WMSCI 2014

The 18th World Multi-Conference on Systemics, Cybernetics and Informatics

July 15 - 18, 2014 – Orlando, Florida, USA

PROCEEDINGS

Volume I

Edited by:

Nagib Callaos
Natalja Lace
William Lesso
Belkis Sánchez
Michael Savoie

COPYRIGHT

PROGRAM COMMITTEE
Chairs: Michael Savoie (USA)
C. Dale Zinn (USA)

Adamopoulou, Evgenia	National Technical University of Athens	Greece
Aksoy, Mehmet S.	King Saud University	Saudi Arabia
Al Obaidy, Mohaned	Gulf College	Oman
Aldred, J. R.	US Air Force Academy	USA
Ariwa, Ezendu	University of Bedfordshire	UK
Avery, Paul A.	Southwest Research Institute	USA
Ayuga Téllez, Esperanza	Technical University of Madrid	Spain
Barkstrom, Bruce R.	Unaffiliated	USA
Bönke, Dietmar	Reutlingen University	Germany
Brown, Anson	Univiersity of Houston-Victoria	USA
Chen, Jingchao	University Donghua	China
Chen, Lisa Y.	I-Shou University	Taiwan
Chen, Yil	Huafan University	Taiwan
Cherinka, R.	The MITRE Corporation	USA
Cunha, Idaulo J.	Intellectos	Brazil
Danjour, Miler Franco	Federal Institute of Education, Science and Technology from Rio Grande do Sul	Brazil
Dantas de Assunção, Marcus V.	Federal Institute of Education, Science and Technology from Rio Grande do Sul	Brazil
Dawoud, Dawoud S.	Victoria University	Uganda
Demestichas, Konstantinos	National Technical University of Athens	Greece
Doma, Salah	Alexandria University	Egypt
Dosi, Vasiliki	University of Ioannina	Greece
Erina, Jana	Riga Technical University	Latvia
Erins, Ingars	Riga Technical University	Latvia
Erkollar, Alptekin	University of Applied Sciences Wiener Neustadt	Austria
Figueroa-Sandoval, Benjamín	Colegio de Postgraduados	Mexico
Flórez Fernández, Héctor A.	Francisco Jose of Caldas District University	Colombia
Fox, Kevin	Harris Corporation	USA
Frank, Timothy	US Air Force Academy	USA
Fries, Terrence P.	Indiana University of Pennsylvania	USA
Gaile-Sarkane, Elīna	Riga Technical University	Latvia
Ganapathi, Nanthini	Research Triangle Institute	USA
Gardezi, A. K.	Colegio de Postgraduados	Mexico
Geppert, Luigi	Catholic University of Milan	Italy
González García, Concepción	Technical University of Madrid	Spain

Guedes, Andre F. S.	Intellectos	Brazil
Guedes, Vilmar P.	Intellectos	Brazil
Hahn, Thomas	University of Arkansas at Little Rock	USA
Ham, Chan	Southern Polytechnic State University	USA
Hashimoto, Shigehiro	Kogakuin University	Japan
Hendel, Russell J.	Towson University	USA
Hepdogan, Seyhun	University of Central Florida	USA
Jerabek, Michal	Czech Technical University in Prague	Czech Republic
Jung, Seul	Chungnam National University	South Korea
Khan, Ata	Carleton University	Canada
Kocmanova, Alena	Brno University of Technology	Czech Republic
Koke, Tatjana	University of Latvia	Latvia
Kukielka, Leon	Koszalin University of Technology	Poland
Kung, C. M.	Shih Chien University	Taiwan
Lace, Natalja	Riga Technical University	Latvia
Lahlouhi, Ammar	University of Biskra	Algeria
Loffredo, Donald	University of Houston-Victoria	USA
Lokaj, Zdenek	Czech Technical University in Prague	Czech Republic
López Román, Leobardo	University of Sonora	Mexico
Lu, Shen	Soft Challenge LLC	USA
Márquez-Berber, S. R.	Chapingo Autonomous University	Mexico
Mbobi, Aime Mokhoo	Centennial College	Canada
Meyer, Heiko	Gefasoft AG	Germany
Mostafaeipour, Ali	Yazd University	Iran
Naddeo, Alessandro	University of Salerno	Italy
Nagai, Yasuo	Tokyo University of Information Sciences	Japan
Nemecek, Petr	Tomas Bata University in Zlin	Czech Republic
Nishigaki, Yasuyuki	Ryukoku University	Japan
Nkhoma, Mathews	RMIT University Vietnam	Vietnam
Oganisjana, Karine	University of Latvia	Latvia
Oh, Joon-Yeoul	Texas A&M University	USA
Pfeifer, Michael	Technical University of Dortmund	Germany
Pham, Hiep Cong	RMIT University Vietnam	Vietnam
Pratt, Dee	Durban University of Technology	South Africa
Radich, Joseph	Center for Health & Wellness Royal Palm Beach	USA
Rahmes, Mark	Harris Corporation	USA
Rivza, Baiba	Latvia University of Agriculture	Latvia
Romanowicz, Agnieszka	European Environment Agency's	Denmark
Rossmann, Jürgen	RWTH Aachen University	Germany
Ruan, Tongjun	New Mexico Institute of Mining and Technology	USA
Russo, Dario	Bank of Italy	Italy
Ščeulovs, Deniss	Riga Technical University	Latvia
Sears, Alfred	Center for Health & Wellness Royal Palm Beach	USA
Segall, Richard	Arkansas State University	USA
Shiraishi, Yoshiaki	Nagoya Institute of Technology	Japan
Šimberová, Iveta	Brno University of Technology	Czech Republic

Soares da Silva, Renata Lissa	Federal Institute of Education, Science and Technology from Rio Grande do Sul	Brazil
Sohn, Han Suk	New Mexico State University	USA
Souza, Mônica L.	Intellectos	Brazil
Sriratanaviriyakul, Narumon	RMIT University Vietnam	Vietnam
Stasytyte, Viktorija	Vilnius Gediminas Technical University	Lithuania
Stotyr, Martin	Czech Technical University in Prague	Czech Republic
Straub, Jeremy	Jacksonville State University	USA
Sun, Baolin	Wuhan University	China
Sun, Degui	University of Ottawa	Canada
Tartari, Simone	Intellectos	Brazil
Tavakkoli, Alireza	University of Houston-Victoria	USA
Thompson, Laura	Keiser University	USA
Trzaska, Mariusz	Polish Japanese Institute of Information Technology	Poland
Tsaur, Woei-Jiunn	Da-Yeh University	Taiwan
Tsiligaridis, John	Heritage University	USA
Tsubaki, Michiko	University of Electro-Communications	Japan
Vasinek, Vladimir	Technical University of Ostrava	Czech Republic
Vrána, Stanislav	Czech Technical University in Prague	Czech Republic
Wei, Xinzhou	New York City College of Technology	USA
Whebber Souza de O., Patrícia	Federal University of Rio Grande do Norte	Brazil
Wilder, Kathy	Harris Corporation	USA
Yates, Harlan	Harris Corporation	USA
Zaretsky, Esther	Givat Washington Academic College of Education	Israel
Zelinka, Tomas	Czech Technical University in Prague	Czech Republic
Zeverte-Rivza, Sandija	Latvia University of Agriculture	Latvia
Zhang, Qingyu	Arkansas State University	USA
Zhang, Xiaozheng Jane	California Polytechnic State University	USA

ADDITIONAL REVIEWERS

Aavani, Amir	Simon Fraser University	USA
Abdel Hafez, Hoda	Suez Canal University	Egypt
Aboalsamh, Hatim	King Saud University	Saudi Arabia
Abramchuk, George (Heorhi)	Natural Science Engineering Research Council	Canada
Acharya, Sushil	Robert Morris University	USA
Adamkó, Attila	University of Debrecen	Hungary
Adel, Miah	University of Arkansas at Pine Bluff	USA
Alba-Flores, Rocio	Georgia Southern University	USA
Alexik, Mikulas	University of Zilina	Slovakia
Aliguliyev, Ramiz	Azerbaijan National Academy of Sciences	Azerbaijan
Allasia, Walter	Eurix Research	Italy
Almulla, Mohammed	Kuwait University	South Korea
Al-Yasiri, Adil	University of Salford	UK
Anita, Mary	Anna University	India
Aoki, Toru	Shizuoka University	Japan
Arous, Najet	National Engineering School of Tunis	Tunisia
Aruga, Masahiro	Tokai University	Japan
Arya, Ali	Carleton University	Canada
Augusciuk, Elizabeth	Warsaw University of Technology	Poland
Aveledo, Marianella	Simon Bolivar University	Venezuela
Azariadis, Philip	University of the Aegean	Greece
Baek, Jonghun	Kyungpook National University	South Korea
Bahrami, Mehdi	Cloud Lab,University of California, Merced	USA
Bandera, Antonio	University of Malaga	Spain
Bangert, Patrick	Algorithmica Technologies	USA
Bangyal, Waqas	Iqra University Islamabad	Pakistan
Bani-Salameh, Hani	The Hashemite University	Jordan
Baniulis, Kazys	Kaunas University of Technology	Lithuania
Bargiotas, Dimitrios	Technological Education Institution of Chalkida	Greece
Batos, Vedran	University of Dubrovnik	Croatia

Baudin, Veronique	Laboratory for Analysis and Architecture of Systems	France
Beg, Azam	United Arab Emirates University	United Arab Emirates
Behzadan, Amir	University of Central Florida	USA
Belardo, Salvatore	University at Albany	USA
Belguith, Lamia Hadrich	University of Sfax	Tunisia
Benamar, Abdelkrim	University of Tlemcen	Algeria
Benazzouk, Amar	University of Picardie Jules Verne	France
Bidarra, José	Aberta University	Portugal
Bilbao, Javier	University of the Basque Country	Spain
Bilbao, Josu	Ikerlan Technological Research Center	Spain
Boger, Zvi	Ben-Gurion University	Israel
Boonjing, Veera	King Mongkut's Institute of Technology Ladkrabang	Thailand
Bouhadada, Tahar	University of Annaba	Algeria
Butrous, Nasir	Australian Catholic University	Australia
Camilleri, Mario	University of Malta	Malta
Campeanu, Andrei	Politechnical University of Timisoara	Romania
Cardoso, Eduardo	Monterrey Institute of Technology and Higher Education	Mexico
Castellini, Horacio	National University of Rosario	Argentina
Chadwick, Steve	Aeronautical University	USA
Challa, Radhakumari	Sri Sathya Sai Institute of Higher learning	India
Chamilothoris, George	Technological Institute of Piraeus	Greece
Chang, Yu-Ju	National Chi Nan University	Taiwan
Chen, Jingchao	DongHua University	China
Chen, Yuh-Wen	Da Yeh University	Taiwan
Ching, Wee Soon	Nanyang Polytechnic	Singapore
Chorikavil Thomas, Jose	Central Plantation Crops Research Institute	India
Chuvakin, Anton	Gartner	USA
Cipolla-Ficarra, Francisco	Asociación Latina de Interacción Persona-Ordenador	Italy
Cirella, Jonathan	RTI International	USA
Corcuera, Pedro	University of Cantabria	Spain
D`Ulizia, Arianna	National Research Council	Italy
De la Puente, Fernando	University of Las Palmas de Gran Canaria	Spain
Dean, Jeff	Cleveland State University	USA
Debono, Carl James	University of Malta	Malta

Dediu, Adrián Horia	University Rovira i Virgili	Spain
Deep, Kusum	Indian Institute of Technology	India
Deng, Hepu	RMIT University	Australia
Diaconescu, Stefan	Softwin SRL	Romania
Dizdaroglu, Bekir	Karadeniz Technical University	Turkey
Dolotov, Alexander	National Institute of Justice	USA
Domínguez-Dorado, Manuel	University of Extremadura	Spain
Dutta, Sourav	IBM	India
Dvornik, Josko	University of Split	Croatia
Eby, Gulsun	Anadolu University	Turkey
El Bakkali, Hanan	National School of Computer Science and Systems Analysis	Morocco
Elmahboub, Widad	Hampton University	USA
Elsas, Alexander	Frankfurt University	Germany
Erkollar, Alptekin	University of Applied Sciences	Austria
Eynard, Benoit	University of Technology of Compiegne	France
Félez, Jesús	Technical University of Madrid	Spain
Fiorini, Rodolfo	Polytechnic University of Milan	Italy
Florescu, Gabriela	National Institute for Research and Development in Informatics	Romania
Floyd, Raymond	Innovative Insights, Inc.	USA
Foglia, Pierfrancesco	University of Pisa	Italy
Fox, Richard	Northern Kentucky University	USA
Fratini, Antonio	University of Naples Federico II	Italy
Furukawa, Susumu	University of Yamanashi	Japan
Fúster-Sabater, Amparo	Institute of Applied Physics	Spain
Galatchi, Dan	Polytechnic University of Bucharest	Romania
Gamero Inda, Eduardo	Durango Institute of Technology	Mexico
Gani, Abdullah	University of Malaya	Malaysia
García Marco, Francisco Javier	University of Zaragoza	Spain
García-Zambrana, Antonio	University of Malaga	Spain
Genser, Robert	Institute for Handling Devices and Robotics	Austria
Georgopoulos, Voula	Technological Educational Institute of Patras	Greece
Giampapa, Joseph	Carnegie Mellon University	USA
Gini, Giuseppina	Polytechnic of Milan	Italy
Gnatyuk, Volodymyr	V.E. Lashkaryov Institute of Semiconductor Physics of the National Academy of Sciences of Ukraine	Ukraine
Golayoglu Fatullayev, Afet	Baskent University	Turkey

Gómez Moreno, Hilario	University of Alcala	Spain
Gonçalves, Ricardo	New University of Lisbon	Portugal
Goureau, Pascal	École Nationale Supérieure de l'Électronique et de ses Applications	France
Grediaga, Ángel	University of Alicante	Spain
Gubaidullin, Gail G.	University of Applied Sciences Koblenz	Germany
Guedes de Souza, Sergio	Federal University of Rio de Janeiro	Brazil
Gutiérrez Tornés, Agustín Francisco	Monterrey Institute of Technology and Higher Education	Mexico
Hanakawa, Noriko	Hannan University	Japan
Hartley, Roger	University of Leeds	UK
Haruvy, Nava	Netanya Academic College	Israel
Hashimoto, Shigehiro	Kogakuin University	Japan
Hasnaoui, Salem	Technologies National School of Engineering of Tunis	Tunisia
Hetzer, Dirk	T-Systems	Germany
Hicks, William	IMET Corporation	USA
Hong, Sungbum	Jackson State University	USA
Hu, Hong	Hampton University	USA
Hung, C. K. (Patrick)	University of Ontario Institute of Technology	Canada
Ilkova, Tatiana	Bulgarian Academy of Sciences	Bulgaria
Intakosum, Sarun	King Mongkut's Institute of Technology Ladkrabang	Thailand
Ishimura, Naoyuki	Hitotsubashi University	Japan
Jamilah, Ahmad	Science University of Malaysia	Malaysia
Janota, Ales	University of Zilina	Slovakia
Jaoua, Ali	University of Qatar	Qatar
Jedra, Mohamed	Laboratoire Conception et Systemes	Morocco
Jia, Xiuping	University of New South Wales at the Australian Defence Force Academy	Australia
Jiménez Rodríguez, Lourdes	University of Alcala	Spain
Kalpic, Damir	University of Zagreb	Croatia
Kamrani, Ehsan	University of Lorestan	Iran
Kaneko, Itaru	Tokyo Polytechnic University	Japan
Karamat, Parwaiz	The Open Polytechnic of New Zealand	New Zealand
Karpuhin, Oleksandr	Kharkiv National University of Radio and Electronics	Ukraine
Kawaguchi, Masashi	Suzuka National College of Technology	Japan
Kawarazaki, Noriyuki	Kanagawa Institute of Technology	Japan

Kher, Shubhalaxmi J.	Arkansas State University	India
Kim, Do-Hoon	Kyung Hee University	South Korea
Kim, Tae Yong	Chung-Ang University	South Korea
King, Roger	University of Colorado	USA
Kinser, Jason	George Mason University	USA
Klett, Fanny	German Workforce ADL Partnership Laboratory	Germany
Knipp, Tammy	Florida Atlantic University	USA
Kobti, Ziad	University of Windsor	Canada
Kopparapu, Sunil Kumar	Tata Consultancy Services Limited	India
Kreisler, Alain	University Paris 06	France
Krothapalli, Sreenivasa Rao	Indian Institute of Technology Kharagpur	India
Kroumov, Valeri	Okayama University of Science	Japan
Kung, Hsu-Yang	National Pingtung University of Science and Technology	Taiwan
Kushwaha, Ramesh	University of Michigan	USA
Lara, Teodoro	University of the Andes	Venezuela
Lara Palma, Ana Mª	University of Burgos	Spain
Laudani, Antonino	University of Roma Tre	Italy
Law, Rob	Hong Kong Polytechnic University	Hong Kong
Ledesma Orozco, Sergio E.	Guanajuato University	Mexico
Lee, Chang Won	Hanyang University	South Korea
Lee, DoHoon	Pusan National University	South Korea
Lee, Jong Kun	Changwon National University	South Korea
Lee, Yih-Jiun	Chien Kuo Technology University	Taiwan
Li, Di	South-China University of Technology	China
Lin, Chin-Feng	National Taiwan Ocean University	Taiwan
Lipikorn, Rajalida	Chulalongkorn University	Thailand
Litvin, Vladimir	California Institute of Technology	USA
Liu, Hong	Embry-Riddle Aeronautical University	USA
Liu, HsinChih Frank	Shih Chien University	Taiwan
Llosas, Yolanda	Eastern University	Cuba
Lopes da Silva, Paulo Afonso	Military Engineering Institute	Brazil
Lorenz, Pascal	University of Haute Alsace	France
Lou, Shi Jer	National Pingtung University of Science and Technology	Taiwan
Mainguenaud, Michel	Institut National des Sciences Appliquées	France
Marra, Cirley Barbosa	University Center of Cerrado Sponsorship	Brazil
Martín, Francisco	University of Malaga	Spain
Martínez Carreras, Mª Antonia	University of Murcia	Spain

Mascari, Jean-François	Italian National Research Council	Italy
Masini, Barbara	University of Bologna	Italy
Matsuda, Michiko	Kanagawa Institute of Technology	Japan
Mauri, Marco	Polytechnic of Milan	Italy
Mekras, Nikolaos	ANTER Ltd.	Greece
Mestre, Pedro	University of Trás-os-Montes and Alto Douro	Portugal
Milani, Mattia	Intercom R. R. L.	Italy
Mohsen, Ghribi	University of Moncton	Canada
Moody, Scott	The Boeing Company	USA
Moraes, Altino J. Mentzingen	Federal District University	Brazil
Morshed, Ahmed Hesham	Ain Shams University	Egypt
Mozar, Stefan	Electrical Testing Services Pty Ltd.	Australia
Mtika, Chatonda	Mantaro Product Development Services	USA
Munoz, Humberto	Southern University and A and M College	USA
Muñoz García, Ana Celina	University of the Andes	Venezuela
Myers, Susan	RTI International	USA
Nagy, Endre L.	Society of Instrument and Control Engineers	Japan
Nahm, In Hyun	Sun Moon University	South Korea
Naziha, Ahmad Azli	University Technology of Malaysia	Malaysia
Newsome, Mark	Hewlett-Packard Company	USA
Niegemann, Helmut M.	University of Erfurt	Germany
Nieuwenhuysen, Paul	Free University of Brussels	Belgium
Occelli, Sylvie	Economic Social Research Institute of Piemonte	Italy
Odetayo, Michael	Coventry University	UK
Olson, Patrick C.	National University	USA
Ong, Sim-Heng	National University of Singapore	Singapore
Park, Se Hyun	Chung-Ang University	South Korea
Parrilla Roure, Luís	University of Granada	Spain
Pérez, Diego	University of Vigo	Spain
Pisu, Pierluigi	Ohio State University	USA
Popentiu, Florin	University of Oradea	Romania
Priest, John	University of Texas Arlington	USA
Pruski, Alain	Paul Verlaine University – Metz	France
Radová, Vlasta	University of West Bohemia in Pilsen	Czech Republic
Ramírez Barrios, José Luís	University of Carabobo	Venezuela
Reggiani, Luca	Polytechnic Institute of Milan	Italy
Reis, Arsénio	University of Trás-os-Montes e Alto Douro	Portugal
Ríos Valledepaz, Javier M.	Metropolitan University	Venezuela

Riquelme, Bibiana	Physics Institute of Rosario	Argentina
Rivera, Ángel	Utica College	USA
Rizki, Mateen	Wright State University	USA
Robinson, Jean	Reseach Triangle Institute International	USA
Rodríguez, Mª Dolores	University of Alcala	Spain
Rosete, Juan	Technological Institute of Queretaro	Mexico
Rot, Artur	Wroclaw University of Economics	Poland
Sahyun, Steven	University of Wisconsin	USA
Saitta, Francesco	University of Palermo	Italy
Salahddine, Krit	LESSI	Morocco
Santagati, Cettina	University of Catania	Italy
Santos de Sá Lisboa, Flávia	University of São Paulo	Brazil
Sarukhanyan, Hakob G.	Institute for Informatics and Automation Problems	Armenia
Savorski, Victor	Russian Academy of Sciences	Russian Federation
Saxena, Ashutosh	Infosys Technologies Limited	India
Seyed Abbassi, Behrooz	University of North Florida	USA
Shapira, Joseph	Comm & Sens Ltd.	Israel
Shi, Lei	University of Maryland, Baltimore County	USA
Shklyar, Benzion	Holon Academic Institute	Israel
Singh, Vijander	Netaji Subhas Institute of Technology	India
Sinkunas, Stasys	Kaunas University of Technology	Lithuania
Snae Namahoot, Chakkrit	Naresuan University	Thailand
Soeiro, Alfredo	University Porto	Portugal
Sokolov, Sergey	Keldysh Institute for Applied Mathematics	Russian Federation
Sousa, António	Institute of Biomedical Engineering	Portugal
Srivastava, Anurag	Indian Institute of Information Technology and Management	India
Stoleru, Neculai Daniel	Military Technical Academy in Romania	Romania
Štork, Milan	University of West Bohemia	Czech Republic
Strugar, Ivan	University of Zagreb	Croatia
Suárez Sarmiento, Álvaro	University of Las Palmas de Gran Canaria	Spain
Suárez-Garaboa, Sonia Mª	University of A Coruña	Spain
Sutherland, Trudy	Vaal University of Technology	South Africa
Swart, William	East Carolina University	USA
Szabo, Raisa	Nova Southeastern University	USA
Szygenda, Stephen A.	Southern Methodist University	USA

Tam, Wing K.	Swinburne University of Technology	Australia
Teixeira, Márcio	Uberlândia Federal University	Brazil
Teshigawara, Yoshimi	Soka University Engineering	Japan
Thulasiraman, Parimala	University of Manitoba	Canada
Torres, Ana	Polytechnic Institute of Santarém	Portugal
Travieso González, Carlos M.	University of Las Palmas de Gran Canaria	Spain
Trofimov, Vyacheslav	Moscow State University	Russian Federation
Tsai, Chang-Lung	Chinese Culture University	Taiwan
Tsaur, Woei-Jiunn	Da-Yeh University	Taiwan
Tuzikov, Alexander V.	United Institute of Informatics Problems National Academy of Sciences of Belarus	Belarus
Ucal, Meltem	Kadir Has University	Turkey
Valova, Iren	University of Massachusetts	USA
Vemuru, Srinivasa	Ohio Northern University	USA
Vincent, Nicole	Paris Descartes University	France
Vlad, Radu-Constantin	Technical University of Cluj-Napoca	Romania
Wada, Shigeo	Tokyo Denki University	Japan
Wang, Fuchung	National Chengchi University	Taiwan
Wang, Xiao Hui	University of Pittsburgh	USA
Wang, Yin-Tien	Tamkang University	Taiwan
Wataya, Roberto Sussumu	Adventist University Center of Sao Paulo	Brazil
Wen, Feng	Southwest Jiaotong University	China
Wilson, Hentie	University of South Africa	South Africa
Woo, Hong S.	Middlesex University	UK
Yahiaoui, Azzedine	Eindhoven University of Technology	Netherlands
Yin, Jianwen	Dell Computer Corporation	USA
Yoshida, Eri	Toyohashi University of Technology	Japan
Yu, Chien-Chih	National ChengChi University	Taiwan
Zemliak, Alexander	Autonomous University of Puebla	Mexico
Zhang, Qinyu	Harbin Institute of Techology	China
Zhang, Wei Bin	Sun Microsystems	China
Zhang, Yanlong	Manchester Metropolitan University	UK

ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

Abdullah, H.	University of South Africa	Zambia
Achimas Cadariu, Andrei	Iuliu Hatieganu University of Medicine and Pharmacy	Romania
Adji, Teguh Bharata	Gadjah Mada University	Malaysia
Afullo, Thomas Joachim	University of KwaZulu Natal	South Africa
Ahmad, Azlin	MARA University of Technology	Malaysia
Amayeh, Gholamreza	Magic Leap Inc.	USA
Andrea, Bencsik	Széchenyi István University	Hungary
Askraba, Vedran	Monash University	Australia
Azar, Jimmy	Uppsala University	Sweden
Baek, Jinsuk	Winson-Salem State University	USA
Balakrishnan, Prasanna	Sura Systems Pvt. Ltd	India
Bayat, Majid	Kharzmi University	Iran
Bengtsson, Ewert	Uppsala University	Sweden
Berleant, Daniel	University of Arkansas at Little Rock	USA
Bernardi, Ansgar	German Research Center for Artificial Intelligence	Germany
Bethke, Albert	RTI International	USA
Brandejsky, Tomas	Czech Technical University in Prague	Czech Republic
Bucciero, Alberto	Italian Research Council	Italy
Burita, Ladislav	University of Defence	Czech Republic
Chou, Te-Shun	East Carolina University	USA
Chu, Yu-Ling	Taipei Municipal Huajiang Senior High School	Taiwan
Contreras, Leonardo	Simon Bolivar University	Venezuela
Costa, Marly	Federal University of Amazonas	Brazil
Courte, Dale	University of Dayton	USA
Cui, Jinrong	Harbin Institute of Technology	China
Daneshmand Malayeri, Amin	Malayer Azad University	Iran
Drevin, Gunther	North West University	South Africa
Dvoøák, Zdenìk	University of Zilina	Slovakia
Elaydi, Hatem	Iugaza	Palestine

Ellabidy, Magdy	Wentworth Institute of Technology	USA
Erdogan, Eylem	Kadir Has Universitesi	Turkey
Falahati, Abolfazl	Iran University of Science and Technology	Iran
Florescu, Gabriela	National Institute for Research and Development in Informatics	Romania
Gamess, Eric	Central University of Venezuela	Venezuela
Gardezi, A. K.	Colegio de Postgraduados	Mexico
Gregus, Michal	Comenius University	Slovakia
Guo, Hongyu	University of Houston-Victoria	USA
Gutiérrez, Ángel	Montclair State University	USA
He, Haibo	University of Rhode Island	USA
Hoag, John	Ohio University	USA
Huang, Chih Wei	National Central University	Taiwan
Hussein, Mohammed	Iugaza	Palestine
Ilkova, Tatiana	Bulgarian Academy of Sciences	Bulgaria
Janota, Ales	University of Zilina	Slovakia
Kaabouch, Naima	University of North Dakota	USA
Kapley, Atya	National Environmental Engineering Research Institute	India
Karigar, Chandrakant	Bangalore University	India
Kim, Cheolhong	Chonnam University	South Korea
Kim, Jung H.	North Carolina A&T State University	USA
Kim, Sang Chul	Hankuk University of Foreign Studies	South Korea
Koolagudi, Shashidhar G.	National Institute of Technology Karnataka	India
Kopp, Carlo	Monash University	Australia
Kouhi Jelehkaran, Bahram	Islamic Azad University	Iran
Kumar S., Vijaya	Indian Institute of Information Technology, Design & Manufacturing	India
Kwon, Joon-Hee	Kyonggi University	South Korea
Lee, Changsu	The University of Suwon	South Korea
Lee, Moon-Kyu	Keimyung University	South Korea
Li, Peng	East Carolina University	USA
Li, Ping	University of the Health Sciences	USA
Li, Zheng	Tsinghua University	China
Liang, Wen-Yau	National Changhua University of Education	Taiwan
Lin, Chih Hsing	Chip Implementation Center	Taiwan
Lin, Xiaogong	Harbin Engineering University	China
Liu, Fei	The University of New Mexico	USA
Liu, Jinxing	Qufu Normal University	China

Lohichi, Borhen	2iE	France
Mahenthiran, Sakthi	Butler University	USA
Majer, Tomáš	University of Žilina	Slovakia
Malý, Vlastimil	University of Defence	Czech Republic
Mastacan, Lucian	Gheorghe Asachi Technical University	Romania
Mesaric, Josip	Josip Juraj Strossmayer University in Osijek	Croatia
Mihály, Görög	Univerity of Corvinus	Hungary
Mikulski, Jerzy	Silesian University of Technology	Poland
Mohapatra, Pratap	Indian Institute of Technology Kharagpur	India
Mtsweni, Jabu	University of South Africa	Zambia
Murugesan, Karuppasamy	Atilin University	India
Nakashima, Kenichi	Kanagawa University	Japan
Oblitey, William	Indiana University of Pennsylvania	USA
Osi, Albert	Oakwood University	USA
Paul, Joseph	National Institute of Technology Calicut	India
Pencek, Thomas	University of South Florida	USA
Pozar, Josef	The Police Academy of the Czech Republic in Prague	Czech Republic
Pribyl, Pavel	Czech Technical University in Prague	Czech Republic
Rajan, Kanjirodan	Indian Institute of Science	India
Ramesh, Hariharan	Indian Institute of Technology Guwahati	India
Razavizadeh, S. Mohammad	Iran University of Science & Technology	Iran
Rebok, Sandra	Spanish National Research Council	Spain
Rodionov, Alexey	Institute of Computational Mathematics and Mathematical Geophysics	Russian Federation
Rosman, Pavel	Tomas Bata University	Czech Republic
Ryu, Daehyun	Hansei University	South Korea
San Martín Salas, César E.	University of the Frontier	Chile
Sharmah, S. P.	Indian Institute of Technology Kharagpur	India
Shi, Lei	University of Maryland, Baltimore County	USA
Shih, Sheng-Wen	National Chi Nan University	Taiwan
Sim, Khe Chai	National University of Singapore	Singapore
Sodomka, Petr	Brno University of Technology	Czech Republic
Spinola, Mauro	University of São Paulo	Brazil
Stang, David	American University	USA
Starek, Tomas	INTENS Corporation S.R.O.	Czech Republic
Stefanizzi, Maria Laura	University of Salento	Italy
Strefezza, Miguel	Simon Bolivar University	Venezuela
Sun, Fuchun	Tsinghua University	China

Tan, Tan-Hsu	National Taipei University of Technology	Taiwan
Tollenaere, Michel	G-SCOP	France
Trawiński, Bogdan	Wroclaw University of Technology	Poland
Tseng, Bill	The University of Texas at El Paso	USA
Turner, Linda	Langara College	Canada
Vuppala, Anil Kumar	International Institute of Information Technology, Hyderaba	India
Wang, Jian	Shanghai University	China
Wang, Yan	Beihang University of China	China
Williams, Quinton	Jackson State University	USA
Wongsuwarn, Hataitep	Kasetsart University	Thailand
Xu, Hongjun	University of KwaZulu Natal	South Africa
Yang, Hung-Jen	National Kaohsiung Normal University	Taiwan
Yao, Zhongmei	University of Dayton	USA
You, Cheolwoo	Myongji University	South Korea
Yukhimets, Dmitriy	Institute of Automation and Control Processes	Russian Federation
Zhang, Guoping	Morgan State University	USA
Zhao, Hua-An	Kumamoto University	Japan
Zhirabok, Alexey	Far-Eastern Federal University	Russian Federation

HONORARY PRESIDENTS OF PAST CONFERENCES
Bela Banathy
Stafford Beer
George Klir
Karl Pribram
Paul A. Jensen
Gheorghe Benga

HONORARY CHAIR
William Lesso

PROGRAM COMMITTEE CHAIRS
Michael Savoie
C. Dale Zinn

GENERAL CHAIR
Nagib Callaos

ORGANIZING COMMITTEE CHAIRS
Belkis Sánchez
Andrés Tremante

CONFERENCE PROGRAM MANAGER
María Sánchez

HARDCOPY PROCEEDINGS PRODUCTION CHAIR
María Sánchez

TECHNICAL CONSULTANT ON COMPUTING SYSTEM / CD PROCEEDINGS PRODUCTION CHAIR
Juan Manuel Pineda

META REVIEWERS SUPPORT
Dalia Sánchez

SYSTEM DEVELOPMENT, MAINTENANCE AND DEPLOYMENT
Dalia Sánchez
Keyla Guédez

OPERATIONAL ASSISTANT
Marcela Briceño

HELP DESK
Juan Esqueda

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	100	100.00
United States	19	19.00
Japan	13	13.00
Czech Republic	9	9.00
Latvia	9	9.00
Lithuania	5	5.00
Taiwan	5	5.00
Brazil	4	4.00
Slovakia	4	4.00
India	3	3.00
Russian Federation	3	3.00
South Korea	3	3.00
Turkey	3	3.00
Canada	2	2.00
China	2	2.00
Italy	2	2.00
South Africa	2	2.00
Thailand	2	2.00
Chile	1	1.00
Ecuador	1	1.00
Finland	1	1.00
Germany	1	1.00
Iran	1	1.00
Israel	1	1.00
Mexico	1	1.00
Sweden	1	1.00
United Kingdom	1	1.00
Vietnam	1	1.00

Foreword

Our purpose in the 18th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2014) is to provide, in these increasingly related areas, a multi-disciplinary forum, to foster interdisciplinary communication among the participants, and to support the sharing process of diverse perspectives of the same transdisciplinary concepts and principles.

Systemics, Cybernetics and Informatics (SCI) are being increasingly related to each other in almost every scientific discipline and human activity. Their common transdisciplinarity characterizes and communicates them, generating strong relations among them and with other disciplines. They work together to create a whole new way of thinking and practice. This phenomenon persuaded the Organizing Committee to structure WMSCI 2014 as a multi-conference where participants may focus on one area, or on one discipline, while allowing them the possibility of attending conferences from other areas or disciplines. This systemic approach stimulates cross-fertilization among different disciplines, inspiring scholars, originating new hypothesis, supporting production of innovations and generating analogies; which is, after all, one of the very basic principles of the systems' movement and a fundamental aim in cybernetics.

WMSCI 2014 was organized and sponsored by the International Institute of Informatics and Systemics (IIIS, www.iiis.org), member of the International Federation of Systems Research (IFSR). The IIIS is a *multi-disciplinary organization for inter-disciplinary communication and integration*, which includes about 4500 members. Consequently, a main purpose of the IIIS is to foster knowledge integration processes, interdisciplinary communication, and integration of academic activities. Based on 1) the transdisciplinarity of the systemic approach, along with its essential characteristic of emphasizing *relationships* and *integrating* processes, and 2) the multi-disciplinary support of cybernetics' and informatics' concepts, notions, theories, technologies, and tools, the IIIS has been organizing multi-disciplinary conferences as a platform for fostering inter-disciplinary communication and knowledge integration processes.

Multi-disciplinary conferences are organized by the IIIS as support for both **intra-** and **inter-disciplinary** communication. Processes of intra-disciplinary communication are mainly achieved via traditional paper presentations in corresponding disciplines, while conversational sessions, regarding trans- and inter-disciplinary topics, are among the means used for inter-disciplinary communication. Intra- and inter-disciplinary communications might generate *co-regulative cybernetic loops*, via negative feedback, and *synergic* relationships, via positive feedback loops, in which both kinds of communications could increase their respective effectiveness. Figure 1 shows at least two cybernetic loops if intra- and inter-disciplinary are adequately related. A necessary condition for the effectiveness of Inter-disciplinary communication is an adequate level of **variety** regarding the participating disciplines. *Analogical thinking and learning processes* of disciplinarians depend on it; which in turn are potential sources of the creative tension required for cross-fertilization among disciplines and the generations of new hypothesis. An extended presentation regarding this issue can be found at www.iiis.org/MainPupose.
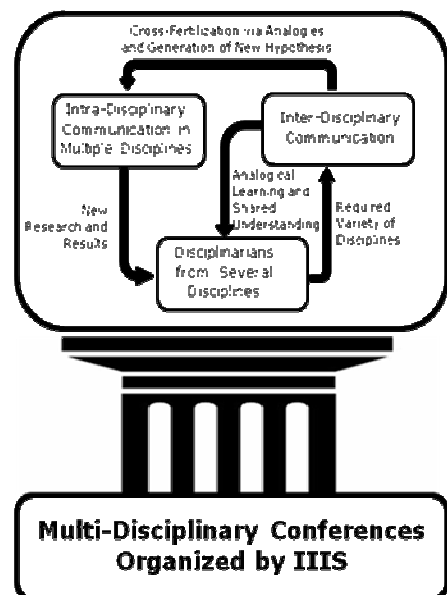
Figure 1

In the specific case of Systemics, Cybernetics and Informatics (SCI), the IIIS is an organization dedicated to contribute to the development of the Systems Approach, Cybernetics, and Informatics potential, using both: knowledge and experience, thinking and action, theory and practice, for:

a) the identification of synergetic relationships among Systemics, Cybernetics and Informatics, and between them and society;
b) the promotion of contacts among the different academic areas, through the transdisciplinarity of the systems approach;
c) the identification and implementation of communication channels among the different professions;
d) the supply of communication links between the academic and professional worlds, as well as between them and the business world, both public and private, political and cultural;
e) the stimulus for the creation of integrative arrangements at different levels of society, as well as at the family and personal levels;
f) the promotion of transdisciplinary research, both on theoretical issues and on applications to concrete problems.

These IIIS objectives have directed the organizational efforts of yearly WMSCI/ISAS conferences since 1995.

On behalf of the Organizing Committee, I extend our heartfelt thanks to:

1. the 746 members of the Program Committee from 59 countries (including the PC members of the events organized in its context and jointly with it);
2. the 477 additional reviewers, from 63 countries, for their **double-blind peer reviews**; and
3. the 172 reviewers, from 41 countries, for their efforts in making the **non-blind peer reviews**. (Some reviewers supported both: non-blind and double-blind reviewing for different submissions).

A total of 1006 reviews made by 649 reviewers (who made at least one review) contributed to the quality achieved in WMSCI 2014. This means an average of 6.45 reviews per submission (156 submissions were received). ***Each registered author had access, via the conference web site, to the reviews that recommended the acceptance of their respective submissions***. Each registered author could also 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 the 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.

In the organizational process of WMSCI 2014, about 156 papers/abstracts were submitted. These pre-conference proceedings include about 100 papers that were accepted for presentation from 27 countries (44 countries taking into account the presentations in collocated events). I extend our thanks to the invited sessions' organizers for collecting, reviewing, and selecting the papers that will be presented in their respective sessions. The submissions were reviewed as carefully as time permitted; it is expected that most of them will appear in a more polished and complete form in scientific journals.

This information about WMSCI 2014 is summarized in the following table, along with the other collocated conferences:

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
WMSCI 2014	156	649	1006	1.55	6.45	100	64.10%
IMSCI 2014	90	316	556	1.76	6.18	45	50.00%
IMETI 2014	61	291	565	1.94	9.26	25	40.98%
CISCI 2014	120	509	1128	2.22	9.40	49	40.83%
TOTAL	427	1765	3255	1.84	7.62	219	51.29%

We also extend our gratitude to the invited sessions and special track organizers, as well as to the co-editors of these proceedings, for the hard work, energy and eagerness they displayed preparing their respective sessions. 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 the Program Committee Chair of past conferences, and as Honorary President of WMSCI 2014, 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 Sánchez, who brilliantly managed the organizing process.

Our gratitude to Professors Bela H. Banathy, Stafford Beer, George Klir, Karl Pribram, Paul A. Jensen, and Gheorghe Benga who dignified our past WMSCI conferences by being their Honorary Presidents. Special thanks to Dr. C. Dale Zinn and Professor Jorge Baralt for co-chairing WMSCI 2014 Program Committee and to professors Andrés Tremante and Belkis Sánchez for co-chairing the Organizing Committee. We also extend our gratitude to the following scholars, researchers, and professionals who accepted to deliver plenary workshops and/or to address the audience of the General Joint Plenary Sessions with keynote conferences.

Plenary Workshop, more details (abstracts and short bios) were included in the Conference Program booklet and at http://www.iiis.org/summer2014plenaryevents/

Professor Thomas Marlowe, Seton Hall University, USA, Department of Mathematics and Computer Science, Program Advisor for Computer Science, Doctor in Computer Science and Doctor in Mathematics

Dr. Susu Nousala, Aalto University, Finland, Researcher in Sustainable Design, and Research Fellow at the (Australasian Centre for the Governance and Management of Urban Transport) Faculty of Architecture- Buildi

Plenary Keynote Speakers, more details more details (abstracts and short bios) were included in the Conference Program booklet and at http://www.iiis.org/summer2014plenaryevents/

Professor Leonid Perlovsky, Harvard University and The Air Force Research Laboratory, USA

Professor Shigehiro Hashimoto, Kogakuin University, Japan, Associate to the President and Dean of Admissions Center, Doctor of Engineering and Doctor of Medicine Biomedical Engineering

Professor T. Grandon Gill, University of South Florida, USA, Director of the Doctorate of Business Administration

Dr. Jeremy Horne, President-emeritus, Southwest Area Division, American Association for the Advancement of Science (AAAS), USA

Dr. Karl H. Müller, Director of The Steinbeis Transfer Center New Cybernetics, Vienna, Austria and Professor at the University of Ljubljana, Slovenia

Professor Richard Segall, Arkansas State University, USA, Computer Information Technology

Professor Thomas Marlowe, Seton Hall University, USA, Department of Mathematics and Computer Science, Program Advisor for Computer Science, Doctor in Computer Science and Doctor in Mathematics

Dr. Ronald Styron, University of South Alabama, USA, Director of the Quality Enhancement Plan

Dr. Heidi Ann Hahn, Los Alamos National Laboratory, USA, Director of the Engineering Capability Development, Past President of the International Council of Systems Engineering (INCOSE) Enchantment Chapter

Dr. Robert Cherinka, MITRE Corporation, USA, Senior Principal Information Systems Engineer

Mr. Joseph Prezzama, MITRE Corporation, USA, Lead Communications Engineer, leads the program in support of the Chief Information Officer (CIO) for the United States Special Operations Command (USSOCOM)

Professor Mario Norbis, Quinnipiac University, USA, School of Business

Dr. Marta Szabo White, Georgia State University, USA, Associate Professor, Strategic Management, Director of the Robinson College of Business Honors Program, Director of Study Abroad Programs: Transition Economies and Business Mediterranean Style

Dr. Susan S. Nash, University of Oklahoma, USA, Director of Education and Professional Development, American Association of Petroleum Geologists

Dr. Matthew E. Edwards, Alabama A&M University, USA, Professor of Physics and Former Dean, School of Arts and Sciences

Dr. Russell Jay Hendel, Towson University, USA, Dept of Mathematics

Professor Tomas Zelinka, Czech Technical University in Prague, Czech Republic, Vice-Dean of the Faculty of Transportation Sciences

Dr. Fred Maymir-Ducharme, Executive Architect at IBM, USA, Defense & Space, and Adjunct Professor, University of Maryland University College (UMUC), USA

Dr. Jennifer Styron, University of South Alabama, USA, College of Nursing, Former Research Specialist and Distance Education Coordinator at The University of Southern Mississippi, USA

VOLUME I

CONTENTS

Communication and Network Systems, Technologies and Applications

Computer Science and Engineering

Computing Techniques

Control Systems, Technologies and Applications

Ethics, Cybernetics and Informatics

Image and Multidimensional Signal Processing

Image, Acoustic, Speech and Signal Processing

Information Systems Development

Information Systems, Technologies and Applications

Optical Systems, Technologies and Applications

Authors Index

Corporate Governance as a Factor for Investment Decision Making on CEE Equity Markets

Julija BISTROVA, Natalja LACE
Faculty of Engineering Economics and Management, Riga Technical University
6 Kalnciema Str., Riga, LV-1048, Latvia

and

Manuela TVARONAVICIENE
Faculty of Business Management, Vilnius Gediminas Technical University
Saulėtekio al. 11, Vilnius, LT–10223, Lithuania

ABSTRACT

In the developed stock markets the corporate governance aspect is crucial in the stock portfolio selection process for investor seeking to achieve shareholder value sustainability. In the emerging markets the importance of the corporate governance role just starts to be realized by the investors and by the corporate managers.

The present research, looking at the stock performance leaders and laggards, analyzes whether the corporate governance system matters to achieve long-term shareholder value within the Central and Eastern European stock markets universe. Corporate governance quality was assessed and compared among the out- and underperformers. Additionally, the financial results plausibility and the ownership structure were considered.

The obtained results provide the proof that the corporate governance does matter as the market outperformers have above average corporate governance quality and provide trustworthy financial results more often than the underperforming companies.

Keywords: Corporate governance, Ownership structure, Earnings quality, stock performance, CEE equity markets

1. INTRODUCTION

The topic of the corporate governance is being thoroughly discussed withtin the Central and Eastern Europe (CEE) stock markets, and obviously the companies succeed in their efforts bringing their corporate governance systems to the world-class quality standards, complying with the best practice principles (Bistrova, 2014). But the question is about the pay-off and the trade-off between the investments in the corporate governance system establishment and the decent economic and market performance.

It is a well-known fact that the „evil" stock indices significantly beat the stock indices comprised of only SRI complying companies (Lobe, Walkshäusl, 2008). However, there are also the studies proving that it is worth investing in the companies sticking to the high level of CSR as they deliver market-neutral or above market average performance (Arx & Ziegler, 2008; Wang, 2011).

However, being SRI compliant and having excellent management team and supervisory board organization is just one side of the corporate governance system. The other and the most influential stock performance wise is the ethical side of the earnings management as well as the ownership structure particularly in the regions, where the concentrated shareholding is widely spread phenomenon.

Central and Eastern European companies quite often have major owner in the capital structure, who is also being very active in the routine company management (Lace, Bistrova & Kozlovkis, 2013). Besides, the creative accounting practices tend to emerge on the corporate landscape of CEE countries quite often (Bistrova, 2014). All these factors undoubtedly affect the performance of the CEE investors' stock portfolios.

Moreover, low level of corporate transparency and low media coverage typical for the emerging markets could lead to the non-normal return distribution (Connover, 2011). Therefore, the majority of investors in CEE companies appreciate very good information disclosure, which could positively influence the stock performance of the company in the long-term.

Thus, the aim of the present study is to find out to what extent the above mentioned corporate governance system elements affect the stock performance.

The study's principal hypothesis is the following:

The companies having good corporate governance quality and high earnings plausibility are able to deliver higher shareholder value in the long-term.

Additionally, the authors checked to what extent the shareholder structure influences stock performance. For these purposes the ownership classisifaction was developed offering to distinguish major owners by the type (the aim of the shareholding) and by the location. Besides, the authors verified whether high ownership concentration structure adds more value when investing in the companies.

2. LITERATURE REVIEW

Content analysis on the shareholder value sustainability factors mentioned in the scientific articles proves that the corporate governance factor gains importance (Bistrova, 2014). It enjoys more attention than it used to both from the corporates and from the active investors' side, who often see the necessity to integrate this factor into the portfolio selection process.

A number of studies conducted on the developed markets state that the corporate governance has strong influence on the stock market returns. Gompers, Ishii and Metrick (2003) constructed "Governance Index", which covered the assessment of shareholders' rights at 1,500 companies in 1990s. Based on the index, they have modeled the portfolio strategy that would consider 'long' companies with strongest rights (lowest decile) and 'short' companies with weakest rights (highest decile). As a result, the investor could earn 8.5% outperformance. Similar study was done by Drobetz et al. (2003) in Germany showing the monthly difference in performance of well and poorly governed firms of 1.73%.

