t_r \\ 1 & \text{dist}(\mathbf{v}, \mathbf{v}') \leq t_r, \end{cases} \tag{8}$$

Equation (8) represents the fact that a message emitted in location $\mathbf{v}'$ can be perceived by an MA in location $\mathbf{v}$ only if the distance between the two locations is less that t_r.

In the following optimization experiment we have assumed a rectangular grid with $n = 100$ hence with $100 \times 100 = 10,000$ cells, and we have randomly scattered 50 sinks in the grid. The grid is represented in Figure 6, where the sinks are drawn as black spots. Since each cell is represented by an MA with $P = 20$ states, the product state space of the overall system has $N = 20^{10,000}$ states!

At time $t = 0$ the sinks start emitting their pheromone message with the highest intensity level while all the other nodes have a pheromone level equal to 0. This situation is represented in the upper part of Figure 7. Then, the pheromone

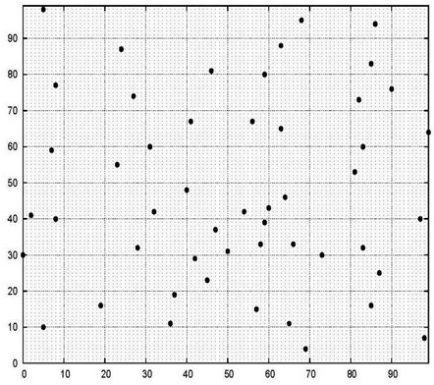


Fig. 6. The 100×100 grid with $10,000$ cells and 50 randomly scattered sinks

messages spread into the grid according to the described MAM model, until the pheromone intensity gradient reaches a steady condition, that depends on the balance between the emission rate λ and the evaporation rate μ, as represented in the bottom part of Figure 7 . The steady gradient intensity in Figure 7 provides the required optimal solution.

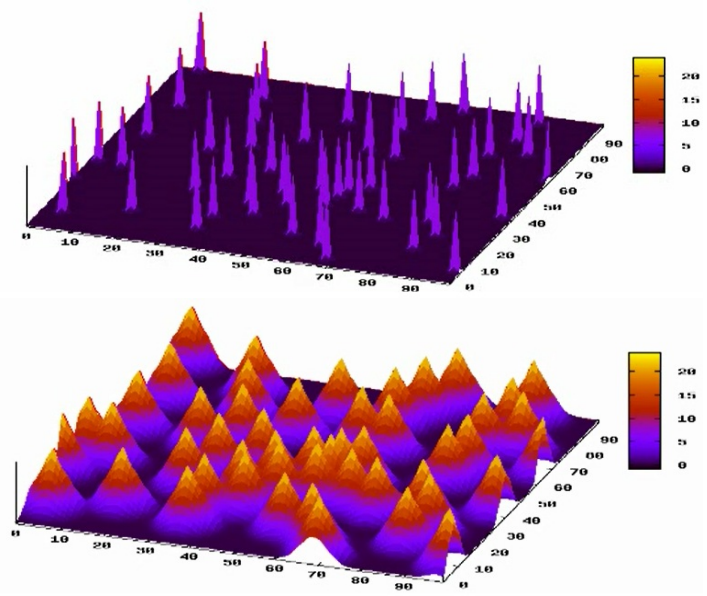


Fig. 7. The pheromone gradient at $t = 0$ (upper part) and in the final steady condition (bottom part)

Each cell, following the steepest gradient line in Figure 7, reaches the closest sink with the minimum number of hops. It is noteworthy to observe that the optimal solution is searched among $N = 20^{10,000}$ states, and that all the N states are in principle reachable depending on the position of the sinks. Furthermore, the algorithm is self-adapting to any topological modification like a variation in the position or in the number of sinks.

The steady gradient configuration is reached in few minutes on a standard laptop, since we have strongly limited the

interaction term by means of the perception function defined in Equation 8. Nevertheless, an optimal solution is obtained with any configuration and number of sinks.

5. CONCLUSIONS

The analytical MAM model has been presented to show how it can provide a flexible framework to model complex stochastic systems made of many interacting parts that have a local behaviour that can be modified by the global interaction. Moreover, the model is a location sensitive model, in the sense that the geographical position of the objects and their mutual distances are part of the model. This spatially dependent characteristic is reflected both in the structure of the model (for instance MAs of different classes) and in the values of the model parameters. The analytical MAM model is solved by resorting to numerical techniques.

The illustrative example shows how a MAM model can search for an optimal solution in a huge state space for which any other techniques (both analytical and simulative) will fail.

REFERENCES

[1] F. Ball, R.K. Milne, I.D. Tame, and G.F. Yeo. Superposition of interacting aggregated continuous-time Markov chains. *Advances in Applied Probability*, 29:56–91, 1997.

[2] M. Benaim and J.Y. Le Boudec. A Class Of Mean Field Interaction Models for Computer and Communication Systems. *Performance Evaluation*, 65(11-12):823–838, 2008.

[3] A. Bobbio, M. Gribaudo, and M. Telek. Analysis of large scale interacting systems by mean field method. In *QEST 08, IEEE Computer Society*, pages 215–224, 2008.

[4] J.Y. Le Boudec, D. McDonald, and J. Mundinger. A generic mean field convergence result for systems of interacting objects. In *4th International Conference on Quantitative Evaluation of Systems - QEST2007*, Edinburgh, 2007.

[5] D. Bruneo, M. Scarpa, A. Bobbio, D. Cerotti, and M. Gribaudo. Analytical modeling of swarm intelligence in wireless sensor networks. In *Fourth International Conference on Performance Evaluation Methodologies and Tools (Valuetools 2009)*, 2009.

[6] D. Bruneo, M. Scarpa, A. Bobbio, D. Cerotti, and M. Gribaudo. Adaptive swarm intelligence routing algorithms for WSN in a changing environment. In *IEEE Sensors 2010 Conference*, pages 1813–1818, 2010.

[7] D. Bruneo, M. Scarpa, A. Bobbio, D. Cerotti, and M. Gribaudo. Markovian agent modeling swarm intelligence algorithms in wireless sensor networks. *Performance Evaluation*, In Press, Corrected Proof:–, 2011.

[8] P. Buchholz. Hierarchical Markovian models -symmetries and aggregation. *Performance Evaluation*, 22:93–110, 1995.

[9] P. Buchholz. Hierarchical structuring of superposed GSPNs. *IEEE Transactions Software Engineering*, 25:166–181, 1999.

[10] P. Buchholz and T. Dayar. Comparison of multilevel methods for Kronecker based Markovian representations. *Computing*, 73:349–371, 2004.

[11] P. Buchholz and P. Kemper. Kronecker based matrix representations for large Markov chains. In M. Siegle B. Haverkort, H. Hermanns, editor, *Validation of Stochastic Systems*, pages 256–295. Springer Verlag - LNCS, Vol 2925, 2004.

[12] D. Cerotti, M. Gribaudo, and A. Bobbio. Presenting dynamic markovian agents with a road tunnel application. In *MASCOTS09*, pages 621–624. IEEE-CS, 2009.

[13] D. Cerotti, M. Gribaudo, A. Bobbio, C.T. Calafate, and P.Manzoni. A Markovian agent model for fire propagation in outdoor environments. In *Computer Performance Engineering (EPEW2010)*, pages 131–146. Springer Verlag - LNCS, Vol 6342, 2010.

[14] A. Bobbio D. Cerotti, M. Gribaudo. Disaster propagation in inhomogeneous media via markovian agents. In *Critical Information Infrastructure Security*, pages 328–335. Springer Verlag - LNCS, Vol 5508, 2009.

[15] M. Gribaudo and A. Bobbio. Performability analysis of a sensor network by interacting markovian agents. In *Proceedings 8-th International Workshop on Performability Modeling of Computer and Communication Systems (PMCCS-8)*, 2007.

[16] M. Gribaudo, D. Cerotti, and A. Bobbio. Analysis of on-off policies in sensor networks using interacting Markovian agents. In *4-th Int Workshop on Sensor Networks and Systems for Pervasive Computing - PerSens 2008*, pages 300–305, 2008.

[17] M. Gribaudo, C.-F. Chiasserini, R. Gaeta, M. Garetto, D. Manini, and M. Sereno. A spatial fluid-based framework to analyze large-scale wireless sensor networks. In *IEEE International Conference on Dependable Systems and Networks, DSN2002*, 2005.

[18] J. Hillston. Fluid flow approximation of PEPA models. In *2nd International Conference on Quantitative Evaluation of Systems - QEST*, pages 33–43, 2005.

[19] M.G. Hinchey, R. Sterritt, and C. Rouff. Swarms and swarm intelligence. *IEEE Computer*, pages 111–113, April 2007.

[20] J.M. Kelif and E. Altman. Downlink fluid model of CDMA networks. In *IEEE 61th Vehicular Technology Conference (VTC 2005)*, 2005.

[21] P. Kemper. Transient analysis of superposed GSPNs. *IEEE Transactions on Software Engineering*, 25:182–193, 1999.

[22] B.D. Plateau and K. Atif. Stochastic automata network for modeling parallel systems. *IEEE Transactions on Software Engineering*, 17:1093–1108, 1991.

[23] B.D. Plateau and J.M. Fourneau. A methodology for solving Markov models of parallel systems. *Journal of Parallel and Distributed Computing*, 12:370–387, 1991.

[24] K. Trivedi. *Probability & Statistics with Reliability, Queueing & Computer Science applications*. Wiley, II Edition, 2001.

Gain Scheduling Control Experiment of Balancing Transformer Robot using LEGO Mindstorms

Kentaro Hirata, Mayumi Tomida, and Kazuyoshi Hatada
Graduate School of Information Science,
Nara Institute of Science and Technology
Ikoma, Nara, 630-0192, Japan

ABSTRACT

In this paper, we consider a Hands-On experiment with LEGO Mindstorms to demonstrate advanced control theory, especially the gain scheduling method. Such a learning style is extensively studied in the field of engineering education these days. Our goal is to create a mobile robot which transforms its posture while maintaining the balance. First, we derive the equation of motion of our robot. Since the inertia matrix contains the scheduling parameter related to the posture, we remove it through a redundant descriptor expression. Based on the obtained LPV (Linear Parameter Varying) model in LFT (Linear Fractional Transformation) form, we design gain scheduling controllers via the dilated LMIs (Linear Matrix Inequalities). Finally, we show the simulation results and experiments.

keywords: Gain Scheduling, LPV System, Dilated LMI and Descriptor Form

1 INTRODUCTION

To prevent the recent tendency among young generations to move away from the scientific field and to fill the gap between the theory and practice in engineering discipline, growing attention is focused on hands-on education [1]. In this paper, we describe our attempt to introduce this hands-on approach into the study of advanced theories in the control engineering discipline. Specifically, advanced control experiment using LEGO Mindstorms [2] is considered[1]. As is widely understood, Mindstorms is a suitable tool for hands-on experiences on robotics. Actually, our institute has been providing kids robot school (Fig. 1) using Mindstorms for local community and elementary schools more than five years as one of our Academic Volunteer Education activities. It is used not only in elementary education [4], but also in control experiments for undergraduates [9, 7], in Lab experiments [3] and even in introductory education for engineers in industries [8].

Fig. 1 Kids Robot School

[1]LEGO and Mindstorms are trademarks of the LEGO Group.

From control engineering perspective, various attempts including anti-sway control experiment [9] (Fig. 2) based on classical and modern control theory, or stabilization of balancing robot (NXTway-GS) [10] (Fig. 3) based on modern and robust control theory are reported.

Fig. 2 Anti-sway experiment

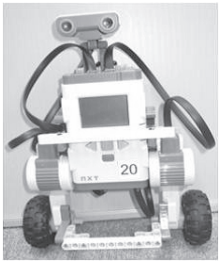


Fig. 3 NXTway-GS

Our target here is a demonstration of the gain scheduling control on Mindstorms platform. This control scheme is classified as one of the advanced robust control theory. In general, the discrepancy between the theory and practice becomes larger as the theoretical advance goes further. This is why we selected the gain scheduling control. Fig. 4 shows our LEGO based robot and the target behavior. Such an acrobatic motion with LEGO robot is appealing and can emphasize the importance of the control theory behind the modern high-tech products even for newcomers. Usually, the construction of laboratory-level control experimental facilities by ourselves costs time and money. In contrast, the experiment here *only* requires with one LEGO Mindstorms kit, one gyro sensor (Fig. 5) sold by a third party company [6] and a laptop PC.

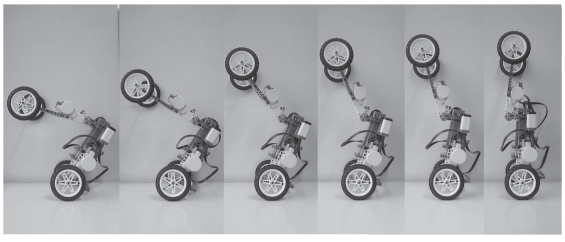


Fig. 4 Balancing Transformer Robot

Fig. 5 Gyro Sensor

The software development environment nxtOSEK [5] with real time OS for LEGO Mindstorms is provided as a free and open source software. Another potential merit to use LEGO Mindstorms is that we can avoid the time-consuming system identification process inevitable for usual lab experiment with custom-made plants. Since LEGO Mindstorms is produced under its industrial standard, the dynamics of the robots made from the same parts are exactly the same.

The following notations are used. For a square matrix A, $\mathrm{He}[A] = A + A^T$. Symmetric part in LMI condition is expressed by $*$.

2 MODELING

Balancing Transformer Robot

Our robot maintains the balance by driving two wheels on the ground. The motor installed at "waist" is only used for transforming. This is a kind of two-link under-actuated robot. However, its dynamical behavior is not the same as famous relatives like Pendubot [11] or Acrobot [12]. Thus we need to derive its mathematical model first.

Fig. 6 shows a schematic diagram of our robot. The variables θ_w, θ_b and ϕ denote the wheel rotation angle, the lower link rotation angle and the angle between the upper and lower links (posture angle), respectively in [rad]. This plant can be regarded as a parameter dependent system in terms of the posture angle. The control input is the voltage command for the DC motor denoted by v. The model parameters are summarized in Table 1.

For simplicity, let us assume uniform mass distributions for both links. Then the moment of inertia of the whole body around the rotation axis of the wheels is given by

$$I_b(\phi) = \frac{M_1 l_1^2 + M_2 l_2^2}{12} + \frac{M_1 M_2 \{(l_1^2 + l_2^2)/2 + l_1 l_2 \cos\phi\}}{2(M_1 + M_2)}.$$

The moment of inertia of one wheel is given by $I_w = mr^2/2$. Using the following notations

$$\begin{aligned}
&M_\ell = \left(\frac{1}{2}M_1 + M_2\right) l_1,\\
&\alpha_1 = (M_1 + M_2 + 2m)r^2 + 2(I_w + I_m),\\
&\alpha_2(\theta_b, \phi) = rM_\ell \cos\theta_b + \frac{1}{2} rM_2 l_2 \cos(\phi - \theta_b),\\
&\alpha_3(\phi) = \frac{M_\ell^2 + M_2^2 l_2^2/4 + M_\ell M_2 l_2 \cos\phi}{M_1 + M_2} + I_b(\phi),\\
&\alpha_4 = 2\left(\frac{K_\tau K_b}{R_m} + c_m\right)\\
&\alpha_5(\theta_b, \dot\theta_b, \phi) = -rM_\ell \sin\theta_b \dot\theta_b^2 + \frac{r}{2} M_2 l_2 \sin(\phi - \theta_b)\dot\theta_b^2,\\
&\alpha_6(\theta_b, \phi) = -gM_\ell \sin\theta_b + \frac{1}{2} gM_2 l_2 \sin(\phi - \theta_b),\\
&\alpha_7 = 2K_\tau / R_m\\
&\boldsymbol{\theta} = \begin{bmatrix} \theta_w & \theta_b \end{bmatrix}^T,
\end{aligned}$$

Lagrange's equation of motion of the robot is given by

$$\begin{bmatrix} \alpha_1 & \alpha_2(\theta_b,\phi) \\ \alpha_2(\theta_b,\phi) & \alpha_3(\phi) \end{bmatrix} \ddot{\boldsymbol{\theta}} + \alpha_4 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \dot{\boldsymbol{\theta}} + \begin{bmatrix} \alpha_5(\theta_b, \dot\theta_b, \phi) \\ \alpha_6(\theta_b, \phi) \end{bmatrix} = \alpha_7 \begin{bmatrix} 1 \\ -1 \end{bmatrix} v. \quad (1)$$

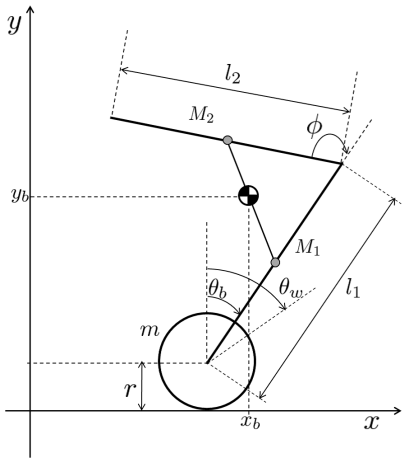


Fig. 6 Schematic diagram of our robot

Table 1 Parameters of our robot

M_1 = 0.565 [kg]	upper link weight
M_2 = 0.170 [kg]	lower link weight
m = 0.030 [kg]	weight of a wheel
l_1 = 0.170 [m]	upper link length
l_2 = 0.255 [m]	lower link length
r = 0.040 [m]	wheel diameter
g = 9.807 [m/s^2]	gravity acceleration
R_m = 6.690 [Ω]	internal resistance
K_b = 0.468 [V· s/rad]	DC motor speed constant
K_τ = 0.317[Nm/A]	DC motor torque constant
$I_m = 1 \times 10^{-5}$[kg · m^2]	rotor moment of inertia
c_m =0.0022[Nm· s/rad]	viscous friction coefficient

LPV Representation in Descriptor Form

For controller design, we need linearized model of (1) around the equilibrium point determined by the posture angle ϕ. Let $\bar\theta_b$ denote the lower link rotation angle corresponding to the equilibrium state. From the balancing condition that the center of gravity of the link system is located on a vertical line crossing the rotational center of the wheels, explicite expression of $\bar\theta_b$ as a function of ϕ can be derived as

$$\bar\theta_b(\phi) = \arctan \frac{M_2 l_2 \sin\phi}{2M_\ell l_1 + M_2 l_2 \cos\phi}. \quad (2)$$