The significant correlation of such factors as CG index, CEO-Chairman separation and independence of board members with stock performance was found by Bhagat and Bolton (2008). But they did not find any evidence to prove the assumption that the quality of CG is a proxy for future stock performance. The findings of their study also show that given the low quality of corporate governance of a certain entity and given its poor performance, there is a high probability of management turnover.

Positive correlation between the firm value and the quality of corporate governance in case of 300 largest European companies (FTSE Eurotop 300) has been indicated by Dutch scientists Bauer, Guenster, Otten (2004). But when adjusting for country difference, the relationship is weakening.

The contrary situation was discovered in Japan by Aman and Nguyen (2007), who discovered that poorly governed firms outperform well-governed firms. The results were statistically insignificant, but the study clearly showed that significantly higher risk is attributed to the poorly governed firms.

Some research was conducted considering separate factors, which determine the quality of corporate governance. Strong relationship was identified between equity performance and board independence (Hermalin & Weisbach, 1998, 2003; Bhagat & Black, 2002), stock ownership of board members (Bhagat et al., 1999), separation of the CEO and Chairman positions (Brickley et al., 1997).

The story in the emerging markets is a bit different: due to ownership structure, which is often concentrated, the companies have rather low motivation to disclose information to outsiders. The need in the minority shareholders is obviously less, compared to the situation in the developed markets. The regulations regarding corporate governance are less strict than they are in the developed stock markets. In most cases these are just recommendations imposed by the local stock exchanges. Anyway, the question of the influence of corporate governance becomes more topical. In case of favourable outcome (positive correlation with stock returns), the obtained results proved by the empirical research can be used to persuade the companies to stick to the best practice.

The available related literature provides the evidence of outperformance of the well-governed companies also in the emerging markets. Roy Kouwenberg (2006) states that the corporate governance matters with regard to Thai public companies: stock return of the best 20% companies according to the CG score in the period 2003-2005 was by 19% p.a. better than the stock return of the weakest 20% companies.

The Indian market represented as NIFTY 50 was studied by Samontaray (2010), who found significant relationship between the share price and such independent variable as EPS, sales, net fixed assets as well as corporate governance factors.

Another important consituent of the corporate governance system is the ethical management, which in parts can be detected by the earnings quality analysis. A number of scientists proved that (Dechow & Dichev, 2001; Mahedy, 2005; Sloan, 1996) the accruals as a measure of earnings plausibility negatively affect equity performance. The US scientists Sloan (1996) and Houge and Loughran (2000) empirically proved that the companies having high accruals perform worse than the companies that do not have any accruals. The analysts of an asset management company Bernstein (Mahedy, 2005) proved that accruals are a powerful tool, which can be used for predicting future earnings and share performance of the company.

3. RESEARCH DESIGN

The expanded notion of the corporate governance includes not only the relation to the system of the governing bodies, but it also covers the ethical approach to the company management as well as the ownership structure.

Therefore, the author focused on the key dimensions of the corporate governance in its expanded notion: quality of the management team and supervisory board organization, quality of the earnings management and the ownership structure specifics. These corporate governance elements were evaluated separately for the market leaders (outperformers) and the market laggards (underperformers) in terms of the total shareholder return. These groups of companies were compared, according to the following criteria:

- Overall rating of corporate governance quality is based on the model developed by Bistrova and Lace in 2011, which includes the assessment of the board of directors, management team, quality of the investor relations and the information disclosure;
- Earnings quality assessment was based on the evaluation of the accruals level relative to the net operating assets;
- Owership specifics was evaluated based on the major shareholder type (family, government, financial, strategic owneror free float), on the location of the major holder (local or foreign) and on the ownership concentration level (dispersed or concentrated). The holder was classified as a major holder if it owns more than 10% of the total capital, while the ownership structure was considered to be concentrated if a major holder owns 25% of the capital or more.

The research sample is limited by the quoted CEE companies, which were the components of the local stock exchanges main lists in the financial year 2010. The static sample composition was considered in order to avoid the survivorship bias. The sample list includes 117 companies, the components of the main indices of the CEE stock exchanges: 15 companies traded in the Czech Republic (PX Index), 10 companies traded in Croatia (CROBEX), 12 companies traded in Hungary (BUX), 20 companies traded in Poland (WIG), 10 companies traded in Romania (BET), 7 companies traded in Slovakia (SAX), 6 companies traded in Slovenia (SBI TOP), 13 companies traded in Estonia (OMX BBGI), 5 companies traded in Latvia (OMX BBGI), 19 companies traded in Lithuania (OMX BBGI). These are the largest companies traded in the CEE countries with relatively good liquidity and above average market capitalization compared to other companies traded on the CEE market. One company, Czech electricity company CEZ, has dual listing on Prague Stock Exchange and on Warsaw Stock Exchange, so it was considered only once leading to the overall sample of 116 companies.

The selection of the companies, which managed to deliver sustainable out- or underperformance was made on the annual basis, classifying the company if it managed to beat the equally-weighted performance of the market as the winning company (outperformer), and as the laggard (underperformer) if it delivered TSR below the market return. The timeline of the analysis included 8 annual periods: from 2005 till 2012, which also covered the financial crisis becoming an important milestone for the CEE equity market development. The company was classified as a sustainable outperformer if it delivered above average result for 5 years and more; and it was classified as a sustainable underperformer if it delivered below average result for 5 years and more. The analysis shows that the companies underperforming the market 5 times and more (48 firms) significantly exceed the number of the companies, outperforming the market for 5 years and more (15 firms). Therefore, it was decided to consider also the expanded sample of the outperformers, i.e. the companies delivering alpha for 4 years and more (41 firm).

4. RESEARCH RESULTS

A. Corporate Governance Quality Evaluation

Corporate governance analysis, made according to the proposed corporate governance model states that the companies, which outperformed for 5 periods and more, are better managed and are more transparent than the average company, while the underperformers have an opposite situation, but the difference, however, is not very substantial (Figure 1).
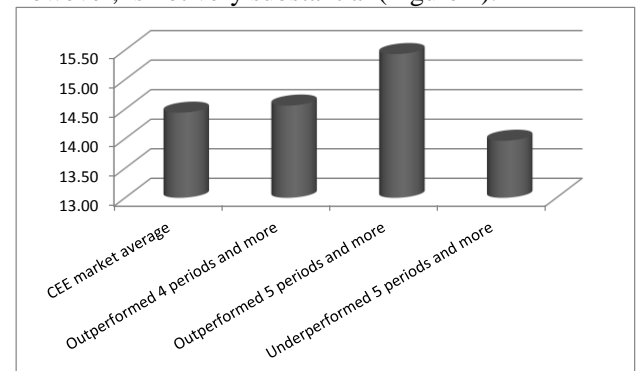
Figure 1 Average corporate governance quality levels of out- and underperformers.

Further deeper analysis of the difference of corporate governance quality between the sustainable out- and underperformers proves that the major discrepancy between these two groups of the companies is seen in the quality of the information disclosure and the level of investor relations (Table 1). In the conditions of the limited information availability in the emerging markets investors obviously value more the quality of the information provided by the listed companies.

Table 1

Average corporate governance levels of sustainable out and underperformers

	Super-visory Board	Mana-gement Team	Investor Rela-tions	Informa-tion Disclosure
CEE market average	4.91	2.91	1.84	3.90
Outperformed 4 periods and more	4.90	2.89	1.86	4.11
Outperformed 5 periods and more	5.18	2.98	2.09	4.59
Underperfor-med 5 periods and more	4.97	2.85	1.69	3.65

The quality of the management team organization is not very important aiming to achieve sustainable outperformance. Although the leading companies (outperformed for 5 periods and more) have the highest rating for the supervisory board, also the lagging companies can boast of well-established structures of the board of directors, earning higher rating than the market in general and higher than the expanded sample of the outperformers.

B. Earnings Quality Assessment

Accrual level as a proxy of the earnings quality, which allows defining plausible earnings forecast and, therefore, market expectations, was tested with the two year lag to the equity performance. Two year lag turned to be the optimal for the accruals to have an influence on the stock performance. Obviously data manipulation is negatively reflected in the financial results during the period of more than one year (Bistrova, 2014).

The chart (Figure 2) demonstrates that basically in each year the outperformers had lower accrual level, calculated according to the cash flow method, than the companies, which are classified as consistent underperformers.
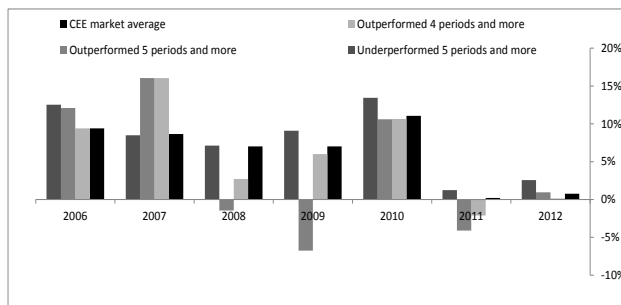
Figure 2 Median accruals level (based on C/F calculation methodology) of sustainable out- and underperformers.

The exception is 2007, when the investors obviously didn't pay attention to the financial state of the companies, which delivered high alpha. The statement was proved also in case of the profitability (many outperformers had lower than average profitability in 2007). Besides, the accruals level of outperformers and underperformers was calculated according to the data extracted primarily from the corporate balance sheets. The results obtained from this method (proposed by Richardson & Tuna), which assumes net operating assets being in the centre, are not perfectly comparable with the results obtained from the calculations according the cash flow method, when the cash flows from operations and investments are the key elements. The difference between the leading and the lagging companies is not obvious and the logical pattern of the relationship between the sustainable outperformance and very good earnings quality, indicating ethical corporate management, cannot be spotted.

C. Ownership Specifics

According to the data disclosed in Table 2, there is no particular bias to a certain ownership type as the percentage share in every group of the sustainable outperformers is not too much different from the general market. Still the concentrated outperformers sample (5 periods and more) is a bit more widely represented by the companies with the financial and strategic ownership, while a bit less by the companies with the governmental and family ownership. Obviously, CEE family-owned enterprises as opposed to their successful counterparts in the developed markets are not distinguished by the higher performance alpha and even some out of the researched sample filed for bankruptcy. So, the practice of the "built to last" family enterprises in the CEE financial field is not established yet.

Table 2

Sustainable out- and underperformers classified according to the ownership type

	Outperformed		Underperformed	CEE market average
	4 periods and more	5 periods and more	5 periods and more	
Share of total number of companies in the segment (%)				
Financial	20	23	20	19
Strategic	35	37	34	35
Government	20	15	20	20
Family	25	25	26	27
Share of total number of companies classified according to the particular shareholder type (%)				
Financial	25	13	28	100
Strategic	24	11	26	100
Government	24	8	26	100
Family	22	10	25	100

A third of the companies, which were classified as the underperformers, had foreign investors as major shareholders, while leading companies more often than the general market had local ownership as indicated in Table 3.

Table 3

Sustainable out and underperformers classified according to the ownership domicile

	Outperformed		Underper-formed	CEE market average
	4 periods and more	5 periods and more	5 periods and more	
Share of total number of companies in the segment (%)				
Local	69	69	63	68
Foreign	31	31	37	32
Share of total number of companies classified according to the particular shareholder type (%)				
Local	23	10	25	100
Foreign	22	9	32	100

Ownership concentration level obviously has more impact on the performance quality than the previous classifications of the ownership type. The companies, which were consistently underperforming the market, are more likely to have a dispersed ownership as proved by the results in Table 4. The companies delivering sustainable performance alpha in 92% of cases had concentrated shareholding structure, when an institutional or a private investor holds more than 25% of the total capital.

Table 4

Sustainable out- and underperformers classified according to the ownership concentration level

	Outperformed		Underper formed	CEE market average
	4 periods and more	5 periods and more	5 periods and more	
Share of total number of companies in the segment (%)				
Concentrated	83	92	63	75
Dispersed	17	8	37	25
Share of total number of companies classified according to the particular shareholder type (%)				
Concentrated	35	14	22	100
Dispersed	21	4	39	100

Possibly concentrated shareholding is the optimal ownership structure for the emerging market at its current development stage. The companies operate more efficiently if the business management decisions are influenced by one party, while the dispersed ownership does not provide the background for the successful enterprise development. It can occur due to the weak management team being not able to deliver sustainable TSR outperformance and, therefore, is the cause of high agency costs.

5. CONCLUSIONS

The obtained research results on the relationship between the corporate governance systems and equity performance robustness within the CEE stock markets partially prove the hypothesis, indicating that not every corporate governance aspect is crucial in search of the long-term performance alpha.

The study on the corporate governance environment discrepancies between the good and the bad companies in terms of the performance consists of the three parts: a) the general evaluation of the governance, where the primary focus was put on the management and supervisory board; b) the quality of earnings assessment, which alludes to a certain extent to the ethics of the management approach; c) the research on the ownership influence, considering the typology of the major owner, its location and the degree of the ownership concentration.

Though the leading companies can boast of the higher than average quality of the corporate governance in contrast to the lagging companies, the difference is not significant. Detailed analysis of the corporate governance elements evidences that the difference in the CG rating primarily is achieved by the very good information disclosure of the outperforming companies. This observation indicates high importance of the transparency of the company in the emerging European, where the media coverage of the listed companies is rather limited.

Earnings quality, which in the present research paper was evaluated based on the accruals level, has a potential to become an important factor to distinguish between the sustainable outperformers and the sustainable underperformers. The accruals calculated according to the cash flow method clearly are lower for the companies delivering sustainable outperformers basically through the whole observation period.

Ownership structure analysis evidences that the type of the major shareholder does not matter when selecting the stocks for the equity portfolio. The location of the major owner obviously has a more significant effect than the type of the shareholding. The most significant discrepancy between the market outperformers and market underperformers though appears when considering the shareholder structure concentration level. The investors are advised to give preference to the concentrated ownership, when selecting the companies for the "long" strategy equity portfolio in the emerging CEE equity market.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] U. Arx, A. Ziegler, (2008), **The Effect of CSR on Stock Performance: Net Evidence for the USA and Europe**. Working paper 08/85, CER-ETH, Center of Economic Research at ETH Zurich.

[2] J. Bistrova, N. Lace, (2011), Evaluation of Corporate Governance Influence on Stock Performance of CEE Companies. **The 15th World Multi-Conference on Systemics, Cybernetics and Informatics Proceedings**, Vol. 1, USA, Orlando, pp. 59-64.

[3] J. Bistrova, (2014), **Achieving Shareholder Value Sustainability on Central and Eastern European Equity Markets**. PhD Thesis. Riga: RTU, 203 p.

[4] H. Aman, P. Nguyen, (2007), Do Stock Prices Reflect The Corporate Governance Quality Of Japanese Firms? **Journal of The Japanese and International Economies**, Vol. 22, No. 4, pp. 647-662.

[5] R. Bauer, N. Gunster, R. Otten, (2004), Empirical Evidence on Corporate Governance in Europe. **Journal of Asset Management**, Vol. 5, No. 2, pp. 91-104.

[6] S. Bhagat, B. J. Bolton, (2008), Corporate Governance and Firm Performance. **Journal of Corporate Finance**, Vol.14 (Special Edition on Corporate Governance), pp. 257-273.

[7] S. Bhagat, D. Carey, Ch. Elson, (1999), Director ownership, corporate performance, and management turnover. **The Business Lawyer**, Vol. 54, pp. 1999.

[8] J. A. Brickley, J. L. Coles, G. Jarrell, (1997), Leadership structure: Separating the CEO and chairman of the board. **Journal of Corporate Finance**, Vol. 3, pp. 189-220.

[9] P. Dechow, I. Dichev, (2001), **The Quality of Accruals and Earnings: The Role of Accrual Estimation Error**. Stephen M. Ross School of Business/ Haas School of Business.

[10] W. Drobetz, A. Shillhofer, H. Zimmermann, (2003), Corporate governance and expected stock returns: Evidence from Germany. **European Financial Management**, Vol. 10, pp. 267–293.

[11] P. Gompers, J. Ishii, A. Metrick, (2003), Corporate governance and equity prices. **The Quarterly Journal of Economics**, Vol. 118, No. 1, pp. 107-155.

[12] B. Hermalin, M. Weisbach, (1998), Endogenously Chosen Boards of Directors and Their Monitoring of the CEO. **American Economic Review**, Vol. 88, No. 1, pp. 96-118.

[13] T. Houge, T. Loughran, (2000), Cash flow is king: Cognitive errors by investors. **Journal of Psychology and Financial Markets**, Vol. 2, No. 1, pp. 161-175.

[14] R. Kouwenberg, (2006), **Does Voluntary Corporate Governance Code Adoption Increase Firm Value in Emerging Markets? Evidence from Thailand.** Research paper, Mahidol University, College of Management.

[15] N. Lace, J. Bistrova, K. Kozlovskis, (2013), Ownership Type Influence on Dividend Payments in CEE Countries**. Business: Theory and Practice Journal**, Vol. 14, No. 3, pp. 259-266.

[16] S. Lobe, Ch. Walkshäusl, (2008), **Vice vs. Virtue Investing**. Retrieved on May 12, 2014 from http://www.fma.org/.

[17] J. Mahedy, (2005), **Taking Earnings Quality into Account: A Practitioner's Research into Balance-Sheet Accruals.** Retrieved on December 3, 2007 from http://www.alliancebernstein.com/.

[18] D. P. Samontaray, (2010), Impact of Corporate Governance on the Stock Prices of the Nifty 50 Broad Index Listed Companies. **International Research Journal of Finance and Economics**, Vol. 41, pp. 7-18.

[19] R. Sloan, (1996), Do stock prices fully reflect information in accruals and cash flows about future earnings? **Accounting Review**, Vol. 71, pp. 289-315.

[20] Y.G. Wang, (2011), Corporate Social Responsibility and Stock Performance—Evidence from Taiwan. **Modern Economy**, Vol. 2, No. 5, pp. 788-799.

Managers Capabilities and Competences in Changing Environment: Empirical Study in Latvia

Janis CAUNE, Inga LAPINA, Elina GAILE-SARKANE, Modris OZOLINS and Indra BORKUS
Riga Technical University, Faculty of Engineering Economics and Management
6 Kalnciema Street, Riga, LV-1048, Latvia

ABSTRACT

Last decade of high economic turbulence and significant fluctuations has tested managers' ability to adapt and apply the relevant skillset and competences to ensure success of their companies in various market conditions. The aim of research is to analyse if managerial competencies required to successfully lead a company differ in times of rapid growth or crisis and normal market conditions. Research methods such as review of the literature and organized focus groups with managers of SMEs representing various service and manufacturing industries are applied. As a result, authors developed a customized list of managerial competencies and identified individual importance of different competencies in periods of growth and crisis. Whilst research is limited to SMEs in Latvia, this is a first step in a series of studies to explore managers' capabilities and competence models.

Keywords: Competence, Capabilities, Changing environment.

1. INTRODUCTION

Entrepreneurs` ability to generate new ideas and commercialize them successfully is the main driver of economic growth. This applies to all industries and organizations. Consequently, country`s economic growth depends highly on proper development and use of each entrepreneur`s capabilities and competencies, which is the object of this research.

Country`s economy is constantly changing and recent history of turbulent economic changes in Latvia – rapid growth followed by sudden and steep decline (crisis), brings to the table a wealth of information and experience that has so far been unexplored and undervalued and needs to be accumulated and analysed. This recent experience can form a solid base for future modelling of manager`s competences. In light of economic cycles it is important for companies and their managers to be prepared for the next wave by drawing the appropriate conclusions from the recent experience. However, while the need is evident, there is a clear lack of empirical research to provide companies and managers with understanding and tools to adapt to expected changes.

Authors found ample research indicating existence of significant competence gaps between the competences required and those actually applied in management [27], [15], [35], [26]. Managerial competence gaps are caused by changes in different internal or external aspects (job content, industry requirements, etc.). This supports the notion that during the time of change previous experience and competences are no longer applicable and new competences are needed [26]. However, while competence gaps are analysed authors found no research on the competence gaps in relation to change of economic environment. Similarly, there is ample research on economic cycles, including aspects of performance management. For example, Meyer and Stensaker [23] have explored issue how organizations can develop capacity for change and Chau *et al.* [8] have explored issue of performance management in time of crisis.

Interviews with applicants conducted as part of admissions process for Master of Business Administration studies at Riga Technical University, revealed managers' (i.e., applicants) concern with the need for different skill-set and capabilities to successfully manage a company in severe market conditions, such as rapid growth or decline. These interviews served as the basis for starting the research.

Managerial competencies required to successfully lead a company in crisis or rapid growth environment are different. The individual importance of these competencies also varies according to the market environment. In this paper authors are investigating managers` competencies and changes of their importance in different development stages of economic cycles. For purposes of this research the impact of individual organizational lifecycles that each company passes through alongside more global economic changes, will be omitted. Following the scientific approach of analysis – moving from more general to more particular – we focus solely on analysis of small and medium enterprise (SME) managers` competencies and capabilities in different economic cycles as the first step in the series of research on management of companies in growth and crisis situations.

2. MANAGERS` COMPETENCES, COMPETENCE MODEL AND CAPABILITIES

Many authors across the globe have been writing about the competence theory, yet the development of it has taken different routes in different geographies. In the USA, development of competence theory-based management approach begun in 1976, when McClelland [20] created tests to assess enterprise managers. Boyatzis [4], [5] continued similar studies and empirically determined manager's characteristics. Spencer and Spencer [30] are also among the greatest contributors to the development of the competence theory. In Europe, Mansfield and Mitchell [18] were among the first ones to start the discussion about competences within the context of labour law. Their considerations were based on studies conducted by Frank [14]. Meanwhile, inclusion of competences in management standards had already been described by Cockerill [9] and Mathewman [19].

The concept of competence has been developed further by many well-known scientists and researchers, entrepreneurs and representatives of public organisations. Mulder [24], for example, analysed competence development in organisations and compared over 40 definitions of the concept of competence, distinguishing differences on a number of dimensions.

Armstrong [2] defines the concept of competence, taking into account two aspects: behavioural competences (soft skills),

which are characterised as an opinion on how to act in order to properly perform one's work; and technical and functional competences (hard skills), which must be known and implemented by individuals in order to perform one's work well.

Another frequently used term is competency. United Nations Industrial Development Organization – UNIDO [34] suggests that a competency is a set of skills, related knowledge and attributes that allow an individual to successfully perform a task or an activity within a specific function or job.

Latvian professor Rauhvargers [28] defines competence as totality of knowledge, skills and attitudes, which qualifies one for the performance of tasks of a specific type or level.

The meaning of competence is linked to an individual's ability to learn, communicate and cooperate in a changing environment. Based on this approach and considering the rapidly changing environment, it can be inferred that competence has in fact become one of the key characteristics of a person and thus is also closely related to a company's life cycle [17].

According to the studies done, authors conclude that, when speaking of competence in Latvia, three elements are implied – abilities, knowledge and skills, which are then further split into several groups of elaborately defined competencies.

The development of the competence concept reveals a new research direction – competence models. The link between competence and performance is identified as highly important. In Anglo-Saxon countries the competence models are focused on management competences. The widely accepted definition of the competence model used by managers of Anglo-Saxon countries is the set of desired competencies – skills, knowledge, attitudes, underlying characteristics or behaviour – that differentiate effective performers from ineffective ones [4], [5], [21]. Delamare le Deist and Winterton [10] stressed that every country employs an individual approach to evaluation of the competencies and development of the competence models. According to them, the behavioural approach is the USA tradition, the functional approach is the UK tradition, while France, Germany and Austria belong to countries with a multi-dimensional and holistic approach to evaluation of competencies. We believe the multi-dimensional holistic approach to be the most relevant for Latvia, yet the overwhelming dominance of SMEs in our economy combined with the small internal market results in having some specific features to the competence models used here.

Competence model consists of multiple competence groups, each of which contains numerous competencies – individual skills and abilities. Various sources and studies show different groupings of competencies, forming various competence models suitable for any particular case and situation. For example, I. Lapiņa and D. Aramina [17] have organized competencies into four basic groups, each encompassing qualities defined in greater detail: Professional competencies; Social and communication competencies; Personal and responsibility competencies; Innovative, learning and leadership competencies.

Competencies and competence groups are considered in the context of various sciences. Competence models are widely used in strategic management, human resource management, education etc. In the field of education, for example, competencies are grouped according to three criteria [32]:

– to know and understand (theoretical knowledge, the ability to understand),
– to know how to behave (practical skills and behaviour skills and their use in practice);
– to know how to live (values, attitudes, responsibility).

Another widely used competence model developed by UNIDO [34] suggests that all competencies can be grouped into 3 categories: Managerial; Generic; Technical and Functional. Authors in the field of management services often talk about competence models in the aspect of talent management, leadership and other management areas.

The concept of capabilities is commonly found in literature on competitive advantage, changing environment, assets and resources, processes and activities, learning processes, and specificity and commonality of dynamic capabilities [3], [7], [11], [22], [29], [31], [36].

However, most of the discussion of capabilities focuses around the organizational capabilities rather than manager's competences and abilities in running an organization – which is the focus of our research. Of all the definitions we found, Teece *et al.* [31] and Mulders and Romme [25] come closest to our understanding of capabilities within the context of managers' capabilities.

According to Teece *et al.* [31], the term capabilities emphasizes the key role of strategic management in appropriately adapting, integrating and reconfiguring internal and external organizational skills, resources and functional competences to match the requirements of a changing environment. Meanwhile, Mulders and Romme [25] define dynamic capability as capabilities that convey deliberate knowledge, invoked on a repeated basis, on how to question purpose and effectiveness of the resource base; this deliberate knowledge then serves to generate and modify operating routines and processes to address changing environments and/or create market change.

No single opinion exists about what is a competency or competence neither in general definitions, nor in managing a company or organisation, or in performing job responsibilities.

For the purposes of this research authors define the following concepts: ***capability*** is the ability to use the competence and apply competence models in various situations; ***competence*** (plural – competences) is a set of skills, knowledge and attitude or set of competencies (single – competency) that include also personal behaviour or intent.

3. EMPIRICAL STUDY OF MANAGERS` COMPETENCES IN CHANGING ENVIRONMENT

All market economies periodically move through four development stages of expansion, peak, recession, and recovery. Similar stages can often be observed in an individual company's development. Authors analysed manager's competencies and capabilities in changing environment, focusing on extreme market conditions.

Crisis and rapid growth are considered the extreme phases of a cyclical development. Over the past years, companies in Latvia have experienced both situations: the steep economic growth of the mid-2000s followed by the rapidly evolving global crisis in 2008~2010. Adaptability has been crucial in both situations – the crisis and rapid growth. Yet the question which managerial competencies are necessary to ensure successful company management during each of these periods remains largely unanswered, especially for SMEs.

Statistics illuminate the dominance of medium, small and even micro companies in Latvian economy: according to Eurostat Statistical database [13], SMEs comprise 99,8% of all enterprises in Latvia. Since SMEs form such a crucial part of the economy in Latvia, we limited the scope of our research to small and micro enterprises.

The research methodology is based on research design and methods as developed by Eisenhardt [12] and further described

by Yin [37]. Research is based on eight main steps: defining research question(s), selecting cases, combining qualitative and quantitative data, entering the field, analysing data, extending existing theory, comparing emerging theory with existing literature and closing the research.

In this research, interviews were not conducted as deliberate part of the research, but served more indirectly – to prove the topicality of research before it was undertaken, as authors already mentioned earlier.

We used focus group method in our research for its highly interactive and dynamic character; diversity of opinions present in a focus group leads to a more creative and productive discussion, when compared to individual interviews. Also, Research by Griffin and Hauser [16] shows that one two-hour focus group reveals about the same number of results as two one-hour interviews while Carson *et al.* [6] indicate that "focus groups concentrate clearly on a specific topic and involve interactive discussion among its participants".

The total number of focus groups respondents was 34, aged 24 to 63, both sexes and representing 34 different SMEs covering a variety of service and manufacturing industries. Focus groups were composed in such a way as to ensure maximum diversity of participants from both academic and business environment, and several methods and procedures were applied and tested in order to achieve the planned results of the focus groups.

In total, 3 focus groups of 8–14 persons each (size determined according to the methodology described by Griffin and Hauser [16] were conducted.

We used the following process in conducting focus group discussions: at the beginning, focus group participants were asked to individually name 4–5 managerial competencies of a successful SME manager in each of the competence groups defined by Lapina and Aramina [17] – Professional competencies, Social and communication competencies, Personal and responsibility competencies and Innovative, learning and leadership competencies. These were then combined in a common list, creating certain sets of competencies within each group. As a result of the focus group discussions, 30 competencies that a manager should have to successfully manage an SME in crisis or rapid growth conditions were identified.

Customer need identification and analysis methodology of Quality Function Deployment method [1] was used to gather the results of the focus groups.

To ensure the model reflects managerial competencies, authors renamed some of the groups and relocated some competencies to different competence groups – resulting in a new competence model with the following four competence groups: Professional; Social and communication; Personal, responsibility and leadership; and Innovative and learning competences. As a result, a modified competence model for an "ideal" SME manager was developed, combining all managerial competencies into four competence groups.

Then, participants spent 15–25 minutes individually ranking the competence groups and competencies within the groups according to their importance for Latvian SMEs under regular market conditions. Afterwards, the participants ranked the competencies according to their importance in various market conditions – crisis and rapid growth.

The weight of each specific competency within a competence group is depicted – see Figures 1-4. The indicator is calculated for normal market conditions and various market conditions – crisis and rapid growth. The results of focus groups are described in detail below.

When ranking specific competencies according to their importance in managing a company under regular market conditions within each competence group, members of the focus groups agreed on the following:

- the most important *Professional* competencies are analysing and evaluating, and only then come planning and managing. Presentation skills were named the least important and motivation skills came third from the bottom. However, the ability to persuade and motivate is the third most important in the group of Social and communication competencies;

- in the next group – the *Social and communication* competencies group, the ability to form relationships within and outside the company were identified as the most important ones. Surprisingly, respondents considered language skills to be of minor importance. This, however, could be explained in two different ways: first, knowing at least two foreign languages has now become "a must have" for managers, especially if they operate in the export market, and second, great part of SMEs in Latvia work with local customers and their daily business does not require speaking foreign languages. It is likely that the absence of information technology proficiency on the list can be explained in the same way;
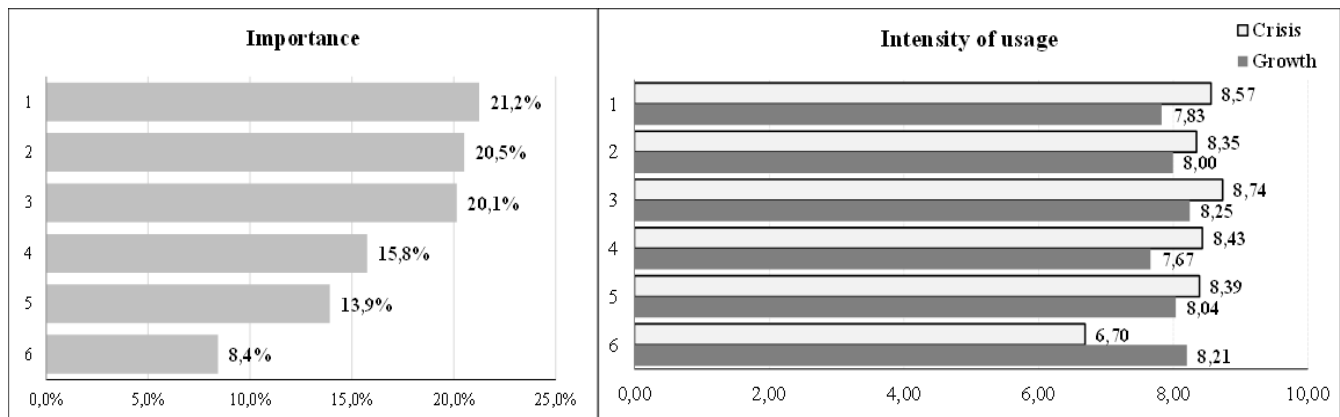
1. Ability to analyse and evaluate; 2. Ability to plan;
3. Ability to manage; 4. Ability to motivate;
5. Ability to react, delegate, and allocate risks; 6. Ability to present

Figure 1. Usage intensity of *Professional* **competencies** and their importance [created by authors]

Importance

1	22,7%
2	22,4%
3	19,5%
4	18,4%
5	10,5%
6	6,5%

Intesity of usage ☐ Crisis ■ Growth

	Crisis	Growth
1	7,96	7,92
2	8,04	8,54
3	8,39	7,83
4	7,61	7,58
5	6,70	7,17
6	6,39	7,58

1. Ability to form relationships within the company;
2. Ability to form relationships outside the company;
3. Ability to persuade and motivate;
4. Ability to form and organize the team work;
5. Ability to find compromise, diplomacy;
6. Ability to communicate in a foreign language
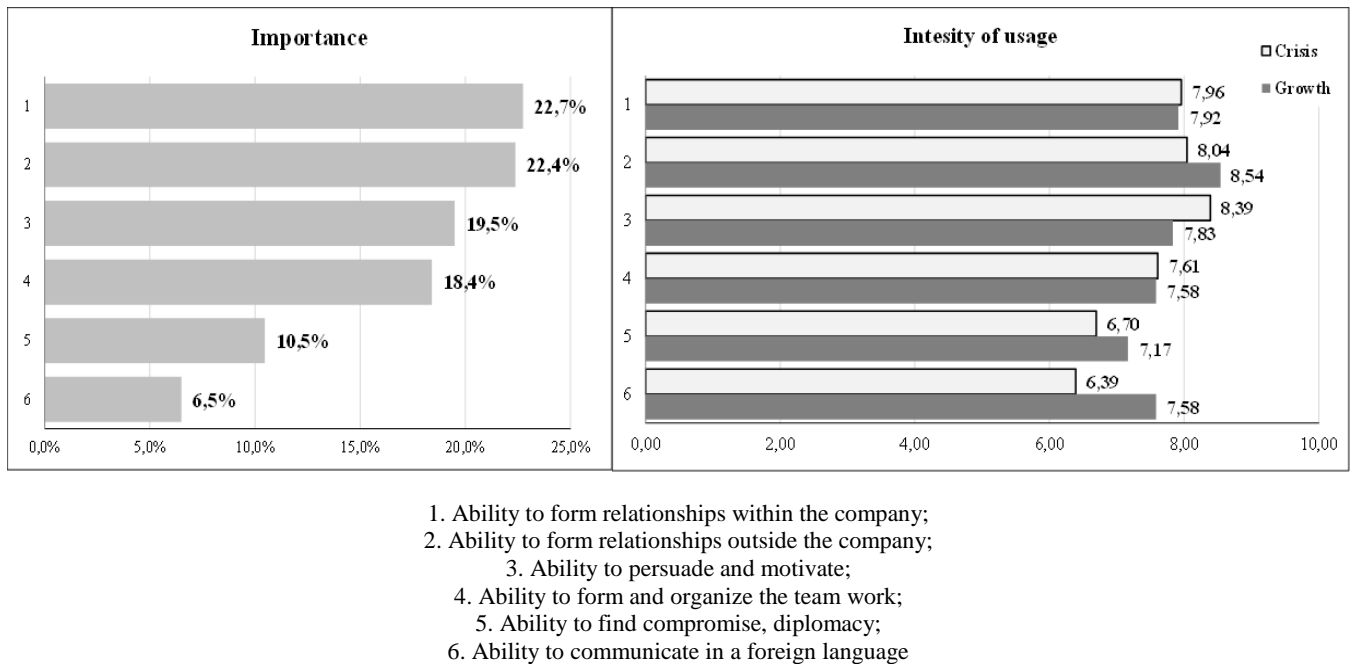
Figure 2. Usage intensity of *Social and communication* **competencies** and their importance [created by authors]

- the most important *Personal and responsibility, leadership* competencies were leadership, self-organisation, self-development and responsibility, while self-criticism was considered of low importance. Since large part of these competencies can be thought of as personal qualities, it is typical to have a wide variety of opinions regarding the importance of any specific competency in this group; the rankings vary significantly while simultaneously attaching similar level of importance to several competencies. This is not

unexpected nor extraordinary, since this competence group is highly dependent on the manager's personality, specific organizational culture, business sector, customers, and other factors;
- creativity and ability to spot and cease opportunities are ranked first and second, respectively, within the group of *Innovative and learning* competencies. The interesting and rarely mentioned ability to notice illogical issues (also errors) came in last.

Importance

1	14,0%
2	12,4%
3	12,4%
4	10,9%
5	9,8%
6	9,5%
7	9,2%
8	9,0%
9	8,7%
10	4,1%

Intensity of usage ☐ Crisis ■ Growth

	Crisis	Growth
1	7,30	7,38
2	7,18	7,33
3	5,96	7,04
4	7,91	7,46
5	7,00	7,42
6	7,91	7,96
7	7,65	7,67
8	8,00	7,88
9	7,77	7,83
10	6,57	6,79

1. Leadership; 2. Self-organisation and self-development; 3. Responsibility;
4. Collaboration; 5. Erudition; 6. Respect; 7. Trustworthiness, loyalty;
8. Intelligence; 9. Intuition; 10. Self-criticism (reasonable)

Figure 3. Usage intensity of *Personal, responsibility and leadership* **competencies** and their importance [created by authors]

1. Ability to be creative; 2. Ability to spot and cease opportunities;
3. Ability to generate ideas; 4. Ability to take risk;
5. Willingness to learn; 6. Ability to promote employee development;
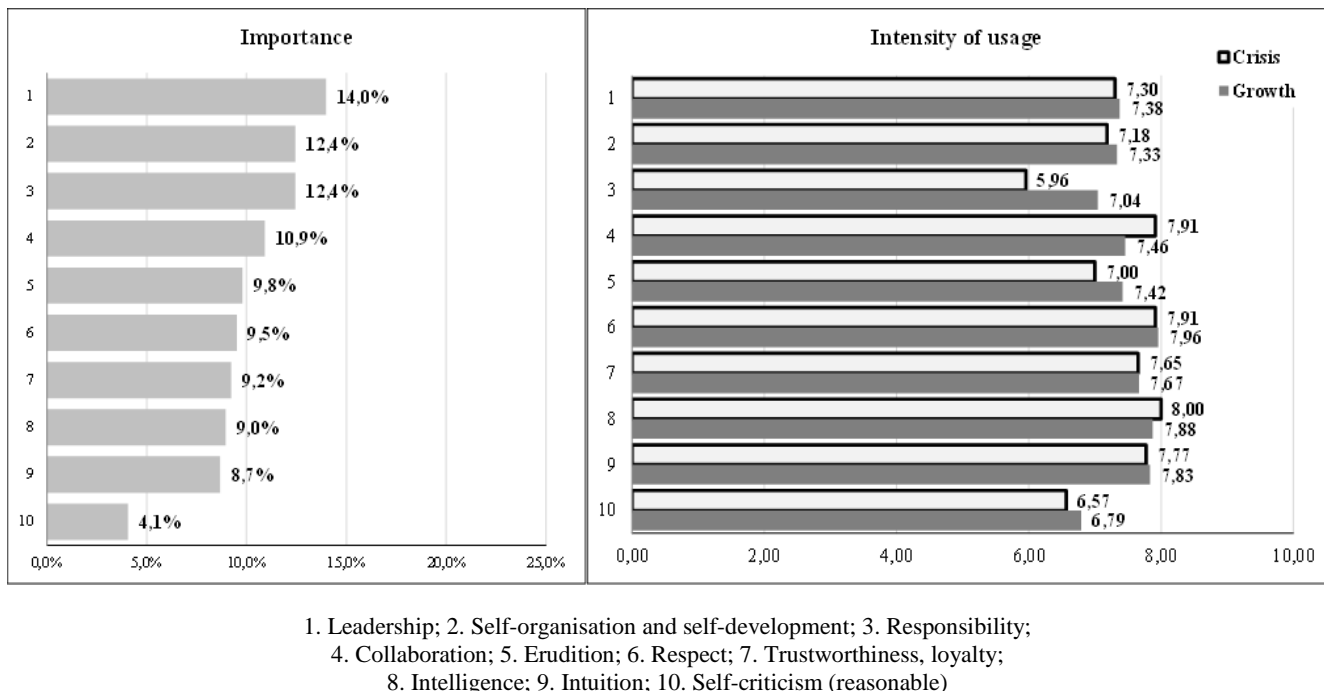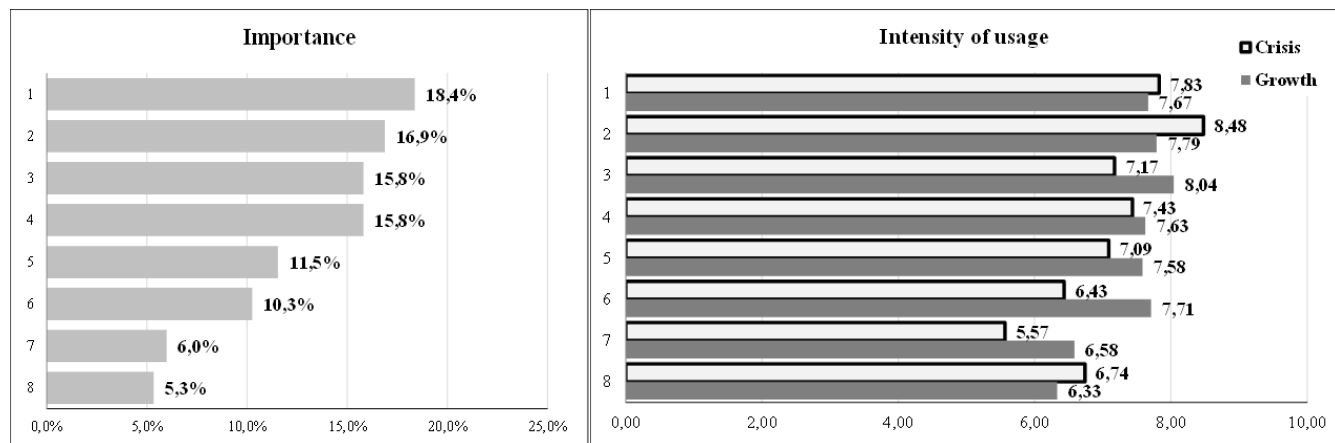7. Ability to pass on knowledge and skills; 8. Ability to notice illogical issues

Figure 4. Usage intensity of *Innovative and learning* **competencies** and their importance [created by authors]

Analysing the arrangement and assessment of individual competencies under extreme market conditions, it can be observed that rapid growth does not ensure a more balanced use of all competencies of a manager, while in crisis predominance of certain competencies can be observed. For example, during crisis the importance of presentation skills reduces, when compared to periods of rapid growth (from 8.21 to 6.7). Significant drop in importance is also observed regarding responsibility (from 7.04 to 5.96). At the same time increased importance of analysis and evaluation abilities (+ 0.74 points), along with the ability to motivate (+ 0.76 points), and ability to spot and cease opportunities (+ 0.69 points) is observed in periods of crisis.

The most important competencies during crisis in each of the groups are ability to manage (among Professional competencies), ability to persuade and motivate (Social and communication competencies), intelligence (Personal and responsibility, leadership competencies group), and ability to spot and cease opportunities (Innovative and learning competences).

In rapid growth circumstances, however, the priority competencies are: ability to manage (within Professional competencies), ability to form relationships outside the company (Social and communication competencies), respect (Personal and responsibility, leadership competencies), and ability to generate ideas (Innovative and learning competencies).

The actual environmental conditions that a manager faces affect his/her ability to use these competencies. This ability to use specific competencies determines the set of capabilities a manager possesses. Thus, the set of capabilities a manager applies in everyday life is much narrower than the set of competencies this same manager actually holds.

competences. They particularly examined differences in competences required during economic downturn and growth.

When analysing the varying importance of competencies in light of different economic conditions, authors concluded that importance of competencies does vary along with economic changes – in times of economic crisis professional competence group plays a more significant role than the other groups, while during periods of growth importance of non-professional competence groups increases more significantly. When exploring in more detail the importance of individual competencies, high variance was observed: i.e., during crisis the importance of presentation skills and responsibility falls (a drop from 8.21 to 6.7 and from 7.04 to 5.96, respectively); while increased importance of analysis and evaluation abilities (+ 0.74 points), the ability to motivate (+ 0.76 points) and ability to spot and cease opportunities (+ 0.69 points) is observed.

The analysis of individual responses showed that importance of certain competences varies depending on the size of the company – even within the same statistical group of SMEs. The research revealed that managers of relatively larger companies – the medium sized enterprises within the SME group – were less involved in daily functional activities and had more space for the actual management function. In this case, the managers were assigning significantly different importance to various competencies in times of crisis and growth. Therefore, we conclude that the competences needed in small and micro enterprises differ from those required in managing medium size companies – and the differences in crucial managerial competences with regard to the company size need further exploration.

Based on the results of this research we are convinced further research on capabilities and competencies in SMEs will allow gaining a more thorough understanding of their role in ensuring sustainability in changing business environment.

4. CONCLUSIONS

Challenges that managers face regarding the competencies and capabilities needed to survive and succeed in a dynamic environment have been investigated in this research. The authors investigated importance of various managerial

5. REFERENCES

[1] American Supplier Institute (ASI). **Quality Function Deployment** (QFD). Kvalitetin for 2000 – talet. Norway: ASI Scandinavia, 1992.

[2] M. Armstrong, **A Handbook of Human Resource Management Practice 10th ed.**, London/Philadelphia: Kogan Page Ltd, 2006.

[3] C. Bowman, V. Ambrosini, "How the Resource-based and the Dynamic Capability Views of the Firm Inform Corporate-level Strategy", **British Journal of Management**, Vol. 14, 2003, pp. 289-303.

[4] R.E. Boyatzis, **The Competent Manager: A Model for Effective Performance**, New York: J. Wiley, 1982.

[5] R.E. Boyatzis, "Competencies in the 21st century", **Journal of Management Development**, Vol. 27, No. 1, 2007, pp. 5–12.

[6] D. Carson, A. Glimore, C. Perry, K. Gronhauigh, **Qualitative Marketing Research**, London: Sage, 2001.

[7] E. Cavusgil, S.H. Seggie, M.B. Talay, "Dynamic Capabilities View: Foundations Research Agenda", **Journal of Marketing Theory and Practice**, Vol. 15, No. 2, 2007, pp. 159-166.

[8] V.S. Chau, H. Thomas, S. Clegg, A. Leung, "Managing Performance in Global Crisis", **British Journal of Management**, Vol. 23, Special Issue, Supplement S1, 2012, pp. S1-S5.

[9] T. Cockerill, "The kind of competence for rapid change", **Personnel Management**, September, 1989, pp. 52–56.

[10] F. Delamare Le Deist, J. Winterton, "What Is Competence?", **Human Resource Development International**, Vol. 8, No. 1, 2005, pp. 27-46.

[11] M. Easterby-Smith, M.A. Lyles, M.A. Peteraf, "Dynamic Capabilities: Current Debates and Future Directions", **British Journal of Management**, Vol. 20, Special Issue, Supplement S1, 2009, pp. S1-S8.

[12] K.M. Eisenhardt, "Building theory from case study research", **Academy of Management Review**, Vol. 14, No. 4, 1989, pp. 532-550.

[13] **Eurostat Statistical database**, Small and medium-sized enterprises, 2013, http://epp.eurostat.ec.europa.eu/portal/page/portal/european_business/special_sbs_topics/small_medium_sized_enterprises_SMEs [electronic source]

[14] E. Frank, "The UK's Management Charter Initiative: the first three years", **Journal of European Industrial Training**, Vol. 17, No. 1, 1991, pp. 9–11.

[15] P. Galagan, "Bridging the Skills Gap - Part I" **Public Manager**, Vol. 38, No. 4, 2010, pp. 61-67.

[16] A. Griffin, J.R. Hauser, "The Voice of the Customer", **Marketing Science**, Vol. 12, No. 1, 1993, pp. 1-27.