The change of $\bar\theta_b$ versus ϕ is depicted in Fig. 7. If we restrict

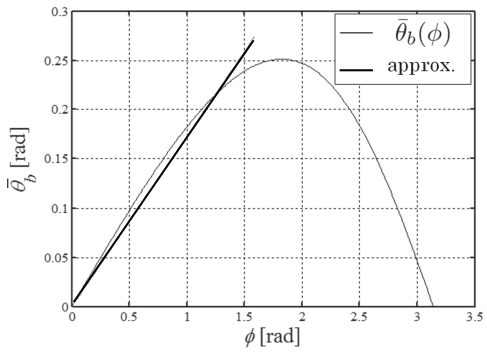


Fig. 7 Relationship between $\bar{\theta}_b$ and ϕ

the transforming range as $\phi \in [0, \pi/2]$, the least square approximation of $\bar{\theta}_b(\phi)$ in this interval is given by $\bar{\theta}_b(\phi) = k\phi$, $k = 0.1726$. Let $\Delta\theta_b$ denote the deviation from the equilibrium, i.e., $\theta_b = \bar{\theta}_b + \Delta\theta_b$. Based on this, we linearize the nonlinear equation of motion (1). We regard $\bar{\theta}_b$ as a small angle since the range corresponding to the interval $\phi \in [0, \pi/2]$ is $\bar{\theta}_b \in [0, 0.2745]$. Taylor expansion of $\cos\bar{\theta}_b$ around $\bar{\theta}_b = 0$ yields $\cos\bar{\theta}_b \simeq 1 - \bar{\theta}_b^2/2$. By choosing $\rho := \bar{\theta}_b^2$ as the scheduling parameter, nonlinear functions in (1) can be expressed as

$$\cos\bar{\theta}_b = 1 - \frac{\rho}{2}, \quad \cos\phi = 1 - \frac{\rho}{2k^2},$$

$$\cos(\phi - \bar{\theta}_b) = 1 - \frac{\rho}{2}\left(\frac{1}{k} - 1\right)^2.$$

By substituting these parameterizations into (1), we obtain the following LPV system representation in the descriptor form:

$$\begin{bmatrix} I & 0 \\ 0 & E_1 + \rho E_2 \end{bmatrix} \begin{bmatrix} \dot{\theta} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & I \\ A_1 + \rho A_2 & A_3 \end{bmatrix} \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ B_2 \end{bmatrix} v, \tag{3}$$

where

$$e_1(1) = rM_\ell + \frac{1}{2}rM_2l_2,$$

$$e_1(2) = \frac{(M_\ell + M_2l_2/2)^2 + M_1M_2(l_1 + l_2)^2/4}{M_1 + M_2} + \frac{1}{12}(M_1l_1^2 + M_2l_2^2),$$

$$e_2(1) = -\frac{1}{2}rM_\ell - \frac{1}{4}rM_2l_2\left(\frac{1}{k} - 1\right)^2,$$

$$e_2(2) = -\frac{M_2l_1l_2}{2k^2},$$

$$E_1 = \begin{bmatrix} \alpha_1 & e_1(1) \\ e_1(1) & e_1(2) \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & e_2(1) \\ e_2(1) & e_2(2) \end{bmatrix},$$

$$a_1 = M_\ell g + \frac{1}{2}M_2gl_2,$$

$$a_2 = -\frac{1}{2}M_\ell g - \frac{1}{4}M_2gl_2\left(\frac{1}{k} - 1\right)^2,$$

$$A_1 = \begin{bmatrix} 0 & 0 \\ 0 & a_1 \end{bmatrix}, \; A_2 = \begin{bmatrix} 0 & 0 \\ 0 & a_2 \end{bmatrix},$$

$$A_3 = \alpha_4 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \; B_2 = \alpha_7 \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

$$\theta = \begin{bmatrix} \theta_w & \Delta\theta_b \end{bmatrix}^T.$$

Conversion into LFT Form

Since the coefficient matrix in the left-hand side of (3) also contains the scheduling parameter ρ, it is difficult to apply existing LMI-based analysis and synthesis techniques directly. To overcome this situation, the redundancy of the representation of the descriptor form can be used [12]. Specifically, ρ-dependent terms are moved to the left-hand side by adding an algebraic constraint. Then we derive an LPV model in LFT form by eliminating the redundant state as shown below.

Multiplication of E_1^{-1} from the left to the second row of (3) yields

$$\ddot{\theta} = \bar{A}_1\theta + \bar{A}_3\dot{\theta} + \rho(\bar{A}_2\theta - \bar{E}_2\ddot{\theta}) + \bar{B}_2v, \tag{4}$$

where

$$\bar{A}_1 = E_1^{-1}A_1, \; \bar{A}_2 = E_1^{-1}A_2, \; \bar{A}_3 = E_1^{-1}A_3,$$

$$\bar{E}_2 = E_1^{-1}E_2, \; \bar{B}_2 = E_1^{-1}B_2.$$

Let $\xi_1 = \theta$, $\xi_2 = \dot{\theta}$. Introduce the third descriptor variable

$$\xi_3 = \bar{A}_2\theta - \bar{E}_2\ddot{\theta},$$

which corresponds to an algebraic constraint. Then (4) is written as

$$\dot{\xi}_2 = \bar{A}_1\xi_1 + \bar{A}_3\xi_2 + \rho\xi_3 + \bar{B}_2v.$$

By denoting

$$\xi = \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \end{bmatrix}^T, \; \tilde{E} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix},$$

$$A_0 = \begin{bmatrix} 0 & I \\ \bar{A}_1 & \bar{A}_3 \end{bmatrix}, \; A_L = \begin{bmatrix} 0 \\ -I \end{bmatrix}, \; A_R = \begin{bmatrix} \bar{E}_2\bar{A}_1 - \bar{A}_2 & \bar{E}_2\bar{A}_3 \end{bmatrix},$$

$$B_0 = \begin{bmatrix} 0 \\ \bar{B}_2 \end{bmatrix}, \quad B_R = \bar{E}_2\bar{B}_2, \quad D = -\bar{E}_2,$$

$$\tilde{A}(\rho) = \left[\begin{array}{c|c} A_0 & \rho A_L \\ \hline A_R & I - \rho D \end{array}\right], \; \tilde{B} = \left[\begin{array}{c} B_0 \\ \hline B_R \end{array}\right],$$

one can derive a redundant descriptor system equivalent to (3) as

$$\tilde{E}\dot{\xi} = \tilde{A}(\rho)\xi + \tilde{B}v.$$

By eliminating ξ_3, an LPV model in LFT form is given as

$$\begin{aligned} \dot{\theta}_a(t) &= A(\rho)\theta_a(t) + B(\rho)v(t), \\ \theta_a &= \begin{bmatrix} \xi_1 & \xi_2 \end{bmatrix}^T, \end{aligned} \tag{5}$$

where

$$\begin{bmatrix} A(\rho) & B(\rho) \end{bmatrix} = \begin{bmatrix} A_0 & B_0 \end{bmatrix} + A_L\rho(I - \rho D)^{-1}\begin{bmatrix} A_R & B_R \end{bmatrix}.$$

3 GAIN SCHEDULING CONTROLLER DESIGN

Modern gain scheduling is a scheme to use variable controller structures when the value of the time-varying parameter of an LPV system can utilized online [13].

Dilated LMI Approach

Here we consider the problem of designing a gain scheduling state feedback law

$$v(t) = K(\rho)\theta_a(t), \tag{6}$$

against the plant (5). We use dilated LMI approach [14], [15] to decouple the parameter-dependent variables to reduce the computational burden. In [12], the authors are designing a gain scheduling controller for Acrobot by combining an H_∞ specification and a constraint on the pole assignment region. We consider here an LQ-type specification with pole assignment since it is better for tuning based on the trial designs for frozen systems.

Lemma 1 (Dilation Lemma [14], [15]) *Let the matrices $\mathcal{A}_{11} \in \mathrm{R}^{n\times m}$, $\mathcal{A}_{12} \in \mathrm{R}^{n\times l}$, $\mathcal{A}_{21} \in \mathrm{R}^{l\times m}$, $\mathcal{A}_{22} \in \mathrm{R}^{l\times l}$ and a symmetric matrix $\mathcal{P} \in \mathrm{R}^{n\times n}$ are given. Suppose that $\mathcal{A}_{22}$ is nonsingular. Then, the following two conditions are equivalent.*

(i) *There exists $\mathcal{X} \in \mathrm{R}^{m\times n}$ satisfying the* LMI

$$\mathcal{P} + \mathrm{He}\{(\mathcal{A}_{11} + \mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21})\mathcal{X}\} < 0.$$

(ii) *There exist $\mathcal{X} \in \mathrm{R}^{m\times n}$, $\mathcal{V} \in \mathrm{R}^{l\times n}$ and $\mathcal{W} \in \mathrm{R}^{l\times l}$ satisfying the* LMI

$$\begin{bmatrix} \mathcal{P} + \mathrm{He}[\mathcal{A}_{11}\mathcal{X}] & * \\ -\mathcal{A}_{21}\mathcal{X} & 0 \end{bmatrix} + \mathrm{He}\left[\begin{bmatrix} \mathcal{A}_{12} \\ \mathcal{A}_{22} \end{bmatrix} \begin{bmatrix} \mathcal{V} & \mathcal{W} \end{bmatrix}\right] < 0. \tag{7}$$

Potential merits of using a dilated LMI condition like (7) are

- No cross term between $\mathcal{X}$ and $\mathcal{A}_{12}$ or $\mathcal{A}_{22}$ appears.
- It does not contain the inverse of $\mathcal{A}_{22}$.