[17] I. Lapiņa, D. Aramina, "Changing of Topicality of Human Competencies within Companies' Life Cycle", **The 15th World Multi-Conference on Systemics, Cybernetics and Informatics**, Proceedings, Vol. I, USA, Florida, Orlando, 2011, pp. 106–111.

[18] B. Mansfield, L. Mitchell, **Towards a Competent Workforce**, London: Gower, 1996.

[19] J. Mathewman, "Trends and developments in the use of competency frameworks", **Competency**, Vol. 1995/1, No. 4, 1995, pp. 23-34.

[20] D.C. McClelland, **A Guide to Job Competency Assessment**, Boston: McBer, 1976.

[21] P. McLagan, Great Ideas Revisited. **Training and Development**, Vol. 50, No. 1, 1996, pp. 60–66.

[22] A.G. Menon, "Revisiting Dynamic Capability", **IIMB Management Review**, 2008, pp. 22-33.

[23] C.B. Meyer, I.G. Stensaker, "Developing capacity for change", **Journal of Change Management**, Vol. 6, No. 2, 2006, pp. 217-231.

[24] M. Mulder, "Competence – the essence and use of the concept in ICVT", **European journal of vocational training**, Vol. 40, No. 1, 2007, pp. 5-21.

[25] D.E.M. Mulders, A.G.L. Romme, **Unpacking Dynamic Capability: A Design Perspective**, Chapter 4, in *New Approaches to Organization Design, Information and Organization Design*, Series 8, New York: Springer Science+Business Media, 2009, pp. 61-78.

[26] S.H. Munkeby, **Middle manager competencies required for success in the 21st century**, Dissertation, Arizona: University on Phoenix, 2007.

[27] C. Rappe, T. Zwick, "Developing leadership competence of production unit managers", **Journal of Management Development**, Vol. 26, No. 4, 2007, pp. 312-330.

[28] A. Rauhvargers, "Establishing the European Higher Education Area - progress made in the period of 2005-2007", **Humanities and social Sciences in Latvia**, Vol. 3, No. 52, 2007, pp. 5–35.

[29] G. Schreyögg, M. Kliesch-Eberl "How dynamic can organizational capabilities be? Towards a dual-process model of capability dynamization", **Strategic Management Journal**, Vol. 28, 2007, pp. 913–933.

[30] L.M. Spencer, S.M. Spencer, **Competence at work: Models for Superior Performance,** New York: John Wiley & Sons, 1993.

[31] D.J. Teece, G. Pisano, A. Shuen, "Dynamic Capabilities and Strategic Management", **Strategic Management Journal**, Vol. 18, No. 7, 1997, pp. 509–533.

[32] Tuning Educational Structure in Europe. Pilot Project Final Report, Phase 2, **TUNING Educational Structures in Europe**, 2005, http://www.tuning.unideusto.org/tuningeu/ [electronic source]

[33] Ulrich, K.T. and Eppinger, S.D. (2008). Product Design and Development, 4th edition. New York, NY: McGraw-Hill.

[34] UNIDO competencies, **United Nations Industrial Development Organization** (UNIDO) (2012), www.unido.org [electronic source]

[35] V. Wickramasinghe, N. De Zoyza, "A comparative analysis of managerial competency needs across areas of functional specialization", **Journal of Management Development**, Vol. 28, No. 4, 2009, pp. 344-360.

[36] S.G. Winter, "Understanding dynamic capabilities", **Strategic Management Journal**, Vol. 24, No. 10, 2003, pp. 991–995.

[37] R.K. Yin, **Case Study Research: Design and Methods**. London: SAGE Publications, 2009.

Bank Customer Profitability Management

Jana ERINA
Faculty of Engineering Economics and Management, Riga Technical University
6 Kanciema Str., Riga, LV-1048, Latvia

and

Ingars ERINS
Faculty of Computer Science and Information Technology, Institute of Applied Computer Systems, Riga Technical University
1/4 Meža Str., Riga, LV-1048, Riga

ABSTRACT

The goal of the present article is to investigate and assess the directions of customer profitability management of the Latvian commentarial banks, which are aimed at promoting bank customer centricity.

To achieve the goal, the authors initially conducted content analysis of scientific literature on customer centricity. Based on the results of the content analysis, the following tasks were completed: 1) a survey questionnaire was developed, which was used in polling the customers of Latvian commercial banks with regard to bank customer centricity; 2) a model refocusing product-centric customer segmentation to customer-centric segmentation was elaborated; 3) customer bank account statements were analyzed; 4) customer ecomap was developed.

As a result of the survey conducted, it was concluded that customer centricity of the Latvian commercial banks is at the average level. In turn, the analysis of customer bank account statements, which was focused on customer centricity, demonstrated that bank customers can be classified according to these principles. As a result, three customer segment groups were identified. In the conclusion, the authors suggest elaborating ecomaps for the profitable customer segment group.

Keywords: Customer centricity, Banks, Segmentation, Profitability, Efficiency, Ecomaps, Management

1. INTRODUCTION

Banks play a decisive role in any sector of the national economy. With the increase of global competition, the issues of customer profitability management become more and more topical, which is attested by numerous studies conducted by scholars in many countries throughout the years [1; 2; 3].

Recently, bank customer centricity and customer-centric customer segmentation are mentioned as being among the most essential aspects of customer profitability management, which form the framework of business ecosystem concept. Customer centricity is defined as *"company attitude that helps to shape business processes, which is based on the recognition of the customer needs, thus creating long term financial value for the customers and bank"* [4]. Also Yükçü and Karakelleoğlu (2013) studied the directions of customer profitability management in the context of customer centricity [5].

The goal of this article is to investigate and assess the directions of customer profitability management of the Latvian commentarial banks, which are aimed at promoting bank customer centricity.

To achieve the goal, the authors applied various research methods, such as qualitative and quantitative economic research methods and qualitative and quantitative data processing methods including statistical methods, as well as content analysis.

2. LITERATURE REVIEW

Customer profitability management based on the increase of customer value as a financial asset is mainly influenced by four factors: customer life cycle management, enterprise data base, customer quantitative value, and optimization of customer acquisition and retention costs [3]. Blattberg et al (2001) also stress that customer profitability management is based on mutual relations, however, customer needs differ depending on the time each client remains a customer [3]. In this regard, Blattberg et al (2001) distinguish five life cycle phases: 1) prospects – are not customers yet, but still reflect a potential equity; 2) first time buyers – customers, who purchased one product and service; this phase is characterized by the lowest retention rate; 3) early repeat buyers – customers, who purchased one product and service, they are more sensitive to change than first time buyers; 4) core customers – customers, who do not use products and services regularly any more, yet they are characterized by the highest retention rate and the highest sales rate per customer; 5) core defectors – are ready to switch to rival suppliers [3].

Customer acquisition stage is also considered as a transition to customer retention, which is one of the most important factors for creation of customer value [6], cost saving [7], and relationship building [8]. Considering relationship building stage, Berger and Bechwati (2001) state that in this phase if there are good mutual relationships, customers are willing to purchase additional products and services [9]. In turn, Bolton

(2004) point out that at this stage product and service price level and satisfaction can also positively affect relationships [10]. In the last phase customers may decide to switch to rival suppliers if they are not satisfied [11].

Theory and practice of relationship marketing stress that it is necessary to concentrate on development and maintaining of long-term relations [12]. However, it is important to remember that not all customers can ensure high profitability to an enterprise, therefore, it is necessary to assess customer life time value, in such a way estimating all costs and revenues associated with the customers to be able to calculate current value cash flow [13]. Fader (2012) defines CLV as the future value of the current cash flow, which is associated with a particular client, in such a way identifying the value of each customer and determining the total value of enterprise customer profitability [14]. It also helps to group the customers into segments (profitable and less profitable), improve selling techniques, forecast future tendencies, and use the resources more efficiently.

Customer segmentation, which is based on the life cycle values, can be divided into three categories: segmentation, which is based only on life cycle values; on life cycle value components; on life cycle values and other information [15; 3]. Segmentation focused on life cycle values reflects all existing customers and their relative value in the descending order, the one based on life cycle value components – existing and future value, customer loyalty, in turn, the third category takes into account also demographic data and customer transaction records [16].

There is no consensus among the scientists with regard to the concept of customer segmentation; however, they share the basic perception that it is based on the structuring of the customer base into groups, the participants of which have similar or same characteristics, ensuring effective and unique offer to the customers.

Banks continuously stress that it is necessary to change the existing strategy for a strategy of offering products and services based on customer needs. However, the existing situation in the field of financial services demonstrates opposite tendencies, banks still focus their attention on products rather than customer needs. As stated by Tseng and Piller (2003), banks should adjust products and services to customer needs instead of adjusting the customers to products and services [17]. Caselli (2005) expresses similar opinion, pointing out that banks have focused on the quality of the services offered rather than on customer centricity needs [18].

Customer-centric entrepreneurship provides additional value to the bank [19], it also ensures customer loyalty. Selden and Colvin (2003) in their research discovered that customer centricity does not only ensure profitability [2003], but also increases customer portfolio. In turn, Seybold (1998) considered that higher profitability was ensured by loyal customers [21], whereas in Seybold (2001) – by mutual relations [22]. Bolton (2004) stresses that customer-centric business processes also comprise mutual relations that ensure customer retention, satisfaction and value [10]. Hamilton (2004) has concluded that in the future, while selecting loans or other types of individual financial solutions the customers will ascribe value to trust [23].

In his turn, Fader (2012) discovered that customer centricity is the strategy that is based on reconciling of products and services with the needs of the most profitable customers, and that allows generating higher profit in longer term [14]. Galbraith (2011) has come to a conclusion that customer centricity is not an option, but rather a necessity that is based on building mutual relations [24].

Customer-centric approach includes not only customers and their needs analysis, but also a range of other strategies. For example, Fader (2012) studies also product- and service-centric model and compares it with customer-centric model. Mello (2003) also uses the same approach as Fader (2012) demonstrating that within service-centric model initially certain products are developed and only then customer needs are considered, whereas within customer-centric model products and services are developed based on customer needs [25; 14]. Mello (2003) identified also market- and customer-centric strategy, distinguishing four main stages: preparation for customer visit, customer data processing, customer needs analysis and search for solutions [25].

3. METHODOLOGY

In order to assess customer profitability management directions, which are aimed at promotion of bank customer centricity and segmentation of the existing customers, at the first stage of research the authors developed survey questionnaires and polled the customers of the Latvian commercial banks. In the scientific literature, bank customer centricity in the view of bank customers is referred to as customer centricity index.

The questionnaires used in the research were elaborated based on the research conducted abroad by other authors [26; 27; 28; 29; 30]. The questionnaires consisted of five question blocks: from 1 till 5 (demographic data), from 6 till 20 (home pages, customer visits to bank branches), from 21 till 23 (distributive justice), from 24 till 26 (procedural justice), from 27 till 29 (interactional justice).

Justice is based on the level of customer satisfaction and future behavior, which depends on the consumer perception of the business integrity in its treatment of the customer [31]. Distributive justice derives from social exchange theory, which determines future deals; the customers will be satisfied if the invested capital is equal to the ratio of the source [32]. Procedural justice refers to fairness policy procedures and criteria; customers feel satisfied if in their view a fair decision has been made. Interactional justice shows the attitude of the service providers to the customers and problem solution skills [33].

Measurements were made on the five-point Likert scale (from 1 – strongly disagree till 5 – strongly agree), excluding the questions extracting demographic data.

The research was conducted from August to November, 2013. Customers including both natural and legal persons took part in the survey.

The research results were processed with SPSS statistical data analysis methods and were represented graphically.

At the following stages of research on the basis of content analysis of scientific literature, a model was elaborated that reflects the information on the customer demonstrating whether the bank is centered on products and services or it is customer-centric. Based on the model elaborated by the authors, customer bank account statements of ten commercial banks for the period from 1 November, 2012, till 1 November, 2013 were analyzed considering customer centricity. The analysis of bank customer

account statements was performed using NVivo software, classifying the customers into segments.

In the conclusion, the authors developed ecomaps for the profitable customer group. The essence of ecomap principles lies in representing mutual relations among the stakeholders [34].

4. RESEARCH RESULTS

In total, 187 respondents have taken part in the surveys, 59% are female and 41% – male. 43% of the respondents are aged 20-29; 36% are aged 30-39 and 21% are aged above 40. 61% of the respondents have higher education, 20% hold Master Degree, and 19% have secondary education. The responses to the questions regarding bank home page information and bank branch visit experience are summarized in Table 1.

Table 1

Customer-centric Index Measurement at the Banks by Customers

Customer-centric index metrics	Natural persons	Legal persons
Home pages, customer visits to bank branches	3.21	3.31
ATM availability	4.02	3.65
Bank location availability	3.87	3.93
Waiting time at the bank	2.67	2.98
Convenience at the bank	2.78	3.03
The bank knows my needs	2.12	2.84
Call handling quality	4.04	3.89
Transaction quality and speed	4.23	4.08
Home page quality	4.21	4.30
Checking costs	2.68	2.91
Home page information is easy to understand	3.11	3.19
Transparency	2.56	2.77
Home page search	2.94	3.01
Home page navigation	2.86	3.03
Home page speed	4.13	4.17
Other customer thoughts and discussions	2.02	2.01
Distributive justice	2.76	3.00
The outcome I have received is fair	2.25	2.67
I have not obtained what I expected	3.85	3.91
Provision of the required products and services	2.19	2.43
Procedural justice	3.18	3.17
The problem was resolved correctly	2.34	2.57
Customer complaints are not considered promptly enough	4.21	3.93
The bank can be flexible in solving the problems	2.99	3.02
Interactional justice	2.76	2.83
Employees focused on my problem	2.96	3.07
Employees did not strive to solve my problem	2.87	2.91
Employees did not show any interest in trying to solve my problem	2.45	2.52
Total customer centricity index	2.97	3.07

Considering the data summarized in Table 1 on home pages and customer visits to bank branches, it can be noticed that in the

opinion of natural persons, the most important factors are ATM availability, call handling quality, transaction quality and speed, as well as home pages quality and speed. At the same time, legal persons ascribe the highest value to transaction quality and speed and home page quality.

The summarized data on distributive justice demonstrates that in the opinion of bank customers (natural persons and legal persons), they did not receive the expected results from the bank (Q22). The respondent rating of Q21 is 2.25 (natural persons) and 2.67 (legal persons), but the rating of Q23 is 2.19 (natural persons) and 2.43 (legal persons), which is a sign of low rating. In this regard, the banks should consider the questions that are connected with customer perception of distributive justice.

The summarized answers of respondents regarding procedural justice demonstrate that the banks do not consider complaints filed by the customers promptly enough – 4.21 (natural persons) and 3.93 (legal persons). Bank customers also consider that the problem that had been solved were not the ones they wanted to be addressed (Q24). Bank flexibility concerning solution of various issues was evaluated with average rates (Q26 – from 2.99 till 3.02).

Considering the results summarized on interactional justice it can be noticed that respondents gave very low rating to almost all questions (lower than 3 according to the Likert scale), with the exception of Q27 rated by legal persons at 3.07.

Survey results regarding bank customer centricity demonstrated that customer centricity of the Latvian commercial banks is on the average level (3.02 according to the Likert scale), thus reflecting customer profitability management problems persisting at the banks analyzed. Therefore, considering the problems persisting at the commercial banks, and based on scientists abroad [35; 16; 36; 37], the authors developed a scheme demonstrating the transfer from bank product and service centricity to customer-centric segmentation, as it is shown in Figure 1.
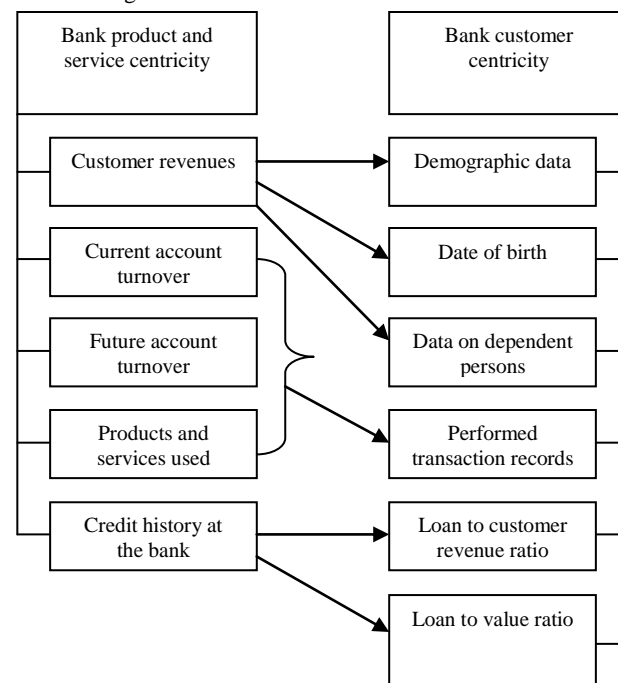
Figure 1 Customer evaluation criteria, transition from the existing centricity on products and services to customer centricity

Based on the analysis of the scientific literature and Figure 1, the author analyzed ten natural person customer account statements from several banks for the last year (from 1.11.2012. till 1.11.2013.) using NVivo software. The data obtained are summarized in Table 2.

Table 2

Analysis of customer bank account statements of natural persons

No	Demographic data	Transaction description
1.	Woman, 37 years old	Monthly wage income (1000.00 EUR) Cash withdrawals <u>Standing orders</u>: Loan repayment; Transfer of life insurance premiums; Utility bills; Service bills (Latvenergo, Latvijas gāze, LMT, Internets). <u>Irregular payments</u>: Trip once a year (airplane ticket purchase, hotel bills); Payments for goods at non-food stores (mainly in August, September and December); Payments for beauty services; Payments for purchases at food stores.
2.	Woman, 23 years old	Monthly wage income (EUR 426.00) <u>Standing orders</u>: Internet bills; <u>Irregular payments</u>: Payments for goods at food and non-food stores; Payments for purchases at leisure facilities.
3.	Man, 31 years old	Monthly wage income (EUR 1138.00) Cash withdrawals <u>Standing orders</u>: Car leasing; Utility bills; Service bills (Latvenergo, Internets). <u>Irregular payments</u>: Fuel purchase; Payments for goods at food and non-food stores; Payment for purchases at restaurants and cafes; Airplane ticket purchase; Hotel booking and bill payments; Annual transfer of a life insurance premium.
4.	Woman, 52 years old	Monthly wage income (EUR 398.00) Monthly cash withdrawal. <u>Irregular payments</u>: Payments for goods at food stores.
5.	Woman, 28 years old	Monthly wage income (EUR 569.00) Cash withdrawals <u>Standing orders</u>: Utility bills; Service bills (Latvenergo, Bite). <u>Irregular payments</u>: Fuel purchase; Payments for goods at food and non-food stores; Airplane ticket purchase.
6.	Man, 68 years old	State pension (EUR 355.00) Monthly cash withdrawal
7.	Man, 43 years old	Monthly wage income (EUR 1200.00) Cash withdrawals <u>Standing orders</u>: Loan repayment; Car leasing. <u>Irregular payments</u>: Payments for goods at food and non-food stores.
8.	Man, 63 years old	Monthly wage income (EUR 284.00) State pension (EUR 199.00) Monthly cash withdrawal. <u>Irregular payments</u>: Payments for goods at food stores.
9.	Woman, 35 years old	Monthly wage income (EUR 640.00) Cash withdrawals <u>Standing orders</u>: Loan repayment; Utility bills; Latvenergo service bills. <u>Irregular payments</u>: Payments for goods at food and non-food stores; Hotel booking and bill payments.
10.	Woman, 19 years old	Scholarship (EUR 99.00) <u>Irregular payments</u>: Payments for goods at food and non-food stores.

The data summarized in Table 1 demonstrate that it is possible to distribute bank customers into several segments according to their current account turnover based on customer centricity. Thus, based on the analysis conducted, the authors distributed bank customers into three segment groups: most profitable customers, not profitable so far, but in the future can become profitable, and non-profitable customers. This kind of typology was also used by other scientists, such as Global Information & Research Adkit (2011), Rust et al. (2000) and Fader (2012) [35; 38; 14].

Based on the research results presented in Table 2, at the next stage of the research the authors elaborated an ecomap for the profitable customer segment. The principles of ecomap introduction at the bank are shown in Figure 2.

The authors conclude that the changes in customer segmentation should be introduced in grouping the customers – natural persons, but it can be also applied regarding legal persons. Introducing and using ecomap principles in their daily work to evaluate the most profitable customers, banks will gain stable customer equity as well as will better understand their customer needs.
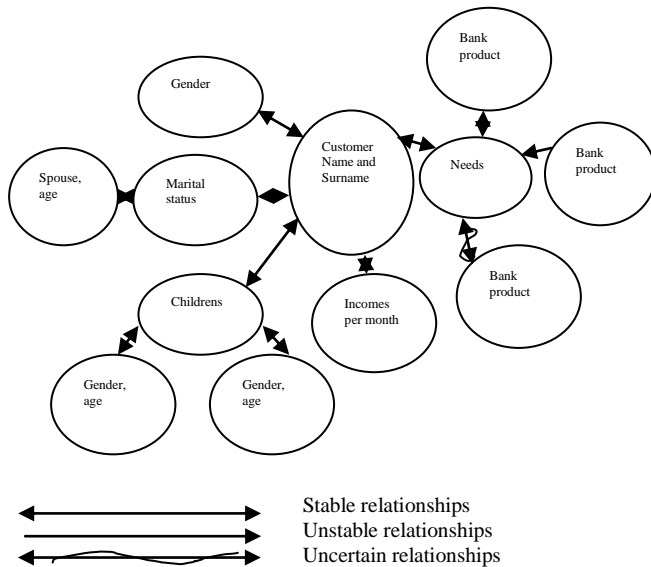
Figure 2 Example of bank customer ecomap

5. CONCLUSIONS

Conducting the research on customer profitability management at the Latvian commercial banks, the authors mainly considered the direction aimed at promoting bank customer centricity and customer-centric customer segmentation.

Using the questionnaires elaborated by the authors concerning bank customer centricity, the customers of the Latvian commercial banks – both natural and legal persons – were polled. Having conducted the survey, it has been concluded that the banks demonstrate an average level of customer centricity, as well as customer profitability management.

The next stage of the research that comprised customer bank account statement analysis considering customer centricity has demonstrated that it is possible to classify bank customers into three groups according to the level of profitability. It has also been concluded that customers, who use similar products and services as the customers analyzed, can be included into the following segment groups: most profitable customers, not profitable so far, but in the future can become profitable, non-profitable customers.

In order to ensure that the banks conduct their activities on the basis of customer-centric customer profitability management, the authors propose elaborating ecomaps for the profitable customer group. Using ecomaps, the banks will be able to understand their customer needs as comprehensively as possible, in such a way maximizing not only customer, but also bank profitability.

6. REFERENCES

[1] G. Foster, M. Gupta, and L. Sjoblom, "Customer profitability analysis: Challenges and new directions", **Journal of Cost Management (Spring)**, Vol. 5, No. 17, 1996.

[2] G. Foster, and S. M. Young, "Frontiers of management accounting research", **Journal of Management Accounting Research**, Vol. 9, 1997, pp. 63-77.

[3] R. Blattberg, G. Getz, and J. Thomas, **Customer Equity: Building and Managing Relationships as Valuable Assets**, Harvard Business School Press, 2001.

[4] J. Eriņa, & I. Eriņš, "Assessment of Bank Customer Equity" **The 17th World Multi-Conference on Systemics, Cybernetics and Informatics**, Vol. 2, 2013, pp. 7-12.

[5] S. YÜKÇÜ, & İ KARAKELLEOĞLU, "Customer-driven Profitability Analysis: a Management Accounting Approach Through Customer Portfolio Applications, **World of Accounting Science**, Vol. 15, No. 1, 2013.

[6] A. W. Grant, & L. A. Schlesinger, " Realize your customers' full profit potential", **Harvard Business Review**, Vol. 73, No. 5, 1995, pp. 59-72.

[7] F. F. Reichheld, & W. E. Sasser, "Zero defections: quality comes to services", Harvard **business review**, Vol. 68, No. 5, 1990, pp. 105-111.

[8] F. J. Mulhern, "Customer profitability analysis: Measurement, concentration, and research directions", **Journal of interactive marketing,** Vol. 13, No. 1, 1999, pp. 25-40.

[9] P. D. Berger, & N. N. Bechwati, "The allocation of promotion budget to maximize customer equity", **Omega**, Vol. 29, No. 1, 2001, pp. 49-61.

[10] M. Bolton, "Customer centric business processing", **International Journal of Productivity and Performance Management**, Vol. 53, No. 1, 2004, pp. 44-51.

[11] R. T. Rust, & A. J. Zahorik, "Customer satisfaction, customer retention, and market share", **Journal of retailing**, Vol. 69, No. 2, 1993, pp. 193-215.

[12] R. M. Morgan, & S. D. Hunt, "The commitment-trust theory of relationship marketing", **The journal of marketing**, 1994, pp. 20-38

[13] P. D. Berger, & N. I. Nasr, "Customer lifetime value: marketing models and applications", **Journal of interactive marketing**, Vol. 12, No. 1, 1998, pp. 17-30.

[14] P. Fader, **Customer Centricity: Focus on the Right Customers for Strategic Advantage**, Wharton Digital Press, 2012.

[15] C. C. H. Chan, "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer", **Expert Systems with Applications**, Vol. 34, No. 4, 2008, pp. 2754-2762.

[16] S. Y.Kim, T. S. Jung, E. H. Suh, & H. S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study", **Expert Systems with Applications**, Vol. 31, No. 1, 2006, pp. 101-107.

[17] M. M. Tseng, & F. T. Piller, **The customer centric enterprise**, Springer Berlin Heidelberg, pp. 3-16, 2003.

[18] S. Caselli, **Corporate banking strategies: products, markets and channels**, In Strategy and Organization of Corporate Banking, Springer Berlin Heidelberg, pp. 37-62, 2005.

[19] F. T. Piller, & K. Moeslein, "From economies of scale towards economies of customer integration", **Technical University of Munich, Department of General and Industrial Management**, 2002, pp. 0942-5098.

[20] L. Selden, & G. Colvin, **Turn Unprofitable Customers into Profitable or Former**, The American Banker, 2003.

[21] P. B. Seybold, & R. T. Marshak, **Customers. com: how to create a profitable business strategy for the Internet and beyond**, Times Business, 1998.

[22] P. B. Seybold, R. T. Marshak, & J. M. Lewis, **The customer revolution**, New York, NY: Random House, 2001.

[23] B. A. Hamilton, **The Customer-Centric Organization**, Booz Allen Hamilton, 2004.

[24] J. R. Galbraith, **Designing the customer-centric organization: A guide to strategy, structure, and process**, Jossey-Bass, 2011.

[25] S. Mello, **Customer-Centric Product Definition: The Key to Great Product**, PDC Professional Publishing, 2003.

[26] Customer Carewords. Retrieved from http://www.customercarewords.com/customer-centric-index.html

[27] M. Deutsch, **Distributive justice: A social-psychological perspective**, New Haven, CT: Yale University Press, 1985, pp. 122-155.

[28] J. A. Colquitt, & J. B. Rodell, "Justice, trust, and trustworthiness: A longitudinal analysis integrating three theoretical perspectives", **Academy of Management Journal**, Vol. 54, No. 6, 2011, pp. 1183-1206.

[29] H. Nadiri, & C. Tanova, "An investigation of the role of justice in turnover intentions, job satisfaction, and organizational citizenship behavior in hospitality industry", **International Journal of Hospitality Management**, Vol. 29, No. 1, 2010, pp. 33-41.

[30] R. Searle, D. N. Den Hartog, A. Weibel, N. Gillespie, F. Six, T. Hatzakis, & D. Skinner, "Trust in the employer: the role of high-involvement work practices and procedural justice in European organizations" **The International Journal of Human Resource Management**, Vol. 22, No. 05, 2011, pp. 1069-1092.

[31] J. R. McColl-Kennedy, & B. A. Sparks, "Application of fairness theory to service failures and service recovery", **Journal of Service Research**, Vol. 5, No. 3, 2003, pp. 251-266.

[32] J. S. Adams, "Towards an understanding of inequity", **The Journal of Abnormal and Social Psychology**, Vol. 67, No. 5, 1963, pp. 422.

[33] J. G. Blodgett, D. J. Hill, & S. S. Tax, "The effects of distributive, procedural, and interactional justice on postcomplaint behavior", **Journal of Retailing**, Vol. 73, No. 2, 1997, pp. 185-210.

[34] A. Hartman, "Diagrammatic assessment of family relationships", **Families in Society**, Vol. 76, No. 2, 1995, pp. 111-122.

[35] Global Information & Research Adkit. (2011). Retrieved from http://thefinanser.co.uk/files/adkit---retail-customer-segmentation-in-worldwide-banking-1.pdf

[36] K. Tsiptsis, & A. Chorianopoulos, Data **mining techniques in CRM: inside customer segmentation**, John Wiley & Sons, 2011.

[37] A. Machauer, & S. Morgner, "Segmentation of bank customers by expected benefits and attitudes", **International Journal of Bank Marketing**, Vol. 19, No. 1, 2001, pp. 6-18.

[38] R. T. Rust, V. A. Zeithaml, & K. N. Lemon, **Driving Customer Equity: How Customer Lifetime Value Is Reshaping Corporate Strategy**, 2000.

Evaluation of Latvian Commercial Banks Performance

Ingars ERINS
Faculty of Computer Science and Information Technology, Institute of Applied Computer Systems, Riga Technical University
1/4 Meža Str., Riga, LV-1048, Riga

and

Jana ERINA
Faculty of Engineering Economics and Management, Riga Technical University
6 Kanciema Str., Riga, LV-1048, Latvia

ABSTRACT

The aim of this article is to assess performance efficiency of the Latvian commercial banks in the period from 2006 till 2012.

In order to achieve research results, the authors initially conducted content analysis of the scientific literature on the most frequently used methods of bank performance efficiency assessment. Having performed content analysis, it has been concluded that in the scientific literature bank performance is estimated using parametric and non-parametric methods. One of the most popular methods is Data Envelopment Analysis (DEA). Therefore, the authors analyzed the performance efficiency of the Latvian commercial banks on the basis of the data from Bankscope data base, primarily having determined bank financial indicators that are most frequently analyzed using DEA.

Assessing performance efficiency of the Latvian commercial banks, the authors have come to the conclusion that frequently large banks are comparatively a lot more efficient than small banks. One of the main reasons might be the fact that large banks have lower administrative expenses, and that gives them the opportunity to increase performance efficiency and competitiveness in comparison with small banks. The data obtained demonstrate that small banks need to look for new solutions to improve their performance efficiency.

Keywords: Banks, DEA, Performance, Latvia, CCR, BCC

1. INTRODUCTION

Banking system plays a vital role in the development of the national economy of any country; therefore, economic development of each country can be ensured providing the savings are effectively channeled for investment. Topicality of the transition economy is also attested by strategic investors, who pay more and more attention to bank performance efficiency and the measures to improve it [1]. Thus, the issues of bank performance efficiency are important considering not only macroeconomic, but also microeconomic aspects.

For example, Lensink and Hermes (2004) in their research discovered that the entrance of foreign banks to the local market is highly dependent on the level of development of the national economy of each country in general and its banking sector in particular [2]. Rapid development of the banking sector of the European countries, which started after the expansion of the EU, proves the validity of this statement.

Bank performance efficiency assessment methods can be conditionally divided into two categories: parametric and no-parametric methods, which are based on Stochastic Frontier Approach (SFA) and Distribution Free Approach (DFA).

Data Envelopment Analysis (DEA) is a deterministic methodology, which allows determining relative performance that is based on input and output data. It is also called Decision Making Unit (DMU).

The research conducted and the available literature on bank performance efficiency assessment is mainly dedicated to the analysis of the USA and Asian country banks, comparatively less attention has been paid to the European country banks, and there are only few studies dedicated to the Latvian banks.

Using DEA method in the analysis of different countries it has been discovered that foreign banks work with lower performance efficiency than local banks [3; 4; 5]. In turn, the studies of other scientists demonstrate the opposite: foreign banks are more efficient than the local.

The goal of this article is to evaluate performance efficiency of the Latvian commercial banks in the reporting period of 2006 – 2012.

To achieve the goal, the following research methods were used: qualitative overview of the scientific literature on DEA, including the method monographic and descriptive method, as well as deterministic, nonparametric approach.

2. LITERATURE REVIEW

The issues of bank performance efficiency assessment not only form an inherent part of the daily operations of financial institutions, they also become particularly topical in the period of transition economy or on the onset of the financial crisis.

Bank performance efficiency assessment methods discussed in the scientific literature are mainly based on the assumption that all banks conduct their activities according to the same model, disregarding a range of influencing factors, such as economic

situation in a country and the market system within which the banks operate. For example, Dietsch and Lozano-Vivas (2000) consider that not including specific indicators of each country into the assessment of bank performance efficiency may give erroneous results [6]. Grigorian and Manole (2002) expressed similar views, stressing that comparing the performance efficiency of the banks across different states specific environment indicators of each country should be taken into account [2].

One of the first studies on non-parametric performance efficiency assessment method was performed in 1957. In this study technical production performance efficiency was determined using one input and output variable [7]. Slightly later this method was elaborated by Banker et al. (1984) and by Färe et al. (1985) [8; 9]. Only in 1978 Charnes et al. named the previously used non-parametric method such as CCR model [10]. The model of Banker et al. (1984) is called BCC in the scientific literature [8].

The principle of the CCR model is based on the performance efficiency assessment concept elaborated by Farrell (1957), which uses either many input variables and just one output variable, or many input and output variables [7]. Charnes et al. (1978) have applied linear combinations in this model, assuming that yield value is fixed [10]. BCC model, in its turn, is based on variable output indicators on the scale. The relative DMU performance efficiency mathematically is expressed as the ratio of the sum of the output variables and the sum of the input variables.

Rossi et al. (2005) analyzed performance efficiency of the banks of nine CEE countries in the period from 1995 to 2002 using the stochastic frontier analysis [11]. The obtained data demonstrated that banks with low cost performance efficiency show higher profitability. In contrast, Semih Yildirim and Philippatos (2007), Rossi et al. (2005) discovered that there is no strong correlation between performance efficiency and loan quality; however, performance efficiency is indeed influenced by external factors [12; 11].

Weill (2007) performing research from 1996 to 2000 on the banks of CEE and Western European countries using stochastic frontier approach has come to the conclusion that there are differences in performance efficiency of Western and CEE countries banks [13]. For example, the banks of the Czech Republic and Hungary are as efficient as the banks of Western European states, except the banks in Greece and Portugal. A more significant increase in performance efficiency has been observed in the banks of the CEE states rather than in the Western European banks. Differences in performance efficiency of the analyzed countries are explained by the fact that CEE banks have been influenced by the transition economy as well as government performance.

Semih Yildirim and Philippatos (2007) having analyzed performance efficiency of the banks of 12 CEE countries in the period from 1993 to 2000 have concluded that the banks of Poland and Slovenia operate most efficiently, whereas the banks of Russia and the Baltic States (Latvia, Lithuania and Estonia) are least efficient [12]. The authors have also discovered that higher performance efficiency is characteristic of larger banks and the banks that mainly concentrate on deposits; large proportion of bad loans and high concentration of banks decease performance efficiency; higher performance efficiency is demonstrated by foreign banks rather than local [12]. Fang et al

(2011) have come to a similar conclusion that foreign banks work by far more efficiently than the local banks [14].

In the later studies on the CEE countries banks, it has been discovered that the banking sectors of the Czech Republic and Romania work most efficiently, and that performance efficiency has grown since these countries joined the EU [15].

Having analyzed scientific papers from several data bases on the research on bank performance efficiency assessment applying the DEA methodology within the period of 2005 – 2012, which was conducted in the USA and Asian countries, the authors have learned that the employees and capital are considered as input variables, while loans, profitability, and related revenues – as output variables [16; 17; 18; 19; 20; 21]. But having summarized the studies on the European states banks within the period of 2005 – 2012, the authors of the paper have noticed that labor and capital are most frequently used as the input variables, while deposits and net loans – as the output variables [22; 23; 24; 25; 12; 26].

Development of the Latvian Commercial Banking Sector

The development of the financial system in Latvia started in 1988, when the banking sector was reorganized [27]. A new dual financial system and the re-establishment of Latvia's independence promoted rapid development of the banking sector. Starting with 1992 till 1993, 61 banks in Latvia received a license for provision of financial services [28].

But according to the data of the Association of Commercial Banks of Latvia, in 4th quarter of 2012 in Latvia banking services were provided by 20 banks and 9 branches of foreign banks, as well as lending institutions or their branches registered in the countries of the European Economic Area, which submitted a respective application to the Financial and Capital Market Commission [28].

Assessing the development of the Latvian banking sector, the authors have concluded that the following financial performance criteria are considered as being the most important: changes in product and service sales volume, operating profit and revenues, as well as loan repayment rate and potential profit per share.

3. METHODOLOGY

Taking into consideration the results of the scientific paper analysis using NVivo software and the data available, in order to assess the performance efficiency of the Latvian commercial banks the authors have used two different approaches.

The 1st approach is based on two inputs (x_1, x_2) and two outputs (y_1, y_2), as it can be seen in Figure 1.
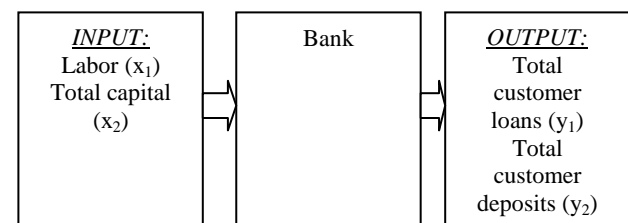
Figure 1 Variables for bank cost performance efficiency assessment (1st approach)

The 2nd approach is based on Rossi et al. (2005). The authors replaced the number of bank employees with total assets of each bank in the reporting period (see Figure 2).
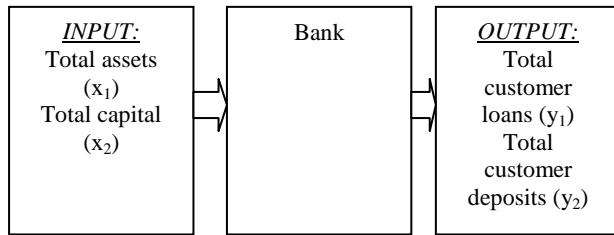
Figure 2 Variables for cost bank performance efficiency assessment (2nd approach)

To assess performance efficiency of the banks, the authors have employed the most frequently used models – CCR DEA and BCC DEA input and output oriented. The authors have used the following formula to calculate the overall bank performance efficiency rate (see formula 1):

$$Total\ bank\ performance\ efficiency = \frac{CCR\ Efficiency}{BCC\ Efficiency} \quad (1)$$

In order to determine performance efficiency of the Latvian commercial banks, the data extracted from home pages of the Latvian commercial banks, the Association of Commercial Banks of Latvia and Bankscope database for the time period from 2006 to 2012 have been used.

Having studied the scientific literature on bank performance efficiency, it can be stated that banks are considered efficiently operating if the performance efficiency index exceeds or is equal to 1, and the banks are considered inefficient, if the performance efficiency index is below 1.

4. RESEARCH RESULTS

To analyze performance efficiency of the commercial banks employing the first approach, the authors have used the data from eight commercial banks, which provided all the information necessary for analysis. Performance efficiency indicators of the analyzed Latvian commercial banks estimated according to CCR model approach are summarized in Table 1.

Table 1

Performance Efficiency of the Latvian Commercial Banks Based on CCR Model (1st approach)

Banks	Efficiency by years							
	2006	2007	2008	2009	2010	2011	2012	Mean
SEB bank	1.000	0.937	0.877	0.823	0.738	0.798	0.758	0.847
ABLV Bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DNB bank	1.000	0.928	0.820	0.989	0.967	1.000	1.000	0.958
PrivatBank	1.000	0.729	1.000	1.000	0.755	0.862	1.000	0.907
ALTUM	0.719	N/a	1.000	0.696	0.589	0.710	0.276	0.665
Trust Commercial Bank	1.000	0.793	0.556	0.577	0.424	0.472	0.462	0.612
Baltic International Bank	0.526	0.844	1.000	0.846	0.451	0.485	0.636	0.684
SMP Bank	0.361	0.557	0.841	0.699	0.830	1.000	1.000	0.755

Performance efficiency analysis data summarized in Table 1 show that in 2006 highest performance efficiency was demonstrated by the banks – market leaders (based on the market place (MP): SEB bank – 16%; ABLV Bank – 5.1%; DNB bank – 8.3%) and relatively smaller banks (PrivatBank – 0.7% MP and Trust Commercial Bank – 1.7% MP). In 2007 as throughout the entire research period high performance efficiency was demonstrated by ABLV Bank, whereas other commercial banks showed low performance efficiency. In 2011

and 2012, DNB bank showed high performance efficiency indicator, similar situation could be observed at SMP bank (market share -0.2% MP). PrivatBank demonstrated the highest performance efficiency from 2008 to 2009, and in 2012. ALTUM (4.3% MP) showed low performance efficiency in the entire research period, except year 2008.

Performance efficiency of the Latvian commercial banks according to BCC model approach is shown in Table 2.

Table 2

Performance Efficiency of the Latvian Commercial Banks Based on BCC Model (1st approach)

Banks	Efficiency by years							
	2006	2007	2008	2009	2010	2011	2012	Mean
SEB bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ABLV Bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DNB bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PrivatBank	1.000	0.941	1.000	1.000	0.821	0.941	1.000	0.957
ALTUM	0.746	N/a	1.000	0.706	0.589	0.710	0.366	0.686
Trust Commercial Bank	1.000	0.978	0.895	0.861	0.526	0.796	0.720	0.825
Baltic International Bank	0.878	1.000	1.000	1.000	0.656	0.911	0.804	0.893
SMP Bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The results of the analysis summarized in Table 2 demonstrate that using BCC performance efficiency model, the most efficient performance was demonstrated by the following commercial banks: SEB bank, ABLV Bank, DNB bank and

Analyzing performance efficiency indicators of commercial banks according to the second approach, the authors used the data obtained from fourteen Latvian commercial banks to perform the estimates.

SMP Bank. At the same time, PrivatBank, ALTUM, Trust Commercial Bank, Baltic International Bank (0.5% MP). Performance efficiency problems were identified in the performance of other banks (research period 2006, from 2010 till 2012).

In order to assess performance efficiency of the Latvian commercial banks, the authors have initially analyzed the input and output variables applying the CCR model (see Table 3).

Table 3

Performance Efficiency of the Latvian Commercial Banks Based on CCR Model (2nd approach)

Banks	Efficiency by years							
	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**Mean**
Swedbank	1.000	1.000	0.957	0.939	0.904	0.970	0.877	0.950
SEB bank	1.000	0.989	0.946	0.965	0.901	0.930	1.000	0.961
ABLV Bank	1.000	1.000	0.985	1.000	1.000	1.000	0.985	0.996
DNB bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Rietumu Bank	0.914	0.899	0.842	0.891	0.992	0.950	1.000	0.927
Norvik Bank	0.914	0.897	0.821	1.000	0.962	0.946	1.000	0.934
PrivatBank	1.000	0.977	1.000	1.000	0.896	0.951	0.968	0.970
ALTUM	0.894	0.993	1.000	0.902	1.000	1.000	0.490	0.897
Baltikums Bank	0.720	0.828	0.813	0.828	0.873	0.975	0.993	0.861
Trust Commercial Bank	1.000	0.903	0.863	0.926	0.898	0.908	0.920	0.917
Expobank	0.985	1.000	0.966	0.964	1.000	0.897	0.878	0.956
Baltic International Bank	0.947	0.910	1.000	1.000	0.903	0.939	0.969	0.953
SMP Bank	0.981	0.996	0.968	0.999	0.974	1.000	1.000	0.988
Bank M2M Europe	0.751	0.750	0.662	1.000	0.954	0.050	0.048	0.602

Having employed the second approach to commercial bank performance efficiency assessment, the authors concluded that in 2006 highest efficiency was demonstrated by the banks – market leaders (Swedbank – 24% MP, SEB bank, ABLV Bank and DNB bank) and smaller banks (PrivatBank and Trust Commercial Bank). However, the situation started to change in 2007, when performance efficiency of SEB bank, PrivatBank and Trust Commercial Bank dropped, the improvement was

observed at Expobank (1.7% MP). Stable high performance efficiency in the period from 2008 to 2012 was maintained by DNB bank, whereas other banks went through the growing and falling stages.

The findings obtained using the DEA BCC model approach are summarized and presented in Table 4.

Table 4

Performance Efficiency of the Latvian Commercial Banks Based on BCC Model (2nd approach)

Banks	Efficiency by years							
	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**Mean**
Swedbank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SEB Bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ABLV Bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DNB Bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Rietumu Bank	1.000	1.000	1.000	0.891	1.000	1.000	1.000	0.984
Norvik bank	0.915	0.897	0.893	1.000	0.985	1.000	1.000	0.956
PrivatBank	1.000	0.994	1.000	1.000	0.904	0.953	0.968	0.974
ALTUM	0.895	1.000	1.000	0.905	1.000	1.000	0.509	0.901
Baltikums Bank	0.865	0.901	0.847	0.877	1.000	0.987	0.998	0.925
Trust Commercial Bank	1.000	0.910	0.865	0.927	0.899	0.926	0.925	0.922
Expobank	0.989	1.000	1.000	0.978	1.000	0.909	0.887	0.966
Baltic International Bank	0.999	0.956	1.000	1.000	0.958	0.951	0.979	0.978
SMP Bank	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Bank M2M Europe	0.867	0.795	0.788	1.000	1.000	1.000	1.000	0.921

The results of the assessment of performance efficiency of the Latvian commercial banks using BCC approach demonstrated that such banks as Swedbank, SEB bank, ABLV Bank, DNB Bank and SMP Bank remained efficient in the research period,

whereas the performance of other banks was characterized by increase and reduction of performance efficiency indicators.

Based on the results of performance efficiency analysis of the Latvian commercial banks, the authors determined the total

commercial banks performance efficiency according to general efficiency formula (1) in Methodology.

Latvian commercial banks performance efficiency indicators are summarized in Table 5.

Table 5

Total Performance Efficiency of the Latvian Commercial Banks

Banks	1st approach	2nd approach
Swedbank	N/a	0.950
SEB bank	0.847	0.961
ABLV Bank	1.000	0.996
DNB bank	0.958	1.000
Rietumu Bank	N/a	0.942
Norvik Bank	N/a	0.977
PrivatBank	0.947	0.996
ALTUM	0.969	0.996
Baltikums Bank	N/a	0.931
Trust Commercial Bank	0.741	0.995
Expobank	N/a	0.990
Baltic International Bank	0.766	0.974
SMP Bank	0.755	0.988
Bank M2M Europe	N/a	0.654

Using two approaches to estimating performance efficiency at the Latvian commercial banks, the authors conclude that the majority of Latvian commercial banks have to look for new solutions with an aim to increase their performance efficiency. The authors also conclude that in order to obtain objective results on bank performance efficiency, it is necessary to use the first approach, as there are considerable differences in the results obtained using the two approaches to data analysis, as well as these approaches are based on several studies conducted abroad, but the 2nd approach has been developed only by one author.

Having obtained the given research results, the authors have attested the results of research conducted abroad [14], for example, the conclusion that banks – market leaders perform more efficiently. In the context of Latvia, these are DNB bank with a market share in terms of assets 9.51% and ABLV Bank with the market share of 12.28% (based on Bankscope data for 2012).

5. CONCLUSIONS

Within the framework of the research on the performance efficiency of the Latvian commercial banks, the authors mainly considered bank customers as lenders and debtors.

Employing two types of models within DEA approach – BCC and CCR, as a result of analysis the existing performance efficiency problems persisting in the Latvian commercial banking sector have been identified.

In the course of research the authors have come to the conclusion that the highest performance efficiency is demonstrated by the banks – market leaders, while comparatively smaller banks are inefficient with few exceptions. It has also been concluded that performance efficiency of the majority of banks is variable, which is reflected by growing and falling phases.

The research performed has provided information on the performance efficiency rate of the banks operating in Latvia, which allows making decisions considering future perspectives

of each bank not only from the standpoint of a shareholder, but also from the standpoint of a client and bank personnel.

Taking into consideration how important the efficiently developed banking system is, especially in the states with a high competitive dynamics as Latvia, the banks should change the present strategy to constantly improve their performance efficiency rate.

The authors believe that it would be required to perform additional investigations applying other input and output data, which could influence the bank economic indicators as well.

6. REFERENCES

[1] A. N. Berger, I. Hasan, & M. Zhou, "The effects of focus versus diversification on bank performance: Evidence from Chinese banks", **Journal of Banking & Finance**, Vol. 34, No. 7, 2010, pp. 1417-1435.

[2] R. Lensink, & N. Hermes, "The short-term effects of foreign bank entry on domestic bank behaviour: Does economic development matter?", **Journal of Banking & Finance**, Vol. 28, No. 3, 2004, pp. 553-568.

[3] D. A. Grigorian, & V. Manole, "Determinants of commercial bank performance in transition: An application of data envelopment analysis", **International Monetary Fund**, No.2002-2146, 2002.

[4] I. Hasan, & K. Marton, "Development and efficiency of the banking sector in a transitional economy: Hungarian experience", **Journal of Banking & Finance**, Vol. 27, No. 12, 2003, pp. 2249-2271.

[5] O. T. San, L. L. Theng, & T. B. Heng, "A Comparison on Efficiency of Domestic and Foreign Banks in Malaysia: A DEA Approach", **Business Management Dynamics**, Vol. 1, No. 4, 2011, pp. 33-49.

[6] M. Dietsch, & A. Lozano-Vivas, "How the environment determines banking efficiency: A comparison between French and Spanish industries", **Journal of Banking & Finance**, Vol. 24, No. 6, 2000, pp. 985-1004.

[7] M. J. Farrell, "The measurement of productive efficiency", **Journal of the Royal Statistical Society**, Series A (General), Vol. 120, No. 3, 1957, pp. 253-290.

[8] R. D. Banker, A. Charnes, & W. W. Cooper, W. W., "Some models for estimating technical and scale inefficiencies in data envelopment analysis", **Management science**, Vol. 30, No. 9, 1984, pp. 1078-1092.

[9] R. Färe, S. Grosskopf, & C. K. Lovell (Eds.), **The Measurements of Efficiency of Production**, Vol. 6, Kluwer Academic Pub, 1985.

[10] A. Charnes, W. W. Cooper, & E. Rhodes, "Measuring the efficiency of decision making units", **European journal of operational research**, Vol. 2, No. 6, 1978, pp. 429-444.

[11] S. P. Rossi, M. Schwaiger, & G. Winkler, **Managerial behavior and cost/profit efficiency in the banking sectors of Central and Eastern European countries**, Oesterr Nationalbank, 2005.

[12] H. Semih Yildirim, & G. C. Philippatos, "Efficiency of banks: recent evidence from the transition economies of Europe, 1993–2000", **European Journal of Finance**, Vol. 13, No. 2, 2007, pp. 123-143.

[13] L. Weill, "Is there a gap in bank efficiency between CEE and Western European countries?", **Comparative Economic Studies**, Vol. 49, No. 1, 2007, pp. 101-127.

[14] Y. Fang, I. Hasan, & K. Marton, "Bank efficiency in transition economies: recent evidence from South-Eastern Europe", **Bank of Finland Research Discussion Paper**, Vol. 5, 2011.

[15] A. M. Andries, "The determinants of bank efficiency and productivity growth in the Central and Eastern European banking systems", **Eastern European Economics**, Vol. 49, No. 6, 2011, pp. 38-59.

[16] M. Ariff, & L. Can, "Cost and profit efficiency of Chinese banks: a non-parametric analysis", **China Economic Review**, Vol. 19, No. 2, 2008, pp. 260-273.

[17] X. Chen, M. Skully, & K. Brown, "Banking efficiency in China: Application of DEA to pre-and post-deregulation eras: 1993–2000", **China Economic Review**, Vol. 16, No. 3, 2005, pp. 229-245.

[18] N. Thoraneenitiyan, & N. K. Avkiran, "Measuring the impact of restructuring and country-specific factors on the efficiency of post-crisis East Asian banking systems: Integrating DEA with SFA", **Socio-Economic Planning Sciences**, Vol. 43, No. 4, 2009, pp. 240-252.

[19] F. Sufian, "Determinants of bank efficiency during unstable macroeconomic environment: Empirical evidence from Malaysia", **Research in International Business and Finance**, Vol. 23, No. 1, 2009, pp. 54-77.

[20] S. C. Ray, & A. Das, "Distribution of cost and profit efficiency: Evidence from Indian banking", **European Journal of Operational Research**, Vol. 201, No. 1, 2010, pp. 297-307.

[21] A. Azadeh, S. F. Ghaderi, M. Mirjalili, & M. Moghaddam, "A DEA approach for ranking and optimisation of technical and management efficiency of a large bank based on financial indicators", **International Journal of Operational Research**, Vol. 9, No. 2, 2010, pp. 160-187.

[22] A. S. Camanho, & R. G. Dyson, "Cost efficiency measurement with price uncertainty: a DEA application to bank branch assessments", **European Journal of Operational Research**, Vol. 161, No. 2, 2005, pp. 432-446.

[23] O. Havrylchyk, "Efficiency of the Polish banking industry: Foreign versus domestic banks", **Journal of Banking & Finance**, Vol. 30, No. 7, 2006, pp. 1975-1996.

[24] B. Casu, & C. Girardone, "Bank Competition, Concentration and Efficiency in the Single European Market", **The Manchester School**, Vol. 74, No. 4, 2006, pp. 441-468.

[25] S. Fries, & A. Taci, "Cost efficiency of banks in transition: Evidence from 289 banks in 15 post-communist countries", **Journal of Banking and Finance,** Vol. 29, No. 1, 2005, pp. 55-81.

[26] C. Siriopoulos, & P. Tziogkidis, "How do Greek banking institutions react after significant events?—A DEA approach", **Omega**, Vol. 38, No. 5, 2010, pp. 294-308.

[27] **The Bank of Latvia**. Retrieved from www.bank.lv

[28] **The Association of Latvian Commercial Banks**. Indices of bank activities [Data file]. Retrieved from http://www.bankasoc.lv/lv/statistika/

Entrepreneurial Development in Countries with High Share of Informal Economy

Gulnara KHADIULLINA, Anna SVIRINA, Alfia ZABBAROVA
Economics and Management Department, Kazan National Research Technical University
10 Karla Marksa street, Kazan, 420126, Russian Federation

ABSTRACT

The goal of the research is to define interrelation between informal economy and entrepreneurial activity and major factors that influence people's incentives to become nascent entrepreneurs in the countries with high share of informal economy and reveal what types of incentives would be more efficient to increase rate of early entrepreneurship activity. The object of the study is entrepreneurial development trends in the countries with high share of informal economy. The authors analyze interrelation between the share of informal economy and entrepreneurial activity, define the influence that level of economic freedom and quality of doing business have on the level of early entrepreneurial activity, suggest that there exist of different types of government regulation of entrepreneurship in economies with high share of informal sector, define the most significant factors affecting share of informal economy and create a multifactor model for prognosis of informal economy share in relation to the state of factors that shape quality of doing business. Calculations and data processing were carried out by Microsoft Excel and SSPS Statistics software.

Keywords: entrepreneurial development, informal economy, entrepreneurial activity.

1. INTRODUCTION

Entrepreneurial activity is considered the basement of post-crisis economic recovery by a number of scholars [2] who recognize it as a key tool for providing sustainable growth. It has also been revealed that high nascent entrepreneurship growth rates are usually found in the countries with high share of informal economy [23, 11], where such activity is driven not only by necessity, but also due to the desire to provide better living [12] and henceforth entrepreneurial activity in this countries has high impact on both national and global trends of economic development.

It is shown in the existing literature, which entrepreneurial development trends under influence of informal economy are different from the ones observed within economic environment with well-developed institutions [8]. While tendencies of entrepreneurial activity within institutionally developed environment had frequently been the subject of research, such trends for nascent and existing entrepreneurship in case of polarized economies with high share of informal economy are currently underdeveloped. The difference between those two types of entrepreneurial development environment does not allow efficient implementation of entrepreneurial incentives, created in institutionally developed environments, within economies with high share of informal part.

The above mentioned reasons reveal the necessity to define main specific features of entrepreneurial development in the economies with high share of informal segment. Previously conducted research had not defined a set of problems affecting entrepreneurs under informal economy since it was focused on some special issues of enterprises' functioning, including dealing with corruption (both internal and external), development under weak property rights, governmental regulation of entrepreneurial activities etc. However, it seems necessary to define tendencies of entrepreneurial development in countries with high share of informal economy on the basis of holistic approach that would allow understanding of the input of such environments into global entrepreneurship development.

2. INTERRELATION BETWEEN INFORMAL ECONOMY SHARE AND ENTREPRNEURIAL ACTIVITY

Existing qualitative research suggests the average rate of entrepreneurial activity in informal sector of the economy is relatively higher than in the formal part [16] and finds the reasoning for this in exponential growth of urban population and relative inability of the cities to provide jobs for those moving from rural areas. Other scholars argue that entrepreneurial development within informal economy has to be defined by geographical factors [18], level of institutional development which is influenced heavily by historical precedence [17, 19], ethno linguistic fractionalization [3], historical impact and sociocultural factors [22, 13]. Though some of these findings were also supported by quantitative data [22, 13], but still the mentioned research does not explain whether high share of informal economy is related to the level of entrepreneurial activity. Existing literature also does not provide enough evidence on the fact of interrelation between entrepreneurial activity and share of informal economy.

Within this paper we look for interrelation between share of informal economy in gross domestic product and the level of entrepreneurial activity. The data on sizes of informal economy was obtained from report by Schneider et al [20], and henceforth we adopt the definition of informal economy that is used in the mentioned research: we claim that informal economy includes all the economic activities in production of goods and services that "are deliberately concealed from public authorities" [20]. In order to estimate the level of entrepreneurial activity we use the early stage entrepreneurial activity rate introduced by Global entrepreneurship monitor (GEM) (TEA rate) [12]. This exact factor was chosen due to the fact that current research shows that nascent entrepreneurs are the ones who provide maximum economic effect in the countries with a big share of informal economy, since they create maximum amount of jobs [9], and this effect is the most important for the countries in question.

For the purpose of this research we have stated the following null-hypothesis: (1) the level of early entrepreneurial activity (including improvement-driven entrepreneurial activity) is not influenced by the level of economic freedom defined by Heritage Foundation [10] in case of countries with high share of informal economy; (2) there is no relation between the level of imposed taxes and entrepreneurial activity (including improvement-driven entrepreneurial activity) in countries with

high share of informal economy, (3) there is no relation between the share of informal economy and level of early entrepreneurial activity (including improvement-driven entrepreneurial activity), and (4) there is no relation between dynamics of informal economy and dynamics of entrepreneurial activity. These four hypotheses were chosen due to the fact that qualitative research carried out in economic literature tends to reveal the above mentioned ties. These hypotheses were tested with the data from above mentioned sources by using SSPS Statistics 22.0.

The results for testing first two hypotheses for the sample of 29 countries with high share of informal economy (the sample was created on the basis of Schneider's average informal economy size estimation, and included only countries with share of informal economy above 22% of gross domestic product; the second filter for the sample was availability of other data – as a result the sample is 18% from original data set) can be seen in Table 1 (Pearson correlation).

Table 1. Pearson correlation between entrepreneurial activity

Indicator		TR	EF	IDE	ATR
TEA rate (TR)	Pearson correlation	1	,018	,478**	-,246
	Value		,927	,009	,197
	N	29	29	29	29
Economic Freedom (EF)	Pearson correlation	,018	1	-,085	-,195
	Value	,927		,661	,311
	N	29	29	29	29
Share of improvement driven entrepreneurship (IDE)	Pearson correlation	,478**	-,085	1	-,255
	Value	,009	,661		,182
	N	29	29	29	29
Average Tax Rate (ATR)	Pearson correlation	-,246	-,195	-,255	1
	Value	,197	,311	,182	
	N	29	29	29	29

**. Correlation is significant at 0,01.

As it can be seen from the Table 1, two first null hypothesis are supported with the analytical data – the only correlation found (TEA rate with IDE rate) arises from original GEM consortium methodology, and there is no statistically significant relationship between the level of economic freedom (which is in this case a relatively low one) or imposed tax rates (despite the fact that average level of taxation in this sample is 52.23%). The same findings were also supported by non-parametric correlation analysis. Both Kendall's tau-b test and Spearman test had found to statistically significant relationship between analyzed factors (except for interrelated within global entrepreneurship monitoring methodology TEA and IDE criteria). Thus, the first two null hypothesis were supported, and there is no statistically significant relationship between level of imposed taxes, economic freedom and early entrepreneurial activity for the countries with high share of informal economy.

The third stated hypothesis was tested by using the data from Schneirder's report and GEM analysis [20, 12]. In order to provide statistical testing for this relationship, we have formed the following sets of data: share of informal economy in sample countries (unlike the first two hypothesis testing, in this case a sample was created to include countries with different share of informal economy: from 8.4% for US to 53.7% in Peru) and TEA rates in 2001, 2004 and 2007. Total sample included 35 countries with factor-driven, efficiency-driven and innovation-driven economies. On the first stage we have tested the influence of informal economy on TEA rates both in the

same year and three years later (on the basis of suggestion that effect of informal economy might be more important for shaping future entrepreneurial environment), and it was proven that the relationship is stronger in case the same period of time (both correlations are statistically significant, but for the same year to value is 0.667 while value for the period ahead is 0.454), so later both current and future state of TEA ratios were analyzed to address the third hypothesis. Statistical analysis had shown there is a strong relationship between share of informal economy and entrepreneurial activity, so the third null hypothesis was rejected.

Finally for the research of relationship between informal economy and entrepreneurial activity we have tested for statistical significance relationships between informal economy share absolute and comparative values and entrepreneurial activity ratio. In order to carry out this analysis the sample used for testing hypothesis 3 was modified (the countries with share of informal economy below 20% of gross domestic product were excluded, and this left a sample of 18 countries). Pearson correlation analysis had shown that there is a strong positive correlation between dynamics of informal economy (defined as difference in informal economy share to GDP in 2007 and 2004). The same hypothesis was also tested by using non-parametric correlations, and results can be seen in Table 2.

Table 2. Non-parametric correlation between entrepreneurial activity

	Indicator		TEA2007	TEA2010
Kendall's coefficient	Share of informal economy in 2004	Correlation coefficient	,163	,349*
		Value	,344	,044
		N	18	18
	Share of informal economy in 2007	Correlation coefficient	,131	,317
		Value	,448	,068
		N	18	18
	Informal economy share dynamics	Correlation coefficient	,157	,502**
		Value	,363	,004
		N	18	18
Spearmen coefficient	Share of informal economy in 2004	Correlation coefficient	,251	,448
		Value	,316	,062
		N	18	18
	Share of informal economy in 2007	Correlation coefficient	,232	,404
		Value	,354	,096
		N	18	18
	Informal economy share dynamics	Correlation coefficient	,343	,689**
		Value	,164	,002
		N	18	18

**. Correlation is significant at 0,01.
*. Correlation is significant at 0,05.

As it can be seen from the Table 2, the strongest relationship in case of non-parametric correlation analysis exists between dynamics of informal economy share and future level of entrepreneurial development. This rejects the fourth null hypothesis, which can be also illustrated by Figure 1, derived from the same sample.
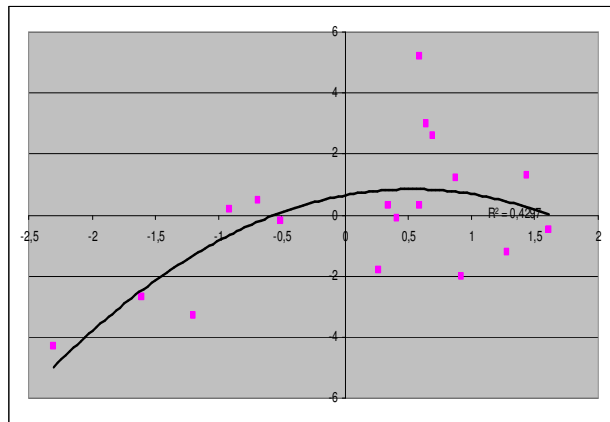
Figure 1. Relation between informal economy
dynamics and TEA dynamics

As it can be seen from the figure, the highest rate of entrepreneurial development is acquired in case the dynamics of informal economy is low. The influence of informal economy share on the level of early entrepreneurial activity was defined in earlier research [14], and can be described with an U-shaped curve where TEA rate is relatively high before share of informal economy increases up to 20 percent, than from 20 to approximately 32-33 entrepreneurial activity decreases, and starting from 33 percent it starts increasing again, and the highest level of growth is achieved when share of informal economy is above 40%. However, a significant share of countries has the level of informal economy between 20 and 33 percent with relatively low dynamics – and these are the countries that suffer from relatively low entrepreneurial activity.

Provided analysis allows concluding that entrepreneurial activity is influenced more by informal economy rather than economic freedom or financial restrictions imposed by the regulatory authorities. This became the basement for testing the next set of hypothesis concerning reverse relationship: how informal economy is influenced by the factors defining entrepreneurial environment.

3. TRENDS OF ENTREPRENEURIAL ACTIVITY UNDER WEAK STATES

Within the second part of our research we evaluated relationship between quality of doing business and level of entrepreneurial activity in case of economies with high share of informal sector on the basis of data sample of 29 countries created with Schneider's ranking and Doing business ranking [7]; the sample countries have a share of informal economy higher than 22.2% of GDP.

At this stage of research we tested the relationship between doing business quality level as defined by the World Bank, and level of nascent entrepreneurial activity, and found no statistically significant correlation of those two variables. This means nascent entrepreneurial activity to be influenced by some individual factors that define the quality of doing business. The results of correlation analysis are presented in Table 3.

As it can be derived from the Table 3, the level of entrepreneurial activity within economies with high share of informal sector is influenced only by two of fifteen analyzed factors: there is a strong positive relation between early-entrepreneurial activity and legal rights enforcement level (+0.458) and a strong negative relation with the average numbers of days required for performance of construction (-0.391). The other factors, including tax rates, duration of property registration period, administrative procedures for starting business does not affect people's entrepreneurial intentions in case entrepreneurs are dealing with high share of informal economic activities.

At the same time it can be derived from Table 3 that difficulties of registering and taxpaying procedures are interrelated (correlation coefficient +0.531) as well as costs of starting business and construction costs (+0.431) and a number of procedures required to register business in relation and average length of tax payment period (+0.464).

Table 3 Correlation of entrepreneurial activity level and doing business quality factors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	-,246	-,206	-,078	,050	-,052	-,168	-,155	,295	,458*	-,328	,025	-,064	,331	-,391*
2	-,246	1	,733**	,107	,244	,422*	,009	,628**	-,072	-,257	,225	-,045	,308	,013	,301
3	-,206	,733**	1	,371*	,062	,443*	,366	,464*	,018	-,197	,105	,073	,531**	,083	,332
4	-,078	,107	,371*	1	-,387*	,167	,173	,510**	-,098	,163	,037	,314	,289	,243	,344
5	,050	,244	,062	-,387*	1	,196	,158	-,142	,191	-,049	,175	-,299	,070	-,164	-,208
6	-,052	,422*	,443*	,167	,196	1	-,021	,256	,170	-,295	-,073	,080	-,116	-,194	,173
7	-,168	,009	,366	,173	,158	-,021	1	-,088	-,094	-,093	-,065	-,033	,401*	-,053	-,022
8	-,155	,628**	,464*	,510**	-,142	,256	-,088	1	-,252	,114	,052	,014	,186	,101	,366
9	,295	-,072	,018	-,098	,191	,170	-,094	-,252	1	,371*	,080	-,010	,141	,088	,028
10	,458*	-,257	-,197	,163	-,049	-,295	-,093	,114	,371*	1	,009	,123	,057	,265	,025
11	-,328	,225	,105	,037	,175	-,073	-,065	,052	,080	,009	1	-,088	,084	,197	,580**
12	,025	-,045	,073	,314	-,299	,080	-,033	,014	-,010	,123	-,088	1	-,110	,437*	-,033
13	-,064	,308	,531**	,289	,070	-,116	,401*	,186	,141	,057	,084	-,110	1	,215	,168
14	,331	,013	,083	,243	-,164	-,194	-,053	,101	,088	,265	,197	,437*	,215	1	-,178
15	-,391*	,301	,332	,344	-,208	,173	-,022	,366	,028	,025	,580**	-,033	,168	-,178	1

**. Correlation is significant at 0,01. *. Correlation is significant at 0,05.

1 – TEA rate 2 – average tax rate, 5, 3 – number of procedures to start business, 4 – costs to restore insolvency, 5 – insolvency restoration rate, 6 – average number of days needed to prepare taxation documentation, 7 – average number of days needed to enforce contract, 8 – number of days needed to pay taxes, 9 – level of investors disclosure, 10 – legal rights enforcement level, 11 – duration of construction procedures, 12 – costs of starting business, 13 – the number of procedures needed to register property, 14 – cost of construction, 15 – average number of days needed for construction.

This allows us to propose the following hypothesis (taking into account that the sample includes only countries with high share of informal economy): informal economies usually develop under weak states that can take either a form of "absent state" or "extra governmental regulation". By "absent state" we mean the environments where formal regulations are not insured in a country, and contract enforcement is mainly implemented by informal institutions. By "extra government regulation" we mean environments where responsibility for enforcement of formal contracts is distributed between a number of governmental controlling and regulatory institutions, which in the end of the day are unable to insure these contacts. According to our cluster analysis, the latter type is a more common one. Henceforth, it is necessary to reveal some entrepreneurial behaviour trends that appear under weak states.

On this stage of analysis we have also found out, that the share of informal economy is strongly related to a number of days required to prepare taxation documents and pay taxes (+0.262, significant at 0.05), duration of period required to register owner's property rights (+0.297, significant at 0.05) and costs of construction (+0.318, significant at 0.05). This supports findings of both de Soto and Schneider, who emphasized the importance of legal procedures duration and influence of this factor onto informal economy. At the same time it is clear that the factors that influence nascent entrepreneurial activity under weak states are different from the ones defining share of informal economy – henceforth it can be concluded that entrepreneurial activity level can be quite high in economies with high share of informal part, if certain measures are taken.

In order to define the disproportion between the two above mentioned types of weak states in terms of their influence on early entrepreneurial activity ratio and share of informal economy in gross domestic product we have insured graph analysis of the situation (see Figure 2).
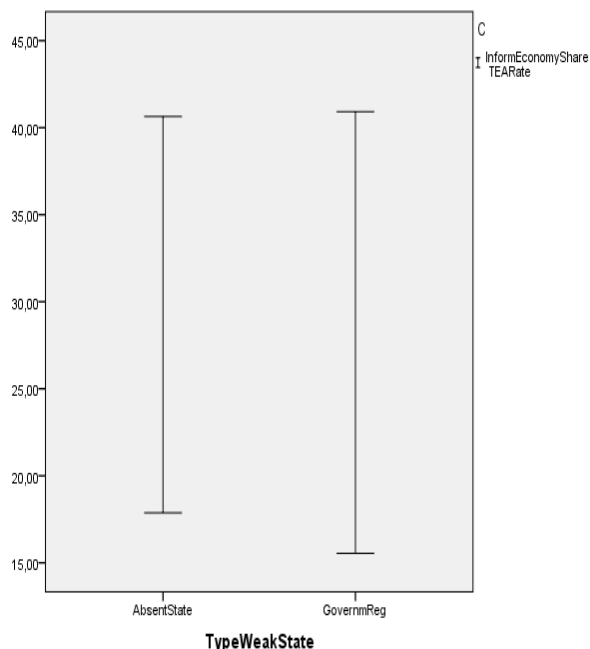

Figure 2. Minimum level of TEA and maximum level of informal economy share in case of absent states and governmental regulation

As it comes clear from the figure, absents states come to be relatively more efficient than over-regulated ones, and this allows to make a conclusion that in case of weak absent states there appear different types of regulations: the ones that lie out of formal sector on the one hand, and are entrepreneurial in the nature on the other.

In our opinion, this explains the phenomena of social entrepreneurship which appeared in countries with high share of informal economy and high level of entrepreneurial activity. Social entrepreneurship research studies social entrepreneurs as a special phenomena and is aiming to figure out the reasons for social entrepreneurs' economics success. But analysis of social entrepreneurship definitions [5] shows that one can not clearly define what a social entrepreneur is: it is stated that social entrepreneur addresses social needs, but in general all entrepreneurs are addressing certain societal needs [21]. The same thing was as well mentioned by Mair [15] who insists that all successful entrepreneurs are generating some social value. Mentioned research results lead to a conclusion that by nature all entrepreneurship is social, the trend to stop being socially active arises not from the nature of entrepreneurship but out of the state of environment. In case of weak states such entrepreneurial structures arise to cover weaknesses of the state – as it is shown by an example of the most famous social entrepreneurial structures such as Aravind or Grameen.

However, the scholars have revealed that social entrepreneurship has a number of specific features [6]: (1) mission focus (adopting and pursuing a mission to create and sustain social value; (2) ambition (recognizing and relentlessly pursuing new opportunities to serve the chosen mission and/or enlarge it), (3) learning (engaging in a process of continuous innovation, adaption, and learning); (4) efficiency and efficacy (acting boldly without being limited by resources currently in hand); (5) sustainability (exhibiting heightened accountability to the constituencies served and for the outcomes created – market cuts out unsustainable social entrepreneurs faster than regular ones). The latter two features are in our opinion the most important ones since they provide understanding on the nature of social entrepreneurship and appearance of the phenomena under weak states: social entrepreneurs took over where weak state failed to create institutions to solve certain problems (sustainability) and provided higher level of efficiency and efficacy than average commercial enterprise, which is also a specific feature of entrepreneurial performance in economies with high share of informal sector (efficiency and efficacy).

4. MODEL OF INFORMAL ECONOMY DEVELOPMENT

Finally, revealed interrelation between informal economy and early entrepreneurial activity (including the factors influencing it) allowed us to propose a multi-factor model for defining the share of shadow economy as a dependent variable. This model was developed using described above sample of 29 countries with high share of informal economy by using SSPS Statistics software to develop a multifactor model that includes as independent variables factors describing the quality of doing business environment. Descriptive statistics of the model is the following: R square and corrected R square equals 1, approximate value of mistake equals 0, which means the model demonstrates high level of reliability. Dependent variable of the model is share of informal economy in comparison to country's gross domestic product, independent variables left after statistical testing in the linear regression model are the following: costs of construction (coefficient 0.663), duration of property rights registration (coefficient 0.096), insolvency recovery rate (coefficient 0.172), insolvency recovery costs (coefficient 0.305), investors' protection (coefficient -0.291), amount of procedures for contract enforcement (coefficient

0.421), average corporate tax rate (coefficient 0.422), managers' liability (coefficient 0.642), duration of taxation documents preparation and payment (coefficient 0.174), amount of procedures for tax payment (coefficient 0.702), number of procedures for property rights registration (coefficient -0.539), time required to enforce formal contracts (coefficient 0.523), extent of investors disclosure (coefficient 0.662), percentage of improvement driven early entrepreneurial activity (coefficient 0.225), average time required for construction (coefficient 0.514), amount of legal procedures necessary to start business (coefficient -0.682), average time required to start business (coefficient 0.579) and average cost of starting new business (coefficient -0.136).

Table 4 Model statistics: correlation and significance of independent variables

	Correlation			
	Null	Partial	Component	Significance
Construction costs	,273	1,000	,273	,181
Duration of property registration procedure	,184	,998	,050	,018
Insovency recovery rate	-,182	,999	,079	-,031
Recovering insovency costx	,265	1,000	,179	,081
Level of investor protection	-,102	-1,000	-,180	,030
Contract enforcement procedure	,097	1,000	,316	,041
Tax rate	,272	1,000	,227	,115
Director liability	,128	1,000	,313	,082
Duration of tax payment perions	,034	,999	,087	,006
Taxes payment procedures	,321	1,000	,338	,225
Property regulation procedures	-,223	-1,000	-,376	,120
Contract enforcement duration	-,037	1,000	,299	-,019
Extent disclosure for investors	-,115	1,000	,321	-,076
Improvement driven early entrepreneurship	,014	1,000	,152	,003
Average construction ieriod	,213	1,000	,369	,110
Start vusiness procedures	,011	-1,000	-,367	-,007
Duration of starting business period	,254	1,000	,224	,147
Costs of starting business	,185	-,999	-,061	-,025

Table 4 shows other data defining model's reliability: correlation and significance of independent variables in the developed model of informal economy share estimation which was estimated within SSPS Statistics analysis.

As it can be seen from Table 4, costs of construction, average taxation rate, total amount of tax payment procedures, average construction period and duration of staring business process are the most significant factors defining level of informal economy. Some of the support existing research findings (for example, significance of business registration supports de Soto's and Schneider's findings) while the others, such as amount of tax payment procedures, reveal the major politics that are to come into focus of attention if governments wish to reduce share of informal economy. It is also important that the most significant factors revealed within the model seem

to be the major obstacles for entrepreneurial development under weak states that impose high level of government regulation, and the mentioned elements of doing business environment are to become the primary issues to reduce government regulation aiming to decrease share of informal economy.

5. CONCLUSIONS

The above described research findings allowed us to come to the following set of conclusions concerning entrepreneurial development in countries with high share of informal economy:

1. There is a strong relationship between the share of informal economy and early entrepreneurial activity which is u-shaped, and the minimum of activity is found in case informal economy share is within the range from 20 to 33% according to Schneider's methodology; at the same time the level of economic freedom or average level of taxation do not influence early entrepreneurial activity, which means that creation on entrepreneurship incentives is possible for both states with low and high share of informal economy.

2. There is no statistically significant correlation between the quality of doing business and early entrepreneurial activity, and it supports the finding that nascent entrepreneurship is relatively high both in case of high and low share or informal economy; at the same time some of the factors that define quality of doing business, such as legal rights enforcement level and duration of construction procedure, also influence entrepreneurial activity.

3. High share of informal economy, according to acquired quantitative results, arises in case state governments are weak; weakness of government can perform either in a form of absent state or overwhelming government control – and it is proven that the second type in general leads to lower entrepreneurial activity on the one hand and to higher share of informal economy in comparison with gross domestic product on the other.

4. Regulation of informal economy can be achieved if costs of construction, average taxation rate, total amount of tax payment procedures, average construction period and duration of staring business process (the most significant factors defining level of informal economy) would be changed, while other incentives for doing business have relatively low influence on the level of informal economy.

On the basis of research several proposals on increasing efficiency and efficacy of entrepreneurial activity regulation in economies with high share of informal sector can be made:

1. In order to increase entrepreneurial activity steps should be taken to reduce informal sector of the economy by changing procedures that define average costs of construction, total amount of tax payment procedures, average construction period and duration of staring business process, as well as measures that aim to decrease average taxation rate; at the same time improvement of economic freedom index, or only decreasing average tax level will not lead to increase of early entrepreneurial activity.

2. Entrepreneurs show high level of activity both in case informal sector of economy is relatively small (less than 20%), and in case it becomes relatively large (over 33% of economy), but different attitudes are needed to achieve high rate of nascent entrepreneurship in these cases: in the first case it is necessary for the government to create special incentives for entrepreneurial development, while

in the latter case entrepreneurs require minimum regulation procedures in order to be active on the early stage.

3. Taking into account the structure of informal economy share model it is necessary to take measures to reduce costs of construction, duration of property rights registration, insolvency recovery rate, insolvency recovery costs, amount of procedures for contract enforcement, average corporate tax rate, duration of taxation documents preparation and payment, amount of procedures for tax payment, number of procedures for property rights registration, time required to enforce formal contracts, average time required for construction, amount of legal procedures necessary to start business and average cost of starting new business – in order to decrease share of informal economy by affecting entrepreneurial behavior.

6. REFERENCES

[1] D.Acemoglu, S.Johnson, J.Robinson. Institutions as a Fundamental Course of Long-Run Growth. Handbook on Economic Growth. Vol. 1A. Amsterdam: North-Holland, 2005. pp. 3585-472.

[2] D.Ahlstrom. How Business Contributes to Society. Academy of Management Perspectives. Vol. 24, No. 8, 2010. pp. 11-24.

[3] A.Alesina, A.Devleeschauwer, W.Easterly, S.Kurlat, R.Wasciarg. Fractionalization. Journal of Economic Growth. Vol. 8, No. 2, 2003. pp. 94-155.

[4] Q.Ashraf, O.Galor. The "Out of Africa" Hypothesis, Human Genetic Diversity and Comparative Economic Development. American Economic Review. Vol. 103, No. 1, 2013. pp. 1-46.

[5] P.Dacin, T.Dacin, M.Matear. Social Entrepreneurship: Why We Don't Need a New Theory and How Do We Move Forward from Here. Academy of Management Perspectives August, 2010. pp. 37-58.

[6] J. Dees. The meaning of "social entrepreneurship". www.caseatduke.org/documents/dees_sedef.pdf

[7] Doing Business World Bank Ranking. http://www.doingbusiness.org/

[8] H. de Soto. The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else. New York: Harper Collins, 2000. pp. 146-201.

[9] R. Fairlie. Kauffman Index for Entrepreneurial Activity 1996-2010. Retrieved from http://www.kauffman.org/uploadedFiles/KIEA_2011_report.pdf on April 12, 2011

[10] Heritage Foundation. 2007 Index of Economic Freedom. http://www.heritage.org/index/

[11] M.Hitt, D.Ireland, D.Sirmon, C.Trahms. Strategic Entrepreneurship: Creating Value for Individuals, Organizations, and Society. Academy of Management Perspectives. Vol. 25 No. 5, 2011. pp. 57-75.

[12] Global Entrepreneurship Report. Available at http://www.gemconsortium.org/

[13] .L.Guiso, P.Sapienza, L.Zigales. Cultural Biases in Economic Exchange? Quartely Journal of Economics. Vol. 124, No. 3. 2009. pp. 95-131.

[14] G.Khadiullina, E.Shurkina, A.Svirina. Informal Economy Dynamics as an Important Factor Defining Level of Social Unease. Volume 2013 (2013), Article ID 720741, Journal of Economics Studies and Research, DOI: 10.5171/2013.720741

[15] J. Mair. Introduction to Part II: Exploring the Intentions and Opportunities behind Social Entrepreneurship. New York: Palgrave Macmillan, 2006. pp. 64-89.

[16] B. Marx, T.Stocker, T.Suri. The Economics of Slums in the Developing World. Journal of Economic Perspectives. Vol. 27, No. 4, 2013, pp. 187-210.

[17] D.North, R.Thomas. The Rise of Western World: The New Economic History. Cambridge, UK: Cambridge University Press, 1973. pp. 370-379.

[18] K.Pomeranz. The Great Divergence: China, Europe, and the Making of the Modern World Economy. Princeton: Princeton University Press, 2000. pp. 106-144.

[19] D.Rodrick, A.Subramanian, F.Trebbi. Institutions Rule: the Primacy of Institutions over Geography and Integration of Economic Development. Journal of Economic Growth. Vol. 9, No. 2, 2004. pp. 131-165.

[20] F.Schneider, A.Buehn, C.Montenegro. Shadow Economies all over the World: New Estimates for 162 Countries from 1999 to 2007. http://www.econ.jku.at/members/Schneider/files/publications/LatestResearch2010/SHADOWECONOMIES_June8_2010_FinalVersion.pdf

[21] C.Schramm. All Entrepreneurship is Social. Stanford Social Innovation Review Spring 2010. http://www.ssireview.org/articles/entry/all_entrepreneurship_is_social

[22] G.Tabellini. Institutions and Culture: Presidential Address. Journal of European Economic Association. Vol. 6, No. 2-3, 2008. pp. 255-294.

[23] J.Webb, L.Tihanyi, D.Ireland, D.Sirmon. You Say Illegal, I Say Legitimate: Entrepreneurship in the Informal Economy. Academy of Management Review. Vol. 34, 2009. pp. 492-510. doi:10.5465/AMR.2009.40632826

Sustainable Corporate Performance Index for Manufacturing Industry

Alena KOCMANOVA, Marie PAVLAKOVA DOCEKALOVA
Brno University of Technology, Faculty of Business and Management,
Department of Economics, Brno, Czech Republic

and

Petr NEMECEK
Tomas Bata University in Zlín, Faculty of Management and Economics,
Department of Management, Zlín, Czech Republic

ABSTRACT

The article presents a methodological approach consisting in the application of a discriminant analysis in evaluations using financial and non-financial ESG and economic indicators, including the establishment of the sustainable corporate performance index of Czech companies from the manufacturing industry. The proposed methodology for comprehensive performance evaluation is adjusted to allow for an assessment of financial and non-financial indicators and their weights. It is particularly necessary to focus on such financial and non-financial indicators that provide to some extent information about the potential risk to companies' sustainability. This will not only allow for an ex post evaluation of companies, but also of their sustainable development. The aim of this paper is to establish sustainable corporate performance index for Czech companies from the manufacturing industry based on defined financial and non-financial ESG and economic performance indicators. Environmental, social, corporate governance and economic indicators are often characterized by a large number of variables that are partly qualitative in nature and can only be evaluated on the basis of a subjective assessment, which inevitably includes various types of knowledge. The sustainable corporate performance index was determined by means of a discriminant analysis. The use of discriminant analysis to measure the comprehensive performance of the company can also improve the accuracy of qualitative aspects captured in a non-deterministic manner. Indicators selected for discriminatory analysis included those relevant financial indicators that relate to sustainability and non-financial indicators that also meet these conditions for Czech company from the manufacturing industry. The result of discriminant analysis is the sustainable corporate performance index, which should reflect the sustainability of companies from the manufacturing industry.

Keywords: environmental, social, corporate governance, economic performance, sustainable corporate performance index, discriminant analysis

1. INTRODUCTION

Corporate Sustainability can be considered a corporate response to sustainable development represented by strategies and procedures that focus on key issues associated with the global sustainable development. Sustainable development involves the simultaneous pursuit of economic prosperity, environmental quality and social equity. Companies aiming for sustainability need to perform not against a single, financial bottom line but against the triple bottom line [1].

Currently, corporate sustainability can be defined as the integration of environmental, social, corporate governance as well as economic performance. Corporate sustainability is one of the facets in the demand for companies that publish sustainability results in areas such as environmental, social and corporate governance. Sustainability results are required from stakeholders such as investors wishing to know whether to engage in sustainable business practices [2], [3]. In addition, environmental and social performance is an important part of many companies' business strategy for the control of processes (Ferreira et al 2009), and a sufficient incentive to monitor financial and non-financial, environmental, social and corporate governace (ESG) and economic indicators. Sustainability assessment using financial and non-financial, ESG and economic performance indicators leads to the establishment of sustainable corporate performance (SCP) in many companies. And this also relates to the determination of the sustainable corporate performance index (SCPI).

At the research institute of the Faculty of Business and Management of BUT, sustainable corporate performance has been studied in the grant project "Development of Methods for Multi-factor Performance Measurement of Companies in a Selected Industry", and an SCP model has been proposed which includes financial and non-financial ESG and economic performance indicators for companies in the manufacturing industry. These financial and non-financial ESG and economic performance indicators were designed using multi-dimensional statistical methods of cluster and principal

component analysis (PCA) and factor analysis (FA). Research results were published both in journals and in books [4], [5], [6], [7], [8], [9]. In a follow-up grant project entitled "Measuring Corporate Sustainability in Selected Sectors", research efforts focus on determining the SCPI and sustainable value.

The aim of this paper is to establish the SCPI for Czech companies in the manufacturing industry. The SCPI could then be defined as a composite indicator that includes financial and non-financial, ESG and economic performance indicators and their weights. The inclusion of the corporate governance indicator together with environmental, social and economic performance indicators into the SCPI can globally characterize the problems and risks which investors consider in the context of corporate behaviour.

2. SUSTAINABLE CORPORATE PERFORMANCE INDEX: CONCEPTUAL AND THEORETICAL ANALYSIS

Companies seek to achieve long-term benefits by integrating sustainability-related activities into their core business strategy [10], [11]. In general, we can say that companies integrate sustainability methods, tools and practices because they feel obliged to do so, because they themselves want to it, or are forced to do so [12]. In the context of sustainable development concept, it is apparent that there is a need for performance indicators that are able to express the interdependence between ESG and economic performance. Composite indicators have recently become more widespread and popular for that purpose. This is primarily due to their ability to accumulate larger amounts of information in a single indicator.

A comprehensive method of evaluating performance of companies by means of financial and non-financial indicators may certainly become a specific set of instruments for assessing corporate sustainability. This is particularly true about multi-dimensional models employing several indicators that are assigned specific weights. The company's comprehensive performance is then expressed by a composite indicator, which can assess the company's sustainability. There are essentially two approaches to the evaluation of companies, namely evaluation using a set of indicators which includes the so-called key indicators, and evaluation using a single indicator (a composite indicator), which is a synthesis of partial indicators.

The currently most frequently used models in the Czech Republic are especially those that focus either on the determination corporate performance in terms of value creation, i.e. credit scoring indicators, or they assess the company based on its ability to pay its obligations, and these are bankruptcy indicators. These bankruptcy models are an example of composite indicators that focus primarily on financial indicators, such as the Altman Z_score, Credibility Index IN, Taffler Model, etc.

The first significant research was conducted by Altman and reported in an article published in 1968, where the author examined the determination of company credit scoring using a discriminant analysis and a single composite indicator. Thanks to its high prediction accuracy, the article met with great response and served as a basis for many articles that either extended the model or verified it on data from other markets or periods [13]. In 1978, Altman and Eisenbeis [14] came up with the final Z-score model, which took into account, among other things, new standards of financial reporting. The resulting model is called "Zeta Analysis".

In the Czech Republic, several prominent researchers studied the issue of corporate performance evaluation [15]. Their models were produced with the use of multiple discriminant analysis. The data in their studies, however, come exclusively from Czech enterprises and performance evaluations were made in Czech enterprises only. In their performance evaluations, financial indicators and their sets are predominant. Most financial indicators are associated with financial accounting based on the accounting standards; these are important indicators for measuring the financial performance of companies for tax purposes as well as for capital markets.

Performance evaluation based solely on the basis of financial indicators suffers from many shortcomings because of historical data, its exclusive focus on short-term goals, but its major problems are its apparent disconnectedness from strategy and its frequent lack of transparency and reliability. These deficiencies of financial indicators should be eliminated by non-financial indicators and models that are formed by such non-financial indicators, e.g. the Czech Benchmarking Index, Enterprises Evaluation Model created by H. Pollak, Argenti's Model, etc. The most common and most elaborate enterprise evaluation system is the Balanced Scorecard (BSC) system.

The theme of composite indicators has also been picked up by a number of other authors [16], [17]. For environmental sustainability, for instance, several composite indicators have been developed, e.g. the environmental sustainability index (ESI) compiled by the World Economic Forum for 142 countries, the wellbeing index (WI) aggregating 36 indicators, the FTSE4Good Index for social issues, the Natur-Aktien-Index or the Dow Jones Sustainability Indexes (DJSI), etc.

3. REASERCH METHODOLOGY

The proposed methodology of comprehensive assessment of sustainability of companies employs a method of new approach to the assessment of companies in the Czech context, in particular by means of examining the potential of Multiple Discriminant Analysis (MDA) which was also used by Altman for the design of bankruptcy models [18], [19], [20]. Altman applied the Multiple Discriminant Analysis to data from 66 bankrupt and non-bankrupt

companies listed at the New York Stock Exchange at the time. MDA was the dominant method for the design of models at that time, and was later replaced with less demanding statistical techniques, such as logit analysis (LA) which was applied by [21], [22], probit analysis (PA), which was used by [23], [24] for the assessment of companies.

Using the logit model and logistic regression in 2010, the latest version of the prediction model of E.I. Altman was published in 2010 - the Z-metrics model. Altman participated in its development with the RiskMetrics group in response to the economic crisis and the prediction capabilities of Z-models. The model is designed for both small and large companies from Canada and USA, and elsewhere. Further, models of linear probability (LPM) can also be used, Tamari's risk index - a scoring method, or the neuron network method, all of which have been researched since the early 1990s, can also be used.

The basis for empirical research for comprehensive assessment of companies, and potentially also the SCPI is comprised of a set of financial and non-financial ESG and economic performance indices for companies in the manufacturing industry. A questionnaire was developed in order to determine same, voluntary instruments for the environmental area, ISO 14 000 and EMAS, for the social area, CSR and ISO 26 000, for the corporate governance area, OECD Principles of Corporate Governance and the Green Paper - the EU corporate governance framework, serving as a basis therefor, together with other international and domestic resources (GRI, IFAC, ASSET, EFFAS-DVFA and Czech Statistical Office); for the economic area, annual reports of companies and data from AMADEUS was used. Of the database compiled, 79 companies in the manufacturing industry with the number of employees over 250 were investigated. The following manufacturing companies were involved: electrical engineering, engineering, medical products, metallurgy, textile and leather production, chemical production and food processing.

Sustainability of the company is also shown by the introduction of voluntary instruments in a company in the manufacturing industry, as such tools encourage considerate treatment of the environment, social responsibility and good corporate governance.

Table 1 Voluntary instruments in companies in the manufacturing industry (in %)

Use	ISO 9000	MRP	ISO 14000	EMAS	OHSAS 18000	EMA	CSR	Codex Corporate governace	Cleaner production	LCA	Enviro-mental labelling of products
No	10,1	51,9	44,3	84,8	51,9	91,1	74,7	88,6	69,6	78,5	69,6
Yes	89,9	48,1	55,7	15,2	48,1	8,9	25,3	11,4	30,4	21,5	30,4

The selection of non-financial indices is related to the company's objective - sustainability. There is a causal nexus between these non-financial indicators and financial indicators. They are structured in such a way so as to make it possible to assess whether any change therein occurred or not, and if so, whether desirable or undesirable. Environmental performance indicators tend to be quantitative, i.e., tangible indicators, and are expressed in various units. Social and corporate indicators are mainly qualitative, i.e., intangible indicators. There was an effort to replace such intangible indicators with a hard indicator, a change in which reflected a change in the soft indicator as a substitute indicator.

Indexes ESG and economic performance indicators were set using the Principal Component Analysis (PCA) and Factor Analysis (FA) for companies in the manufacturing industry [25], [26]. indexes (I_{ENVI}) for environmental performance indicators (EN) were set for three areas capable of measurement: *Emissions, Consumption of resources, Waste,* which contain 7 environmental indicators (EN1 - costs of environmental investments, EN2 - total atmospheric emissions, EN3 - total greenhouse gas emissions, EN4 - total consumption of renewable energy, EN5- total annual consumption of water, E6 - total annual production of waste, EN7 - total annual production of hazardous waste), and their metrics.

Index (I_{SOC}) for social performance indicators (SOC) were also set for four areas capable of measurement: *Society, Human rights, Labor relations, Product liability*, which contain 8 social factors (SOC1 - Community, SOC2 - Funding for municipalities, SOC3 - Discrimination, SOC4 - Equal opportunities, SOC5 - Employment fluctuation level, SOC6 - Mortality rate, SOC7- Marketing communication, SOC8 - Customer health and safety), and their metrics.

Index (I_{CG}) for corporate governance indicators (CG) were also set for two areas capable of measurement: *Monitoring, Efficiency of CG*, which contain 6 corporate governance indicators (CG1- Corporate information for aim, CG2- Accountability, CG3 - Remuneration, CG4 - Composition, CG5 - Equal opportunities, CG6 - Legal compliance), and their metrics.

The methodological approach to comprehensive assessment of company performance using financial indicators follows, just like one based on non-financial indicators, the same objective: sustainability. The financial indicators were designed in line with ESG performance indicators, with GRI 2006, 2011, 2013 and annual reports of companies serving as a basis.