LQ-type Design via Dilated LMI

Let us consider the performance index

$$J = \int_0^\infty [\theta_a(t)^T Q \theta_a(t) + v(t)^T R v(t)]dt,$$

with $Q = H^T H \geq 0$ and $R > 0$ for the plant (5) and the initial condition $\theta_a(0) = \theta_a^0$. When there exist $\gamma(\rho) > 0$, $X(\rho)$ and $F(\rho)$ satisfying the following LMI

$$\begin{bmatrix} \mathrm{He}[A(\rho)X(\rho) + B(\rho)F(\rho)] & * & * \\ HX(\rho) & -I & * \\ -F(\rho) & 0 & -R^{-1} \end{bmatrix} < 0, \tag{8}$$

$$\begin{bmatrix} \gamma(\rho)I & * \\ I & X(\rho) \end{bmatrix} > 0, \tag{9}$$

then the state feedback (6) with the gain

$$K(\rho) = F(\rho)X^{-1}(\rho), \tag{10}$$

stabilizes the closed-loop system and the performance criterion

$$J < \gamma(\rho)\|\theta_a^0\|^2$$

is satisfied [16]. With

$$\left[\begin{array}{c|c} \mathcal{A}_{11}^O & \mathcal{A}_{12}^O(\rho) \\ \hline \mathcal{A}_{21}^O & \mathcal{A}_{22}^O(\rho) \end{array}\right] = \left[\begin{array}{ccc|c} A_0 & 0 & B_0 & A_L\rho \\ H & 0 & 0 & 0 \\ 0 & 0 & -I & 0 \\ \hline A_R & 0 & B_R & I - \rho D \end{array}\right],$$

$$\mathcal{X}^O(\rho) = \begin{bmatrix} X(\rho) & 0 & 0 \\ 0 & 0 & 0 \\ F(\rho) & 0 & 0 \end{bmatrix},$$

$$\mathcal{P}^O = \begin{bmatrix} 0 & * & * \\ 0 & -I & * \\ 0 & 0 & -R^{-1} \end{bmatrix},$$

the LMI (8) is written as

$$\mathcal{P}^O + \mathrm{He}\{(\mathcal{A}_{11}^O + \mathcal{A}_{12}^O\mathcal{A}_{22}^{O-1}\mathcal{A}_{21}^O)\mathcal{X}^O\} < 0. \tag{11}$$

From the dilation lemma, this is equivalent to

$$\begin{bmatrix} \mathcal{P}^O + \mathrm{He}[\mathcal{A}_{11}^O\mathcal{X}^O] & * \\ -\mathcal{A}_{21}^O\mathcal{X}^O & 0 \end{bmatrix} + \mathrm{He}\left[\begin{bmatrix} \mathcal{A}_{12}^O \\ \mathcal{A}_{22}^O \end{bmatrix} \begin{bmatrix} \mathcal{V} & \mathcal{W} \end{bmatrix}\right] < 0. \tag{12}$$

Thus our design problem is now reduced to a feasibility problem to find $\mathcal{X}^O$, $\mathcal{F}^O$, $\mathcal{V}$, $\mathcal{W}$ satisfying (12) and (9) simultaneously. We restrict the variables $X(\rho)$ and $F(\rho)$ to be affine in ρ whereas $\mathcal{V}$ and $\mathcal{W}$ are supposed to be constant. Then the whole condition (12) becomes affine in ρ. Consequently, (12) can be expressed by a convex combination of the conditions corresponding to the maximum and the minimum values of ρ. If one can find the solutions at two endpoints, $X(\rho)$ and $F(\rho)$ can also be obtained from a convex combination of these endpoint solutions.

Pole Assignment Region Constraint via Dilated LMI

The procedure to assign the closed-loop poles of (5) with the state feedback (6) inside a prescribed circle is shown in [12]. Given a circle centered at $(-q, 0)$ with radius r, the desired assignment is achieved if there exist $X(\rho)$ and $F(\rho)$ satisfying

$$\begin{bmatrix} (q^2 - r^2)X(\rho) & 0 \\ 0 & -X(\rho) \end{bmatrix} + \mathrm{He}\left[\begin{bmatrix} -qA_F(\rho) & 0 \\ A_F(\rho) & 0 \end{bmatrix}\right] < 0, \tag{13}$$

$$A_F(\rho) = A(\rho)X(\rho) + B(\rho)F(\rho).$$

Under the notations

$$\mathcal{A}_{11}^D = \begin{bmatrix} -qA_0 & -qB_0 & 0 & 0 \\ A_0 & B_0 & 0 & 0 \end{bmatrix},$$

$$\mathcal{A}_{12}^D(\rho) = \begin{bmatrix} A_L\rho & 0 \\ 0 & A_L\rho \end{bmatrix},$$

$$\mathcal{A}_{21}^D = \begin{bmatrix} -qA_L & -qB_R & 0 & 0 \\ A_L & B_R & 0 & 0 \end{bmatrix},$$

$$\mathcal{A}_{22}^D(\rho) = \begin{bmatrix} I - \rho D & 0 \\ 0 & I - \rho D \end{bmatrix},$$

$$\mathcal{X}^D(\rho) = \begin{bmatrix} X(\rho) & 0 \\ F(\rho) & 0 \\ 0 & X(\rho) \\ 0 & F(\rho) \end{bmatrix},$$

$$\mathcal{P}^D(\rho) = \begin{bmatrix} (q^2 - r^2)X(\rho) & 0 \\ 0 & -X(\rho) \end{bmatrix},$$

the LMI condition (13) is rearranged into

$$\begin{bmatrix} \mathcal{P}^D + \mathrm{He}[\mathcal{A}_{11}^D \mathcal{X}^D] & * \\ -\mathcal{A}_{21}^D \mathcal{X}^D & 0 \end{bmatrix} + \mathrm{He}\left[\begin{bmatrix} \mathcal{A}_{12}^D \\ \mathcal{A}_{22}^D \end{bmatrix} \begin{bmatrix} \mathcal{V} & \mathcal{W} \end{bmatrix} \right] < 0. \tag{14}$$

Similarly to the case of (12), one can reduce this problem into two endpoint conditions by an adequate choice of the order of the variables in ρ.

4 SIMULATION AND EXPERIMENT

Simulation

Via the synthesis procedure given in the previous section, we design a gain scheduling controller for our robot. The minimum and the maximum values of the scheduling parameter ρ corresponding to the posture angle range $\phi \in [0, \pi/2]$ are $\rho_{\min} = 0$, $\rho_{\max} = 0.0735$. We use the following design parameters:

$$Q = \mathrm{diag}[1 \times 10^3, 1 \times 10^3, 1, 1], \quad R = 1 \times 10^3,$$
$$\gamma(\rho) = (1-\eta)\gamma_1 + \eta\gamma_2, \quad \eta = \frac{\rho - \rho_{\min}}{\rho_{\max} - \rho_{\min}},$$
$$\gamma_1 = 4.03 \times 10^5, \quad \gamma_2 = 3.22 \times 10^5,$$
$$q = 0, \quad r = 80.$$

We denote the endpoint solutions by

$$X_1 := X(\rho_{\min}), \quad X_2 := X(\rho_{\max}),$$
$$F_1 := F(\rho_{\min}), \quad F_2 := F(\rho_{\max}).$$

Under the setting described above, we solve a feasibility problem consisting of 6 LMIs, i.e., (9), (12) and (14) for $\rho = \rho_{\min}$ and $\rho_{\max}$, in terms of the variables X_1, X_2, F_1, F_2, $\mathcal{V}$ and $\mathcal{W}$. From these solutions, we determine the feedback gain $K(\rho)$ by (10) where

$$X(\rho) = (1-\eta)X_1 + \eta X_2, \quad F(\rho) = (1-\eta)F_1 + \eta F_2.$$

The simulation result of the time response of the obtained closed-loop system under transformation is shown below. The initial state is given by $\theta_a(0) = [0, 0.1, 0, 0]^T$. For the first 5 seconds, the posture angle ϕ is set to be 0. Then it is increased $\pi/30$ [rad] per second (Fig.8). The transition of the angle of the lower link θ_b (the body angle) is plotted in Fig. 9. The balancing stability is maintained under the time-varying posture angle resulting in a shape transformation.

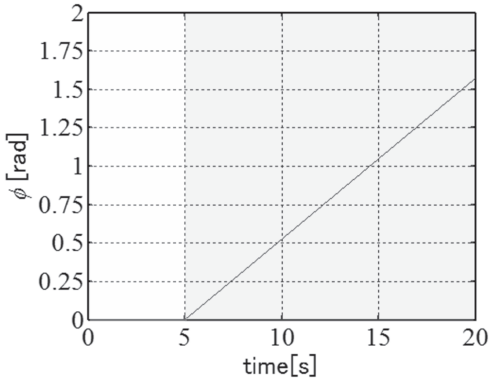


Fig. 8 Angle of posture ϕ

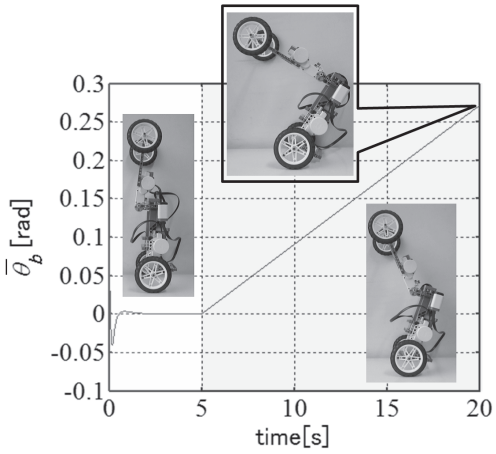


Fig. 9 Angle of body $\bar{\theta}_b$

Experiment

In the experiment, we experience large spillover vibration due to the backlash of the "waist" gear. So we modified the structure to reduce the amount of vibration (Fig. 10).

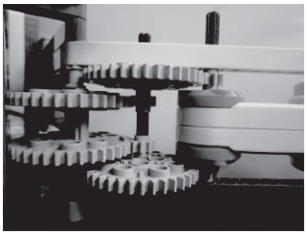


Fig. 10 Improved Gear Structure

The second idea for the implementation is to reduce the complexity of the online computation. If we implement the control law (10) "as is", one must write the source code for matrix inversion. Due to the limited computational resource of Mindstorms, it is rather difficult to run such codes on the embedded CPU. By using a matrix inverse formula, one can rewrite $K(\rho)$ as

$$K(\rho) = \frac{F(\rho)\ \mathrm{adj}\ (X(\rho))}{\det(X(\rho))}.$$

Since one can compute the numerator and the denominator offline as a function of ρ by using some symbolic math softwares, the computation of the feedback gain (10) can be realized by simple arithmetic operation given ρ.