Indexs (I_{ECO}) economic indicators (ECO) were set for three areas capable of measurement: *Profitability, Economic results, Cash flow*, which contain 8 economic indicators (ECO1 – ROE, ECO2 – ROA, ECO3- ROI, ECO4 – ROCE, ECO5 - Profit (EAT), ECO6 - Turnover size, ECO7 - Asset turnover, ECO8- Cash-flow), and their metrics.

Empirical research of integrated (comprehensive) performance of companies in the manufacturing sector using financial and non-financial ESG and economic indicators will mean that such aggregated indicators will be able to provide the most comprehensive picture of sustainability in the companies concerned. The assessment is based on a multi-criteria evaluation aggregating various indicators, whereby it is subsequently possible to determine the SCPI using those results.

4. RESULTS AND DISCUSSION

The aim of the sustainable corporate performance index is to assess the sustainability of companies from the manufacturing industry.

The sustainable corporate performance index is based on a set of Indexes I_i , i = {ENVI, SOC, CG, ECO},

which are produced by aggregating financial and non-financial, ESG and financial performance indicators.

The performance Indexes I_{ENVI}, I_{SOC}, I_{CG} and I_{ECO} do not act independently but, rather, are interacting with each other, which means that the effect of some indicators may be cancelled out by action of some other indicators. To determine how the indicators act together, the Multiple Disciminant Analysis was applied.

The general equation of discriminant analysis:
$$Z = a_1 X_1 + a_2 X_2 + \ldots + a_p X_p$$
where $a_1, \ldots a_p$ are discrimination coefficients and $X_1, \ldots X_p$ are selected independent variables that best explain the classification into groups.

The most common statistical method for predicting the financial failure is the multi-dimensional discriminant analysis (MDA) based on the categorization of individual quantities into several characteristic groups.

The discriminant analysis is used to identify the Indexes I_{ENVI}, I_{SOC}, I_{CG} and I_{ECO} that help enhance sustainability, or otherwise. An analysis of variance showed that the Indexes I_{soc} *Labour relations*, I_{soc} *Product labelling*, I_{CG} *Monitoring* , I_{ECO} *Economic result and* I_{ECO} *Profitability* were statistically insignificant. These Indexes were dropped and the discriminant analysis was performed again. Tab. 2 below shows that when the five statistically insignificant were left out, the others were statistically significant.

Table 2 ANOVA - reduced discriminant analysis

Indexes	Wilks' Lambda	F	df1	df2	Sig.
I_{ENVI} EMISSIONS	0.874	11.150	1	77	0.001
I_{ENVI} CONSUMPTION OF RESOURCES	0.943	4.661	1	77	0.034
I_{ENVI} WASTE	0.897	8.853	1	77	0.004
I_{SOC} SOCIETY	0.904	8.213	1	77	0.005
I_{SOC} HUMAN RIGHTS	0.899	8.649	1	77	0.004
I_{CG} EFFICIENCY OF CG	0.872	11.317	1	77	0.001
I_{ECO} CASH-FLOW	0.934	5.404	1	77	0.023

Box's M conformity of covariance matrices indicated to us that the null hypothesis of covariance matrices conformity must be rejected. That assumption can, nevertheless, be usually circumvented by comparing the natural logarithms of the determinants of the covariance matrices. The value of Wilks' lambda is an analogue to R2 in the regression analysis. In our case, too much variance is left unexplained, the value of 1-lambda is the proportion of variance explained with the help of discriminated groups. In our case, too much variance remained unexplained (31.8%), and it can therefore be assumed that we failed to find all relevant indicators (see Tab. 3).

Table 3 Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0.682	28.148	7	0.000

Using the 7 listed sub-indices, we nevertheless managed to correctly classify almost 76.3% of the cases, which is quite a good result. It can therefore be assumed that we did not find all the relevant cases (19.5%), see Tab. 4.

The SCPI discriminant equation employs Standardized Canonical Discriminant Function Coefficients:

$SCPI$
$= 0.432\, I_{SOC} HUMAN\ RIGHTS$
$+ 0.426\, I_{CG} EFFICIENCY\ OF\ CG + 0.386\, I_{ENVI} WASTE$
$+ 0.280\, I_{ECO} CASH\ FLOW + 0.263\, I_{ENVI} EMISSIONS$
$- 0.30\, I_{ENVI} CONSUMPTION\ OF\ RESOURCES$

SCPI values of < -0.041 indicate an unsustainable company, SCPI values of > 0.070 indicate the company's sustainability and SCPI values in the <-0.041; 0.049> interval do not give clear information in relationship to sustainability.

Table 4 Classification Results[a,c]

		Sustainability	Predicted Group Membership		Total
			no	yes	
Original	Count	no	33	8	41
		yes	9	29	38
	%	no	80.5	19.5	100.0
		yes	23.7	76.3	100.0
Cross-validated[b]	Count	no	26	15	41
		yes	12	26	38
	%	no	63.4	36.6	100.0
		yes	31.6	68.4	100.0

a. 78.5% of original grouped cases correctly classified.
b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
c. 65.8% of cross-validated grouped cases correctly classified

The sustainable corporate performance index is a composite indicator that makes it possible to assess sustainability development of companies from the manufacturing industry. The sustainable corporate performance index may be important for decision making of managers, investors, etc., as well as an initial composite indicator for *Integrated Reporting*. An accurate selection of appropriate indicators is, however, difficult, and it is therefore necessary to identify suitable financial and non-financial performance indicators for each of the sectors in the Classification of Economic Activities (CZ-NACE) so that they fulfilled, and were grouped into, a single *Sustainable corporate performance index.*

5. CONCLUSIONS

A comprehensive corporate performance measurement using financial and non-financial indicators consisted of three stages: identification of financial and non-financial ESG and economic indicators affecting the company's performance and their transformation into measurable indicators, the measurement of the indicators, the use of the indicators in practice. The basis for the design of the SCPI were financial and non-financial performance indicators that were determined by factor analysis (including indices I_{ENVI}, I_{SOC}, I_{CG} and I_{EKO}). SCPI was produced using discriminant analysis. All calculations were analysed by the SPSS program, version 22, for Windows.

The aim of the article was therefore to test the discriminant analysis method for creating the SCPI. For the creation of the sustainable corporate performance index and its subsequent improvements, data obtained in questionnaire surveys of companies in the manufacturing industry were used.

The designing of a SCPI is one of the ways how to create a necessary tool for sustainability evaluation of companies in the manufacturing industry.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] N. Yakovieva, "Sustainable benchmarking of food supply chains", **George Perkins Marsh Institute Working Paper,** 2009-02, 2009.

[2] A. Huggins, W. J. Green, R. Simnett. "The Competitive Market for Assurance Engagements on Greenhouse Gas Statements: Is There a Role for Assurers from the Accounting Profession?", **Current Issues in Auditing**, Vol. 5, No. 2, 2011, pp. A1-A12.

[3] G. Pflugrath, P. Roebuck, R. Simnett, "Impact of assurance and assurer's professional affiliation on financial analysts' assessment of credibility of corporate social responsibility information", **Auditing: A Journal of Practice & Theory** Vol. 30, No. 3, 2011, pp. 239-254.

[4] A. Kocmanová, P. Nemecek, "Economic, environmental and social issues and corporate governance in relation to measurement of company performance", In **Liberec Economic Forum Liberec**: Technical university in Liberec, Faculty of Economics, 2009, pp. 177-187.

[5] A. Kocmanova, M. Docekalova, P. Nemecek, I. Simberova, "Sustainability: Environmental, Social and Corporate Governance Performance in Czech SMEs", In **The 15th World Multi-Conference on Systemics, Cybernetics and Informatics**, 2011, pp. 94-99.

[6] A. Kocmanova et al. **Měření podnikové výkonnosti**, Brno: Littera, 2013.

[7] A. Kocmanová, I. Šimberová, "Modelling of Corporate Governance Performance Indicators",

Engineering Economics, Vol. 23, No. 5, 2012, pp. 485-495.

[8] A. Kocmanová, J. Hornungová, M. Klímková, **Udržitelnost: Integrace environmentální, sociální a ekonomické výkonnosti podniku**, Brno: CERM, Akademické nakladatelství, s.r.o., 2010.

[9] M. Dočekalová, "Construction of corporate social performance indicators for Czech manufacturing industry", **Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis**, Vol. 61, No. 2, 2013, pp. 309-315.

[10] B. R. Chabowski, J. A. Mena, T. L.,Gonzalez-Padron, "The structure of sustainability research in marketing, 1958-2008: a basis for future research opportunities", **Journal of the Academy of Marketing Science**, Vol. 39, No. 1, 2011, pp. 55-70.

[11] L. B. Cruz, E. A. Pedrozo, V. de Fatima Barros Estivalete, "Towards sustainable development strategies: a complex view following the contribution of Edgar Morin", **Management Decision**, Vol. 44, No. 7, 2006, pp. 871-891.

[12] M. Van Marrewijk, "Concepts and definitions of CSR and corporate sustainability: between agency and communion", **Journal of business ethics**, Vol. 44, No. 2-3, 2003, pp. 95-105.

[13] E. I. Altman, R. G. Haldeman, P. Narayanan, "ZETA analysis: A new model to identify bankruptcy risk of corporations", **Journal of Banking and Finance,** Vol. 1, No. 1, 1977, pp. 29–51.

[14] E. I. Altman, R. A. Eisenbeis, "Financial applications of discriminant analysis: a clarification", **Journal of Financial and Quantitative Analysis**, Vol. 13, No. 1, 1978, pp. 185-195.

[15] I. Neumaier, I. Neumaierová, "INFA Performance Indicator Diagnostic System. Central European Business Review", Vol. 3, No. 1, 2014, pp. 35-41.

[16] P. Mederly, P. Novacek, J. Topercer, (2003) "Sustainable development assessment: quality and sustainability of life indicators at global, national and regional level", **foresight**, Vol. 5, No. 5, pp.42 - 49

[17] K. Hrach, J. Mihola, "Metodické přístupy ke konstrukci souhrnných ukazatelů", **Statistika**, No. 5, 2006, pp. 398-418.

[18] E. I. Altman, "Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy", **The Journal of Finance**, Vol. 23, No. 4, 1968, pp. 589-609.

[19] E. I. Altman, E. Hotchkiss, **Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankurptcy, Analyze and Invest in Distressed Debt**, New York: John Wiley & Sons, 2006.

[20] E. I. Altman, **Corporate financial distress: A complete guide to predicting, avoiding, and dealing with bankruptcy**, New York: John Wiley & Sons, 1983.

[21] J. A. Ohlson, "Financial Ratios and the Probabilistic Prediction of Bankruptcy", **Journal of Accounting Research**, Vol. 18, No. 1, 1980, pp. 109-131.

[22] C. V. Zavrgen, "Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis", **Journal of Business Finance & Accounting**, Vol. 12, No. 1, 1985, pp. 19-45.

[23] M. E. Zmijewski, "Methodological Issues Related to the Estimation of Financial Distress Prediction Models", **Journal of Accounting Research**, Vol. 22, 1984, pp. 59-86.

[24] C. S. Lennox, "The Accuracy and Incremental Informatic Content of Audit Reports in Predicting Bankruptcy", **Journal of Business Finance & Accounting**, Vol. 26, No. 5-6, 1999, pp. 757-778.

[25] A. Kocmanová, I. Šimberová, "A Model of Environmental, Social and Governance Performance Indicators of a Company", In **The 17th World Multi-Conference on Systemics,m Cybernetics and Informatics**, 2013, pp. 19-24.

[26] A. Kocmanová, I. Šimberová, "Multidimensional modelling of social performance indicators in processing industry companies in the Czech Republic", **International Journal of Mathematical Models and Methods in Applied Sciences**, Vol. 7, No. 5, 2013, pp. 500-507.

Analysis of the factors influencing investments in intellectual capital

Oksana LENTJUSHENKOVA
Alberta College, 22 Skolas street
Riga, LV-1010, Latvia

and

Inga LAPINA
Riga Technical University, Faculty of Engineering Economics and Management
6 Kalnciema Street, Riga, LV-1048, Latvia

ABSTRACT

The role of intellectual capital investments on company level is increased in last thirty years. According to OECD data tangible assets investments decreased and intangible assets investments increased. Despite of these trends the amount of intellectual capital investments is not significant in Latvia. The aim of the paper is to describe the aspects for the research and survey development in order to prepare sound theoretical basis, as well as to describe and validate solid development of the research. To achieve the aim the previous research and scientific literature are studied and entrepreneurs from certain sectors of national economy are interviewed. In the result of the current research a questionnaire for entrepreneurs was developed and a data base of respondents was created.

Keywords: intellectual capital investments, factors influencing investments, enterprise.

1. INTRODUCTION

In last thirty years companies' investment structure has changed. Tangible assets investments decreased and intangible assets investments increased. For example, in USA between 1972 and 2011 tangible assets investments decreased from 12% till 8% and intangible assets investments increased from 8% till 15%. In Australia, since 1974 -75, the annual growth of investment in intellectual capital has been around 1.3 times bigger than a growth of investments in physical assets such as machinery, equipment and buildings. Also in EU countries the intangible assets investments increased. For instance, in Finland, Denmark, Sweden, France and Netherlands the largest part of enterprise's investments is in intangible assets investments [24]. In some countries the amount of intellectual capital investments is not significant because of different reasons. For example, in Latvia the share of R&D expenditures in GDP is one of the smallest in EU (0.70% in 2011) [11]. R&D intensive sector as % of all sectors was 5.4% in 2008 (9.2%- in Estonia, 7.9% in Lithuania, 12.1% in EU 28 and 20.2% in South Korea) [7]. One of the main funds for these investments is state budget, not private funds.

The main aim of the research is to determine factors influencing investments in intellectual capital at enterprises in Latvia.

The research questions are:

1) What are the factors influencing investments in intellectual capital at an enterprise in Latvia?
2) What is the important weight of each factor?

The following research hypotheses were developed by the authors:

H1: One of the disincentives of intellectual capital investments is lack of funds.
H2: Entrepreneurs will invest in intellectual capital, if they see financial benefits from these investments.

To achieve the research aim the following tasks are determined:

1) To determine most important factors positive or negative influencing intellectual capital investments;
2) To develop a questionnaire for factor determination;
3) To create a data base of respondents;
4) To collect a certain number of statements and interpret results.

The aim of the current paper is to describe the aspects for the research and survey development in order to prepare sound theoretical basis, as well as to describe and validate solid development of the research. The general research structure is described below (see Fig. 1).
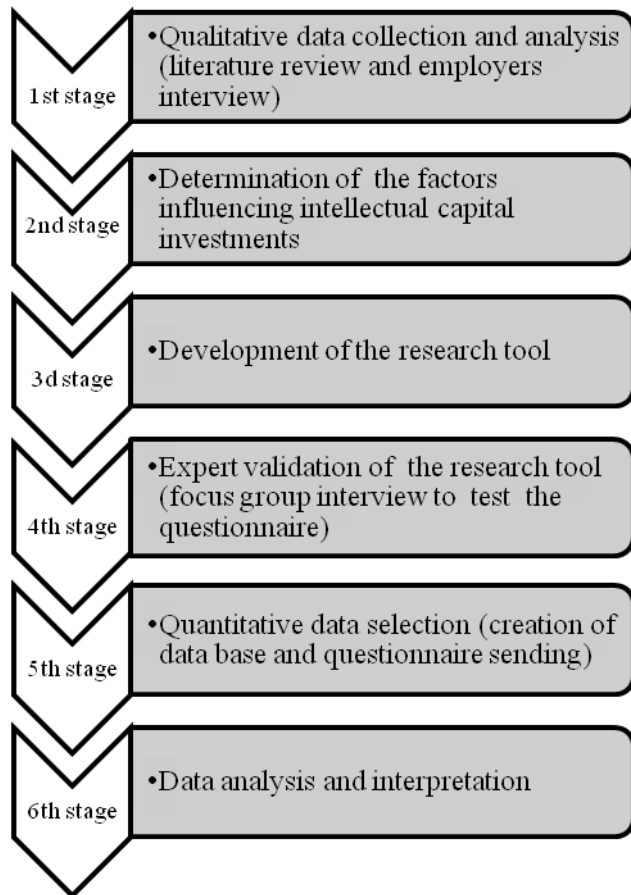

Figure 1. Research structure

The current article includes research stages from the first till the third. The next stages will be presented in the next article. The research will be made at enterprise level: the author will determine main factors, which influence enterprise intellectual capital investments.

2. LITERATURE REVIEW

There are many researches about intellectual capital investments influence on enterprise performance. Some of them disclose positive effects of investments, some research disclose a negative effect. But there are not enough researches about factors influencing intellectual capital investments.

Most of researchers are focused on the intellectual capital measurement methods or performance at the enterprise or national level.

Due to studying literature the authors found that there are several factors influencing an intellectual capital investment, which could be divided in different groups:

1) Internal factors. These factors are under enterprise control and the enterprise could change its;
2) External factors. These factors could influence the enterprise and decision making process, but the enterprise could not control or change them.

The authors determine main internal factors on the basis of previous researches:

1) Human resource qualification. Well educated and qualified employees encourage new knowledge and technology implementation at the enterprise. As a result the enterprise invests more in staff training program and technology modernization [10], [27]. At the same time in some studies it is found that investments in training have no significant effect on enterprise performance [1]. It means, that some enterprise could attract qualified employees not invest in staff training.
2) Employee's low wage. According to employees' point of view this factor could hinder investments.
3) The management and business process organization. An effective resource management, including intellectual resources, could be one of key drivers for value creation at the enterprise. Some researchers distinguish synergy and multiplier effects between intellectual capital components [8], [9], [12], [13], [19]. These effects change intellectual capital investments influence on the enterprise results. For instance, if an enterprise separately invests in technologies, there is no significant positive influence on enterprise performance. Enterprises do not have an optimal assets combination very often. There is no balanced structure of intellectual capital either. Because of these reasons the investments influence on enterprise results is not positive [23]. Some researchers tested the relationship between intellectual capital components in microfinance industry in Uganda. They conclude that positive and strong relationship exists between human capital, structural capital, relational capital and financial performance. This signifies the improvement of intellectual capital components and boosts their association with financial performance [17].

Poor management of intellectual capital decreases the efficiency of intellectual capital investments [16].

4) An enterprise ability to absorb investments. The intellectual capital investments are connected with knowledge flows to enterprise. The knowledge stocks and flows model predicts that competitive advantage depends on the continual accumulation of relevant knowledge stocks from knowledge flows [4], [6]. After a certain point, additional investments and knowledge flows may lead to diminishing returns and, as a result, firm performance. The main task for managers is to make a decision about the type and timing of knowledge flows between potential flows and existing knowledge stocks [21]. At the same time countries with greater accumulated knowledge and human capital stocks tend to see greater productivity gains than those that have smaller knowledge and human capital bases. Countries like the US and Germany achieve higher productivity gains from their pool of R&D stocks and human capital than countries such as Spain and New Zealand [20].

5) Rate of return from investments. There are many different method for calculating rate of return, for example, ROI, profit per employee, Value Added Intellectual Capital Index (VAIC index) etc. [2], [25], [26]. Many researchers use VAIC index for rate of return calculation. For VAIC index calculation simple information from statistic and balance sheet is necessary. Some authors criticize this method: in different researches it is used as intellectual capital measurement and this coefficient shows effectiveness of both capitals: intellectual and physical. For instance, employees need a certain set of resources for work. So their effectiveness depends on these resources effectiveness also.

According to literature review the main external factors are:

1) Partnership and cooperation. According to OECD report cooperation encourages experience and information exchange and declines each partner costs. The established networks increase availability of information, resources and funds. In some countries cooperation is not developed because of different reasons. For example, R&D activities in different countries are done in cooperation at national level or international level (in Estonia approx. 20% R&D activities were done within international partnership in 2006) [24].

Some enterprises develop their own research, but some use created new knowledge and technologies (in Norway the share of enterprises, who used ready R&D results, and the share of enterprises with in-house R&D is quite similar). The biggest part of innovative enterprises has in-house R&D.

2) State support for innovation and knowledge based economy. In different countries governments use direct support and indirect support for R&D. For example, in Canada indirect support through tax policy is bigger, but in Iceland government uses only direct support.

3) Disclosure recommendations about reports preparation on of intellectual capital at the enterprise. There are many different enterprise intellectual capital measurements. For example, MERITUM, EFFAS, InCas, ESG guidelines, GRI guidelines, Danish guidelines etc. These guidelines are voluntary and enterprises do not use them without certain needs. However, since 2009 Danish Commerce and Companies Agency has required the country's largest companies, state-owned enterprises, and institutional investors to state in their annual reports whether they have corporate responsibility policies and how they implement them. Since 2007 Sweden government has required all state-owned companies to produce sustainability reports in accordance with the GRI Guidelines [3]. More than 6000 organizations over the world use these Guidelines In April 2014 the European Parliament adopted the long-awaited Directive on the disclosure of non-financial and diversity information by certain large companies. It means that more than 6000 EU large companies will prepare reports and become more transparent for society.

Availability of information on intellectual capital could encourage investments. Potential investors need this kind of information. For example, investors of 23 countries examined in one study were shown to access ESG metrics provided by Bloomberg an estimated 34 million times in only two quarters of 2011; there has been a significant increase in the number of times this data is accessed. Investors in Switzerland, the United Kingdom, Canada, and Spain are the most interested. In contrast, investors in the Netherlands, Hong Kong, China, Belgium, and Finland are the least interested [5].

4) Accountancy standards. The valuation of intellectual capital investments within

accountancy framework raises several problems relating to their identification, measure and control [28]. According to IASB rules, the accounting treatment of internally generated intangible assets is less rigid but it remains deficient. Indeed, IAS 38 details necessary stages for the creation of an intangible asset by specifying at every stage whether it is possible to predict future economic benefits associated with the asset. The development phase allows an entity to bring proof of existence of an intangible asset's capacity to generate revenue. In this setting, the development costs must be capitalized provided such costs comply with certain conditions. These conditions are: (a) the technical feasibility of completing the intangible asset so that it will be available for use or sale; (b) its intention to complete the intangible asset and use or sell it; (c) its ability to use or sell the intangible asset; (d) how the intangible asset will generate probable future economic benefits; (e) the availability of adequate technical, financial and other resources to complete the development and to use or sell the intangible asset; and(f) its ability to measure reliably the expenditure attributable to the intangible asset during its development [15].

5) An enterprise location. An enterprise location and concentration in one region influence human capital investments decision. If enterprise (competitors) concentration in one region is high, the effectiveness of investments will be lower. The competitors could entice the enterprise employees and the return from investments will be low [22].
6) Protection of intellectual property rights. This factor hinders investments if there is no mechanism for implementation of law in the country and consumer culture and income are at a low level.

The literature review results show, that the factors influencing intellectual capital investment are studied from different points of view. The authors described the factors more often studied in previous researches.

3. FACTORS INFLUENCING INTELLECTUAL CAPITAL INVESTMENTS

According to literature review and employers interview the authors have determined factors influencing intellectual capital investments. All factors are divided into two main groups and subgroups as follows:

1) Factors which are related to financial benefits:
   a. Internal factors: rate of return, profitability, enterprise value, availability of resources, availability of funds;
   b. External factors: state financial support, non-state funds availability, tax policy, accountancy standards (opportunity to include investments into balance sheet)
2) Factors which are related to non-financial benefits:
   a. Internal factors: productivity, availability of infrastructure, business process organization, motivation and loyalty, product quality, competitive advantage;
   b. External factors: protection of intellectual property rights, intellectual capital obligatory reports, economic development in the country, partnership and networks.

These factors will be used for research tool design.

4. DEVELOPMENT OF THE RESEARCH TOOL

The authors select a questionnaire as a research tool for research aim achieving. The questionnaire consists of few sections:

Section A: Respondent profile. The authors used some criteria for respondents. The main criteria are: the sector of national economy; number of employees.

Section B: information about activities in the intellectual capital investments field. The purpose of this section is to disclose the aims of the intellectual capital investment, main investments objects and sources of funds.

Section C: statements about which factors encourage investments.

Section D: statements about which factors hinder investments.

Sections C and D are designed on the basis of literature review. To evaluate each statement respondents will be offered to use 4-point Lykert type scale. The opportunities for evaluation will be: 1 – not important, 2 – relatively important, 3 – average importance, 4 – very important.

In order to test the developed questionnaire the authors will organize a focus group interview. The participants of the interview will be representatives of employers from certain sectors of economy

(managers) and researchers from Riga Technical University and Alberta College.

The authors will research factors influencing intellectual capital investments in two sectors of national economy: industry (B-E according to NACE 2 red.) and wholesale and retail trade (G according to NACE 2 red.).

The respondent's data base is created on the basis of Latvian Chamber of Commerce and Industry (LCCI). Representative number of respondents is calculated as follows:

$$n = \frac{t^2 w(1-w)N}{\Delta^2 N + t^2 w(1-w)}$$

(1)

where n – representative number of respondents; t – 1,96 (if validity is 95%); w – relative frequency 0,5; N – general sample; Δ - sampling error [14].

The estimated general sample is 920 respondents. The calculated representative number of respondents is 271 enterprises. It will be enough for survey results validity [18].

5. CONCLUSIONS

Nowadays the role of intellectual capital on company level has increased. Intellectual capital, its amount and quality become a key driver for sustainable development. For creating and effective usage of intellectual capital investments are needed.

The decision about intellectual capital investments depends on different factors. Most of researchers study intellectual capital investments on enterprise performance. The number of researches about factors influencing investments is limited.

Due to literature review the authors determine internal and external factors. Designing the research tool the authors divided the factors into two main groups: related to financial performance and related to non-financial performance. Each group was divided into two subgroups: internal and external factors (from the enterprise point of view).

The authors developed a theoretical basis for the survey and described the development of the research about factors influencing intellectual capital investments.

The next stage of current research is quantitative data collection and survey preparation. The survey results will be used for decision making about intellectual capital investments methodology development at the enterprise.

6. REFERENCES

[1] S. Boujelben, H. Fedhila, "The Effects of Intangible Investments on Future OCF", **Journal of Intellectual Capital**, Vol. 12, No. 4, 2011, pp. 480 - 494.

[2] M. Clarke, D. Seng, R.H. Whiting, "Intellectual Capital and Firm Performance in Australia", **Journal of Intellectual Capital**, Vol. 12, No. 4, 2011, pp. 505 – 530.

[3] **Corporate Reporting of Intangible Assets: A Progress Report**, OECD, 2012.

[4] D. M. DeCarolis, D.L. Deeds, "The Impact of Stocks and Flows Of Organizational Knowledge on Firm Performance: An Empirical Investigation of the Biotechnology Industry", **Strategic Management Journal**, Vol. 20, No. 10, 1999, pp. 953–968.

[5] R. G. Eccles, G. Serafeim, **The Role of the Board in Accelerating the Adoption of Integrated Reporting**, USA: Conference Board, 2011.

[6] Z. Erden, D. Klang, R.Sydler, G. von Krogh, "Knowledge-flows and Firm Performance", **Journal of Business Research**, Vol. 67, 2014, pp. 2777 - 2785.

[7] **Europe 2020 Targets: Research And Development,** European Commission, 2014.

[8] E. Gaile-Sarkane, "Semi management theory and business sophistication", **WMSCI 2012 - The 16th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings**, Volume 1, 2012, USA, Florida, Orlando, pp. 25 – 30.

[9] E. Gaile-Sarkane, D. Ščeulovs, "Information technologies for sustainable business development: Case of Latvia", **WMSCI 2011 - The 15th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings**, Volume 1, 2011, USA, Florida, Orlando, pp. 76 – 81.

[10] N. Gemmell, "Evaluating the Impacts of Human Capital Stocks and Accumulation on Economic Growth: Some New Evidence", **Oxford Bulletin of Economics and Statistics**, Vol. 58, 1996, pp. 9 – 28.

[11] **Gross domestic expenditure on R&D**, Eurostat, 2014.

[12] C. Huang, C. Liu, "Exploration for the Relationship between Innovation, IT and Performance", **Journal of Intellectual Capital**, Vol. 6, No. 2, 2005, pp. 97 – 119.

[13] C. Huang, M. Wang, M., "The Effects of Economic Value Added and Intellectual Capital on the Market Value of Firms: An Empirical Study, **International Journal of Management,** Vol. 25, No. 4, 2008, pp. 722 – 731.

[14] **Ievads pētniecībā: stratēgijas, dizaini, metodes**, Rīga:RaKa, 2011.

[15] **International Accounting Standards. (IAS 38 Intangible Assets)**, IFRS, 2012.

[16] M. Joshi, D. Cahill, J. Sidhu, M. Kansal, "Intellectual Capital and Financial Performance: an Evaluation of the Australian Financial Sector", **Journal of Intellectual Capital**, Vol. 14, No. 2, 2013, pp. 264 – 285.

[17] N. Kamukama, A. Ahiauzu, J.M. Ntayi, "Intellectual Capital and Performance: Testing Interaction Effects", **Journal of Intellectual Capital**, Vol. 11, No. 4, 2010, pp. 554 – 574.

[18] **KEYS: to Creativity and Innovation: User Guide**, Centre of Creative leadership, 2002.

[19] N. Koleda, N. Lace, G. Kaliyeva, "Assurance of business viability through equity sufficiency", **WMSCI 2013 - The 17th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings**, Volume 2, 2013, USA, Florida, Orlando, pp. 25 – 30.

[20] K.B. Luintel, M. Kan, K. Theodoridis, **How Robust is the R&D Productivity Relationship? Evidence from OECD Countries**, Cardiff Economics Working Paper, No.E2010/7

[21] C.A. Maritan, M.A. Peteraf, "Invited Editorial: Building a Bridge between Resource Acquisition and Resource Accumulation", **Journal of Management**, Vol.37, 2011, pp. 1374 – 1389.

[22] N. Matoushek, F. Robert-Nocoud, "The Role of Human Capital Investments in the Location Decision of Firms", **Regional Science and Urban Economics**, Vol.35, No. 5, 2005, pp. 570 – 583.

[23] I. Naidenova, P. Parshakov, "Intellectual Capital Investments : Evidence from Panel VAR Analysis", **Journal of Intellectual Capital**, Vol. 4, No. 13, 2013, pp. 634 – 660.

[24] **Supporting Investment in Knowledge Capital, Growth and Innovation**, OECD Publishing, 2013.

[25] P.P. Phillips, J.J. Phillips, **Return on investments (ROI) basic,** Aleksandria: ASTD Press, 2005.

[26] A. Pulic, "VAICTM – an Accounting Tool for IC Management", **International Journal Technology Management**, Vol. 20, No. 5,6,7,8, 2000, pp. 702-714.

[27] S. Redding, "The Low- Skill, Low-Quality Trap: Strategic Complementarities between Human Capital and R&D", **The Economic Journal**, 106 (March), 1996, pp. 458 – 470.

[28] D. Zéghal, A. Maaloul, "The Accounting Treatment of Intangibles – A critical Review of the Literature", **Accounting Forum**, Vol. 35, 2011, pp. 262–274.

MACROECONOMIC DETERMINANTS OF GOING PUBLIC: THE EVIDENCE FROM POLAND

Tomáš Meluzín
Faculty of Business and Management, Brno University of Technology
Kolejní 2906/4, 612 00 Brno Czech Republic

and

Marek Zinecker
Faculty of Business and Management, Brno University of Technology
Kolejní 2906/4, 612 00 Brno Czech Republic

and

Justyna Łapińska
Faculty of Economics and Management, Nicolas Copernicus University
ul. Gagarina 13a, 87 - 100 Toruń, Poland

ABSTRACT

The main objective of this study is to indicate the influence of local macroeconomic factors, consequently GDP growth rates, the reference interest rate, industrial production growth rates, Warsaw Stock Exchange Index (WIG) returns and the volume of private equity investments, on the number and value of capital raised by IPOs in an emerging market, Poland, over the period of 2004 to 2012. Our sample includes only local enterprises that conducted an IPO on the Main Market of the Warsaw Stock Exchange. Financial and privatized companies are excluded from the data processing. The Spearman correlation analysis is used to identify the relations between the macroeconomic (external) determinants and the number of listings and the value of capital raised by IPOs. The data is evaluated at the significance level of $\alpha=5\%$. The data is processed in two ways. Firstly, dependent variables ($X_1…X_4$) and explanatory variables ($Y_1…Y_5$) are analysed simultaneously, next, the explanatory variables are lagged by one year. The findings of the simultaneously data analysis can be summarized as follows: A significant difference was found to exist between real GDP growth rates and the number of IPOs and between industrial production growth rates and the number of listings. GDP growth rates have also a significant positive influence on the value of capital raised by IPOs. In addition to that, there exists a statistically significant positive relationship between the industrial production growth rates and median IPO value. Surprisingly, there is just a little empirical evidence that a relationship between explanatory a dependent variables exists when the macroeconomic indicators are lagged by one year. Only the pattern of volatility of WIG index returns lead the median value of capital raised IPOs. The correlation coefficient between real GDP growth rates and median IPO value is even negative when we shift dependent variables relative to the explanatory variables by one year.

Key Words: Initial Public Offerings (IPOs), Macroeconomic Factors, External Determinants, Central and Eastern Europe (CEE), Poland.

1. INTRODUCTION

Based on the extensive literature review there are some key internal and external factors that have significant effect on the financing choice of enterprises via IPOs.

Relations between the number of IPOs and the macroeconomic factors (i.e. external conditions) have been investigated in a relatively limited number of studies. Loughran et al. (1994) investigate the timing of IPOs in fifteen countries in relation to inflation-adjusted stock price indexes and GDP growth rates. The results suggest a positive relationship between the number of IPOs and stock price levels, however no positive correlation with the cycle movements. Rydqvist and Högholm (1995) compare the data for a sample of family-owned enterprises in Sweden (1970–1991) and eleven European countries (1980-1989). They find that "most going public activity took place after an exceptionally sharp stock price increase, and that going public activity is not related to the business cycle". Ljungqvist (1995) suggests that high number of IPOs is positively correlated with both high stock index levels and good business conditions and tends to follow phases of extensive IPO underpricing. Breinlinger and Glogova (2002) investigate the explanatory power of selected macroeconomic factors influencing IPOs by analysing a data set of annual IPO volumes for six developed continental European countries over a time period of 18 years. The authors followed the question if there are stable indications that IPOs depend on stock index returns for what they termed consolidated periods. The results show that a "logarithmic transformation of IPO volumes (representing authors´ supposition of a nonlinear relationship between IPO volumes and stock index returns) leads to persistently

significant estimates for both pooled and individual country regressions". The hypothesis that percentage changes in savings, GDP growth and interest rates have explanatory power for IPO volumes could not be supported by empirical evidence.

A paper by Ameer (2012) shows a significant negative relationship between the interest rate and the number of IPOs and a significant positive relationship between the industrial production and the number of IPOs in the emerging market of Malaysia. Bilson et al. (2002) find a moderate evidence to support the connection between local macroeconomic factors and stock returns in emerging markets.

Microeconomic studies deal first of all with the motivation of going public. The main factors taking influence on decision to go public in a firm are divided in both academic and professional literature into four groups.

Authors in the first group point out the situation of a firm which needs to obtain external funds to undertake external net present value projects. The primary securities market offers an opportunity to raise capital from a large number of previously unknown investors. As a result, more capital is accumulated than a single investor, or a limited number of investors, would be able or willing to provide. Raising funds through an IPO should be an alternative to borrowing, particularly in companies with high investments (now or in the foreseeable future), high proportion of debt in the capital structure, and high potential for growth (Chemmanur and Fulghiery, 1999; Ritter and Welch, 2002). Pagano et al. (1998) investigated a comprehensive data set of Italian companies. They conclude that U.S. companies usually undergo a considerable growth process after listing while the decision of Italian independent companies can be interpreted as "an attempt to rebalance their balance sheet after large investments and growth. Brau and Fawcett (2006) investigated 336 nonfinancial U.S. companies that had successfully completed an IPO or attempted and subsequently withdrew an IPO. More than a half of the interviewed CFOs strongly supported the notion that "an IPO serves to create public shares for use in future acquisitions".

Pagano et al. (1998) followed by Black and Gilson (1998) connect IPOs with another reason. Public trading of shares provides a great advantage to both the issuers, to whom the shares issued provide a long-term source of financing, and to the investors, who can sell the shares purchased at any time on secondary markets and thus recover the desired liquidity, i.e. the money they invested. The short-term financial funds of individual investors are thereby transformed into long-term sources, which then make it possible to implement large-scale investment projects. Raising the stock capital through a public issue also eliminates the difference between the large shareholders, who tend to take a long-term approach to investing, and the small ones, who value liquidity and are therefore inclined to take a short-term view of their investment.

Zingales (1995) and Black and Gilson (1998) argue that going public is the way how the majority shareholder's desire to reduce his stake in the company. Thereby, an IPO allows venture capitalists to cash out and resolves the problem of generational succession in a family-run enterprise. From the shareholder's perspective, the option to cash out his co-ownership by selling the shares of stock at any time on the secondary market is an advantage that imparts a great deal of

flexibility to his financial decisions. On the other hand, Brau and Fawcett (2006) findings do not support the public statement of many venture capitalists that an IPO is an integral part of their harvest strategy because "firms with VC presence rank four motivations higher than the opportunity to allow VCs to cash-out".

The last reason for going public is to gain a non-financial advantage from IPO implementation (Maksimovic and Pichler, 2001; Ježek, 2004). Going public is therefore associated with positive effects in the area of marketing. First of all, an IPO can increase the publicity or reputation of the firm. The prestige can be very advantageous in recruiting key employees as well as marketing products and services. Brau and Fawcett (2006) point out that the motivation for going public in the form of firm reputation enhancing and analysts' attention attracting is significant for "smaller, younger, high-tech, and VC-backed firms".

The motivation of this study is to make a contribution to the academic literature by addressing the question if local macroeconomic factors have any influence on the number of IPOs in Poland and the value of capital raised by enterprises in this market. The number of variables and the time period covered in this paper are greater than those considered in previous Polish IPO studies (Sejkora,2013; Brzeszczynski, 2014; Jargot, 2006 and Dudko – Kopczewska, 2004).

This study is organized as follows. The second part represents the problem formulation and deals with research design, i.e. data and methodology. The third part presents the empirical research results. The last section summarizes and provides concluding remarks.

2. MATERIAL AND METHODOLOGY

This paper addresses the issue whether local macroeconomic factors have any influence on the number of IPOs and the value of capital raised by IPOs in the emerging market of Poland. The nature of this study is based on the theory and previous empirical research. All macroeconomic indicators analyzed in this paper have sufficient support in the finance academic studies (Ameer, 2012, Breininger and Glogova, 2002, Brzeszczynski, 2014, La Porta et al., 1997).

For purposes of this paper the following hypotheses have been outlined:

Hypothesis 1: There is a positive relationship between GDP growth rates and a) the number of IPOs, b) the capital raised by IPOs.

La Porta et al. (1997) assess the influence of economic conditions (namely the legal system) on the number of IPOs using a sample of 49 countries. As reported by Breinlinger and Glogova (2002) the La Portas et al. research results show that "the quality of law enforcement, which is highly correlated with the level of GDP per capita, has a strong positive effect on the number of IPOs". The authors identify a statistically significant relationship between long-term GDP growth rates, i.e. average annual percentage growth of per capita GDP for the period 1970 to 1993, on the number IPOs. On the other hand, the studies conducted by Rydqvist and Högholm (1995) and Loughran et al. (1994) show that the GNP short-term growth rates are no

significant explanatory power for IPO activity across the sample of European countries. Also Breininger and Glogova´s analysis (2002) of annual IPO volumes for six continental European countries over a time period of 18 years could not support the hypothesis that GDP growth rates have explanatory power for IPO volumes.

Hypothesis 2: There is a negative relationship between the reference interest rate and a) the number of IPOs, b) the capital raised by IPOs.

Rees (1997), concentrating on UK data, found no significant link between the number of IPOs and interest rates. Research results by Breininger and Glogova (2002) also indicate that there is no perceivably influence of interest rates (ten-year government bond yields) on demand for raising equity through IPOs. On the contrary, the study published by Ameer (2012) reports the opposite. Ameer´s results (2012) imply that "monetary policy has a direct impact on capital markets and that central bank intervention propagates IPO cycles in Malaysia". Based on a paper by Jovanovic and Rousseau (2004) Ameer (2012) supposed a negative relationship between interest rate and the number of IPOs.

Hypothesis 3: There is a positive relationship between industrial production growth rates and a) the number of IPOs, b) the capital raised by IPOs.

The industrial production index as a measurement of the output of an economy also helps to map structural economy development (Hosley et al., 1985). The industrial production index rate is the indication of business lifecycle and business life cycle affects by its fluctuations the stock market prices (Moody et al., 1993). Besides, authors say that enterprises enter capital markets when other enterprises enter them too, meaning potentially higher overall industrial production (Choe et al. 1993, Lowry and Schwert, 2002).

Hypothesis 4: There is a positive relationship between the stock market index returns and a) the number of IPOs, b) the capital raised by IPOs.

Stock markets around the world are interconnected through the communication channels and information can spread very quickly among investors. Pessimism on stock markets predicts downward pressure on market prices and on the other hand optimism or low amount of pessimism predicts higher stock market trading volume and higher returns (Tetlock, 2007). The pessimism and optimism which affects stock markets is in alignment with investor sentiment theory and the market timing theory. As the stock market index mirrors the investor´s willingness to invest or not, the number of IPOs vary accordingly. Enterprises are more likely to implement IPOs when the stock market promises higher returns and therefore profit for enterprises and also for potential investors. Studies by Loughran et al. (1994), Ljungqvist (1995), Rees (1997) and Rydqvist and Högholm (1995) detect a significantly positive influence of stock index levels and stock index returns on the number of IPOs. Brzeszczynski (2014) analyzed the number of new IPOs and the main stock market index (WIG) returns for the Polish stock market over a period from 1997 to 2008. He detects the correlation coefficient between those two variables 0.0244 when IPOs and stock market index return are analyzed simultaneously. However, the value of this index is 0.5683

when the WIG returns are lagged by one year. Brzeszczynski (2014) concluded that "the number of IPOs in emerging markets and the profitability of the public offers are related to macroeconomic conditions, business cycles and stock market activity. In most emerging market countries there is a time lag between movements of the stock market index and decisions to launch new IPOs".

Hypothesis 5: There is a positive relationship between private equity investments volume and a) the capital raised by IPOs.

Private equity investors view emerging markets as a suitable opportunity to diversify their investment portfolios and to catch excess risk premiums (Errunza, 1983). In the financial literature the going public strategy has been considered for an important channel how venture capitalists can leave investee companies. Cumming and Johan (2012) present the evidence about the number of IPO backed exits from first round (new) investments in seed, early and expansion stage investments in Canada, USA, China, the United Kingdom, France, Israel, India and Germany as a percentage of total worldwide IPO exits in the period 1990 - 2010. While the share of IPOs on the U.S. marked has declined since 1995, China's share on the IPO market in terms of numbers of VC backed IPOs has significantly increased relative to the U.K., France, Israel, India, Germany and Canada in the last decade.

This paper is based on evidence from the Polish capital market over the period of 2004 to 2012. Our sample includes only local enterprises that conducted an IPO on the Main Market of the Warsaw Stock Exchange. Financial and privatized companies are excluded from the data processing. The IPO data were obtained from Warsaw Stock Exchange. Macroeconomic data such as gross domestic product growth rates (GDPGR), reference interest rates (GBY), industrial production growth rates (IPGR), Warsaw stock exchange index (WIG) and private equity investments (PEI) were obtained from the Polish National Bank (NBP), Polish Central Statistical Office (PCSO), OECD Stat Extracts and European Venture Capital Association (EVCA).

Tables 1 and 2 show the source data, i.e. an overview of the variables used for this study.

The following steps were undertaken to analyze the quantitative data:

Firstly, we provide the Spearman correlation analysis in order to investigate the influence of macroeconomic factors on the number of IPOs and value of capital raised by IPOs. The data is evaluated at the significance level of $\alpha=5\%$ and processed in two ways. Dependent variables ($X_1…X_4$) and explanatory variables ($Y_1…Y_5$) are analysed simultaneously, next, the explanatory variables are lagged by one year. The entire statistical evaluation was performed by Statistica.CZ, Version 9.

Secondly, the results of the data analysis are discussed in order to draw some specific issues existing in the Polish primary capital market.

Finally, the research results are compared with the results of similar studies conducted under conditions of developed and emerging capital markets.

3. EMPIRICAL FINDINGS

Table 3 gives an overview of descriptive statistics of annual IPOs´ time series and macroeconomic indicators. A total of 218 IPOs were listed during 2004-2012 period on the Main Market of the Warsaw Stock Exchange including only local enterprises without financials and privatizations. The total value of capital raised by companies on this market was 4,301 million EUR. Prior the financial and economic crisis (2008-2009), there was a remarkable boom in the number of IPOs which increased from 26 in 2004 to 68 in 2007. The sharp decline in the number of IPOs in 2008 was followed by a moderate recover of the market with 16 IPOs in 2010. The average GDP growth rate was 4.32 % and reached its peak level of 6.80 % in 2007 although there were remarkable values of this indicator during the whole period 2006-2008. The average industrial production growth rate was 6.18 % with its highest level of 12.70 % in 2004. Spectacular values of this indicator can be also observed in 2006 and 2007. The only negative industrial production rate was registered in 2009 (-3.80 %). The slight downward movement of the ten-year government bond yields displays the loose monetary policy executed by the Polish National Bank since 2009. Average changes in the WIG index returns, the index of the Main List, were 14.84 % while its maximal negative value of -51.07 % was reached in 2008 and its maximal positive value of 46.85 % was reached in 2009. Poland attracted the highest level of private equity investments within the CEE countries in terms of both the number of deals and the value of invested capital. The average annual value of capital in-flow reached 0.13 % of GDP in the period 2004-2012.

Table 3: IPOs in the Polish Capital Market - Descriptive Statistics, 2004-2012

Note: IPOs on the Main Market of the Warsaw Stock Exchange, without Financials and Privatizations, only Locals

	Mean	Median	Std	Min	Max
IPO Activity Indicators					
Number of Listings/IPOs	24.22	24.00	16.65	8.00	67.00
Value of IPOs, € Million	477.89	425.50	362.69	54.80	1340.60
Macroeconomic Indicators					
% GDPGR	4.32	4.50	1.67	1.60	6.80
% IPGR	6.18	6.70	5.32	-3.80	12.70
% GBY	5.75	5.78	0.56	5.00	6.90
% WIG	14.84	26.24	29.90	-51.07	46.85
% PEI	0.13	0.13	0.06	0.05	0.22

Source: Own processing

Table 4 reports results of the Spearman correlation analysis when dependent variables ($X_1…X_4$) and explanatory variables ($Y_1…Y_5$) are processed simultaneously. Table 5 shows the results of the same analysis when dependent variables ($X_1…X_4$) are lagged by one year. Both tables illustrate the strength and significance of investigated relations.

Using the Spearman correlation analysis a significant difference was found to exist between the number of IPOs and real GDP growth rates at the 5% level of significance (Rs=0.912;

p=0.001); thus H1 is supported by the research results. Second, the Spearman correlation analysis results support also H3, i.e. there is a positive relationship between the number of listings and industrial production growth rates (Rs=0.711; p=0.032). Next, there is a positive relationships between the median IPO value and GDP growth rates (Rs=0.683; p=0.042). Finally, the findings imply that a statistically significant positive relationship exists between the median IPO value and industrial production growth rates (Rs=0.900; p=0.001).

An important feature of the table 5 is the one-year shift of new IPOs relative to the macroeconomic indicators, where the pattern of volatility of WIG index returns obviously lead the median value of capital raised by enterprises on the Main Market of the Warsaw Stock Exchange (Rs=0.867; p=0.002). Surprisingly, the correlation coefficient between median IPO value and real GDP growth rates is negative (Rs=-0.795; p=0.010).

The findings show that the decision about going public and the value of raised capital are dependent on macroeconomic conditions, i.e. that IPOs tend to increase when GDP and industrial production are rising. The relationship is simultaneous and no lagged effects could be observed. The hypothesis that the reference interest rate (H2), the Warsaw Stock Exchange Index WIG (H4) and private equity investments (H5) have explanatory power for IPO numbers and value of capital raised in the Polish capital market could not be supported by empirical evidence.

4. DISCUSSION AND CONCLUSION

In this paper we investigate the influence of macroeconomic factors, consequently GDP growth rates, the reference interest rate, industrial production growth rates, Warsaw Stock Exchange Index (WIG) returns and the volume of private equity investments, on the number and value of capital raised by IPOs in an emerging market, Poland, over the period of 2004 to 2012. Previous investigations of this issue conducted under conditions in terms of both developed and emerging countries show no consistent results regarding the explanatory power of macroeconomic indicators and the number of IPOs and the value of capital raised by this manner of corporate financing.

The research results show that the GDP growth and industrial production have significant impacts on the number of IPOs. Furthermore, Spearman correlation analysis results confirm that there exists a positive relationship between GDP growth rates and the median value of capital raised by IPOs. This study implies that the business cycle has a direct impact on the IPO activity in the Polish capital market. Thus, our research results are similar to those of existing studies (Ameer, 2012; Lowry and Schwert, 2002; Choe et al. 1993) and support the conclusion of Brzeszczynski (2014) that "any decision to launch an IPO should be very carefully analysed using not only past financial data for the company, but also macroeconomic forecasts. Poor timing may results in the loss of capital if stock market prices are too low. "

Table 1: Source Data

Note: IPOs on the Main Market of the Warsaw Stock Exchange, without Financials and Privatizations, only Locals

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012
Explanatory Variables (Yt)									
% Real GDP Growth Rates (Y1)	5.30	3.60	6.20	6.80	5.10	1.60	3.90	4.50	1.90
% Industrial Production Growth Rate (Y2)	12.70	3.60	12.00	9.40	2.70	-3.80	11.10	6.70	1.20
% Ten-Year Government Bond Yields (Y3)	6.90	5.22	5.23	5.48	6.07	6.12	5.78	5.96	5.00
% Change in Stock Market Index (WIG) Returns (Y4)	27.94	33.66	41.60	10.39	-51.07	46.85	18.77	-20.83	26.24
Private Equity Investment as % of GDP (Y5)	0.07	0.05	0.12	0.22	0.18	0.08	0.19	0.18	0.13
Dependent Variables (Xt)									
Number of Listings/IPOs (X1)	26	24	29	67	24	8	16	14	10
Annual IPO Value, € Million (X2)	535.6	378.1	670.7	1340.6	425.5	54.8	174.2	151.3	570.2
Average IPO Value, € Million (X3)	20.6	15.8	23.1	20.3	17.7	6.9	10.9	10.8	57.02
Median Value of IPOs, € Million (X4)	10.0	6.7	10.4	8.2	3.0	2.5	11.1	9.9	3.8

Source: OECD Stat Extracts, Warsaw Stock Exchange Fact Books, European Venture Capital Association (EVCA) Yearbooks, Polish National Bank Statistics

Table 2: Analyzed Variables

	Data Sources	Calculation	Expected Sign
Dependent Variables			
Number of Listings/IPOs	Warsaw Stock Exchange Fact Books	frequency	
Annual IPO Value, Average IPO Value, Median IPO Value	Warsaw Stock Exchange Fact Books	€ Million	
Explanatory Variables			
% Real GDP Growth Rates (GDPGR)	OECD Stat Extracts, Eurostat, Polish Statistical Office	annual growth rates in % (using yearly closing dates)	+
% Industrial Production Growth Rate (IPGR)	OECD Stat Extracts, Eurostat, Polish Statistical Office	annual growth rates in % (using yearly closing dates)	+
% Ten-Year Government Bond Yields (GBY)	Polish National Bank (NBP), OECD Stat Extracts,	in % (using yearly closing dates)	-
% Change in Stock Market Index Returns (WIG)	Warsaw Stock Exchange Fact Books	in % (using yearly closing dates)	+
Private Equity Investment as % of GDP (PEI)	European Venture Capital Association (EVCA)	in % (using yearly closing dates)	+

Source: Own processing

Table 4: Relations between Explanatory and Dependent Variables (Simultaneous Analysis)

Spearman correlation analysis				
Variables	Nr of Observations	Spearman R	t(N-2)	p-value
X1 & Y1	9	**0.912**	5.89	**0.001**
X1 & Y2	9	**0.711**	2.68	**0.032**
X1 & Y3	9	-0.075	-0.20	0.847
X1 & Y4	9	-0.117	-0.31	0.764
X1 & Y5	9	0.130	0.35	0.738
X2 & Y1	9	**0.683**	2.48	**0.042**
X2 & Y2	9	0.400	1.15	0.286
X2 & Y3	9	-0.433	-1.27	0.244
X2 & Y4	9	-0.083	-0.22	0.831
X2 & Y5	9	0.192	0.52	0.620
X3 & Y1	9	0.417	1.21	0.265
X3 & Y2	9	0.317	0.88	0.406
X3 & Y3	9	-0.450	-1.33	0.224
X3 & Y4	9	0.033	0.09	0.932
X3 & Y5	9	-0.059	-0.16	0.881
X4 & Y1	9	0.500	1.53	0.170
X4 & Y2	9	**0.900**	5.46	**0.001**
X4 & Y3	9	-0.067	-0.18	0.865
X4 & Y4	9	-0.033	-0.09	0.932
X4 & Y5	9	0.201	0.54	0.604

Source: Own processing

Table 5: Relations between Explanatory and Dependent Variables (One Year Shift of Dependent Variables Relative to the Explanatory Variables)

Spearman correlation analysis				
Variables	Nr of Observations	Spearman R	t(N-2)	p-value
X1 & Y1	9	0.101	0.27	0.796
X1 & Y2	9	0.393	1.13	0.295
X1 & Y3	9	-0.622	-2.10	0.074
X1 & Y4	9	-0.653	2.28	0.057
X1 & Y5	9	-0.471	-1.41	0.201
X2 & Y1	9	0.134	0.36	0.731
X2 & Y2	9	0.233	0.63	0.546
X2 & Y3	9	-0.653	-2.28	0.057
X2 & Y4	9	0.317	0.88	0.406
X2 & Y5	9	-0.301	-0.84	0.431
X3 & Y1	9	-0.042	-0.11	0.915
X3 & Y2	9	0.050	0.13	0.898
X3 & Y3	9	-0.477	-1.44	0.194
X3 & Y4	9	0.167	0.45	0.668
X3 & Y5	9	-0.268	-0.74	0.486
X4 & Y1	9	**-0.795**	-3.47	**0.010**
X4 & Y2	9	-0.217	-0.59	0.576
X4 & Y3	9	-0.159	-0.43	0.683
X4 & Y4	9	**0.867**	4.60	**0.002**
X4 & Y5	9	-0.619	-2.09	0.075

Source: Own processing

We could not confirm any significant lagged effects concerning the relationship between explanatory and dependent variables. In this context our research results differ from Brzeszczynski (2014) who observed a time lag between movements of the stock market index WIG returns and decisions to launch new IPOs for the Polish capital market over the period from 1997 to 2008.

A further research on the issues addressed in this paper assumes, firstly, a modification of explanatory variables (i.e., saving deposits and pension models), secondly, a modification of the methodology outlined in studies by Ameer (2012) and Breinlinger and Glogova (2002).

ACKNOWLEDGEMENTS

REFERENCES

[1] R. Ameer, "Macroeconomic Factors and Initial Public Offerings in Malaysia", **Asian Academy of Management Journal of Accounting and Finance**, Vol. 8, No. 1, 2012, pp. 41-67.

[2] C. M. Bilson, T. J. Bailsford, V. J. Hooper, "Selecting macroeconomic variables as explanatory factors of emerging stock market returns", **Pacific-Basin Finance Journal**, Vol. 9, No. 4, 2002, pp. 401–426.

[3] B. S. Black, R. J. Gilson, "Venture capital and the structure of capital markets: Banks versus stock markets", **Journal of Financial Economics**, Vol. 47, No. 3, 1998, pp. 243–277.

[4] J. C. Brau, S. E. Fawcett, "Initial Public Offerings: An Analysis of Theory and Practise", **Journal of Finance**, Vol. 61, No. 1, 2006, pp 399–436.

[5] L. Breinlinger, E. Glogova, **Determinants of Initial Public Offerings: A European Time-Series Cross-Section Analysis.** Oesterreichische Nationalbank: Working Paper, Financial Stability Report, 2002.

[6] J. Brzeszczynski, "IPOs in Emerging Markets", **QFinance**, 2014, available at http://www.qfinance.com/financing-best-practice/ipos-in-emerging-markets?full

[7] T. J. Chemmanur, A. Fulghiery, "Theory of the Going-Public Decision", **Review of Financial Studies**, Vol. 12, No. 2, 1999, pp. 249–279.

[8] H. Choe, R. W. Masulis, V. Nanda, "Common stock offerings across the business cycle: Theory and evidence", **Journal of Empirical Finance**, Vol. 1, No. 1, 1993, pp. 3-31.

[9] J. D. Cumming, S. Johan, **The IPO as an Exit Strategy for Venture Capitalists: Regional Lessons from Canada with International Comparisons.** Handbook of Research on IPOs, 2012.

[10] K. Dudko Kopczewska, "Primary-factor analysis of share issues on the Warsaw Stock Exchange", **Bank and Credit**, No. 1, 2004, pp. 31-43 (in Polish).

[11] V. Errunza, "Emerging markets: A new opportunity for improving global portfolio performance", **Financial Analysts Journal**, Vol. 39, No. 5, 1983.

[12] European Venture Capital Association (EVCA), EVCA Yearbooks, available at http://www.evca.eu/knowledgecenter/statisticsdetail.aspx?id=6392

[13] A. Hosley, J. Kennedy, E. James, **Revision of the Index of Industrial Production**. 71 Fed. Res. Bull. 487,1985.

[14] A. Jargot. **Why polish companies go public: an analysis of determinants and consequences of the initial public offering** (MSc in Finance & International Business). Aarhus: Aarhus School of Business, 2006.

[15] B. Jovanovic, P. L. Rousseau, **Interest rates and Initial Public Offerings.** National Bureau of Economic Research, Working Paper, 2004.

[16] V. Maksimovic, P. Pichler, "Technological innovation and initial public offerings", **Review of Financial Studies**, Vol. 14, 2001, pp. 459–494.

[17] R. La Porta, F. Lopez de Silanes, A. Shleifer, R. Vishny, "Legal determinants of external finance", Journal of Finance, Vol. 52, 1997, pp. 1131-1150.

[18] A. P. Ljungqvist, **When do firms go public? Poisson evidence from Germany.** Working Paper, University of Oxford, 1995.

[19] T. Loughran, J. R. Ritter, K. Rydqvist, "Initial public offerings: International insights", **Pacific-Basin Finance Journal**, Vol. 2, No. 2, 1994, pp. 165-200.

[20] M. Lowry, G. W. Schwert, "IPO Market Cycles: Bubbles or Sequential Learning? " **Journal of Finance**, Vol. 67, No. 3, 2002, pp. 1171–1200.

[21] J. Moody, U. Levin, S. Rehfuss, "Predicting the U.S. Index of Industrial Production", **Special Issue of Neural NetworkWorld**, Vol. 3, No. 6, 1993, pp. 791-794.

[22] OECD Stat Extracts, available at http://stats.oecd.org/

[23] M. Pagano, F. Panetta, L. Zingales, "Why do companies go public? An empirical analysis. ", **Journal of Finance**, vol. 53, 1998, pp. 27–64.

[24] S. Paleari et al., **Academic EurIPO Fact Book 2010**. Bergamo: Universoft, 2010.

[25] S. Paleari et al., **Academic EurIPO Fact Book 2009**. Bergamo: Universoft, 2009.

[26] S. Paleari et al., **Academic EurIPO Fact Book 2008**. Bergamo: Universoft, 2008.

[27] Paleari, S. et al., **Academic EurIPO Fact Book 2007**. Bergamo: Universoft.

[28] Paleari, S. et al., **Academic EurIPO Fact Book 2006**. Bergamo: Universoft.

[29] Polish Central Statistical Office, Macroeconomic Indicators, available at http://www.stat.gov.pl/gus/wskazniki_makroekon_ENG_HTML.htm

[30] Polish National Bank, Interest Rate Statistics, available at http://www.nbp.pl/homen.aspx?f=/en/statystyka/oproc/oproc.html

[31] W. P. Rees, "The arrival rate of initial public offers in the UK", **European Financial Management**, Vol. 3, No. 1, 1997, pp. 45-62.

[32] J. R. Ritter, I. Welch, "A review of IPO activity, pricing, and allocations", **Journal of Finance**, Vol. 57, Nr. 4, 2002, 1795–1828.

[33] K. Rydqvist, K. Högholm, "Going public in the 1980s: Evidence from Sweden", **European Financial Management**, Vol. 1, 1995, pp. 287-315.

[34] F. Sejkora, "Analysis of the causes of the low number of IPOs at Prague Stock Exchange compared to Polish Stock Exchange", **Scientific papers of the University of Pardubice (Series D)**, No. 26, 2013, pp. 160-169.

[35] P. Tetlock, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", **Journal of Finance**, Vol. 62, No. 3, 2007.

[36] V. Tomšík. **Komparace makroekonomického vývoje transformačních ekonomik - České republiky, Maďarska a Polska s důrazem na vývoj zahraničního obchodu a platebních bilancí.** Pracovní texty výzkumného projektu „Sociální trendy" WPST 11/1997, Sociologický ústav AV ČR, 1997.

[37] Warsaw Stock Exchange, Fact Books 2004 - 2012, available at http://www.gpw.pl/fact_books

[38] L. Zingales, L., "Insider ownership and the decision to go public", **Review of Economic Studies**, Vol. 60, 1995, pp. 425–448.

Factors Influencing Opportunity Development for New Venture Creation

Karine OGANISJANA
Faculty of Engineering Economics and Management, Riga Technical University
6 Kalnciema Str., Riga, LV-1048, Latvia

ABSTRACT

This paper is based on an interview of twenty Latvian entrepreneurs having the aim to reveal the factors which influence opportunity recognition and evolution in the course of the creation of new companies. With a broad individual-to-environment and environment-to-individual interrelationship view on this issue, it was decided to study who/what initially prompted the entrepreneurs to start their businesses, who/what helped them or hindered in the realization of the identified opportunities and what they would recommend novice entrepreneurs in this context based on their experience acquired. In the result five groups of opportunity development factors covering: personal fulfillment; professionalism; relationship with team, partners and clients; finances and market were determined. It was also revealed that education and experience influence the character and course of the development of the initial business idea.

Keywords: opportunity identification, opportunity development, opportunity development factors, new venture creation.

1. INTRODUCTION

The latest survey Flash Eurobarometer 354 "Entrepreneurship in the EU and beyond" conducted in 2012 by TNS Opinion & Social at the request of the European Commission showed that the Latvian (LV) society is more well-disposed towards entrepreneurship than EU countries on the average. More Latvian respondents were for: feasibility of the idea of self-employment within the next five years (LV – 51%, EU27 – 30% of) and desirability of becoming self-employed (LV – 55%, EU27 – 32%), stating that they were thinking (LV – 21%, EU27 – 13%) and taking steps to start up a business (LV – 29%, EU27 – 23%). Still, all these indicators concern mainly emotions and intentions at the theoretical level. However, the Latvian respondents didn't have higher entrepreneurial experience in operating businesses for more than three years (LV – 20%, EU27 – 26%) and the level of failure of their businesses was higher (LV – 16%, EU27 – 12%) [1]. Therefore it is topical to research the factors which help or hinder the Latvian entrepreneurs and entrepreneurs–to-be to recognize opportunities, develop and utilize them into new products and services.

This research was conducted within January – March 2014. Electronic forms of interviews in Latvian and Russian were sent 20 entrepreneurs (10 female, 10 male) who own production, service or combined production and service companies.

The scopes of the activities of the companies encompass production of: hygiene items; clothes; cardboard playthings for children; wooden toys and souvenirs; paving stones and concrete blocks. The services offered by the companies interviewed are: branding new products; sport tourism; the designing and setting out of gardens; accounting and audit services; hotels; real estate; IT support and trade; disposal of cement in Latvia; international trucking and logistics; trade of metallic constructions; food wholesale; trade of bedding and bedclothes. Two companies realize both cosmetics & body care production and retail and wholesale trade services at the same time. The two youngest among these twenty companies were established in 2010 while the two most experienced ones – in 1994.

2. THEORETICAL PERSPECTIVE OF OPPORTUNITY DEVELOPMENT

Though the field of entrepreneurship research lacks a common platform and is characterized with a great number of competing contradictory theories and approaches, the creation of new economic values serves as a key determinant of the success of entrepreneurial activities [2, 3, 4]. Within this framework, opportunity identification, development and realization become of particular importance and interest.

Opportunities are defined as: 1) the consequence of new ideas, inventions, beliefs and actions which generate new products and services, new ventures, markets, standards and practices [5]; 2) situations in which new goods, services, raw materials, markets and organizing methods can be introduced through the formation of new means, ends, or means-ends relationships [3, 6, 7].

Scholars argue that opportunities are:
- recognized [8];
- perceived [9];
- discovered [10];
- created [5];
- identified [11];
- developed [12, 13].

Opportunity theories focus on various phases and characters of opportunity evolution from the perspective of an individual's creativity [14] and alertness [15], cognitive processes [16], pre-experience and knowledge [17, 18]. Despite the variety of the conceptual approaches, they are complementary as they lay emphasis on different aspects and stages of opportunity evolution, started from its emergence up to its formation, development and utilization. Entrepreneurial opportunities manifest themselves in a variety of ways. According to Eckhardt and Shane, there are three valuable ways of categorizing opportunities: by the locus of the changes that generate the opportunity; by the source of the opportunities themselves; and by the initiator of the change [7].

Origination of opportunities is explained in different ways. Entrepreneurial opportunities can occur as a result of changes in a variety of parts of value creation chain, having different loci of changes which stem from: the creation of new products or services; the discovery of new geographical markets; the creation or discovery of new raw materials; new methods of production; and new ways of organizing processes [19].

Although the sources of opportunities have been proved to be very different, they are mainly explained by 1) individual; 2) environment; or 3) integrated individual and environment sides of entrepreneurship.

The sources of opportunity recognition and utilization conditioned by entrepreneurs' individual characteristics have been argued to be: utilization of knowledge [20]; schemas and mental models [21]; entrepreneurial intention [22]; broad experience and knowledge [17, 18, 23]; the ability to learn and adapt [23]; alertness [15]; personality traits, social networks and prior knowledge [12].

Peter Drucker connected the sources of opportunities with an environment perspective arguing that opportunities are innovations that occur due to changes in industry structure, demand, outside events and demographics [24].

A third group of scholars see sources of opportunities in the combination of individual and environment aspects: underutilized resources and unemployed services combined with effective benchmarking and mimicking in the context of the business [25]; opportunities which are based on the market process: allocative, discovery, & creative [5].

The aim of this research was to explore both the environment and individual conditioned factors which promote or hinder the entire chain of opportunity development and new value creation, covering all the range of individual qualities, personality traits and competencies of entrepreneurs, the environment and their interaction.

3. THE COURSE OF THE RESEARCH

The questions to entrepreneurs were about: the initial goal of creating their companies; how long they worked with the initial idea; how the initial idea was developed; who/what prompted to start the business; who/what helped in the realization of the initial idea of the service/product; what hindrances they had to overcome in that process; based on the acquired entrepreneurial experience what recommendations they would give to novice entrepreneurs.

The social–demographic questions concerned the gender, year of the establishment of the company; character of the activity of the company (service or production); the field of education of the entrepreneur; having or not having economic/business related education and business experience at the moment of the establishment of the company.

The research method for data analysis was qualitative content analysis of the texts of interviews conducted by open coding according to Phillip Mayring's "Step model of inductive category development" [26] for revealing the factors. The coded data were processed with AQUAD 6.0 (Analysis of Qualitative Data) software [27]. In the course of the qualitative content analysis when the meaning of a conceptual code was mentioned with a negative impact, it was coded with an index "neg".

Data Analysis

In order to analyse, what happened to business opportunities in the course of a new venture creation, the responses to the question about "how the initial idea developed" were analysed and grouped according to the meaning of the answers given by entrepreneurs. That revealed five groups of opportunity development scenarios: 1) the idea fully proved out to meet the initial expectations; 2) the idea gave rise to other ideas which brought good results; 3) the idea had to be partly modified; 4) the idea fostered other ideas which didn't bring good results and 5) the idea failed (see fig. 1).
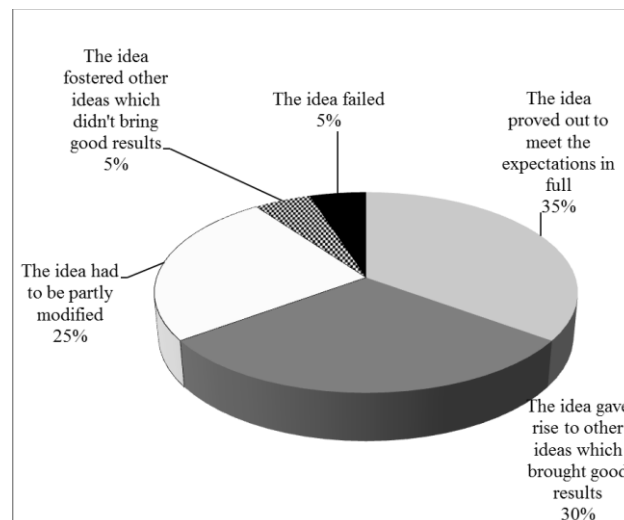

Fig. 1 The scenarios of opportunity development in the course of a new venture creation

The diagram in fig. 1 shows that more than one third of the interviewed entrepreneurs managed to implement the initial ideas into practice and get the expected results. Another third of entrepreneurs stated that their ideas developed and gave birth to new ideas which finally ended in good results. One fourth of the interviewees had to partly modify their business ideas and continued to run their businesses, while 10% of them failed or didn't get good results.

Aiming at exploring whether the character of opportunity development could be connected with having or not having economic/business related education and/or entrepreneurial experience, the corresponding data were organized together with the five scenarios for each interviewee as shown in table 1.

Table 1 Systemization of the data about interviewed entrepreneurs' education and business experience combined with the opportunity development scenarios in their companies

Nr.	Economic and business education	Business experience	The initial idea met expectations	Gave effective ideas	Partly modified	No good results	Failed
1.	1	1	0	1	0	0	0
2.	1	1	1	0	0	0	0
3.	1	1	1	0	0	0	0
4.	1	1	1	0	0	0	0
5.	1	1	1	0	0	0	0
6.	1	1	1	0	0	0	0
7.	1	1	1	0	0	0	0
8.	1	1	1	0	0	0	0
9.	1	1	0	0	1	0	0
10.	0	1	0	1	0	0	0
11.	0	1	0	1	0	0	0
12.	0	1	0	1	0	0	0
13.	0	1	0	0	1	0	0
14.	0	1	0	1	0	0	0
15.	1	0	0	0	1	0	0
16.	1	0	0	1	0	0	0
17.	0	0	0	0	1	0	0
18.	0	0	0	0	0	1	0
19.	0	0	0	0	1	0	0
20.	0	0	0	0	0	0	1

The table was built by assigning a code "1" for the answer "yes" and a code "0" – for "no". The diagram discloses three groups of interrelations: 1) those entrepreneurs (1-9) who had both economic education and experience in business acquired working in another company before mainly managed to achieve the desired result; 2) those who didn't have economic education but had business experience (10-14) on the whole managed to develop the initial idea into new fruitful ideas which brought good results; 3) those who didn't have business experience (15-20) exhibited broad diversity of dealing with opportunity – majority partly modified the initial idea, some developed initial ideas into new ones which brought or didn't bring good results, some – failed.

This shows that the combination of economic/business related education with business pre-experience could have made certain patterns for entrepreneurs based on which they worked with opportunities and succeeded. Comparing separately the effect of business experience with the effect of economic/business related education on opportunity recognition and development, it is seen that experience is more powerful than only education, as entrepreneurial patterns are created only by acting in real life situations but not just within traditional formal education.

Qualitative Content analysis

The other factors which influence opportunity development were determined in the result of qualitative content analysis of the interviewees' texts of answers to the questions on: 1) who/what prompted the idea for starting their businesses; this concerns the opportunity recognition/identification; 2) who /what helped and 3) who /what hindered the realizations of the opportunities; these two questions concern opportunity development; and 4) their recommendations.

The diagram in figure 2 is constructed based on the table of conceptual code frequencies which was created in the result of qualitative content analysis. That shows how entrepreneurs and their perception of processes and values change in different stages of their activity. For example, if intuition was mentioned as a significant prompter of starting a business, it didn't appear in the role of a "helper" in the opportunity development and realization stage. However in the final phase when recommendations were to be given to novice entrepreneurs, intuition again emerges.

Those conceptual codes which were assigned in two or three contexts have many-faceted influences on the course of business opportunity development. For instance, good business ideas, knowledge and competency in the field of the activity of that specific business, Latvian market research, understanding the needs of clients and society, finances, etc. emerged in three contexts. That means that: 1) these factors prompted the entrepreneurs to recognize opportunities; 2) helped in development and realization stages of the opportunities; and 3) therefore, based on their experience the entrepreneurs mentioned them in their recommendations.

The black columns – "Hindrances" are indexed with "neg" (negative). They show that in the course of the realization of the initial idea entrepreneurs faced challenges connected with the lack of: appropriate business idea; full understanding of clients' needs; finances; business competency and had to overcome difficulties in getting clients and good partners; hindrances related to legislation and instability of market. These factors have a serious impact on opportunity recognition and development till it is utilized into values.

Three conceptual codes – good business ideas, understanding needs of clients and society and finances emerged concerning the answers to all the four questions about "prompters", "helpers", "hindrances" and

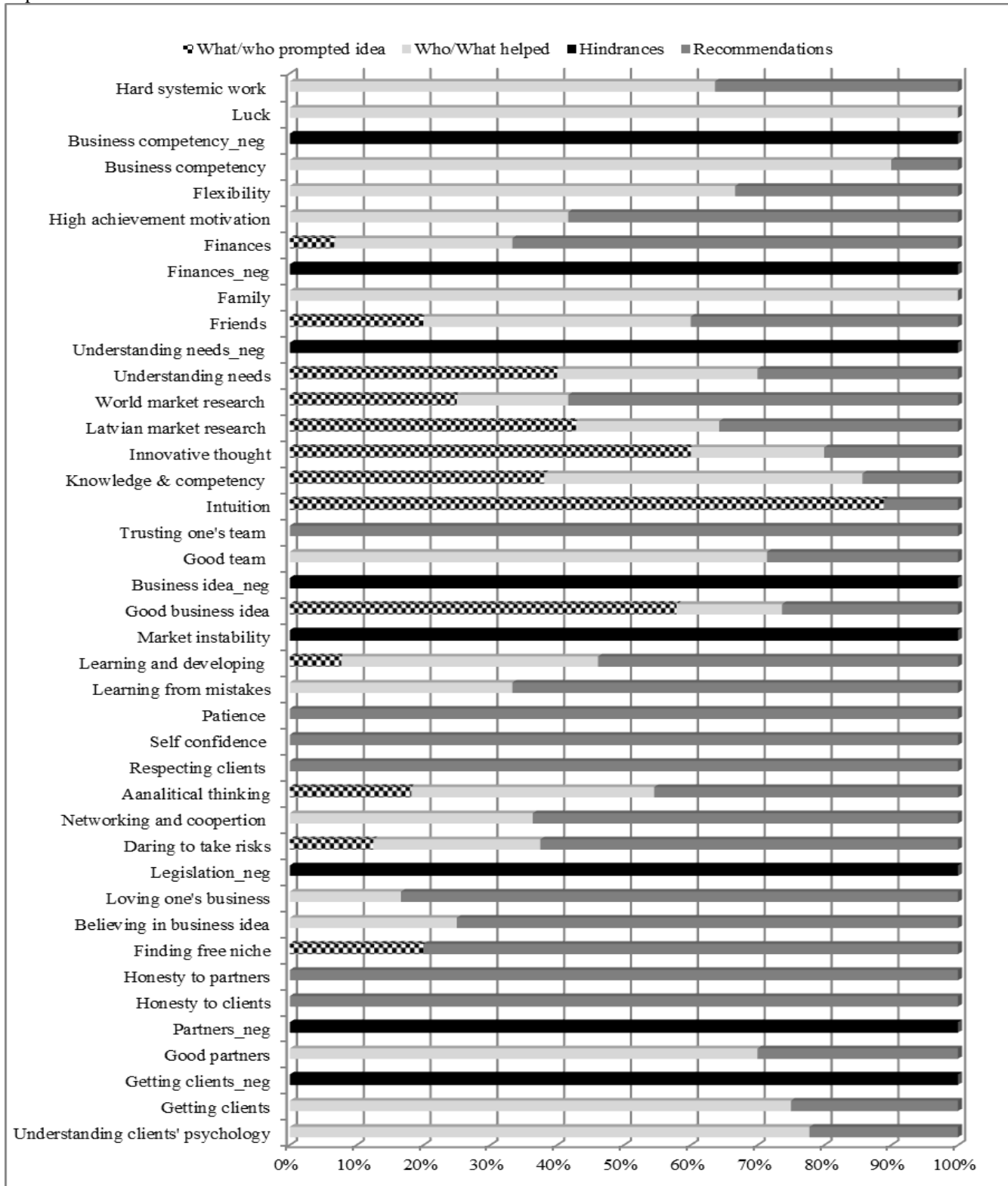"recommendations". That only emphasizes their importance.

Fig. 2 The diagram of factors influencing business opportunity recognition and development; constructed based on the table of conceptual code frequencies

As for cases when the same conceptual codes emerged in the content analysis of the text about who/what helped to realize the business idea and in the text of recommendations, it is quite logical and comprehendible, as entrepreneurs, having reflected their business path and being aware of what or who assisted them, give corresponding advice to others. For example, a good team or hard systemic work helped majority of the respondents; therefore they recommend novice entrepreneurs to pay special attention to the creation of a professional and reliable team and be ready to work hard and purposefully.

However, there are some conceptual codes, such as trusting one's own team, respect and honesty to partners and clients, which appear only in the text of recommendations. That was explained by the entrepreneurs afterwards – answering the question about who/what helped in the realization of the initial idea of the service/product, they focused their attention on the crucial aspects - finances, a good business idea, a good team, etc., which make the pillars of any business. When giving recommendations based on their experience, they mentioned more sophisticated categories like trust, honesty, respect and others which were crystalized in the course of their activities and proved out to be essential success factors.

4. FACTORS WHICH INFLUENCE BUSINESS OPPORTUNITY DEVELOPMENT

Factors which influence opportunity development for new venture creation were determined based on the triangulation and systemization of the qualitative and quantitative results of the qualitative content analysis. The factors present five groups encompassing: personal fulfillment; professionalism; relationship with team, partners and clients; finances and market:

Personal fulfillment
- Education and experience
- Enthusiasm, hard work and purposefulness
- High achievement motivation and high ambitions
- Loving one's work and believing in the idea of the business
- Intuition
- Flexibility
- Creative problem-solving
- Risk taking without being afraid of difficulties
- Never stagnating and always being in development
- Learning from one's own and other people's mistakes and failures

Professionalism
- Core competencies and expertise in the field of one's business
- Experience acquired working in other companies
- Clear understanding of the goal and possibilities for its achievement
- Doing one's work qualitatively

- Keeping abreast with novelties in one's business field
- Mastering of information and using that in business
- Consulting with other professionals for getting advice and help
- Being deeply involved in the realization of the chosen idea without distracting attention on other activities

Relationship with team, partners and clients
- Choosing colleagues and partners very carefully
- Creation and development of a professional team
- Honesty and trust to one's own team
- Honesty to partners and clients
- Development of contacts and cooperating as well with other companies

Finances
- Attraction of finances from different funds which support start ups
- Clear vision and grounded distribution of finances for the utilization of opportunities
- Planning of financial flows
- Readiness to make financial investments which will not be paid back very quickly
- Investing money into enlargement of the business
- Investing money into education and development of the team

Market
- Thorough research of demand and supply in the Latvian and world market concerning the concrete field
- Legislation
- Matching ones intuition to the market situation
- Clear understanding of clients' wishes, needs and psychology
- Getting clients
- Finding a good business idea and a free niche
- Finding opportunities for exporting.

5. CONCLUSION

1. The development of opportunities for a new venture creation may have different characters depending on entrepreneurs' education and experience in business. When economic or business related education is combined with business experience, it is more probable that entrepreneurs are able to realize initial ideas into the desired values. In case if entrepreneurs have either economic/business related education or business experience separately, the latter may have more significant impact on the success of the realisation of opportunities.
2. The factors which influence all the chain of opportunity development, starting from opportunity recognition up to its utilization, encompass an individual-to-environment continuum: personal fulfillment;

relationship with team, partners and clients; professionalism; finances and market.

6. REFERENCES

[1] European Commission, "Entrepreneurship in the EU and beyond", **Flash Eurobarometer 354**, 2012, Retrieved 23 January, 2014 from http://ec.europa.eu/public_opinion/flash/fl_354_en.pdf

[2] B. McKenzie, S. Ugbah, & N. Smothers, **"**Who Is an Entrepreneur?" is still the wrong question?, **Academy of Entrepreneurship Journal**, Vol. 13, No. 1, 2007, pp. 23–43.

[3] S. Shane & S. Venkataraman, "The Promise of Entrepreneurship as a Field of Research", **The Academy of Management Review**, Vol. 25, No. 1, 2000, pp. 217–226.

[4] G. Faltin, **"**Creating a Culture of Innovative Entrepreneurship", **Journal of International Business and Economy**, Vol. 2, No. 1, 2001, pp. 123–140.

[5] S. Sarasvathy, N. Dew, N. Velmuri, & S. Venkataraman, "Three Views of Entrepreneurial Opportunity", in Z. Acs and D. Audretsch (Eds.), **Handbook of Entrepreneurship Research**, Dordrecht: Kluwer Academic Publishers, 2003, pp. 141–160

[6] M. Casson, **The entrepreneur,** Totowa, NJ: Barnes & Noble Books, 1982.

[7] J. T. Eckhardt & S. A. Shane, "Opportunities and Entrepreneurship", **Journal on Management**, Vol. 29, No. 3, 2003, pp. 333–349.

[8] R.A. Baron, "Opportunity Recognition as Pattern Recognition: How Entrepreneurs "Connect the Dots" to Identify New Business Opportunities", **Academy of Management Perspectives**, February, 2006, pp. 104–119.

[9] H. Stevenson & C. Jarillo, "A Paradigm of Entrepreneurship: Entrepreneurial Management", **Strategic Management Journal**, Vol. 11, 1990, pp. 17–27.

[10] I. Kirzner, "Entrepreneurial Discovery and the Competitive Market Process: an Austrian Approach", **Journal of Economic Literature**, Vol. 35, No. 1, 1997, pp. 60–85.

[11] W. Long and W. McMullan, "Mapping the Venture Opportunity Identification process", in J.A. Hornday, F.A. Tardley, J.A. Timmons & K.H. Vesper (Eds.), **Frontiers of Entrepreneurship Research**, Wellesley, MA: Babson College, 1984.

[12] Ardichvili, R. Cardozo & S. Ray, "A Theory of Entrepreneurial Opportunity Identification and Development", **Journal of Business Venturing**, Vol. 18, 2003, pp. 105–123.

[13] S.A. Sanz-Velasco, "Opportunity Development as a Learning Process for Entrepreneurs", **International Journal of Entrepreneurial Behaviour & Research**, Vol. 12, No. 5, 2006, pp. 251–271.

[14] J. Vesalainen & T. Pihkala, "Motivation Structure and Entrepreneurial Intentions", In P. Reyonds et al. (Eds.) **Frontiers of Entrepreneurship Research,** pp. 73-87. Babson Park, MA: Babson College, 1999. J. Ivari, "A Paradigmatic Analysis of Contemporary Schools of IS Development", **European Journal of Information Systems**, Vol. 1, No. 4, 1991, pp. 249-272.

[15] I.M. Kirzner, **Discovery and the Capitalist Process**, Chicago: University of Cicago Press, 1985.

[16] S. Shane, **The Individual – Opportunity Nexus Approach to Entrepreneurship**, Aldershot, United Kingdom: Eward Elgar, 2003.

[17] D.A. Shepherd & D.R. DeTienne, "Discovery of Opportunities: Anomalies, Accumulation and Alertness", In W.D. Bygrave et al. (Eds.), **Frontiers of Entrepreneurship Research,** pp. 138 – 148. Babson Park, MA: Center for Entrepreneurial Studies, 2001.

[18] R.A. Baron, "Opportunity Recognition as Pattern Recognition: How Entrepreneurs "Connect the Dots" to Identify New Business Opportunities", **Academy of Management Perspectives**, February, 2006, pp. 104–119.

[19] J.A. Schumpeter, **The Theory of Economic Development: An Inquiry into Profits, Capital Credits, Interest, and the Business Cycle**, Cambridge: Harvard University Press, 1934.

[20] F. Hayek, "The Use of Knowledge in Society", **The American Economic Review**, Vol. 35, No. 4, 1945, pp. 519–530.

[21] B. Bird & M. Jelinek, "The Operation of Entrepreneurial Intentions". **Entrepreneurship Theory & Practice**, Vol. 13, No. 2, 1988, pp. 21–29.

[22] J. Katz, & W. Gartner, "Properties of Emerging Organizations", **Academy of Management Review**, Vol. 13, No. 3, 1988, pp. 429–441.

[23] G. Hamel & C. K. Prahalad, "Competing in the New Economy: Managing out of Bounds", **Strategic Management Journal**, Vol. 17, No. 3, 1996, pp. 237–242.

[24] P. Drucker**, Innovation and Entrepreneurship: Practice and Principles.** New York: Harper and Row, 1985.

[25] M. Virtanen & S. Mynttinen, "Development of Opportunities in the Value Chain of Rural High Quality Restaurants", prepared to be presented in the **16th Nordic Conference on small Business Research**, Kolding, Denmark, May 20-21, 2010.

[26] P. Mayring, **"**Qualitative content analysis" [28 paragraphs]. **Forum Qualitative Sozialforschung / Forum: Qualitative Social Research Theories** [Online Journal], Vol. 1, No. 2, 2000, June. Retrieved 24 July, 2008 from http://www.utsc.utoronto.ca/~kmacd/IDSC10/Readings/text%20analysis/CA.pdf

[27] G.Huber, & L. Gürtler, AQUAD 6. **Manual for the analysis of qualitative data.** Tübingen, Germany: Ingeborg Huber Verlag, 2004.

SUCCESS FACTORS FOR A COMPANY LONG-TERM DEVELOPMENT: CASE OF LATVIA

Deniss ŠČEULOVS

and

Elīna GAILE-SARKANE

And

Jānis ROZENBERGS

Faculty of Engineering Economics and Management, Riga Technical University, 6 Kalnciema Street, Riga, LV-1048, Latvia

ABSTRACT

The formation of new companies is considered to be one of the most important sources of new workplaces creation. In United States new companies generate half of all new workplaces in the country [1]. Therefore, in Latvia it is particularly important to encourage new businesses' formation and development in order to achieve both economic growth and improvements in employment. It is also questioned by authors of the paper why Latvian many start-ups are short-lived? Why in Latvia not so much success stories compared to other countries - for example Estonia, Scandinavia and United States (in proportion to the size of the economy or population). It is noted several possible causes: lack of access to finance to start-up business, quality of education programs, lack of access to infrastructure and other factors.
Authors of the article will study the factors that positively influenced the development of enterprises, named as „success factors".

Keywords: entrepreneur, success factors, development of a company, growth.

1. INTRODUCTION

The formation of new enterprises is regarded as one of the most important sources of creating new jobs. For instance, US start-ups account for a half of all new jobs [2]. US companies that survive on average by the year seven increase the number of employees by 60 %, whereas companies that survive in Europe increase the number of employees by about 10 %–20 % [3]. The foundation of Latvian and EU economy is formed of micro-, small, and medium enterprises (SME). They account for 98.8 % of all companies [4]. SME face various difficulties in starting entrepreneurship and managing to keep it alive. The biggest challenge for SMEs is the limited availability of various resources (financial, human resources, etc.). Furthermore, if compared with the bigger companies, a smaller number of European SMEs achieve innovations. The situation is deteriorated by structural hardships, such as, lack of managerial and technical skills, and at the state level the labour market is still inflexible. These and other factors adversely affect sustainability of companies. At the same time, there are a range of success factors, which contribute to growth of companies and successful entrepreneurship.

The aim of the study is to determine and group the factors, which positively affect development (hereinafter in the study referred to as "success factors") and contribute to sustainable development of companies in Latvia.

The authors justify this choice with the fact that data about companies, which manage to successfully develop are available less frequently than data about "failure factors", i.e. about companies, which have ceased to exist. The determination and analysis of such success factors will help the developers and implementers of various state education and lifelong learning programmes in their work, as well as assist start-ups and existing companies to analyse and pinpoint their strengths. The same applies to other non-entrepreneurship spheres (such as, non-governmental organisations, etc.), where timely attention must be paid to promoting development.

The authors have set forth **a thesis** to perform the study: irrespective of the sphere and sector of the enterprise's operations and independent of its size (within the context of SMEs), there is a set of factors. The use of these factors promotes successful entrepreneurship and development of the company.

The theoretical and methodological groundwork of the study is formed of scientific articles, monographs, regulatory enactments and researches, conference materials, internet resources, expert opinions published in Latvia and abroad. In the research generally accepted qualitative and quantitative data analysis methods of the economic science were employed, among them, statistical data processing, data grouping, and inductive-deductive data analysis methods. The scientific study employs surveying, observation study method, as well as comparative, and analytical methods, which are used by the authors to compare and analyse facts and assess solutions to specific issues. Microsoft Excel and QDA Miner programs were used for processing and analysing the study results. Authors of the article use tables and figures created with Microsoft Office to ensure visual clarity of the study.

2. LATVIAN ENTREPRENEURSHIP OVERVIEW

Overall, an upwards trend can be observed in Latvia as regards the formation of new companies, as shown by statistics about newly registered economic operators for the time period from 2002 until 2013 (See Table 1).

Table 1

New and eliminated companies in Latvia, 2002-2013 [5]

Year	Number of all new companies	Number of all eliminated companies	Balance
2002	4713	3598	1115
2004	9450	4701	4749
2006	13172	2486	10686
2008	11251	3401	7850
2010	13345	6296	7049
2012	16706	1114	15592

The rate of growth in the number of start-ups companies is positive (see Table 1), excluding the years of global economic recession, which adversely affected also the Latvian micro and macroeconomic situation. The dynamics of liquidated companies has been uneven, however, over the last few years, a positive trend is observed, namely, the number of liquidated companies is reducing. A positive balance remains for the study period - the number of start-ups companies exceeds the number of eliminated companies. It points to a positive trend in Latvian entrepreneurship development.

The authors would like to draw attention to the fact that the table and figures show data about the existing Latvian entrepreneurship forms, therefore, to provide a better outlook, they distinguished between entrepreneurship forms which can bring profits, because according to the laws of the Republic of Latvia, public and non-governmental organisations are non-profit organisations (See Table 2).

Table 2

Economically active statistical units in Latvia 2009- 2012 [6]

Type of entrepreneur ship	Year			
	2009	2010	2011	2012
Self-employed persons	45279	49063	47878	51535
Individual merchants	8242	8142	8000	7657
Commercial companies	63172	65629	72708	79777
Peasant and fishermen farms	11916	13133	13192	12574
Total	128609	135967	141778	151543

Total growth of the number of companies, % compared to the previous year starting from 2009 is 5.7 % in 2010, 4.3 % in 2011, and 6.9 % in 2012. In Latvia, similar to the EU, SMEs make up the biggest part in the total number of companies. Micro enterprises make up the biggest part in the group of micro, medium, and small enterprises, and their number keeps increasing with every year (see Table 3).

Table 3

Economically active statistical units in Latvia by size [7]

Unit by size	Year			
	2009	2010	2011	2012
Micro	115939	123924	129394	138628
Small	10254	9732	9970	10436
Medium	2065	1956	2033	2079
Large	351	355	381	400
Total	128609	135967	141778	151543

The increase in the number of companies in practically all company groups is observed each year, which points to positive growth trends of Latvian economy.

Over the last four years, an annual increase in turnover of companies has been observed, furthermore, in the last two years, companies have generally closed a year with positive profits, which also points to economic growth – common profit after taxes 957.7 milj. EUR in 2011, and 1078.7 milj. EUR in 2012 [8].

Global Entrepreneurship Monitor (GEM) 2012/2013, indicates that, while Latvia has achieved a high early stage entrepreneurship rate, there is still potential for improvement. A gap exists between entrepreneurial intentions and actual partic-ipation in entrepreneurial activity. The Total Early-stage Entrepreneurial Activity) rate (13 %) is only slightly more than half the entrepreneurial intentions rate (22%). According to the GEDI index, Latvians seem to be particularly weak in opportunity perception and recognition and non-fear of failure (despite the positive trend for both of these indicators observed in 2012) [9].

This can be at least a partial explanation for the existing gap. Latvian early-stage enterprises are among those with strong international orientation and high growth ambitions, but the level of innovativeness among Latvian early-stage entrepreneurs remains an area for further improvement. The entrepreneurial gender gap and low rate of participation in entrepreneurial activity of the group aged 55-64 are areas of potential improvement and contribution to entrepreneurial activity and competitiveness of the national economy. One out of every four early-stage entrepreneurs in Latvia still has "necessity motivation". The rate is higher com-pared to the EU GEM average (20%). Moreover, it is still substantially higher compared to the Latvian pre-recession level of necessity-driven entrepreneurship (15% in 2007). In general a strong cyclical component is evident in Latvian entrepreneurial activity and aspirations. Commercial and Physical Infra-structure and Cultural and Social Norms are areas positively evaluated by national experts. National Policy (Regulations), R&D Transfer and Internal Market Dynamics are those requiring immediate attention [10].

Overall, Latvian business development is affected by several factors – starting with the economic, socio-political situation in the country, the global market condition, and ending with the individual knowledge of a company's management and staff, their motivation to start a business, business aptitude, among other factors. The factors that affect the development of entrepreneurship environment can be relatively grouped by their features of impact on environment—macro environment impact factors and micro environment impact factors. These groups, for their part, consist of direct and indirect impact factors. In the article, the authors will not consider the business aptitude of individuals, which has been a very popular object of study and a subject in social studies over the last few years.

3. GROWTH OF A COMPANY

Each company goes through several stages of development—beginning with the establishment of a company and ending with closing of operations. There are multiple theories and opinions as regards a company's life cycle. One of the founders of the company life cycle theory T. Levitt distinguishes between four development stages—introduction, growth, maturity, decline [11]. This theory has been used as the basis and developed by a number of scientists. For instance, I. C. Adizes distinguishes between the following company development stages—Courtship, Infancy, Go-go, Adolescence, Prime, Stability, Aristocracy, Early Bureaucracy, Bureaucracy, Death [12]. Bersin divides the cycle into the stages of formation, rapid growth, maturity, decline, revival or termination of operations [13]. A. M. Ahmed believes that a company transforms from an infant company to a transforming or dying company [14]. Whereas Darrell Zahorsky determines seven company development stages: Seed Stage, Start-Up Stage, Growth Stage, Established Stage, Expansion Stage, Decline Stage, Exit Stage [15].

Regardless of the theoretically described number and name of company's life cycle stages, the authors conclude that company development, similar to the life cycle of goods, can be described with a bell shape line. The study authors consider the second or the Growth stage of company development after a company is established is an important stage to study because in this stage, according to the researchers, the company theoretically "has overcome the start-up issues and the business has made it through the toddler years and it is now a child" [16]. Revenues and customers are increasing with many new opportunities and issues. Profits are strong, but competition is surfacing. Demand begins to accelerate and the size of the total market expands rapidly. It might also be called the "Take off Stage" [11].