Fig. 11 is the snapshots of our experiment. The posture angle is changed in time from $\pi/2$ to 0. The left column shows the result when the stabilizing controller for $\phi = \pi/2$ is used. It loses stability as the posture angle apart from the designed value. The case with the gain scheduling controller is shown in the right column. In contrast to the fixed controller case, the stability is maintained during (and after) the transformation.

5 CONCLUSIONS

In this paper, the gain scheduling control of a balancing transformer robot made from LEGO Mindstorms is considered. After the modeling and conversion into an LPV system representation, we applied the dilation approach to solve a feasibility problem in terms of the parameter dependent LMIs. The obtained design procedure is verified via numerical simulations and experiments.

REFERENCES

[1] M. Sanpei: Miscellaneous Thoughts on Control Education -Theory, Practice, Hands-On-, Measurement and Control, Vol. 46, No. 9, pp. 681/682 (2007) (in Japanese)

[2] http://mindstorms.lego.com

[3] R. Watanabe: Control Experiment using LEGO Mindstorms, Proc. of the 5th Annual Meeting SICE Control Division, pp. 753/758 (2005) (in Japanese)

[4] http://www.legoeducation.jp/mindstorms/teaching/index.html

[5] http://lejos-osek.sourceforge.net/jp/index.htm

[6] http://www.hitechnic.com/

[7] B. S. Heck, N. S. Clements, and A. A. Ferri: A LEGO Experiment for Embedded Control System Design, IEEE Control System Magazine, Vol. 24, Issue 10, pp. 61/64 (2004)

[8] ET Robot Contest, Japan Embedded Systems Technology Association, http://www.etrobo.jp/

[9] P J. Gawthrop, E. McGookin: A LEGO-Based Control Experiment, IEEE Control Systems Magazine, Vol. 24, Issue 10, pp. 43/56 (2004)

[10] Model Based Development of NXTway-GS, The MathWorks, Inc.

[11] H. Kajiwara, P. Apkarian, P. Gahinet: LPV Techniques for Control of an Inverted Pendulum, IEEE Control Systems Magazine, Vol. 19, Issue 2, pp. 44/45 (1999)

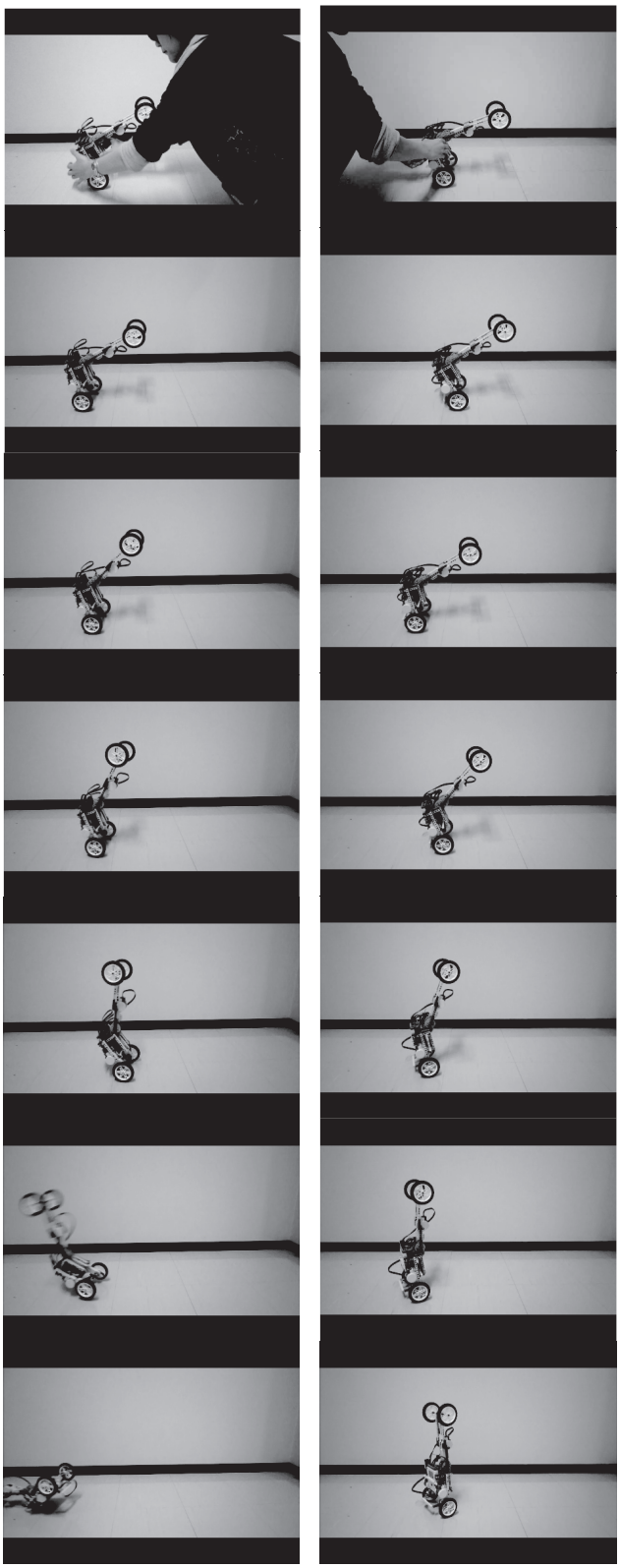
[12] M. Kawata: Gain Scheduling Control System Synthesis of an Acrobot Based on Dilated LMIs, Proc. of the 36th SICE Control Theory Symposium, pp. 405/410 (2007) (in Japanese)

[13] T. Iwasaki: LMI and Control, Shokodo (1997) (in Japanese)

[14] Y. Ebihara and T. Hagiwara: Control System Analysis and Synthesis Using Dilated LMIs, Systems, Control and Information Vol. 48, No.9, pp. 355/360 (2004)

[15] Y. Ebihara, T. Hagiwara: A Dilated LMI Approach to Continuous-Time Gain-Scheduled Controller Synthesis with Parameter-Dependent Lyapunov Variables, Trans. SICE, Vol. 39, No. 8, pp. 734/740 (2003)

[16] A. Fujimori: Robust Control, Coronasha (2001) (in Japanese)



(a) Fixed Controller (b) Gain Scheduling

Fig. 11 Snapshots of Experiment

The sensitivities of the parameters in the WetSpa Extension model for the flood forecasting outputs (with an application to Ve catchment)

Chi PHAM
Department of Civil and Environmental Engineering, Saitama University
255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan
and
Tom DOLDERSUM
University of Twente
Drienerlolaan 5, 7522 NB Enschede, The Netherlands
and
Chiaki T.OGUCHI
Geosphere research Institute, Saitama University
255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan

ABSTRACT

This study focuses on the sensitivity estimation of parameters in WetSpa Extension applying to Ve catchment in Quang Ngai Province, Vietnam. The results show that the groundwater recession coefficient K_g has the strongest sensitivity on the peak runoff and total discharge volume, and strong interaction with other parameters in the model. Surface runoff exponent corresponding to minimum rainfall intensity K_{run} is the parameter noticeably affecting on the time to the peak discharge.

Keywords: Sensitivity analysis, Morris method, WetSpa extension model, flood forecasting, Vietnam

INTRODUCTION

Due to the data shortness, and insufficient perception of the physical processes or the ability of technology to meet the measurement of hydraulic factors, at the moment in Vietnam as well as in the world, scientists have to use many hydrological model to calculate the hydraulic characteristics and to simulate flow distribution in the basin. These models have advantages when the accuracy of the models is not high because of lack of high performance computers. This, the centralized parameter models were often preferred for their simplicity using little number of parameters. However, development of information technology can have the models get higher degree of accuracy using a massive set of parameters.

The reliability of each hydrological model depends on the design of its structure and parameter set. There are many parameters estimated from the topography of the basin, physical properties of soil type, aquifers zones, and land use condition. It is difficult to determine these parameters because the values cannot be measured directly. Therefore, they are often assumed a certain initial values, and then, adjusted to optimal parameters for higher efficiency of models.

So far, Hydrologic Engineering Center (HEC) - U.S. Army Corp of Engineers, MIKE - Danish Hydraulic Institute has been developed and widely used. However, even for these models, validation and simulating processes, which makes it difficult to find a suitable set of parameters for each basin.

There are two methods to determine parameters; try-and-error and optimization. Try-and-error method is more widely used because of its simplicity, although it takes time and subjective to exploit experience. Thus, it is suitable for less parameter models. Optimization method is objective and convenient for exploring the distributed parameter models. In order to reduce calculating steps in optimization method, it is needed to limit the number of adjusted parameters. This process is so-called sensitivity analysis for searching the important parameters in calibration process.

Sensitivity analysis can assess the effects of inputs on outputs of the model by investigating several parameters. What the most important for preliminary assessment of models are to understand the sense of each parameter used in the models. The parameters that are not explicit should not be adjusted since the adjustment may assign inconsistent values with the physical features. Definition of sensitive parameters leads to better estimation of their values and to reduce the operating time for modeling. Some sensitivity analyzing methods have been applied to refine the model parameters before calibration.

Werner et al [23] used Generalized Likelihood Uncertainty Estimation to assess the uncertainty value of the land using distribution in the interaction 1D, 2D hydrodynamic model in Meuse river basin. Bahremand and De Smedt [2] validated and sensitivity analyzed parameters using a model-independent parameter estimator PEST with WetSpa model for the Torysa basin- the quite large area in Slovakia and has achieved advantage results. Ryan Fedak [18] has studied the influence of grid cell size in the two models HEC-1 and TopModel. In addition, we must mention the research of Iman and Helton [10], Campolongo and Saltelli [5], Nguyen and De Kov [16],...