The authors consider the second stage of company development to be very important because during this stage the company grows from "a start-up company" into "a new company" and has already stabilized and is developing rapidly. The organization experiences rapid growth, this growth can be in multiple different directions, founders – over excited by their success – can bring their company to the "brink", many of the new products and ventures being pursued are unprofitable, there are few policies, almost no rules and the founder makes all decisions [11]. During the growth stage, the company develops the need for arranging processes, control, systematization, etc. Adizes sets two main abilities of a company: *flexibility* - ability to effectively flex and adapt to change; controllability: ability to efficiently control organizational efforts. As a company develops, flexibility of the organization falls, whereas controllability grows. Adizes regards the point of intersection of both abilities as the Prime state, at which the company can control the processes without losing flexibility. The company starts achieving this condition at the end of the growth stage, and they must strive to maintain the Prime state for as long as possible [15].

Upon summarizing the case studies performed by the authors, it is concluded that during the growth stage it becomes important for the company to engage in strategic planning, which can help arrange, improve, as well as control processes. The authors assume that the majority of company's development strategies, which are found in scientific writings, are aimed at large companies. Taking into account the fact that, similar to the majority of EU countries, the foundation of the economy is

formed of SMEs, then managerial and decision-making strategies that are suitable for the specifics of these companies are important for them. Therefore, in the study, the authors of the article address specifically this problem.

The authors believe that during the growth stage the manager / owner must pay particular attention to the team. It is proven also by a number of scientific theories. For instance, Sahlman believes that people are just as important as opportunities and deals [18]. Timmons in his model includes three more important cornerstones, which describe the entrepreneurial process – opportunity, resources and team [18].

The authors support the opinion of the scientists, and emphasize the important role of human resources in a company not only during the Growth stage, but also for the overall development of a company. There are undoubtedly a range of other factors that affect a successful development of a company.

4. FACTORS THAT POSITIVELY INFLUECED THE DEVELOPMENT OF COMPANIES

As a result of study of scientific, business literature, and information, incl., studies of leading economists and specialists, the authors have determined a range of factors that influence successful development of companies. Thus, R. Peacock as a result of a study conducted more than ten years ago categorised the following success factor groups: management efficiency, aspects of owner-manager conduct, company's characteristics [19]. The classification of factors includes also features distinguished by other scientists, such as manager's experience in the specific sector or market, education level, experience in entrepreneurship, reasonability of the decision to start a business, goal-oriented nature, etc. Entrepreneurship experience is the factor of influence according to several studies covering the topic. Lack of entrepreneurship experience is considered a risk factor (in the study by the authors Delmar and Shane (2006) about Sweden; in the study by the authors Rotefoss and Kolvereid (2005) about Norway) (references to these important studies are given in the M. Gelderen, R. Thurik, P. Patel study about the Netherlands) [20]. M.K. Mason in his studies about the USA concludes that the success factors are: the owner's attitude, the ability to be objective, willingness to engage help, when necessary, power sharing, and managers' experience. Another very important factor mentioned by the afore enumerated authors is the extent of invested effort, because failures occur, if insufficient work has been invested [21].

Several studies show that the presence of a business plan is a positive factor, however, at the same time A. Osterwalder believes it useful to develop a business model rather than a business plan. According to Osterwalder, a business model describes the reasoning of how an organization creates a value, how it delivers it to clients, and how it acquires a value [22]. Having determined the possible success factors, it is necessary to determine, how successful development of new companies will be measured. The authors A. Cedere and A.McKelvie savā in their study identified the following three key factors: the number of created jobs, attracted investments, value added in a form of services rendered to clients or in a form of goods (turnover), which, at the same time, serves as an indicator describing the value added for the company (though profits would be a more accurate parameter) [23].

William Bygrave believes that a company's success is most directly related to its manager-entrepreneur. He summarized the important characteristics of successful entrepreneurs in 10 D's:

- dream – entrepreneurs have a vision and the ability to implement their dreams;

- decisiveness – they make decisions swiftly, their swiftness is a key factor in their success;
- doers – once they decide on a course of action, they implement it as quickly as possible;
- determination – they implement their ventures with total commitment;
- dedication – they are totally dedicated and work tirelessly;
- devotion – entrepreneurs love what they do;
- details – the entrepreneur must be on top of the critical details;
- destiny – they want to be in charge of their own destiny;
- dollars – getting rich is not the prime motivator, but the measure of success;
- distribute – entrepreneurs distribute ownership of the business with key employees [24].

Jeffrey Timmons believes that successful entrepreneurs feature common traits:

- work hard and are driven by intense commitment and determined perseverance;
- optimistic outlook;
- strive for integrity;
- burn with competitive desire to excel and win;
- dissatisfied with the status quo and seek opportunities to improve almost any situation;
- use failure as a tool for learning;
- eschew perfection in favour of effectiveness;
- believe that they personally can make a difference [18].

The authors of the article point to that several scientists and specialists link a company's success with an entrepreneur's personal competences, experience, various abilities, characteristics, etc. However, there are other important factor groups describing a successful entrepreneurial process. Having analysed sources of scientific and business literature, as well as several studies on the topic of success factors affecting entrepreneurship, the authors of the article reached a conclusion that the Timmons model is best at describing and depicting the success factors influencing the entrepreneurial process and what they entail (See Fig. 4)
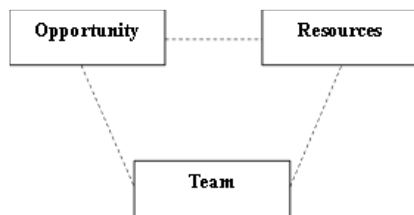
Fig. 4 Timmons Model of the entrepreneurial [25]

Timmons says: "at the heart of the process (Fig 4) is the opportunity. Not all ideas are opportunities, their recognition and evaluation is crucial but the important characteristics of good opportunities are that there is as understanding market demand for the product or service because of its value-added properties, and that it will generate money either as a profit or as a means of creating self-sufficiency for not-for-profit organizations [25].

Resources need to be understood and utilize wisely. For the entrepreneur, at the early stages they are scarce so the entrepreneur needs to minimize and control them as opposed to

maximize and own. The entrepreneurial team is a key ingredient for success. Key elements of resources that need to be understood are tem members. Meanwhile, the team requires determination and persistence, tolerance of risk, ambiguity and uncertainty, creativity, team focus of control, adaptability, opportunity obsession, leadership, communication. The tool which integrates the three together is the business plan" [24].

5. SUCCESS FACTORS OF LATVIAN COMPANIES

Studies conducted in Latvia also produce a similar tendency of a strong link between a company's success and the entrepreneur him/herself. Thus, in a Swedbank entrepreneurship competence study, the main traits were determined, which, according to the opinion by Latvian inhabitants, must describe a person, who wishes to begin a business venture. The courage to take risks is the most important quality for starting a business (48 % respondents), confidence (45 %), and intuition (42 %) were the most frequently mentioned traits. However, the willingness to make profit ranked second to last on the list of desirable qualities in an entrepreneur important for starting a new business.

Another Survey [26] carried out in December 2011 points to the following factors, which could affect business development: an open market (not oversaturated) with relatively low barriers and niche market with development potential; availability of resources that have become scarce in other markets, such as land, timber and low-cost labour; the business culture and mentality congenial to that of the European countries; a multitude of interesting business projects, enterprises that have performed well before investment, well-established production traditions; interesting developments, enterprises with great potential.

In survey mentioned also positive Latvian business environment trends:
- the activity of local companies is increasing, hence the willingness to attract investment and expand business;
- local entrepreneurs are engaging in development of investment projects and founding of their own investment funds;
- the quantity and quality of business ideas offered to investors is increasing;
- cities and municipalities are employing an increasingly proactive approach to attracting investment and strengthening of cooperation with a view to promote regional growth and development.

According survey „foreign investors" views on Latvian entrepreneurs:
- low self-assessment, indulging in self-criticism;
- open to new ideas, motivated to gain new experiences;
- cooperation-oriented, flexible, practical;
- increasingly aware of the need to adopt long-term strategies and to improve the business environment;
- on certain occasions there is a lack of professionalism and knowledge about business relations;
- sometimes attitudes towards foreign investors can be perceived as arrogant;

Foreign investors' perception (the image thereof is dominated by): sound management skills, self-criticism, openness, practical approach, arrogance, lack of knowledge [27]. The authors of the article draw attention of Latvian entrepreneurs to the fact that they must think globally, they must be aware of the fact that the consumer purchasing ability is not restricted to the Latvian market. The companies, which will

take the step to expand their geographic business boundaries, will definitely gain success, because entrepreneurs have all the competences for further development and expansion of their business. Moreover, company founders and owners must take into account that, as a company develops, its team also must develop, developing and improving their business etc. knowledge.

Success factors of Latvian companies

To find out the factors that affect the success of a start-up company in business, the authors of the article have performed an empirical pre-research with the aim of finding out the factors and conditions affecting the creation of a company, as well as their successful start-up and the transition into the stage of "a new company". The pre-research is based on the survey conducted by IDAL "Success code" [28]. The survey dealt with twelve of the most successful companies in Latvia representing various sectors (creative industry, IT fields, representatives of the production and other fields), with the owners and/or managers whereof in-depth interviews were conducted. The surveyed companies have competitive products, which are sought after in Latvian and foreign markets; stable and growth-oriented financial indices pointing to successful entrepreneurship of the companies.

As a result of the survey, several factors were identified related to the role of generation and development of an idea, clients and communication with them in business processes, the importance of the company's founders and team, etc.

To examine data validity, the authors of the article used a data analysis performed with QDA Miner [29] statistical analysis software. *The methodology:* authors explores coding retrieval tools for extracting coded segments associated with specific codes or code patterns and identifying coding co-occurrences, coding sequences and assessing relationships between coding and numerical or categorical properties. Out of specific factors specific key words could be determined (depending on their weight) in the software during the analysis. According to identified key words, factors were found in the whole text corresponding to these key words. As a result, the factors grouped by the authors of the article were grouped and hierarchically arranged depending on their importance. The factor groups are:

1. the idea (factors related to the idea);
2. the team – people (excluding the manager-founder)
3. the manager-founder;
4. clients and the market.

The authors agree to this grouping because in company development it is vitally important to have an idea to begin the business with and to capture the market; the manager (founder), who is able to excite with the idea as well as to convince and lead the team. And the team itself, believing in the product and able to sell it on the market. Finally, a company cannot exist without the clients, their feedback and purchases.

The authors of the article grouped together success factors obtained as a result of the study, based on the Timmons Model [24] dividing them into three groups – *Opportunity, Team and Resources*, because this is the model intended for start-ups.

Factor group "Opportunity".

Totally authors of the paper detected sixty two factors in group "Opportunity", main of them are:

- an idea that renders practical good, an idea at the foundation of entrepreneurship;
- at the basis of good products—synergy of technologies on the basis of overlapping sectors;
- the founder's motivation to create a company derives from the stance - willingness to be independent, make decisions, be responsible for them;

- readiness to renew and transform even a carefully developed initial financial plan;
- awareness that, possibly, the first project might fail and is not profitable, but the obtained experience at least partially make up for the investments and time;
- persistence in obtaining financing and calculating the costs for attracting one client;
- choice of control mechanisms and their adaptation according to the company's specifics, actual needs etc.;
- precaution during the initial stages, because expansion is mostly financed from internal resources;

Factors, related to knowledge about the market:

- market research in all stages of the product life cycle, starting with the appearance of the idea;
- adaptation of the initial planned model of company's operations to market fluctuations and its improvement;
- diversification of the client portfolio;
- knowledge of differences in international market cultures, behaviours, and other aspects;
- it is not the pace that matters in entrepreneurship – it is the rhythm that matters;

Factors, related do entrepreneurial skills and abilities:

- enjoying the entrepreneurship processes despite hardships;
- love is transformed into energy, but energy—into money. If you like it, then energy is automatically generated and you can lead the way forward as the flagship, inspiring people around you;
- sometimes, the most important thing is to decide what "not to do", in order to avoid spoiling what you have;
- continuous improvement of the company turning into a day-to-day work and routine;
- the idea "must be fed"—collect information, think about it.

Factors, related to communication and modern technology using:

- the company's website and free online communication tools providing a quality content about the product;;
- taking advantage of favourable geographical location, which must be emphasized to potential cooperation partners;
- productive and successful communication with stakeholders, for which preparations are due;
- language and other skills important for the sector and the market;
- following the 24-hour rule (reply within 24 hours);
- well-organised communication with export markets to confirm that export and the potential client is a priority for the company;
- regular updates of versions of technologies, thus increasing the value that the client gains by using the products;
- diversification of aim monitoring; ensuring transparency of the control process, using electronisation and process centralisation;
- recording electronic invention methods, client suggestions, etc.

Factor group "Team".

Totally authors of the article identified 31 factors, main of them are:

- the presence of a leader who can motivate his team to achieve a common goal;

- various employee types ensuring effective entrepreneurship processes;
- before starting own business, trying out various roles and gaining entrepreneurial experience and practical knowledge when working at other companies (paid work etc.);
- position of an industry expert—knowing the sector and availability of contacts in the industry;
- theoretical, academic, and practical high level knowledge (higher education—MBA);
- further self-education, by using the acquired academic knowledge base;
- selfless idea generators in the team, sharing their ideas, because not everybody is capable of generating ideas;
- ensuring clear communication in the team and each participant understanding their role in the team;
- during the development stage, the most important functions in a company should be ensured using internal resources instead of outsourcing;
- the team size providing the most important competences to the company, which mutually complement each other;
- the manager in his/her development should not forget sales tricks and methods, because "the bigger and fatter fish" still must be caught by the captain;
- awareness and understanding about the company as a participant in the global market place;
- overcoming obstacles that are not related to skills, but rather to the psychological aspect (the typical Latvian temperament—lack of self-esteem and confidence, fear of trying and risk of failure);
- the founder or idea owner can best define the company's course of development and aims, because the idea and/or the company has been created as a result of specific emotions and needs, they are the initial inspiration and drive for implementation. In a company with a smaller number of employees, it is the owner/founder who also helps others "conjure up" the vision of a bright future;

Factor group "Resources":
- guaranteed positive cash flow;
- use of various financing sources (short-term loans; risk capital funds; "business angels"; awards of business plan competitions; family, relatives, friends, financing for the development of the idea, pre-seed capital, bank credits, stock exchanges);
- being aware of differences between ac-counting and financial planning;
- detailed cash flow planning, by balancing income with expenses and bearing in mind peculiarities of a tax system;
- maintaining as few fixed costs as possible;
- acknowledging the income potential, necessary finances before return on investment, and overall project profitability.

In grouping the factors, the authors of the article would like to maintain that for entrepreneurs, the first two factor groups are important. It could be explained with the fact that all companies, regardless of their considerable achievements on markets, are recently established. During the start-up stage, it is the idea that is the most important, which is then materialized into a product. The factor group "Opportunities" include factors that are related to the idea, a product, market opportunities, etc. Furthermore, all companies are SMEs, in which the team and team-related aspects play an important role. Companies are aware of re-source importance and necessity, but due to the fact that they are in the development stage, precaution is observed when attracting bigger financial resources—companies choose to rely on their own potential during this development stage. The same applies also to other resources.

The article authors point to that resources and financing sources are necessary, which, as a result of analysis, were not estimated as important. Therefore, the authors will perform further factor analyses in the upcoming studies to group and classify success factors affecting long-term development of business.

CONCLUSIONS

The economy of Latvia, just like other EU countries, is based on SMEs. The number of SMEs increases in Latvia each year. Moreover, the financial indices of SMEs over the last few years have improved. All of these facts point to economic recovery after the global crisis.

As start-ups develop, there are several factors that determine their successful operations.

Several scientists and specialists, when performing studies about business success, point to the importance of the company founder's role, characteristics, abilities, knowledge, etc., their mandatory participation in all business processes, in particular during the initial stages. The team also has an important role and only the founder can form that team. Team members must be creative, flexible, and competent, because start-ups have very limited resources, including human resources.

The empirical pre-research survey conducted by the authors allowed identifying several factors affecting a company's long-time development. According to an analysis performed by QDA Miner, the factors were divided into four groups. The authors of the article performed factor grouping based on the Timmons model, because the model is intended for start-up companies. The authors point to that the factors described and grouped in this article will serve as the grounds for further studies, in which the authors are to perform factor classification.

The said factors are mutually related and must be considered as a whole and not separately.

The thesis set forth by the authors is proven with the study on success of Latvian companies, determined by various factors identified by the authors of the article.

6. REFERENCES

[1] SBA Office of Advocacy. Frequently Asked Questions about Small Business, 2012, p.1-4. [Electronic resource], viewed on January 11, 2014, http:<//www.sba.gov/sites/default/files/FAQ_Sept_2012.pdf>

[2] Reinolds, P., Curtin, N. "**New business creation. An international overview**". New York: Springer, 2011. – 379 p.

[3] GEM 2013. Global Report. [Electronic resource], viewed on February 14, 2014, http://www.gemconsortium.org/docs/3106/gem-2013-global-report

[4] SBA Facts 2012. [Electronic resource], viewed on January 13, 2014], http://ec.europa.eu/enterprise/policies/sme/facts-

figures-analysis/performance-review/files/countries-sheets/2012/latvia_lv.pdf

[5] New and eliminated companies in Latvija, 2002-2013. Register of Enterprises of Latvia. New and eliminated companies in Latvija, 2002-2013. Register of Enterprises of Latvia. [Electronic resource], viewed on December 11, 2013], http://www.ur.gov.lv/statistika.html?a=1091

[6] Economically active statistical units in Latvia 2009-2012. Central Statistical Bureau of Latvia. [Electronic resource], viewed on January 13, 2014, Internet:http://data.csb.gov.lv/Table.aspx?layout=tableVieLaout1&px_tableid=SR0011.px&px_path=uzreg__Ikgad%C4%93jie%20statistikas%20dati__01_skaits&px_language=en&px_db=uzreg&rxid=cdcb978c-22b0-416a-aacc-aa650d3e2ce0

[7] Economically active statistical units in Latvia by size group. Central Statistical Bureau of Latvia. [Electronic resource], viewed on January 13, 2014,http://data.csb.gov.lv/Selection.aspx?px_path

[8] Profit or loss of commercial companies 2009 – 2012, mil. eur. Central Statistical Bureau of Latvia. [Electronic resource], viewed on January 13, 2014, http://data.csb.gov.lv/Table.aspx?layout=tableViewLayout1&px_tableid=UF0020_euro.px&px_path=ekfin__Ikgad%C4%93jie%20statistikas%20dati__Uz%C5%

[9] GEDI Index 2012. [Electronic resource], viewed on January 15, 2014, http://www.thegedi.org/research/gedi-index

[10] Krumiņa, M., Paalzow, A. 2012. Global Entrepreneurship Monitor. [Electronic resource], viewed on January, 13 2014, http://www.biceps.org/assets/docs/gem/GEM_2012.pdf

[11] Levitt, T. 1965. Exploit the product life cycle// **Harvard Business Review,** 1965, vol 43, Nov.–Dec. p. 81–94.

[12] Adizes, I. 2013. Understanding the corporate life cycle. [Electronic resource], viewed on January 3, 2014, http://www.adizes.com

[13] Bersin, J. 2008. Make learning part of your business strategy. [Electronic resource], viewed on January 12, 2014,<http://joshbersin.com/2008/07/31/makelearning-part-of-your-business-strategy/>

[14] Ahmed, A.M. Virtual integrated performance measurement// **International Journal of Quality & Reliability Management**, 2002. Vol. 19 Iss: 4: p. 414 – 441. doi: 10.1108/02656710210421580

[15] Zahorsky, D. Find Your Business Life Cycle. [Electronic resource], viewed on January 11, 2014, http://sbinformation.about.com/cs/marketing/a/a040603_2.htm

[16] Adizes, I. 2004. Managing corporate life cycles. 2nd Edition. [Electronic resource], viewed on 4 January, 2014, http://www.adizes.com

[17] Sahlman, W.A., Stevenson, H.H., Roberts, M.J., Bhide, A. **The entrepreneurial venture**. 2nd edition. Harvard: Harvard Business School Press, 1999. – p. 174-175.

[18] Timmons, J. **New Venture Creation: Entrepreneurship for the 21st Century**, 5th Edition, London: McGraw Hill, 1999 – p.46.

[19] Peacock, R. 2000. Failure and assistance of small firms// **Specialist Small Business Consultants.** – 2000 - 20 of Jan. 2013. [Electronic resource], viewed on January 15, 2014, http://www.sbeducation.info/downloads/sbfail.pdf

[20] Gelderen, M., Thurik, R., Patel, P. Encountered problems and outcome status in nascent entrepreneurship// **Journal of small business management,** 49(1), p. 71-91. [Electronic resource], viewed on January 12, 2014, http://people.few.eur.nl/thurik/Research/Articles/Encountered%20problems%

[21] Mason, M.K. What Causes Small Businesses to Fail? [Electronic resource], viewed on January 15, 2014, http://www.moyak.com/papers/small-business-failure.html

[22] Osterwalder, A., Balnk, S., Smith, A. Success Entrepreneurship. [Electronic resource], viewed on January 19, 2014, http://www.slideshare.net/Alex.Osterwalder/successfulentrepreneurship-5747012

[23] Cedere, A., McKelvie, A Creative Center and Beyond. University start-ups in Jonkoping// **JIBS Working paper series,** 2001. No.2001-32001, p. 26.

[24] Bygrave, W. **The Portable MBA in Entrepreneurship**, 2nd Edition. London: Wiley, 1997. – p. 32-33.

[25] Vyakarnam, S., Hartman, N. **Unlocking the Entrepriser Inside**! Singapore: World Scientific Publishing Co. Pte. Ltd., 2011. - p. 28-29.

[26] Survey: Attieksme pret uzņēmējdarbības uzsākšanu. Uzņēmējdarbības kompetences centrs. [Electronic resource], viewed on January 11, 2014], http://www.manambiznesam.lv/lv/2012/03/19/attieksme-pret-uznemejdarbibas-uzsaksanu/

[27] Swedbank Entrepreneurship Competence Centre. [Electronic resource], viewed on January 29, 2014, https://businessnetwork.lv/entrepreneurship-competence-centre

[28] LIAA. 2012. Veiksmes kods. [Electronic resource], viewed on January 13, 2014, http://www.liaa.gov.lv/veiksmes-kods-uznemeju-stasti-un-atzinas

[29] QDA Miner. 2013. [Electronic resource], viewed on January 19, 2014, <http://provalisresearch.com/solutions/>.

USING OF E-ENVIRONMENT AS A TOOL FOR VALUE CREATION IN A COMPANY

Vladimir SHATREVICH
Department of Management, Riga Technical University
Riga, LV-1048, Latvia

and

Deniss ŠČEULOVS
Department of Management, Riga Technical University
Riga, LV-1048, Latvia

ABSTRACT

Abstract. The importance of value creation in small and medium-sized business companies has always been in focus. The changing environment has made a strong impact on all companies all over the world. Nowadays, the value added, which is created by the company, depends not only on tangible assets, but also on intangible assets. It is not enough just to manage internal resources to be efficient, in order to generate high value added. The importance of knowledge and information as a tool to manage external environment brought a new factor to the company. This paper uses the e-environment as a tool to illustrate how the company develops value added and thus increases the shareholder value through focusing on intellectual capital (IC). Since the elements of intellectual capital system are intangible and hardly measurable in company's value added, this paper is aimed to create a model for analysing the company's value added creation through intellectual capital.

Keywords: intellectual capital, Information Communication Technologies, value creation, organizational structure, performance.

1. INTRODUCTION

There is one important question, which has been dominating through history of management, and it is "how do we create value added of company", and thus, enhances shareholder value. Many scientists emphasize the importance of company's value added as the main factor for creating shareholder value. [1-5,33].

Nowadays, the efficiency of value chain as one of the key inputs to added is well understood [6]. Today's knowledge intensive companies have more advantages in more complex environment. The changing environment replaced the perception of company's value added (VA) sources.

Reliance on productive tangible assets such as "raw materials, fixed capital, and even managerial knowledge" no longer account for investments made and wealth created by new and prospering companies [7]. As the primary inputs to organizations' value creation processes are internal resources, but classic economic laws are hardly applicable for knowledge and other intangible resources. These resources traditionally seen as external could make an important contribution to the value creation process of the company. Based on the intellectual capital (IC) approach, the paper begins the research that explores the effect of intangible resource in creation of added value.

Many scientists were analyzing the influence of intangible assets on company's value added [8-11]. Sveiby and O'Regan are assuming that intellectual capital is composed of the following three main parts: external structure, internal structure, and human capital [12,13].

The IC literature draws on aspects of the practical applications, providing a framework for explaining the value creation process as the link, between resources and shareholder value.

Value creation is not only difference between incomes and expenditures. P. F. Drucker accentuates this and emphasizes that "main producers of wealth have become information and knowledge" and also the knowledge productivity [14].

2. INTELLECTUAL CAPITAL APPROACH

The concept of IC started to formalize in the early 1990s by Leif Edvinsson [11].
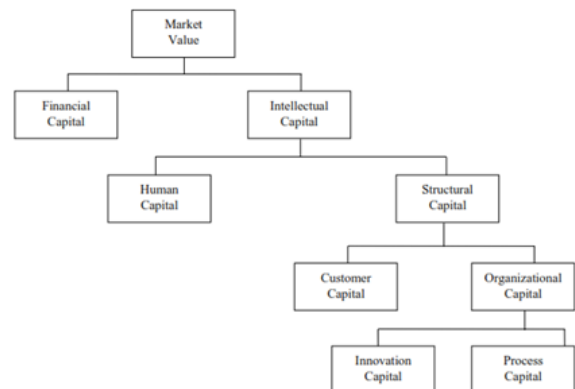
Fig.1. SKANDIA'S value scheme [11]

The work of Skandia (Fig. 1) was presented as a supplement to the annual shareholders report to describe the "true" value of the company's. This new model was created to identify the roots of a company's value by measuring hidden dynamic factors that underlie "the visible company of buildings and products".

By the end of the 1990s, references to intellectual capital in contemporary business publications were commonplace [9,10]. Many scientists started to define IC, having a similar opinion about intellectual capital and its definition [9,10,15,16,18].

Bontis considers that intellectual capital is everything that is in a company: all intangible resources and processes that belong to the company, patents, innovations, and customers, tacit and explicit knowledge [9].

What resources actually make up these generic capital forms is unique to each and every organization, as only those resources that are important for creating value should be included in constructing the distinction tree for an organization [9]. The work of Skandia advised to measure the IC through indexes (Fig. 2), recommended 112 metrics. Later IC-Index practice was created by Roos, Dragonetti and Edvinsson [19].

SAMPLE OF SKANDIA IC MEASURES

Financial Focus	• revenues / employee ($) • revenues from new customers / total revenue ($) • profits resulting from new business operations ($)
Customer Focus	• days spent visiting customers (#) • ratio of sales contacts to sales closed (%) • number of customers gained versus lost (%)
Process Focus	• PCs / employee (#) • IT capacity – CPU (#) • processing time (#)
Renewal and Development Focus	• satisfied employee index (#) • training expense / administrative expense (%) • average age of patents (#)
Human Focus	• managers with advanced degrees (%) • annual turnover of staff (%)

Fig.2. Scandia IC measures [11]

Roos *et al*. propose that the specific measurement of company IC by weightings and indicators can be decided by knowing the company's strategy. Also Roos *et al.* suggests that the main consideration for assigning the weights to indexes should be the relative importance how they help the company achieve its strategic goals [19].

Karl-Erik Sveiby gives his own conceptual framework (Fig. 3) of measuring IC assets based on three families of intangible assets: external structure (brands, customer and supplier relations); internal structure (the organization: management, legal structure, manual systems, attitudes, R&D, software); and individual competence (education, experience) [13].

Visible Equity	Intangible Assets (Stock Price Premium)		
(book value)	**External Structure**	**Internal Structure** (management, legal structure, manual systems, R&D, software)	**Individual Competence**
Tangible assets minus visible debt.	(brands, customer and supplier relations)		(education, experience)

Fig.3. Measuring model of intangible assets [13]

IC approach helps us to develop strategy that focused on intangible resources, allowing them to manage more effective in process increasing in shareholder value.

To conclude, different scientists intellectual capital is understood as the sum of all knowledge in the company that is able to generate company's value added and it is affected by knowledge quality and knowledge productivity.

3. THE CONCEPT OF EENVIRONMENT

The rapid electronic environment development over the last decade has fostered the e-market growth and has provided companies with opportunities that they previously did not have. By employing advantages offered by the e-environment, entrepreneurs can ensure expedient and effective communication with the target audience, by promoting products on the global market. The performed scientific studies show that proper and skilful use of modern technologies can contribute to significant development of companies.

Up to now, no unequivocal studies have been performed about the use of the electronic environment in ensuring development of micro, small, and medium enterprises. Ph. Kotler, D. Tapscott, P. Drucker, and J. A. Pearce maintain that two parallel markets exist and are developing – the traditional and the electronic environment. The electronic environment is used for various needs – for trade, marketing, advertisement, studies, communication, training, etc. Simultaneously, there is an opinion claiming that in future, the majority of transactions will be performed on the electronic market, hence advancing the dominant position of the e-environment in achieving entrepreneurship competitiveness.

The electronic environment already now offers companies practically all the necessary marketing and communication tools for ensuring company development by creating competitive advantages, nevertheless, not all companies can employ the opportunities rendered by the e-environment, in order to increase company competitiveness and productivity.

There are several well-known and popular value theories, such as, the Five forces model [20], Shareholder value model [21], as well as the "Value map" theory, intended for analysing the economic gain for consumers [22].

Various theories were developed many years ago, when the electronic market was not yet developed, and hence are suitable for the conventional market. Due to this reason, the authors of the article suggest that companies use the Alexander Osterwalder's value proposition concept or the approach that is a constituent element of the author's developed business model canvas) [23].

The Osterwalder's business model was formed based on Freeman's stakeholder theory [24]. The model is adapted to today's market needs and conditions, and the importance of the electronic environment, i.e. of the electronic market, in entrepreneurship is taken into account. Osterwalder distinguishes between "value proposition" and "elementary value proposition", which is an element of value proposition.

The authors wish to draw attention to Osterwalder's "value life cycle" consisting of five stages: value creation, appropriation, consumption, renewal, and transfer [25].

All life cycle stages are linked to value consumption, using the electronic environment: value creation process (based on information and communication technologies (ICT) – adaptation of various products for the needs of an individual consumer, e.g., personal computers, footwear, etc. Value appropriation – "a one click purchase" at an internet shop. Value consumption – listening to music, watching a movie, etc. Value renewal – various software updates, value transfer – disposal of old computers and other machinery, giving away unnecessary books and equipment for further use, etc.

Upon combining analyzed models, it can be seen that the information and communication technologies (in the Osterwalder's model) or the information communication technology bear great importance in creating value for consumers and that they undoubtedly affect the company's image. Nevertheless, several empirical studies made by authors in Latvia, show that many Latvian SMEs do not employ ICT and therefore the most suitable way should be sought for how to involve ICT in elaborating business development models.

The value concept is broadly used in various business models, including e-business models. The value forms the basis of several business models.

The e-business model is based on mutual integration of key flows and values and implementation thereof between e-market

participants, through the use of the e-environment. Three main e-business model elements can be distinguished: flows, participants, value. The term e-business model describes a broad spectrum of informal and formal models, which may be used in companies to depict various business aspects, such as operational processes, organisational structures, and financial forecasts [26].

In studying various business model concepts, the authors have come to a conclusion that both business model types (taxonomic and conceptual) can be applied to the Latvian SMEs; however the conceptual business models would still be primary. It is related to the fact that there are many niche and narrow profile companies in Latvia. Moreover, the majorities of companies are operating only on the local market and depend on domestic demand fluctuations.

The conceptual business models enable companies to analyse the current condition more broadly and to evaluate the already existing business. By employing this analysis, companies can develop new business development directions or improve the existing ones, because a modern market demands that companies change and are aware of their global condition. Entering the global market allows companies to reduce their dependency on local market fluctuations.

Taxonomic models, for their part, can serve as a specific type of entrepreneurship. For instance, when developing the conceptual business model, companies will answer the question "How to develop further on?", but the taxonomic model will allow answering the question "What to do in order to develop?"

The use of ICT promotes communication (Fig. 4); moreover, ICT is at the basis of the first stage "value creation" of the value life cycle.
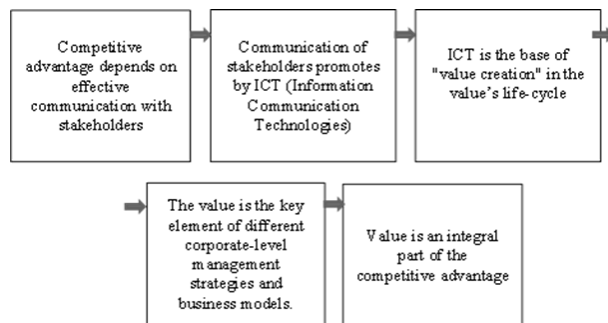
Fig.4. Competitive advantage, ICT and Value intermediation [29]

Based on the authors' performed study about the use of e-environment in Latvian companies [27], having studied value formation theories, having analysed the types and theories of business models, the authors have drawn a conclusion that the most suitable course of action would be to base further development on the Osterwalder's Business Model Canvas. [28]. Forbes has referred to this business model canvas as a simple instrument for creating innovative business models [29]. The model is based on active use of the e-environment in entrepreneurship. There are nine stakeholder groups at the basis of the model. Meanwhile, reciprocal and effective interaction and communication between the stakeholders promotes a company's competitiveness [23].

At the same time, value is an intrinsic part of a competitive advantage. It can be concluded that a competitive advantage depends on effective communication with stakeholders and customers. The previous study done by the authors about competitiveness of Latvian companies' shows that it is the use

of communications networks, being a constituent element of competitiveness of Latvian companies, that the companies are using the least [29]. Thus, the authors of the paper assume that by increasing E-environment element as part of IC system, the competitiveness companies will also increase.

4. CREATION OF COMPANY'S ADDED VALUE

The identification of value-drivers elements in IC system and their subsequent management is seen as the key to value added Authors present the model of IC describing the system how IC resources are used to increase value added. Author`s model of IC composed of mainly three components: human capital, structural capital (organisational capital) and relational capital (social capital).
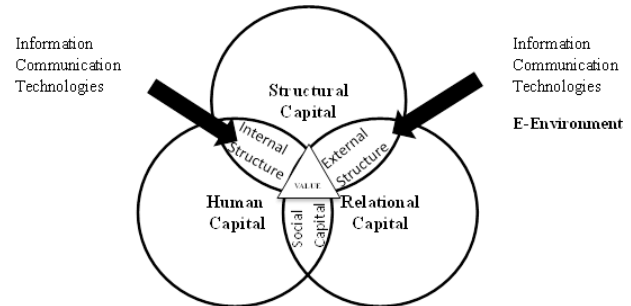
Fig.5. ICT, E-Environment and Value creation intermediation [30]

This model presented three main elements of VA creation - Human Capital is defined as the combined knowledge, skill, innovativeness, and ability of the company's individual employees to meet the task at hand. It also includes the company's values, culture, and philosophy. Structural Capital is the hardware, software, databases, organizational structure, patents, trademarks, and everything else of organizational capability that supports those employees' productivity - in other words, everything that gets left behind at the office when employees go home. Customer capital (Relational Capital) – provided by structural capital, the relationships developed with key customers.

5. MEASURING OF INTELLECTUAL CAPITAL AND INFORMATION COMMUNICATION TECHNOLOGIES

In the context of knowledge, because knowledge itself is invisible, its creation and use are hardly measureable. Nonetheless investing in ICT many valuable outputs are generated (brand, know-ho, etc.). Value generated by knowledge will probably have time lag (long-term) and not always have instant impact on profit (short-term). Promoting investments to ICT and specifically to E-environment, it is possible to evaluate company value.
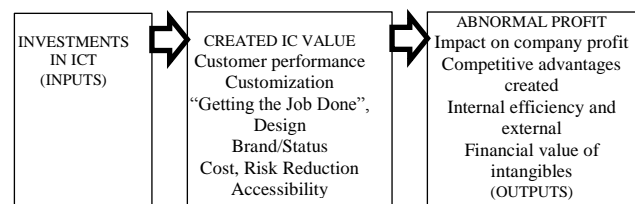

Fig.6. ICT, E-Environment and Value creation mediation [31]

As for beginning should calculate the investment made by company to ICT, comparing to abnormal revenue flow generated by ICT and intangible value created. This model helps to describe the methodology of authors of the paper quantitative evaluation model. Based on the model (Fig.5) puts an emphasis on external efficiency. Current quantitative model concentrates on external reporting, including internet statistics, investment analysis and methods for reporting the nonfinancial value of intangibles. So quantitative model is based mostly on VAIC (Value-Added Intellectual Coefficient) approach.

(1)

$$\mathrm{IICT}_t = \mathrm{HR}_t + \mathrm{SC}_t + \mathrm{RC}_t$$,

Where:

HR_t - Human Resources dedicated to specific ICT project (According to Pulic's concept mainly labour costs) at time t.

SC_t - Investments made to structural capital (maintenance, equipment, R&D costs) at time t,

RC_t is the relational capital expenses (advertising costs – e.g. such systems as Google index etc.) at time t. So we present formulae based on Sveiby (1997) VA creation model and author`s model based on Pulic (2000) (Fig. 3 and Fig 6.):

(2)

$$VA_t = (-\alpha_1 \mathrm{IICT}_t + \beta_1(AP_t)) * \delta_1(IV_t)$$,

Where **t**- value added created at time;

α_1- is correlation coefficient (function of time-series properties);

$IICT_t$ – capital (physical and financial) invested by company to specific ICT project at time t;

β_1- is correlation coefficient (function of time-series properties);

AP_t – abnormal profit generated by company through ICT project per t period;

δ_1- is correlation coefficient (see Table 2);

IV_t - intangible value generated by ICT at time t.

Or it could be seen as:

(3)

$$VA_t = (-\alpha_1 \sum_i^{n-1}(\mathrm{HR}_t + \mathrm{SC}_t + \mathrm{RC}_t) + \beta_1(AP_t)) * \delta_1(IV_t)$$,

Where AP_t - reported abnormal profit are based on a traditional accountant system;

IV_t - Intangible value generated by ICT could be calculated using specific parameters:

(4)

$$IV_t = f(IndX_1 + IndX_2 \dots IndX_{17} + etc.)$$

Authors recommend developing indexes researched in their previous work. (see short description on Table.1).

Table 1
IV indexes in value creation based on conducted research [29]

$IndX_1$	Market share based on unit sold through ICT
$IndX_2$	Relative market share
$IndX_3$	Penetration
$IndX_4$	Gross active customer volume
$IndX_5$	Information diffusion rate

$IndX_6$	Satisfaction
$IndX_7$	"Willing for searching"
$IndX_8$	"Try & Buy"
$IndX_9$	Penetration(t)
$IndX_{10}$	ICT sales forecast
$IndX_{11}$	Repeated sales
$IndX_{12}$	Trial volume
$IndX_{13}$	Opportunities-to-see
$IndX_{14}$	Clickthrough rate
$IndX_{15}$	Session index
$IndX_{16}$	Client behaviour dynamics
$IndX_{17}$	Client time-spent dynamics

Authors agree with Bontis *et.al.* conclusion and emphasise that every company could include or exclude their own indexes based on specific market condition and working profile, that is why formulae could be modified [9].

For better financial statistics we could use Financial Efficiency of VA (the Value Added Capital Coefficient or VACA) [31].

(5)

$$VACA = \frac{VA}{IICT}$$

The Value Added Capital Coefficient (VACA) shows the amount of value added created by a unit of capital (physical and financial) employed.

Human Resource Efficiency of VA (Value Added Human Capital Coefficient – VAHU) [31].

(6)

$$VAHU = \frac{VA}{HR_t}$$

Value Added Human Capital Coefficient – VAHU shows the amount of value added created by a of currency invested in employees during ICT project.

Structural Capital Efficiency of VA (Structural Capital Efficiency – SCE) [31].

(7)

$$SCE = \frac{VA}{\frac{1}{SC_t}} = \frac{SC_t}{VA}$$

Efficiency for structural capital is different. Pulic sustains that structural capital and human capital are reciprocal, the larger the share by human capital, the smaller the share by structural capital.

Structural Capital Efficiency – shows the share of Structural Capital in the value added created.

As soon as original formula of Pulic concept VA sees as [31]:

(8)

$$VAIC = \frac{VA}{CE} + \frac{VA}{HR} + \frac{SC}{VA}$$

Paulic method makes no distinction between flows and inventories [32]. Andriessen explains that Value added is a flow indicator for return on assets, as is capital employed, human capital, and structural capital (which are not inventory indicators). In the original Pulic concept under the VAIC method labour expenses are treated as inventories. Analysing the efficiency mentioned above, authors also conclude that according to Andriessen, this ratio does not mean efficiency, as it provides no information on the contribution by human capital in value creation, but more the share of specific input in VA. But nonetheless, authors insist that these shares in VA are still important for VA analysis as it gives us better understanding of α, β correlation coefficients .

Based on previously conducted research [29], authors of the article made description to the VA creation model. VA creation model regarding financial part (formula 1) - α, β coefficients. Authors also conclude that these values are very sensitive to company strategy. After analyzing financial figures (expenses and profits) coming from ICT, authors of the article suggests (for more information see also Roos et al. (1997) presented conclusions), that α, β coefficients should be based on company strategy [19].

It means that, if company`s strategy is sustainable development, these coefficients should be less sensitive, and, in case, company`s shareholder support speculative strategies - more sensitive. It could be easily understood as soon as one of IC capital fundamentals is long-term value creation.

As for nonfinancial part of our model, using previously conducted research (result are provided by SPSS), authors of the paper found that δ coefficient is more complicated and should be expressed as sum of correlations coefficients (Table 2) [29].

$$\delta = \sum_{i}^{n-1} f_i^j \tag{9}$$

Where δ_1 - is sum of correlation coefficients (Table 2);
$\quad$ i=1 - corresponding IC factor;
$\quad$ j=1 - corresponding IC correlation coefficient.

Table 2
IC factor`s correlation coefficients based on conducted research [29, 177-178]

Nr	Factor description	Factors variables	Correlation coefficient (average)
1	Knowledge about use of e-environment tools	7	0.917
2	Knowledge about ICT tools and its usage	3	0.725
3	Knowledge about e-environment models	1	-0.869
4	Communication with interested party online	4	0.795
5	ICT unit as sale and marketing instrument	1	0.701
6	Usage of e-environment tool for customer and marketing research	3	0.770
7	Knowledge about institutional services	5	0.852
8	E-environment tool acceptation in HR	2	0.825
9	E-environment tool diversity	2	0.811
10	Use of institutional services for business goals.	2	0.736

6. CONCLUSIONS

In this paper our main goal was to reveal the main theoretical and practical aspects of the company's value added. We created the concept model to discover the relation between company's value added and intellectual capital.

The aim is to evaluate the relation between company's value added and intellectual capital. Our model is created to explain the causal relation among these three types of capital in IC model and value added. Only by measuring the causal relation will IC efficiency be determined. Authors declare that Intellectual capital is one of the most relevant intangibles for a company, and the concept with measurement, and the relation with value creation is necessary for modern markets.

We added the variables that will help companies to evaluate contribution of each element of IC.

More empirical researches are needed to investigate the relation effect of Intellectual capital on value creation.

Finally, our findings indicate that the relations between IC elements and VA are complicated. This relationship has different inputs with different evaluating methods and specific impact on VA. The results suggest that companies do not homogeneously receive outputs from IC elements. Nonetheless our presented coefficients and time-series properties have made some contributions to dynamic nature of IC relation and VA.

This paper emphasizes the importance of a company's IC and the positive interaction between them in generating profits for company

As a first step of our research, there is not yet fully approved direct correlation between model factors, and therefore authors are forced to use in future researches more information analysis presented by companies.

7. REFERENCES

[1] Kay, J. 1995. Foundations of corporate success: how business strategies add value. Oxford: Oxford university press.

[2] McLean, G. N. 2006. Organization development: principles, processes, performance. San Francisco: Berrett-Koehler.

[3] Pitelis, C. N. 2009. The co-evolution of organizational value capture, value creation and sustainable advantage, **Organization Studies** 30: 1115-1139. doi:10.1177/0170840609346977

[4] Bowman, C.; Ambrosini, V. 2010. How value is created, captured and destroyed, **European Business Review** 22(5): 479-495. doi:10.1108/09555341011068903

[5] Bang, A.; Molgaard, C. C.; Pia, B. 2010. How to create business value in the knowledge economy. Accelerating thoughts of Peter F. Drucker, **Management Decision** 48(4): 616-627. doi:10.1108/00251741011041382

[6] Porter, M. E. 1979. "How competitive forces shape strategy". **Harvard Business Review**. Retrieved 9 September 2013

[7] OECD. 1996. Measuring what people know. Paris: Organization for Economic Co-operation and Development.

[8] Zeghal, D.; Maaloul, A. 2010. Analysing value added as an indicator of intellectual capital and its consequences on company performance, **Journal of Intellectual Capital** 11(1): 39-60.

doi:10.1108/14691931011013325

[9] Bontis, N. 1999. "Managing Organizational Knowledge by Diagnosing Intellectual Capital: Framing and Advancing the State of the Field", **International Journal of Technology Management**, 18, 5/6/7/8, 433-462.

[10] Stewart, T.A. 1991. "Brainpower: How Intellectual Capital is Becoming America's Most Valuable Asset". **Fortune**. June 3, 1991. pp. 44-60.

[11] Edvinsson, L., and Malone, M.S. 1997. Intellectual Capital: Realizing Your Company's True Value by Finding its Hidden Brainpower. Harper Business, New York.

[12] O'Regan, D.; O'Donnel, P.Coates, 2000. Intellectual capital: a habermasian introduction, **Journal of Intellectual Capital** 1(2): 187-200. doi:10.1108/14691930010377496

[13] Sveiby, K.E. 1997. The New Organizational Wealth: Managing and Measuring Knowledge-based Assets. San Francisco: Barrett-Kohler Publishers

[14] Drucker, P. F. 1993. Post-capitalist society. Harper business, New York, NY.

[15] Brooking, A. 1996. Intellectual Capital: Core Assets for the Third Millennium Enterprise. London: Thomson Business Press, London.

[16] Standfield, K. 1999. Knowcorp, [Electronic resource], viewed on March 22, 2014 <http://www.knowcorp.com>.

[17] Pike, S., Rylander, A. and Roos, G. 2000. Intellectual capital management and disclosure In the Strategic **Management of Intellectual Capital and Organisational Knowledge**. Oxford University Press, New York.

[18] Rylander, A., Jacobsen, K. and Roos, G. 2000. Towards improved information disclosure on intellectual capital. **International Journal of Technology Management** 20(5–8), p. 715–741.

[19] Roos, J., G., Dragonetti, N.C., and Edvinsson, L. 1997. Intellectual Capital: Navigating in the New Business Landscape. Macmillan: London.

[20] Porter, M. 2008. The Five Competitive Forces That Shape Strategy. **Harvard Business Review** (01).

[21] Fruhan, W. 1979. "The NPV Model Strategy – the Shareholder Vale Model" in Financial Strategy: Studies in the Creation, Transfer, and Destruction of Shareholder Value. Homewood: Irwin, p. 122-124.

[22] Kambil, A.; Ginsberg, A; Block, M. Rethinking Values Propositions// **Working Paper, NYU Centre of Research of Information Systems**. 1997, p. 34-36.

[23] Osterwalder, A., Pigneur, Y. 2009. **Business Model Generation**. Hoboken: John Wiley & Sons, p.279.