In Vietnam, the sensitivity analysis has not been adequately attended. Apart from several projects, there is not so many researches concerning on sensitivity analysis. This study should be conducted due to the usefulness for not only model development and adjustment but also to reduce uncertainty in the simulation process.

From the above problem, the objective of this research was to assess the sensitivity of the parameters in the WetSpa model, a relatively new model, which has been applied in Vietnam recently for data collecting, calibration, validation and advanced using in practice.

Selecting the sensitivity analysis methods is usually based on the complexity of the model and analyze target. Morgan et al [15] gave four selection criteria as follows: 1) Uncertainty in model form, 2) The essence of the model, 3) Analyze requirements, and 4) The base conditions. Based on these criterions, Morris which is a global sensitivity analysis method has been shown to be quite effective in previous studies; therefore, the method of Morris was used in this study.

Spatial range and scientific scope of the project was flood simulation for Ve Catchment, An Chi stations, Quang Ngai province, Vietnam.

MORRIS METHOD

The purpose of this method was to improve the economy of a sensitivity analysis. "The economy of a design will be defined to be the number of elementary effects it produces divided by the number of experimental runs." Morris [14].
This method has an economy of:

$$\frac{k}{k+1} \tag{1}$$

Where, *k* is the number of parameters taken to account.
This economical design is based on the construction of a *B** matrix with rows that represent input vectors *x* and columns represent parameters. Each run, only one parameter is in progress, the number of steps is a linear function of the number of parameters. Then the corresponding experiment provides *k* elementary effects from *k+1* runs.

Variation range of parameters

The first step of the Morris method is to set the ranges for the different parameters taken into account. Afterwards the values for *k* and *p* have to be set, where *k* is the number of parameters and *p* is the number of parameter sets (*p* has to be even). In the below *B** matrix, the values of *p* and *k* are 4 and 3, respectively, and then we calculate the value of delta:

$$\Delta = \frac{p}{2(p-1)} = \frac{2}{3} \tag{2}$$

Number of steps:

$$m = k+1 = 4 \tag{3}$$

For each parameter, one base value is randomly chosen from this set:

$$Set = \left\{0, \frac{1}{p-1}, \frac{2}{p-1}, \ldots 1-\Delta\right\} = \left\{0, \frac{1}{3}\right\} \tag{4}$$

For the example *x** can be chosen like this:

$$x^* = \left(0, \frac{1}{3}, \frac{1}{3}\right) \tag{5}$$

Create *B** matrix

To build a *B** matrix, the first step is the selection of a *m*k* matrix *B* with elements that are *0s* and *1s*, such that in every column there are two rows of *B* that differ in only one element. The easiest way to create this matrix is making a triangular of *1s* starting at the second row. Furthermore a $J_{m,k}$ matrix of *1s* and a *k*-dimensional *D** matrix with elements either *+1* or *-1* with equal probability has to be build. At least a *k*k* dimensional *P** matrix which is a random permutation matrix and contains in each column one element equal to *1* and all others to *0*, such that no two columns have *1s* in the same position.

$$B_{m,k} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad J_{m,k} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$D^*_{k,k} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad P^*_{k,k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{6}$$

The *B** matrix would be given by:

$$B^* = \left\{J_{m,1}x^* + (\Delta/2)\left[(2B - J_{m,k})D^* + J_{m,k}\right]\right\}P^*$$

$$= \begin{bmatrix} 2/3 & 1/3 & 1 \\ 0 & 1/3 & 1 \\ 0 & 1/3 & 1/3 \\ 0 & 1 & 1/3 \end{bmatrix} \tag{7}$$

Then the *B** was multiplied by the ranges of the parameters. Since every column in the *B** matrix is standing for a parameter and every row gives a random number for this parameter, every column has to be multiplied by the interval of the corresponding parameter and add the minimum value of the interval. This can be expressed by the following formula:

$$parinput = B^*_{m,1} \times range + min \tag{8}$$

Where, *parinput* is the parameter set with randomly generated values for the parameters.

Elementary effects

Suppose that the output function is $y = f(x_1, x_2, \ldots, x_k)$, the elementary effect of each input has to be defined.
Each row in *B** differs only in one column from the row below, moreover this difference is always equal to $-\Delta$ or $+\Delta$. Therefore, during the simulation of the *parinput*, *n* runs will provide *n-1* elementary effects. After running the model with the *parinput* files, the values of the elementary effects could be calculated. Therefore is used the following statement with corresponding formula; where *j* is the number of row (to calculate all elementary effects of one input file it has the repeat *k* times) and Δ is the same as during the calculation of the *B** matrix:
If $B^*_{j+1} - B^*_j > 0$ then:

$$d_j(parinput_j) = \frac{y(parinput_{j+1}) - y(parinput_j)}{\Delta} \tag{9}$$

Else

$$d_j(parinput_j) = \frac{y(parinput_j) - y(parinput_{j+1})}{\Delta} \tag{10}$$

After simulation of the first *parinput* this action will be repeated *r* times. Therefore, in total there are needed *r*m r*uns of the model.

Sensitivity analysis

From these values, the means and the standard deviations of the different parameters and input variables could be calculated.
Standard deviation

$$\sigma = \sqrt{\frac{1}{1-n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \; with \; \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{11}$$

and mean

$$\mu = \frac{1}{r} \times \sum_{j=1}^{r} d_j \tag{12}$$

These values could be used to analyze the sensitivities of each parameter. The high mean value shows the important global impact when large standard deviation corresponds to the interaction with other factors or nonlinear impact.

WETSPA EXTENSION MODEL

"The WetSpa (extension) model is a GIS based-distributed hydrological model for flood prediction and water balance simulation on catchment scale" Bahremand & De Smedt [2]. The WetSpa extension model used in this research was developed from the WetSpa and WetSpass extension models. WetSpa is an acronym for "Water and Energy Transfer between Soil, Plants and Atmosphere". It is a physically based model and the hydrological processes considered in the WetSpa model are precipitation, depression storage, snowmelt, surface runoff, infiltration, evapotranspiration, percolation, interflow, and groundwater flow. It can simulate runoff and hydrological

characteristics at certain point in the river network and their distribution in grid cell scale.

Twelve global parameters are included in the model: time step d_t (hour), scaling factor for interflow computation K_i, groundwater recession coefficient K_g, initial soil moisture K_{ss}, correction factor for potential evapotranspiration K_{ep}, initial groundwater storage in water depth G_0 (mm), maximum groundwater storage in water depth G_{max} (mm), base temperature for snow melting T_0 (^{0}C), base temperature for snow melting K_{snow} (mm/^{0}C/day), rainfall degree-day coefficient for estimating snowmelt K_{rain} (mm/mm/^{0}C/day), surface runoff exponent when the rainfall intensity is very small K_{run} and the threshold rainfall intensity P_{max} (mm/hour), Liu & De Smedt [11].

These global parameters have physical interpretations that are very important in controlling runoff production and hydrographs at the basin outlet. Nevertheless, it is very difficult to assign them properly over a grid scale. Therefore, it is preferable to calibrate these parameters against observed runoff data in addition to the adjustment of the spatial distributed model parameters.

Among those inputs, the time step d_t is constant, and for the Vietnamese catchments, because ice snow rarely occurs, the three parameters T_0, K_{snow}, K_{rain} can be eliminated from the sensitivity analyzing progress. Furthermore, due to the lack of the evapotranspiration data, K_{ep} would be taken into account in another form that is a part of precipitation (%):

$$x_{\text{model}} = x \times K_r \quad (13)$$

Where, x_{model} is the input precipitation, x is the measured precipitation, K_r is the evaporation recession coefficient (%).

There are some parameters, which values are set in the WetSpa model; therefore, they were also taken into account in this research. The parameter b set to 1.35 in the original model controls the shape of the variation curve for the interception storage. The other parameter m, is put in the groundwater flow equation and get the value of *1* for linear reservoir and *2* for nonlinear reservoir. This parameter can vary between *1* and *2*. Liu & De Smedt [11].

Refering from the previous study about sensitivity analysis of parameters of Liu [12], Bahremand and De Smedt [2] with the condition of Vietnamese area, this research put only seven global parameters K_i, K_g, K_{ss}, G_0, G_{max}, K_{run}, P_{max}, rainfall coefficient K_r and two parameters b and m into sensitivity analysis applied for the Ve catchment.

INPUT DATA

Study area

The study area, Ve river basin is located in the Quang Ngai province in the central region of Vietnam. The total basin has a surface area of 1300 km^2 and the length of 91 km. Within this research, only the upstream part from An Chi was taken into account. This part covers an area of 757.32 km^2.

The basin is rather small and has steep slope, so flood process is very complicated. Located on the biggest rainy center of Quang Ngai Province, heavy rainfall can cause flash flood in the upstream with many serious damages. For Ve basin, the problems need to be solved in flood forecasting are to raise the degree of accuracy and to extent the foresee time of predicting water level in Ve river in order to prevent and reduce damages caused by flood.

Geographic location

Ve river rises from Truong Son - the high mountainous region, with geographic coordinate of 14^{0}32'25" in the North, 108^{0}37'4" in the East. An Chi station is 14^{0}58'15" in the North and 108^{0}47'36" in the East. The study area is totally in Quang Ngai Province, is bounded by Tra Khuc river in the Northern and the Western, by Binh Dinh Province in the Southern and by Dong sea (South-China Sea) in the Eastern.

Topographical characteristics

The topography of the basin can be divided into two types:

The mountainous area is very slope, concentrate water rapidly, and easy to form violent flood with short transform duration.

The plain with quite tableland relief is blocked by sand dunes, which protects flood from abstraction, and causes flooded easily.

Locates in the Eastern side of Truong Son mountain, Ve river basin have a complex topography. It consists of mountainous, midland and plain with many offsets coming from Truong Son Mountain to coastal plain, forms Southwest and Northeast direction valleys. The average height of the basin varies from 100 to 1000m. Land is slope, territory trends lower in the Southwest – Northeast and West – East directions. The midland involves rough low hills with height is from 100 to 500m, quite slope. Plain area is not so flat, and height is about 100m.

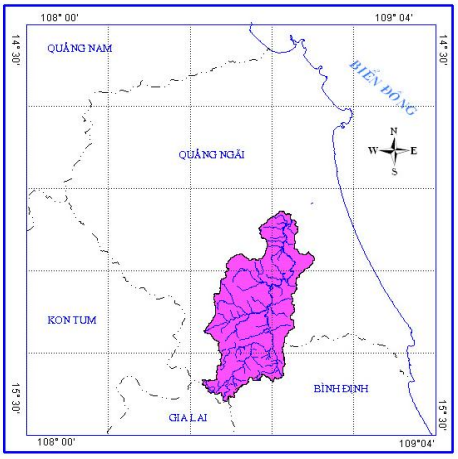


Figure (1). The Ve river basin

Geology

Study area lies along the meridian line, consists of many geological structures with different forming regulations and structural petrologies.

The most common geological characteristic of Ve river basin is rapid change in topographical gradient in profile from the continent to the sea. Therefore, most of the rivers in the region are short and vertical erosion is principal. Deposition and bank erosion occur mainly in coast plain.

Vegetarian cover

Natural forest in the basin, which rarely remains, is mainly medium forest and poor forest which most distributes in the high mountainous. There are a lot of valuable forests and local products. Mountainous and hilly regions have a very little area of forest, most area are shaven hill and industrial soil, or clump. The downstream have cultivated land and inhabitant zone.

Climatic conditions

Ve river basin locates in the Southern of Hai Van pass where has the climate of Mid-Midland climate zone.

In the summer, warm and moist tropical air current of Indian Ocean, equatorial air and the cooler and moister summer Trade winds – tropical air current coming from Pacific Ocean affect the watershed. The air current coming from the Indian Ocean forms rain in the early summer, and becomes hot and dry once it pass the Truong Son mountain.

The winter is not cold, average annual temperature is about 26 – 26.5^{0}C.

There are two main wind seasons in a year: the Northeast monsoon and the Southwest monsoon. Depending on terrain conditions, prevailing wind in each season differs from each area. Winter monsoon is usually in the West, Northwest and Northeast direction. In the summer, prevailing winds are in the West and Southwest direction. Some noticeable weather phenomenon is nimbus, storm and warm dry Western wind.

The number of rainy days is approximately 140 days per year and each year has 1700 sunshine hours.

Average annual air temperature varies between 20 and 22^{0}C in the mountainous (higher than 500m) and from 25 to 26^{0}C in coastal plain.

Average annual absolute air humidity is 23.6 mb. In summer, average monthly absolute air humidity is from 28 to 31 mb in valleys and plain. In winter, average monthly absolute air humidity is from 21 to 28 mb, and the lowest value is about 19-22.5 mb in January.

Average annual evaporation (measured by Picher) varies between 640 and 900 mm.

Precipitation is rather big, especially in the upstream. In the plain, total yearly precipitation is 2000-2200 mm; in the upstream, it exceeds 3000m, even 4000m in Quang Nam mountainous area. Average annual precipitation is spatially highly variable which is from 1600 to 3600 mm and trends rising from the Eastern to the Western.

Annual rainy system has two seasons: rainy season and dry season. The rainy season starts late, usually in September, and ends in December. The amount of rainfall over these four months is about 65 – 85% the total amount of annual precipitation while the dry season lasts in eight months but makes up only 15 – 35% of the total amount of annual precipitation.

Hydrological characteristics

Comparing to different river systems in coastal South-Midland area, the study area (consider from upstream to An Chi station) is quite small, makes up about 64.6% of the total area of Ve river basin. The whole study area lies in Quang Ngai Province. The main stream is 91 km long, sourcing from Nuoc Vo at the height of 1070m to the Dong sea at Long Khe estuary.

Drainage density is quite dense at 0.79m/km^2. Located in the coastal zone, mountainous occupies a very small area. The meandering index of mainstream is 1.3. Having many residual mountains and sand dunes in coastal zone, hydrographic network is interlaced.

Flood season lasts in three months from October to December, occupies 70.6% total annual discharge. Dry season lasts in nine months from January to September and makes up 29.4% total annual discharge. With a plentiful precipitation, in average, there are from six to eight floods occur in a year.

Spatial data

Spatial data for the WetSpa extension includes three digital maps: digital elevation map (DEM), soil type map and land use map. In addition, to compare and calculate basin characteristics, river network and hydrological station network maps also needed. All these maps have the grid cell size of 90x90 m.

Meteorological data

Precipitation data at four stations An Chi, Son Giang, Gia Vuc and Ba To were used to calculate the stream flow in the basin. Among of those, Ba To and An Chi stations are located inside the basin, the two others are outside. The hourly rainfall data calculated from the six-hour data measured in these four stations were used to draw the Thiessen polygons and interpolate data over the entire basin.

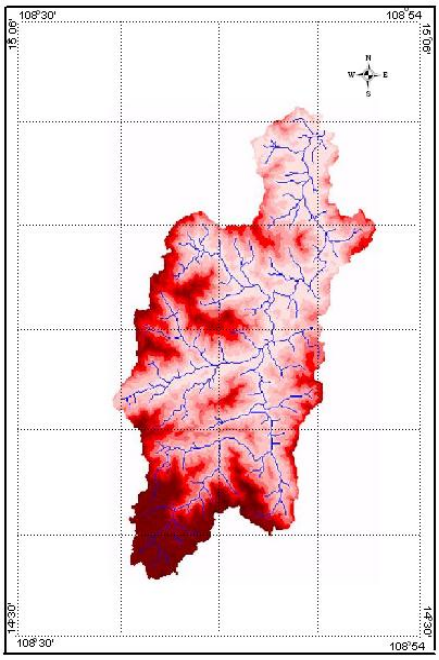


Figure (2). The elevation map

Stream flow data

Flow data series at An Chi station was used to compare to output results from WetSpa model. The hourly data was collected in the storm on November 1999.

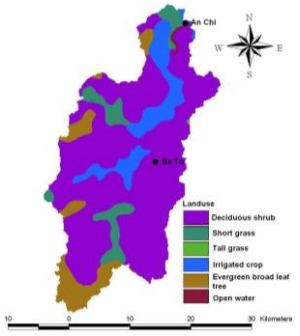


Figure (3). Land use map

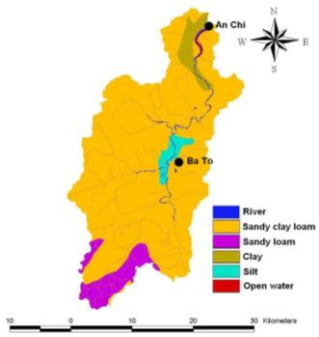


Figure (4). Soil type map

SENSITIVITY ANALYSIS

Calibration process

Initial calibration process to find out variation limits of the parameters in the matrix *B** was done by Tom [20] by the two method Random Sampling and Latin Hypercube Sampling with the below results.

Table (1). Variation range of the parameters

No	1	2	3	4	5	6	7	8	9	10
Parameter	K_r	K_i	K_q	K_{ss}	G_0	G_{max}	K_{run}	P_{max}	b	m
Minimum value	0.9	2	0.002	0	0	50	0	0	0.4	1
Maximum value	1.1	11	0.06	1.5	50	150	10	500	1.6	2

Simulating the output flow

Automatic operation of the model was carried out based on the modified source code in Fortran instead of the original model. Then instead of calculating the output volume for each certain set of parameters one-by-one, the model can execute with all the set (contained in the matrix *B**) in one-time operation. The output will be flow data at downstream corresponding to each set of parameters. In this study, 10,000 calculated parameters gave out 10,000 flooding volume values at An Chi station.

DISSCUSIONS

The sensitivity analysis was done by Morris method for the three output factors peak discharge, total discharge volume and time to the peak discharge.

The Figure (5) presents the graph of the sensitivity of the model inputs and parameters regarding to the highest point in the flow data series. Parameters K_g (3) and K_{run} (7) have high standard deviation, indicating strong interaction with other parameters. K_r (1), K_i (2), P_{max} (8) and b (10) have relatively large standard deviations that demonstrate the ability to interact with each other and with other parameters. K_r (1), K_i (2) and K_g (3) have a high mean value. That represents the influence on peak flow values. P_{max} (8) and m (10) have relatively large mean value, coresponding to significant influence to the output. The remaining factors K_{ss} (4), G_0 (5) and G_{max} (6) are not sensitive to the peak discharge of streamflow.

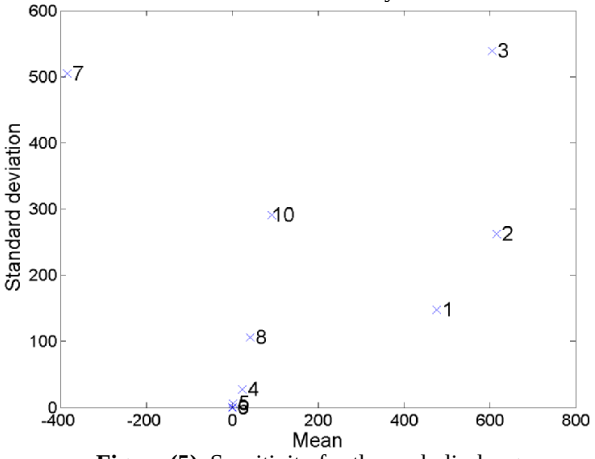


Figure (5). Sensitivity for the peak discharge

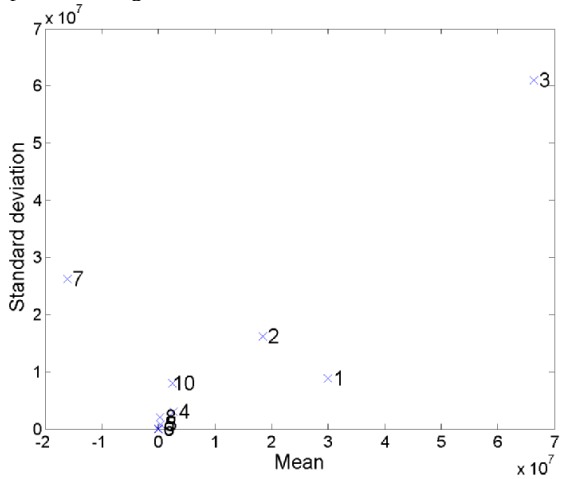


Figure (6). Sensitivity for the total discharge volume

The results of sensitivity analysis for the total value are shown in figure (6). Parameters K_g (3) have a very high standard deviation, indicating a strong interaction with other parameters. Parameter K_r (1), K_i (2), K_{run} (7) and m (10) have relatively large standard deviations demonstrate their interaction ability. Parameter K_r (1), K_i (2) and K_g (3) have a high average value shows the influence on flooding volume. The remaining ones are not sensitive in this case.

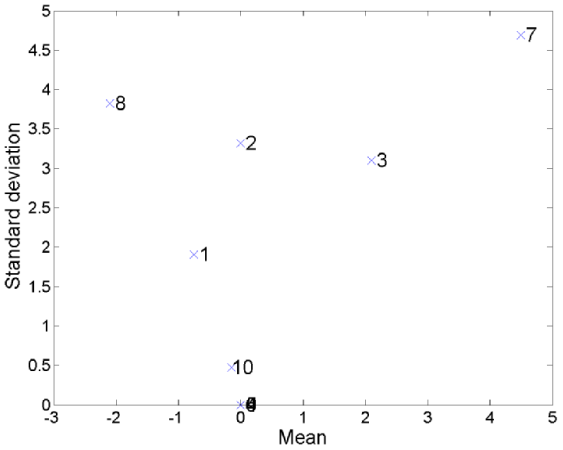


Figure (7). Sensitivity for the time to the peak discharge

It can be seen very clearly from Figure (7) that the parameters all ave very small standard deviation. Therefore, there is almost no interaction between them. Furthermore, only parameters K_g (3) and K_{run} (7) have mean values greater than 1, or ability of changing the delay time an hour. For the hourly flood data, it seems to be like that they are not sensitive for the delay time.

CONCLUSIONS

From the above results, apparently that K_g (3) parameters is the most sensitive one for both the flood peak and the total volume of flooding water amount. Beside of that is the interaction with other parameters in the model. This is the most important factor worth to notice in the calibration period.

The standard deviation of K_{run} is also rather high, reflecting the ability to interact with other parameters. Also this parameter affects the most significant on the delay time.

K_r (1), K_i (2) has strongly influence on the peak as well as the total amount stream flow.

The same test was done for some more storms in the Ve catchment and the obtained results were similar.

From sensitivity analysis by the Morris method, the proposed comclusion is when using WetSpa extension for flood forecasting in Ve river basin, operators need to focus on adjusting values of the groundwater recession coefficient K_g, the surface runoff when the rainfall intensity is very small K_{run}, the evaporation recession coefficient K_r, and the scaling factor for interflow computation K_i.

The Morris method has many advantages in the sensitivity analysis. However, the biggest limitation is that only the sensitivity of each parameter is evaluated, but not the interaction between the parameters. Moreover, the method may neglect the degree of uncertainty of each parameter. In fact, these parameters can be very sensitive but have stable value, or even when the sensitivity is not very large, but uncertainty is quite significant. Therefore, to achieve greater efficiency in calibration process, the further research is needed to assess the sensitivity and uncertainty of the parameters at the same time, or we may use additional methods for sensitivity analysis.

ACKNOWLEDGEMENT

We are grateful for the assistance of Prof. Liu Yongbo at Brussels Free University, as one author of the WetSpa model for providing the latest version of the WetSpa source code, as well as for his instruction in the process of changing the source code model using Fortran programming language. We would like to thank Nguyen Thi Thuy, researcher in institute of Meteorology and Hydrology for the rainfall and flow data as well as the help in the calculation by the model. We would like to thank the two students from the University of Twente, the Netherlands who took part in this study: Daniel Van Puten and especially Tom Doldersum.

REFERENCES

English

[1]. Aronica G., Bates P. D., Horritt M. S., **Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE**, Hydrological processes, Vol.16, pp.2001-2016, 2002.

[2]. Bahremand A., De Smedt F., **Distributed Hydrological Modeling and Sensitvity Analysis in Torysa Watershed, Slovakia**, Water Resources Management, Vol.22, pp.393-408, 2008.

[3]. Beven K., Binley A., **The future of distributed models: model calibration and uncertainty predicition**, Hydrological processes, Vol.6, pp.279-298, 1992.

[4]. Beven Keith, **How far can we go in distributed hydrological modelling?**, Hydrology and Earth System Sciences, Vol.5, pp.1-12, 2001.

[5]. Campolongo F., Saltelli A., **Sensitivity analysis of an environmental model: an application of different analysis methods**, Reliability Engineering & System Safety, Vol.57, pp.49-69, 1997.

[6]. Daniel Van Puten, **Estimating and updating uncertainty with the GLUE methodology**, Bachelor thesis Twente University, Enschede, The Netherlands, 2009.

[7]. FAO, **World reference base for soil resources 2006**, Italia, 2006.

[8]. FAO, **FAO Soil Unit**, Italia, 2006.

[9]. Granger Morgan, Max Herion, Mitchell Small, **Uncertainty**, Cambridge University Press, The United Stated of America, 1990.

[10]. Iman R.L., Helton J.C., **An investigation of uncertainty and sensitivity analysis techniques for computer models**, Risk Analysis Vol.8, pp.71-90, 1988.

[11]. Liu Y.B., De Smedt F., **Documentation and User Manual WetSpa Extension; A GIS based Hydorlogic Model for Flood Prediction and Watershed Management**, Vrije Universiteit Brussel; Department of Hydrology and Hydraulic Engineering, 2004.

[12]. Liu Y.B., Corluy J., **Steps of running WETSPA**, Vrije Universiteit Brussel; Department of Hydrology and Hydraulic Engineering, 2005.

[13]. Morris D.M., **Sensitivity of European Hydrological System snow models. Hydrological aspects of alpine and high mountain areas**, IAHS Publ, Vol.138, pp.122-231, 1982.

[14]. Morris D.M., **Factorial sampling plans for prelimenary computational experiments**, Technometrics, Vol.33, pp.161-174, 1991.

[15]. NSW Department of Commerce Manly Hydraulics Laboratory, **Review and Assesment of Hydrologic/Hydralic Flood Models**, 2006.

[16]. Nguyen, T. G., De Kok J., **Systematic testing of an integrated systems model for coastal zone management using sensitivity and uncertainty analyses**, Environmental Modelling & Software, Vol.22, pp.1572-1587, 2006.

[17]. Nurmohamed, R., Naipal, S., De Smedt, F., **Hydrologic modeling of the Upper Suriname River basin using WetSpa and ArcView GIS**, Journal of spatial Hydrology, Vol.6, pp.1-17, 2006.

[18]. Roberta-Serena Blasone, Jasper A. Vrugt, Henrik Madsen, Dan Rosbjerg, Bruce A. Robinson, George A. Zyvoloski, **Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling**, Water Resources, Vol.31, pp.630-648, 2008.

[19]. Ryan Fedak, **Effect of Spatial Scale on Hydrologic Modeling in a Headwater Catchment**, Master Thesis, 1999.

[20]. Saltelli, A., Chan, K., Scott, E., **Sensitivity Analysis**, Chichester: John Wiley and Sons Ltd, 2000.

[21]. Tom Doldersum, **Global sensitivity analysis of the WetSpa model**, Bachelor thesis, University of Twente, Enschede, The Netherlands, 2009.

[22]. Uhlenbrook, S., Sieber, A., **On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model**, Environmental modelling and software, Vol.20, pp.19-32, 2005.

[23]. V. Vandenberghe, W. Bauwens, P.A. Vanrolleghem, **Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns**, Environmental modelling and software, Vol.22, pp.725-732.

[24]. Werner M.G.F., Hunter N.M, Bates P.D., **Identifiability of distributed floodplain roughness values in flood extent estimation**, Journal of Hydrology, Vol.314, pp.139–157, 2005

[25]. Yu, P., Yang, Y., Chen, S., **Comparison of uncertainty analysis methods for a distributed rainfall-runoff model**, Hydrology, Vol.244, pp.43-59, 2001.

Vietnamese

[26]. Bofu Yu, **Hydrographical and topo-morphological report of Tra Bong, Tra Khuc and Ve river alluvium in Quang Ngai Province**, Vietnam, 2004

[27]. Nguyen Thanh Son, **Simulating the rainfall-runoff process for rational water resources and land use in some upstream watersheds in central region of Vietnam**. PhD thesis, Hanoi University of Science, Vietnam, 2008.

[28]. Nguyen Thi Thuy, **Using the WetSpa extension model for flood simulating of Ca river basin**. Bachelor thesis, Hanoi University of Science, Vietnam, 2008.

[29]. Water Resources Planning Institute, **Usage planning of water resources in Tra Khuc river basin, Quang Ngai Province**, Vietnam, 2003.

AUTHORS INDEX

(Post-Conference Edition)