[24] Freeman, R. 1984. **Edward Strategic Management: a Stakeholder Approach**. Boston: Pitman, 276 p.

[25] Osterwalder, A., Pigneur, Y. 2003. Modeling Value Proposition in E-business// **ICEC'2003 Proceedings**. Lausanne: University of Lausanne.

[26] Laudon, K; Traver, C. 2010. **E-Commerce**. 6th edition. Upper Saddle River: Prentice Hall, p.960-961.

[27] Ščeulovs, D.; Gaile-Sarkane, E. 2010. Identification of factors affecting consumer habits in the e-environment// **Business and Management 2010 selected papers.** Volume 1. Vilnius: VGTU Press "Technika", p. 965-970.

[28] Business Model Foundry: The Business Model Canvas. [Electronic resource], viewed on March 20, 2014,<http://www.businessmodelgeneration.com/canvas>

[29] Ščeulovs, D. 2013. Use of Electronic Environment in Entrepreneurship Development// **Doctoral Thesis.** Riga: RTU Press, p. 80-87.

[30] R. Hermans, I. Kauranen. 2005. Value creation potential of intellectual capital in biotechnology – empirical evidence from Finland. **R&D Management** 35, 2. Blackwell Publishing Ltd.

[31] Pullic, Ante. (2000) VAIC TM: an accounting tool for IC management. **International Journal of Technology Management**, Vol.20, nº5,6,7,8. 2000. MVA.

[32] Andriessen, D. Making sense of Intellectual Capital: designing a method for the Valuation of Intangibles. **Elsevier**: Oxford. 2004.

[33] Lapiņa, I., Maurāne, G., Stariņeca, O. Holistic Human Resource Management Model towards Sustainable and Sophisticated Business// **Proceedings of the 17th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2013)**. p.31-36. [Electronic resource], viewed on April 12, 2014 <http://www.iiis.org/Proceedings/2013Proc/ViewProc2013.asp?id=SA0213>.

Evaluation of Uncertainty Level on the Stages of Business Cycle: Implementation of Quantum Principles

Anna SVIRINA, Elena PARFENOVA, Elena SHURKINA
Department of Economics and Management, Department of Natural Science, Department of Applied Mechanics and Mathematics, Kazan National Research Technical University
10 K.Marksa street, Kazan, 420111, Russia

ABSTRACT

The goal of the research is to propose implementation of quantum principles for evaluation of economic development on stages of business cycle, define the difference between traditional (deterministic) and quantum approaches and to provide quantitative analysis based argumentation for use of quantum economic principles in evaluation of internal and external factors on the stages of business cycle. The object of the study is possibility and reliability of quantum economic principles implementation to evaluate economic system performance. The authors analyze existing approaches towards implementation of deterministic, probabilistic and quantum models for estimating internal and external factors on stages of business cycle, define the benchmark for shift from traditional economy models and principles to quantum principles, describe the stages of business cycle from the quantum economics point of view and provide quantitative analysis of deterministic and quantum models quality on the level of enterprise to prove efficiency and reliability of quantum principles based approach. Calculations and data processing were carried out using Microsoft Excel and SSPS Statistics software.

Keywords: business cycle, quantum principles, quantum economics, model.

1. INTRODUCTION

The global economic crisis had revealed that existing prognosis models that are used for both macroeconomic and microeconomic indicators evaluation have relatively high range of margin, and this does not allow having an accurate idea for creating a strategy. This fact provoked a number of research initiatives and policy changes, and all of them are aiming to estimate the basic principles for internal and external factors evaluation in the modern economy – in order to provide accurate and reliable models for strategic and tactical evaluation of socio-economic environment dynamics.

One of the reasons for relatively low accuracy of existing economic model is the deterministic approach that is being implemented within them: despite the fact that socio-economic environment is a stochastic one, the models introduced to carry out prognosis are mainly determined factor type, or linear equations, and both of them do not allow to evaluate factors (especially the factors of social origin) that have high range of margin. The problem of modelling such processes is solved in natural sciences by introducing probabilistic models, or even models based on quantum principles that allow evaluating uncertainty as a factor defining dynamics of other factors' performance.

The problem of uncertainty as a key factor that affects socio-economic processes was stated by a number of economic scholars [2, 13], but the definition of this phenomena have stayed qualitative; and at the current state of economic system development evaluation of uncertainty seems to become an important element of modelling.

Previously conducted research suggests economic models that evaluate uncertainty as a side, but not a key factor of development; at the same time quantitative analysis, provided by a number of scholars [6, 11, 24] proves that some of the trends are to be evaluated by using uncertainty factor as a key one. In this paper we aim to define the difference between traditional (deterministic) and quantum principles based approach towards economic modelling, and to provide evidence on advantages of using quantum based approach.

2. NATURE OF BUSINESS CYCLES: LITERATURE REVIEW

Theoretical approaches explaining business cycles were created on the basis of economic development retrospective data analysis and evaluation. Summarization of these approaches provided most common definition of business cycles: "Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises; a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximately their own" [4]. According to this definition business cycles are non-linear and reflect diversity of economic development stages, its structure is viewed as a certain sequence of phases on the basis of retrospective data evaluation, duration of cycles and reasons behind changes of cycle stage have objective nature and there is some interrelation between different types of business cycles.

Analysis of existing literature on business cycles shows there are a number of theories that explain cyclic nature of economic development: starting from global level and ending with theories explaining business cycle of a single enterprise [13]. Provided explanations of business cycles' nature was summarized by Rothbard: "It should be recognized that most business-cycle theories – Keynesian, Marxist, Friedmanite, or whatever – and remedies are grounded in the assumption that the cycle stems from some deep flaw in the free-market economy" [21], and a list of reasoning for cyclic development of economic system include at least climate and geopolitics, demographics, social shifts, political shocks, discoveries and innovations, and changes in governmental policies and regulatory procedures, as well as a few other important reasons [13].

At the same time theories explaining business cycles tend to evaluate time as the main independent variable that defines business cycle stage and trend of development; on the basis of this parameter cycles were divided into 3 groups: macrocycles (that are hundreds or even thousands years in length, these are mainly social cycles such as proposed by Gumilev [12], mesocycles (lasting for several years or decades) and microcycles (these cycles last for days, weeks or months). Mesocycles include those of S. Kuznets [17] or the Schumpeterian "innovation lifecycles" [27] that can have duration from a year to several decades; another example of such cycles are Karl Marx's "cycles of main capital turnover" [19] that last for 50-60 years or Kondratiev's waves [16]. On mesolevel there were also a few attempts to estimate the relationship between military activity and business cycles as well as relation of business cycles to the industry dynamics, or to link value profile of society/enterprise and business cycles [9].

On macrolevel one can find Braudel cycles [15], which can be 150 years long, or logistic cycles lasting for 250 years or more [5]. On microlevel we can see a number of different cycles which mainly apply to a single company performance. From mid-XX century economic, social, political, and cultural cycles started to be viewed as a complex phenomena, which meant business cycles of different origin we supposed to complement one another. Henceforth complex synthetic models of cyclical socio-economic development were created [7, 8] to integrate short-term and long-term socio-economic cycles – such approach is illustrated with a graph with multiple sine curves which include crisis, depression, recovery, and growth stages. Each stage has certain features that are repeated over time at the same stage in the next cycle – and thus existing research is imposing deterministic approach towards defining stage and perspectives of economic development on each stage of business cycle.

Existing literature suggests a number of reasons to explain socio-economic system dynamics on different stages of business cycle: some suppose that the sequence of stages depends on the level of systemic risk [14]; others propose that changing of the stages within a business cycle is related to changes in specific monetary policies [3, 27, 28] or define changes as consequence of volatility of money markets [25]. Gersbach and Rochet [10] and Lorenzoni [18] observe a correlation between economic and credit cycles, and Angeletos and La'O [1] found interrelation between informational frictions and stages of business cycle; North et al proves relatively significant institutional influence onto dynamics of countries' economic growth [19]. All of this findings, though shed some light on the nature of business cycles (and especially reasons that provoke crisis), still operate within space of determinism. At the same time existing research shows, that in the long-term period deterministic models that are used to create prognosis of trends of economic development has an average fluctuation of long-term prognosis lies in the range of 24-35%, which means these models are very limited. In our opinion, this is due to the fact that for model development a factor of uncertainty is ignored – and henceforth, creation of models that define socio-economic systems behavior on the stages of business cycle are to be based on the principles proposed by quantum theory in order to allow defining uncertainty as a key factor of economic development.

3. PROPOSITION OF QUANTUM PRINCIPLES FOR MODELLING ECONOMIC DEVELOPMENT ON STAGES OF BUSINESS CYCLE

Analysis of global economic system development in last 40 years shows that a number of fundamental principles of classical political economy, such as, for example, deterministic laws of supply and demand [11] do not explain facts provided by empirical evidence. This was outlined by a number of researches, who tried to develop an alternative model of economic growth on the basis of quantum principles [6, 21, 24, 26]; and on the basis of their research we make the following proposal: classical political economy is based on deterministic principles, while modern economy has a quantum nature – therefore main principles of classical theory are proven in modern world only with a certain probability.

Conversion point from classical to quantum economy is defined in our opinion by the ratio between the volumes of real and financial sectors of the economy: in case there is parity between those, the laws of classical economy operate; if these sectors are highly disproportional – the laws of quantum economics start operating. This can be formalized in the following way:

$$MV = PQ + \hbar * PQ = PQ * (1 + \hbar), \qquad (1)$$

where MV – the volume of stock of money and quasi-money (financial aggregates) corrected by the turnover speed of this total stock, monetary units; PQ – the volume of real sector of the economy, defined as a sum of goods and services produced, corrected by the price, monetary units; $\hbar$ – parameter, characterizing disproportion between financial and real sectors of the economy.

In case $\hbar$ defines excess of quasimonetary mass over real sector volume by certain number of times, the laws of quantum economics come into operation, and replace laws of classical economics, which are to be implemented before reaching this disproportion point. Introduction of $\hbar$ allows to explain contradictions which occur when laws of classical economics are used to structure statistical data, describing contemporary trends of economic growth and, also, the ones that come into focus when theories, explaining behavior of economic subsystems that belong to different levels.

Estimation of conversion point ($\hbar$ level), which characterizes the moment classical economy laws are replaced by laws of quantum economics, would allow to define the range of implementation for both theories: in the range of classical economy the laws explaining behavior of economic agents are determined (in this case economic parameters can be defined both for the current moment and in perspective), while behavior of economic agents in the realm of quantum economy is probabilistic (in this case economic parameters can not be strictly defined since they exist under uncertainty). For example, within classical political economy exists only one equilibrium price defined by determined supply and demand curves – while in quantum economy there would be n probabilistic supply and demand curves, and, as a result, n different probabilistic equilibrium prices. But for the certain time period and certain socio-economic system the factors affecting probability of different equilibrium prices can be defined, and probable combinations of supply and demand can be defined with a certain level of probability, which allows making predictions of economic agents' behavior in certain points of time-space continuum. Analysis of the time and spatial characteristics of the process which occurs within quantum economic system should be carried out by taking into consideration, that if spatial

coordinate is fixed, timing coordinates increase to infinity, and vice versa – henceforth in quantum economics laws of classical economy can be formalized as well with implementation of probability characteristic.

This fact can be explain in the first place by the fact that basic economic equilibrium ($MV = PQ$) within quantum economic space is uncertain, and this uncertainty is provoked mainly by the uncertainty of stock of money and quasi-money in circulation within certain time period. Since volume of quasi-money (especially in case of derivatives market) can be derived from two characteristics – energy one (market value) and timing one (payment period) - fixation of each of the named characteristics within quantum economics space would lead to infinite growth of the other characteristic, and henceforth the main monetarist identity law can not be adhered in the fixed moment of time (due to infinite growth of stock of money and quasi-money). This thesis can be formalized in the following way:

$$\Delta M * \Delta t \geq \hbar \qquad (2)$$

where ΔM is the uncertainty of stock of money and quasi-money, monetary units; Δt – uncertainty of time stock of money and quasi-money is used.

In this case regulation of global economic system can be efficient only within a relatively significant time period, which has a proof in empirical data in terms of anti-cyclic regulation efficiency trends – and indirectly proves existence of both classical and quantum economy. Our qualitative research had shown that current state of global economic development can be considered as quantum one, and it can be derived also from the fact that deterministic models have a very high margin of error, while probabilistic models that are used in some sectors knowledge provide more accurate evaluations.

Within the proposed system the stage of growth in the business cycle becomes longer on each whorl of economic development. During this stage, there is economic space that is not filled, which leads to significant opportunities for entrepreneurs, that can be exploited. The features of this stage are booming economic growth, increased interrelations of the sub-autonomy subsystems that appeared during the recovery stage, and increasing boundary and currency exchanges. The financial sector is growing in accordance with increasing demand from the entrepreneurial sector. New economic formations provide additional opportunities for the use of new types of resources and create the conditions for higher intensity of resource implementation; the new economic formations also generate prerequisites for the creation of new technological solutions, which leads to further growth in entrepreneurial opportunities. During this stage, the growth of the entrepreneurial sector predominates and drags down the growth of the financial sector. By the time a socio-economic system has reached the middle of the growth stage, the economic growth rate begins to decrease. Another important feature is the gradual formation of an institutional environment that is adequate to meet the challenges created by the new economic formation; however, because the entrepreneurial sector is growing rapidly, the level of regulation is relatively low. By the end of this stage, the institutional environment becomes increasingly rigid, and economic space becomes more structured. Governmental interference at this stage should be limited to institutional development, which should stimulate higher intensity of resource use. Entrepreneurial activity occurs mainly in the formal sector, in which relatively loose institutional regulation does not erect significant barriers in terms of exercising

entrepreneurial opportunities. Finally, it is notable that this stage is characterized by an innovation boom.

The stabilization stage begins when growth rates in the entrepreneurial and financial sectors even out. From this point, economic growth is ensured by the prevailing development of the financial sector, as new elements of monetary and/or quasi-monetary mass are developed or when the implementation of new technologies and qualified labor allows an increase in the turnover velocity of financial instruments. The entrepreneurial sector grows during the stabilization stage, but the growth rate is relatively low because the entrepreneurial sector is nearing its limit in this economic formation. The financial sector, by contrast, grows rapidly. Exchange processes within the socio-economic system are intensified, mainly because of increasing speculation, which is largely influenced by the relatively rigid institutional environment that now restricts entrepreneurial activity. Institutions become increasingly harsh; in some cases, institutional regulations appear even before prior to the creation of corresponding entrepreneurial activities. This rigidity leads to an increase in informal economy operations, which, in turn, provokes increasing institutional rigidity. Entrepreneurial activity and resources are driven to the financial sector, in which a high level of basic resource mobility prevents excessive rigidity in the institutional environment. The development of innovations becomes an expensive process, and it restricts entrepreneurial activity in the innovation sector, which results in the prevailing creation of quasi-innovations. The polarization of economic space gradually increases, mainly as a result of the concentration of financial activity at certain points of this space.

When the institutional environment becomes rigid, i.e., when the share of operations in the informal sector of the economy is in excess of one-fourth of the total amount of entrepreneurial operations and the cost of creating innovations begins to exceed the cumulative resources of a large enterprise, the recession stage is about to begin. The specific feature of this stage is a low rate of economic growth. The level of institutional rigidity continues to grow via the activity of multinational companies and supranational regulation. This leads to further growth in the entrepreneurial and financial sectors of the informal economy. By the end of this stage, the informal economy has become larger than the formal economy, which means that the informal economy produces more than half of the services and commodities consumed in the socio-economic system. Innovational activity gradually decreases. The period of innovational development exceeds the visible planning horizon of a single enterprise, and the development of fundamental innovations becomes prohibitively expensive, partly as a result of the concentration of resources in the financial sector. The informal economy begins to produce its own institutional restrictions. This process is provoked by the approach of the maximum activity within the current economic formation, on the one hand, and by rapidly increasing expert power, on the other hand. Approaching the limitations of the current economic formation leads to abrupt growth in innovation-development costs and time consumption. If any fundamental innovations are created at this stage of the business cycle, they are typically products and technologies from the upcoming economic formation, and the development of technologies and materials required for a corresponding mass production becomes prohibitively expensive. The polarization of economic space continues to increase, and by the end of the stage, a tremendous share of the financial sector is concentrated within a few points of economic space. These points of concentration contain a significant share of the socio-economic system's monetary and

quasi-monetary mass and provide conditions to increase the turnover velocity of these assets. Frequently, these points become the spots at which systemic economic crisis begins because the pressure of the financial sector on the entrepreneurial sector is extremely high at these points, which can cause the system to collapse. The economic energy released as a result of this collapse is retranslated to other points in the economic system, and the new recovery stage in the new economic formation begins. The change to a new economic formation is characterized by the collapse of both the institutional and financial systems from the previous business cycle. Recovery begins at the growth points that appeared before the collapse of the previous economic formation.

According to the model described above, the sequence of economic stages is not a function but a projection of the described multidimensional process on a time axis. The growth trend featured in this figure appears as a result of a change in economic formation. Henceforth, time is an independent variable that describes a process of business cycle movement for an observer but is not an independent variable that defines the structure or duration of stages within a business cycle.

4. PROPOSITION OF QUANTUM PRINCIPLES FOR MODELLING ECONOMIC DEVELOPMENT ON STAGES OF BUSINESS CYCLE

Proposition of suggested approach on the level of single enterprise means its external and internal environment are to be evaluated by taking into account uncertainty level. In order to use this approach we propose to evaluate resources as a key internal development factor of a company in the following way:

$$RES_i = res_i \sum_{k=1}^{q} res_{ik} \psi_{ik}, \qquad (3)$$

where RES_i – is amount of i-th resource of an enterprise that is available at a certain time point, monetary or non-monetary units; res_i – is the amount of i-th resource that a company possesses according to deterministic approach; ψ is a wave-type function that describes a state of res_i that is characteristic of state of resource used by a company at each point in time during the its existence; and k $(1\div q)$ is the number of unique states a resource could be found during a certain period. The wave function suggested implements Heizenberg's principle for socio-economic development, and allows to formalize a way to define uncertainty level at each stage of business cycle on the basis of internal and external environment evaluation (the latter means $\hbar$ also affects the type of wave function, and this indicator has a different value on each stage of business cycle – according to the stage of global development).

Implementation of this approach allows viewing enterprises' tangible and intangible assets not as a resource it can use when needed without any limitations, but as a source that can be used with different efficiency – and range of this efficiency can be only defined by implementing indicator that allows estimating level of uncertainty.

For the purposes of this research we have evaluated the data on efficiency of nine companies (both SME's and corporations from different sectors of industry: equipment building, oil processing, oil production, chemistry, telecommunications and services), which included pre-evaluation of resource efficiency – calculated as a ratio of profits (as EBITDA) to the summarized value of resource estimated – by both deterministic and quantum models, and

efficiency that was found in fact three month after the pre-evaluation. For the purposes of this study we estimated efficiency of use for technical, technological, human, informational and financial resources; their value was estimated either on the basis of accounting data acquired within the enterprises, or we used estimations derived from expert opinions that were acquired by both questionnaire and interviewing.

The results of Pearson correlation analysis for technical and human resource efficiency are presented accordingly in Table 1 and Table 2. For all the other factors the results are similar to the ones provided in the Tables.

Table 1. Pearson correlation between deterministic, quantum and factual values of technical resource efficiency

		DTRE	QTRE	FTRE
Deterministic technical resource efficiency (DTRE)	Pearson correlation	1	,919**	,951**
	Value		,000	,000
	N	9	9	9
Quantum technical resource efficiency (QTRE)	Pearson correlation	,919**	1	,985**
	Value	,000		,000
	N	9	9	9
Factual technical resource efficiency (FTRE)	Pearson correlation	,951**	,985**	1
	Value	,000	,000	
	N	9	9	9
**. Correlation is significant at 0,01				

Table 2. Pearson correlation between deterministic, quantum and factual values of human resource efficiency

		DHRE	QHRE	FHRE
Deterministic human resource efficiency (DTRE)	Pearson correlation	1	,902**	,908**
	Value		,000	,000
	N	9	9	9
Quantum human resource efficiency (QTRE)	Pearson correlation	,902**	1	1,000**
	Value	,000		,000
	N	9	9	9
Factual human resource efficiency (FTRE)	Pearson correlation	1,000**	,985**	1
	Value	,000	,000	
	N	9	9	9
**. Correlation is significant at 0,01				

As it can be seen from both tables, and a similar results were acquired for other resources efficiency, use of quantum principle based model using probabilistic approach provides more accurate prognosis. The same set of data was tested using non-parametric correlations, and the results were the following: for quantum model of technical resource efficiency estimation Kendall's tau-b was equal 1.000 (significant at 0.01) for Spearman's coefficient – 1.000; while Kendall's indicator for deterministic model appeared to be only .667, and Spearman's coefficient - .833 (all significant at 0.01). For human resource efficiency we acquired the following results: Kendall's tau-b was equal 1.000 for Spearman's coefficient – 1.000; while Kendall's indicator for deterministic model appeared to be only .556, and Spearman's coefficient - .717 (all significant at 0.01). So in this case the difference is even larger than in case of Pearson coefficient testing. Non-parametric correlation analysis for the variable with minimum range between factual, quantum and deterministic values is shown in Table 3.

As it can be seen from the table, even in case of relatively small range of margin for the estimated variable

(informational resources efficiency), non-parametric correlation for quantum-principle based estimation is higher than for deterministic model, that proves the quality of this type of models to be used for prognosis (even in case of low ranges quantum models have 1.000 correlation instead of .967 shown by Spearman correlation, and instead of .889 for Kendall's indicator). It also seems important that correlation between deterministic results and both quantum and factual results is the same, and shows the difference between the two approaches is significant.

Table 3. Non-parametric correlation between deterministic, quantum and factual values of informational resource efficiency

			DIRE	QIRE	FIRE
Kendall's coefficient	Deterministic informational resource efficiency (DIRE)	Correlation coefficient	1,000	,889[**]	,889[**]
	Quantum informational resource efficiency (QIRE)	Correlation coefficient	,889[**]	1,000	1,000[**]
	Factual informational resource efficiency (FIRE)	Correlation coefficient	,889[**]	1,000[**]	1,000
Spearmen coefficient	Deterministic informational resource efficiency (DIRE)	Correlation coefficient	1,000	,967[**]	,967[**]
	Quantum informational resource efficiency (QIRE)	Correlation coefficient	,967[**]	1,000	1,000[**]
	Factual informational resource efficiency (FIRE)	Correlation coefficient	,967[**]	1,000[**]	1,000
**. Correlation is significant at 0,01					

Within the proposed research visualization of ranges of margin was carried out as well, and the results for financial resource efficiency can be seen in Figure 1.

As it can be derived form that figure, the level of error for quantum principles based estimation is relatively lower than in case of deterministic modeling of the indicator, and that also proves that quantum-based models are more reliable in case of determining the future value of certain internal factors. Similar shapes of the figures were acquired within graph analysis of technical, technological, informational and human resources efficiency evaluation, which allows to generalize the conclusion on higher accuracy of quantum principles based model relative to deterministic approach based.
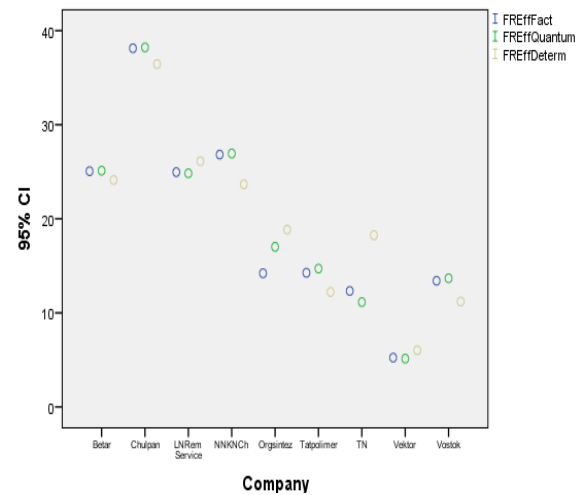
Figure 1. Errorbar for financial resources efficiency (deterministic, factual and quantum estimations)

The results acquired within this research allow proposing the necessity to use quantum principle based models for evaluation of internal and external factors defining behavior of an enterprise in the current economic environment.

5. CONCLUSIONS

During analysis of possibility and reliability of implementing quantum principles to evaluate internal and external environment of an enterprise on the stages of business cycle we have come to the following conclusions:

1. Existing theoretical approaches towards evaluation of business cycles have the major limitation – they view time factor as an independent variable for evaluation of socio-economic systems performance on different stage of business cycle. This leads to wide range of cycles' duration even within one type of a cycle, and that proposes a conclusion that socio-economic system's performance is influenced not only by time factor.

2. Analysis of the contemporary socio-economic system structure (in terms of financial and real economy sector balance) shows that it significantly differs from the structure within which the basic principles of existing deterministic economy are developed, and thus evaluation of development trends in the modern economy are to be based on non-deterministic principles.

3. Investigation of uncertainty influence on socio-economic development drives a conclusion that quantum principles should be implemented for evaluation of its performance, since existing disproportion between financial and real sector do not allow to implement deterministic approach with necessary accuracy.

4. Estimation of enterprise's resource efficiency, carried out on the basis of both quantum principles (using wave function proposed by Heisenberg for evaluation of uncertainty) and deterministic based approach had proven that quantum based models provide more accurate results on micro-level. In case quantum principles are implemented for enterprise's performance evaluation, uncertainty level has to be defined by both macroeconomic uncertainty which is derived from the stage of macroeconomic cycle, and microeconomic uncertainty that has to be defined at the level of enterprise.

Several proposals and limitations of the study were made as the result of this research:

1. In order to improve accuracy of prognosis implemented at the company level a quantum based approach using wave function to evaluate uncertainty, should be used. This approach allows to decrease the range of error (that was proven by analysis of factual and predicted level of resource efficiency, which was carried out for a number of enterprises from different sectors of economy), and henceforth provides sustainable platform for decision-making.

2. Macroeconomic level of uncertainty has to be evaluated by estimating disproportion of financial and real sector of economy volume, certain level of which indicates shift from traditional (deterministic) economy to quantum economy.

3. Taking into consideration the number of cases analyzed to prove efficiency of proposed quantum principles based model for estimation of enterprise's future resource efficiency we need to take into consideration some limitations that arise from a small data sample that was used for the purposes of this study, which means results are to be tested additionally in future.

6. REFERENCES

[1] G.Angeletos, J.La'O. Noisy Business Cycles. NBER Macroeconomics Annual Vol. 24 No. 1, 2009. pp. 319-378.

[2] I.Ansoff. Strategic Management. Classic edition. London: Palgrave Macmillan, 2007. pp.87-114.

[3] F.Bilbiie, F.Ghironi, M.Melitz, V.Midrigan, J.Rotemberg. Monetary Policy and Business Cycles with Endogenous Entry and Products Variety. NBER Macroeconomics Annual Vol. 22, 2007. pp. 69-91.

[4] A.Burns, M.Wesley. 1946. Measuring Economic cycles. National Bureau of Economic Research. 1946. pp. 34-86.

[5] R.Cameron, L.Neal. Concise Economic History of the World: From Paleolithic Times to the Present. Oxford: Oxford University Press, 2003. pp. 69-104.

[6] A.Cencini. Monetary Macroeconomics. A New Approach. New York: Routledge, 2001. pp. 63-84.

[7] D.Comin, M.Gertler. Medium-Term Economic cycles. The American Economic Review Vol. 96 No.3, 2006. pp. 523-551

[8] R.Day. Irregular Growth Cycles. The American Economic Review, 1982. Vol. 72 No. 3, pp. 406-414.

[9] S.Eusepi, B.Preston. *Expectations, Learning and Economic cycle Fluctuations*. American Economic Review Vol. 101, 2011. pp.87-83

[10] H.Gersbach, J.Rochet. Capital Regulations and Credit Fluctuations. CEPR Discussion Paper No. 9077. Retrieved from cepr.org on March 1st, 2014.

[11] C.Goncaves. Chaos and Nonlinear Dynamics in Quantum Artificial Economy. 2012. arXiv:1202.6647v1 29Feb2012.

[12] L.Gumilyov. Ethnogenesis and Biosphere of Earth. Moscow: AST, 2005. pp. 132-146.

[13] A.Hansen. Business Cycles and National Income. NY: Norton publishing, 1951. pp. 23-69.

[14] Z.He, A.Krishnamurthy. A Macroeconomic Framework for Quantifying Systemic Risk. EAE Annual Meeting proceedings. San Diego 4-6 January, 2013.

[15] G.Khakimov. Braudel's „long-distanced time" as a methodological instrument of social sciences cognition. Moscow: Voprosy philosophii, 2009. pp. 46-72.

[16] N.Kondraty'ev. *Problemy Ekonomicheskoy Dinamiki*. Moscow: Economica, 1989. pp. 76-138.

[17] S.Kuznets. Economic Growth and Structure: Selected Essays. New York: Norton, 1965. pp. 82-106.

[18] G.Lorenzoni. Inefficient Credit Booms. Review of Economic Studies Vol. No. 3, 2008. pp. 809-833.

[19] K.Marx. Capital: Critique of Political Economy. London: Penguin Books, 1990. pp. 164-179.

[20] D.North, J.Wallis, B.Weingast. Violence and Social Order: a Conceptual Framework for Interpreting Recorded History. Cambridge: The Syndicate of the Press of the Cambridge University, 2009. pp. 34-42.

[21] S.Rossi. The Theory of Money emissions. A Handbook of Alternative Monetary Economics. Cheltenham and Northampton: Edward Elgar, 2006. pp. 95-112.

[22] M. Rothbard. The Kondratieff Cycle: Real or Fabricated? London: Investment Insights: August, September, 1984.

[23] J.Schumpeter. Economic cycles: A theoretical, historical and statistical analysis of the Capitalist process. Eastford: Martino Publishing, 1939. pp. 43-87.

[24] C.Sobey. A Model of Quantum Economic Development. *2012.* http://mpra.ub.uni-muenchen.de/36422/

[25] J.Taylor,J. Williams. A Black Swan in the Money Market. American Economic Journal: Macroeconomics Vol. 1 No.1, 2009. pp. 58-83.

[26] S.Ternyik. *The Monetary Quantum.* 2012. http://mpra.ub.uni-muenchen.de/41973/

[27] M.Woodford. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton, NJ: Princeton University Press, 2003. pp. 86-97.

[28] M.Woodford. Convergence in Macroeconomics: Elements of the New Synthesis. American Economic Journal: Macroeconomics Vol. 1, No.1, 2009. pp. 267-279. .

Competition and Efficiency in the Banking Sector of EU New Member States

Jelena TITKO
Department of Finance, Riga Technical University
6 Kalnciema str., Riga, LV-1048, Latvia

and

Natalja LACE
Department of Finance, Riga Technical University
6 Kalnciema str., Riga, LV-1048, Latvia

and

Manatzhan TLEUZHANOVA
Kazakh University of Economics, Finance and International Trade
7 Zhubanov str., Astana, Kazakhstan

ABSTRACT

Bank performance and, in particular, bank efficiency is a frequently debated question in the academic environment due to the extremely important role of banks in the national economy. A wide range of studies are devoted to the exploration of bank efficiency drivers. The goal of the current research is to investigate the relationship between competition and efficiency in the banking sector of new member states of the European Union.

Non-parametric frontier technique - Data Envelopment Analysis (DEA) - was used as a tool for measuring bank relative efficiency. Input-oriented DEA model under Variable Returns to Scale assumption was applied. For measuring competition in the banking sector such measures, as Herfindahl-Hirschman Index (HHI) and the share of assets held by the top 5 banks (CR5), were used as proxies.

Study was based on the sample data of 97 banks operating in seven New Member States of EU, covering the period of 2006-2012. Analysis revealed the fact that statistically significant correlation exists between average efficiency and both concentration index and HHI, indicating strong negative relationship between competition and efficiency in the banking sector of NMS.

Keywords: Bank efficiency, Competition, Concentration, DEA, European banks.

1. INTRODUCTION

The overwhelming goal of any bank is shareholder value maximization. One of the main components contributing to the shareholder value is profitability that, in turn, is affected by efficiency. Companies with a greater efficiency are able to lower costs and, consequently, to offer lower prices, gain market share and earn more profit [1]. Thus, the questions of efficiency measuring and enhancing in the banking industry are extremely important both for bank shareholders and managers. Besides, considering the role of banks in the national economy, efficiency-related issues are essential at the government level.

Topicality of the research subject is confirmed also by the results of the authors' conducted research, using such databases, as Scopus, ProQuest and Science Direct. Making a search with the keywords "bank efficiency" more than 50000 articles published since 2012 were found.

A wide range of studies are aimed to detect the factors affecting bank efficiency or to determine the causality between bank efficiency and its drivers. The most frequently investigated issues are regarded to the relationship between bank size and efficiency [2][3][4][5], ownership structure as a factor contributing to the efficiency scores [6][7][8][9][10][11] and the relationship between efficiency and banking structural financial indicators [12][13][14][15][16][17][18][19][20].

The goal of the present research is to investigate the relationship between banking market competition and efficiency. The authors' stated hypothesis is, as follows:

H1: There is a significant relationship between efficiency and competition in the banking sector of the analyzed countries.

Testing of the hypothesis was based on the sample data of banking sector of seven new member states (NMS) of the European Union (EU): Latvia (LV), Lithuania (LT), Estonia (EE), Bulgaria (BG), Malta (MT), Slovakia (SK) and Slovenia (SI). Data set covers the period of 2006-2012.

Data Envelopment Analysis (DEA) was employed in the present study to measure efficiency of individual banks in each particular country. Input-oriented DEA model was constructed based on the intermediation approach. Generalized efficiency scores for banking sectors were estimated with the *median* function.

Competition in the banking sector was measured, using Herfindahl-Hirschman Index (HHI) and the share of assets held by the top 5 banks (CR5).

The present paper contributes to the academic literature scope, extending the information on bank performance in the NMS of the EU, including the Baltic States.

2. BANK EFFICIENCY AND COMPETITION

Bank performance, the main contributing factor to the bank value, can be expressed "in terms of competition, concentration, efficiency, productivity and profitability." [15]. The wide range of closely related concepts causes the necessity to explore the interrelationship between the results of studies on bank performance measuring with application of different methods and ratios.

Hypotheses stated and tested in bank performance-related studies primarily are based on three main theoretical statements: 1) Quiet life hypothesis [20][41][43]; 2) Structure-Conduct-Performance hypothesis [27][42][31]; 3) Efficient Structure Hypothesis [41][42].

1) *Quiet life* hypothesis (QLH) developed by Hicks in 1935 posits that market power will reduce the pressure towards efficiency [21]. Banks with large market share tend to be less efficient, because focus their efforts mostly on risk reduction [22].

2) *Structure-Conduct-Performance* (SCP) hypothesis developed by Bain in 1956 [23] states that higher concentration in the banking market leads to lower competition and hence to higher bank profitability.

3) *Efficient Structure Hypothesis* (ESH) developed by Demsetz in 1973 [24] assumes that more efficient banks increase their market share by pushing less efficient competitors from the market, thus increasing their market shares.

Some examples of the recent studies in regards to the exploration of the link between efficiency and competition in the banking industry are presented in the Table 1.

Table 1.
Studies on relationship between competition and efficiency in the banking industry

Source	Research period	Region/ number of analyzed banks
Castellanos, Garza-García 2013 [25]	2002-2012	Mexico/ n.a.
Fungáčová *et al.* 2012 [26]	2002-2011	China/ 76
Andries, Capraru 2012 [12]	2001-2009	EU27/ 923
Ferreira 2012 [27]	1996-2008	EU27/2124
Tabak *et al.* 2011 [28]	2001-2008	Latin America/ 495
Guillen *et al.* 2014 [29]	1990-2007	Latin America/ 200
Fang *et al.* 2011 [9]	1998-2008	SEE countries/ 171
Abbasoglu *et al.* 2007 [30]	2001-2005	Turkey/ 47
Chortareas *et al.* 2010 [31]	1997-2005	Latin America/ 2500
Bikker, Bos 2008 [15]	1996-2006	30 OECD countries/ 2124
Casu, Girardone 2006 [32]	1997-2003	EU15/ n.a.
Weill 2003 [33]	1994-1999	Eu12/ 1746

The results of studies are controversial that can be explained with specifics of the analyzed region and difference in methods applied in data analysis. There is a range of studies providing support to direct positive relationship between efficiency and competition, i. e., higher competition results in higher efficiency of the banking sector [12][34][25]. Other researchers assert that there is no evidence of correlation between competition and efficiency [32][26][42].

Linking competition in the banking market with concentration level, both positive [28] and negative [41] impact of concentration on efficiency level was proved. Besides, research findings revealed the fact that "degree of concentration is not necessarily related to the degree of competition" [32]. Some researchers do not use concentration indices as proxies for measuring competition.

Thus, one of the crucially important questions for the study on exploring causal link between bank efficiency and competition in the banking sector is the choice of methods and underlying measures applied in the empirical analysis.

The most popular methods for measuring bank efficiency are stochastic frontier approach (SFA) [34][9][12][35][33][8] and Data Envelopment Analysis (DEA) [36][37][38][39][40].

As for measuring competition in the banking sector, commonly used methods and measures are, as follows:

1) Concentration ratios [30][29][41][27];

2) Herfindhal-Hirschman Index [31][28][41][42][27];

3) Lerner index of competition [9][12][20][43];

4) H-statistic developed by Panzar and Rosse [44][32][33][45];

5) Boone indicator [46][25].

The selection of the research methods and underlying measures is based on experts' subjective viewpoint and preferences. However, it should be emphasized that the analysis results differ sufficiently depending on the methods applied. Thus, the choice of model specification and ratios should be substantiated with regards to peculiarities of the analyzed business environment.

3. DATA AND METHODS

Research sample consists of 97 banks (as for 2012) operating in seven European countries. Data was extracted from BankScope database, covering the period of 2006-2012. Such countries, as Poland, Romania and Czech Republic, were excluded from the list, because their banking sector is substantially larger that banking sector in the analyzed countries. For instance, in 2012 the banking sector of Poland was represented by 640 domestic and foreign financial institutions, according to the data of the European Central Bank [47].

To measure competition within the banking sector, the authors used: 1) concentration ratio CR5 - the share of assets held by the top 5 banks, and 2) Herfindhal-Hirschman Index (HHI) – the sum of squared market shares of each bank representing the sector.

Statistics on CR5 in the banking sector of the analyzed countries is presented in the Figure 1 [48].
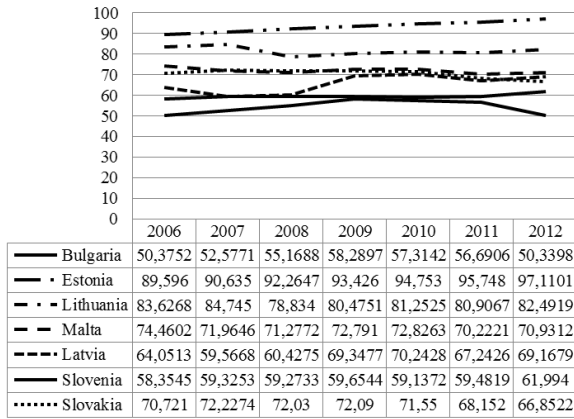
Figure 1. CR5 in the banking sector of NMS, 2006-2012

	2006	2007	2008	2009	2010	2011	2012
Bulgaria	50,3752	52,5771	55,1688	58,2897	57,3142	56,6906	50,3398
Estonia	89,596	90,635	92,2647	93,426	94,753	95,748	97,1101
Lithuania	83,6268	84,745	78,834	80,4751	81,2525	80,9067	82,4919
Malta	74,4602	71,9646	71,2772	72,791	72,8263	70,2221	70,9312
Latvia	64,0513	59,5668	60,4275	69,3477	70,2428	67,2426	69,1679
Slovenia	58,3545	59,3253	59,2733	59,6544	59,1372	59,4819	61,994
Slovakia	70,721	72,2274	72,03	72,09	71,55	68,152	66,8522

Based on the sample statistics, it can be concluded that process of bank capital consolidation led to relatively high concentration within the sector. Concentration ratios in all the countries exceed 50% in all the analyzed period. The most concentrated is Estonian market and the less concentrated is Bulgarian market.

Statistics on HHI in the banking sector of the analyzed countries is presented in the Figure 2 [48].

	2006	2007	2008	2009	2010	2011	2012
Estonia	0,3593	0,341	0,312	0,309	0,2929	0,2613	0,2493
Lithuania	0,1913	0,1827	0,1714	0,1693	0,1545	0,1871	0,1749
Latvia	0,1271	0,1158	0,1205	0,1181	0,1005	0,0929	0,1027
Bulgaria	0,0707	0,0833	0,0834	0,0846	0,0789	0,0766	0,0738
Malta	0,1171	0,1177	0,1236	0,125	0,1181	0,1203	0,1314
Slovenia	0,13	0,1282	0,1268	0,1256	0,116	0,1142	0,1115
Slovakia	0,1131	0,1082	0,1197	0,1273	0,1234	0,1268	0,1221

Figure 2. HHI in the banking sector of NMS, 2006-2012

As for HHI values, in most analyzed countries it does not exceed 2000 points (0,2). The maximum value is equal to 10000 points. The lower the index the closer is the market to monopoly. Estonia demonstrates the highest values of the index, but the decreasing trend indicates the growth of competition in the Estonian banking sector.

For measuring bank efficiency, non-parametric method – Data Envelopment Analysis (DEA) – was used. It was introduced in 1978 by Charnes *et al.* [49] and based on the concept of productive efficiency. The idea is to identify the most efficient companies and to construct the efficient production frontier. Measuring the distance from this frontier, it is possible to evaluate relative inefficiency of other companies within the reference set. Efficiency score is estimated as the ratio of weighted outputs to weighted inputs. To find the weights, optimization task is solved for each company in order to maximize its efficiency score (see formulas 1 and 2).

$$\max h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \tag{1}$$

subject to:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1; \tag{2}$$

Efficiency is estimated relatively to other reference set members. The maximal efficiency score is equal to 1, and the lower values indicate relative inefficiency of analyzed banks.

Initially, the model was applied under constant returns to scale (CRS) assumption followed by the model developed by Banker *et al.* in 1984 [50] that employed variable returns to scale (VRS) assumption. In the present paper input-oriented DEA model under VRS assumption was applied. The input-orientated models are the most frequently used for measuring bank efficiency with DEA [4][39][51]. It is based on the assumption that bank managers have higher control over inputs rather than outputs [52]. DEA model was constructed, based on intermediation approach that emphasizes bank intermediary function. Two variables were employed: bank deposits as inputs and loans as outputs.

4. RESULTS

To evaluate banking sector's efficiency, DEA efficiency scores of individual banks were estimated. The average results for each particular country are presented in the Figure 3.

	2006	2007	2008	2009	2010	2011	2012
Estonia	0,89246	1	0,97506	0,88216	0,33714	0,78531	0,81700
Lithuania	0,83660	0,72381	0,78367	0,68098	0,57668	0,69589	0,76693
Latvia	0,42079	0,47260	0,55526	0,22067	0,26546	0,17932	0,34557
Bulgaria	0,34404	0,67206	0,66200	0,63448	0,54175	0,77141	0,69860
Malta	0,46590	0,49686	0,64539	0,22317	0,09129	0,49654	0,08453
Slovenia	0,42788	0,55367	0,33149	0,33472	0,39937	0,63056	0,64315
Slovakia	0,73096	0,72689	0,17652	0,15730	0,14553	0,35258	0,67055

Figure 3. Efficiency scores of the banking sector of NMS, 2006-2012 [estimated by the authors]

The highest efficiency scores demonstrate Estonia and Lithuania. The similar results were yielded in the previously conducted researches [34]. The period of 2008-2010 is characterized by significant efficiency drop within the sample that was caused by the global financial downturn.

The estimated efficiency scores were used in the correlation analysis that was performed by means of SPSS software. Relationship between DEA scores of each particular country, CR5 and HHI was tested (see Table 2).

Table 2.
Results of the analysis: DEA efficiency vs. CR5 and HHI

	DEA vs. CR5		DEA vs. HHI	
	Pearson correlation	Sig	Pearson correlation	Sig.
EE	0,427	0,339	0,361	0,427
LT	0,527	0,224	0,722	0,067
LV	0,665	0,111	0,655	0,110
BG	0,277	0,547	0,452	0,309
MT	-0,377	0,405	-0,411	0,360
SI	-0,319	0,485	-0,593	0,160
SK	-0,839*	0,018	-0,717	0,069
*. Correlation is significant at the 0.05 level (2-tailed).				

Data provided in the Table 2 indicates the fact that there is no statistically significant correlation either between DEA score and CR5 (except Slovakia) or between DEA score and HHI. It should be emphasized that the values of the correlation coefficients are rather high in some cases (see Latvia and Lithuania), but not statistically significant. However, conducting the same analysis by means of Excel software, the results would point to the significant relationship (considering criterion of 0,5 for Pearson correlation coefficient). It means, in turn, that researchers should perform such kind of analysis with extreme caution, thinking not only on models and ratios, but also on software technical features.

Besides, the correlation coefficients are positive and negative in different cases. In this regard the question arises about the theoretical basis for this relationship.

Based on Quiet Life hypothesis, market power has a negative impact on efficiency due to the slack management. Thus, in highly concentrated market (as in the sample) with several top banks controlling the banking sector efficiency should be lower than in less concentrated market. Consequently, there should be negative correlation between CR5 and DEA scores.

In turn, Structure-Conduct-Performance hypothesis assumes positive relationship between market concentration and overall profitability (efficiency). Banks with a large market share "will face less competition to obtain more output results with less input costs" [27]. Thus, there should be positive correlation between CR5 and DEA scores.

Based on Efficient Structure Hypothesis, efficient companies determine market structure, increasing their market shares and, hence, increasing market concentration. Thus, there should be positive correlation between DEA scores and CR5.

As for HHI relationship with the efficiency scores, the same conclusions can be made from the theory. The higher is HHI (closer to 1 or 10000 points) the closer is the market to

monopoly, the more assets are concentrated in several largest banks.

The authors hold the view that positive relationship between efficiency and concentration is logically to be assumed. Banks with large market share compete against each other and do not spend their resources for competitive struggle with all other market players. Thus, the economy of resources positively influences efficiency.

To finalize the study, correlation analysis was performed using the whole data set (all three variables were included). The results are presented in the Table 3.

Table 3.
Bivariate correlation between DEA scores, CR5 and HHI

		DEA	CR5	HHI
DEA	Pearson Correlation	1	0,406**	0,514**
	Sig. (2-tailed)		0,004	0,000
CR5	Pearson Correlation	0,406**	1	0,904**
	Sig. (2-tailed)	0,004		0,000
HHI	Pearson Correlation	0,514**	0,904**	1
	Sig. (2-tailed)	0,000	0,000	
**. Correlation is significant at the 0.01 level (2-tailed).				

The results based on the whole sample data indicate the significant correlation (**) between DEA scores and both CR5 and HHI (at 99% confidence level). Correlation coefficients are positive in both cases that is aligned with the theory and logical premises. Increase of HHI or CR5 points to the decrease of competition in the market.

The research findings do not allow making an unambiguous conclusion. Results of the analysis of statistics on individual countries (see Table 2) are in the conflict with the results received from the analysis of the whole sample data (see Table 3). The significant contradiction can be explained with a difference in a volume of data set. As for 2012, whole sample included 97 banks, but data of only 8 Estonian and 8 Lithuanian banks was analyzed.

5. CONCLUSIONS

The present paper demonstrates the results of testing the authors' stated hypothesis on the relationship between efficiency and competition in the banking sector of NMS of the EU. Efficiency was measured by means of Data Envelopment Analysis – non-parametric method based on the efficient frontier approach. For measuring competition in the banking sector the authors used concentration ratio (market share of 5 top banks) and Herfindahl-Hirschman Index..

Highest average efficiency was demonstrated by Estonian and Lithuanian banking sector that is aligned with the results of previously conducted studies. Besides, Estonia and Lithuania have the most concentrated banking market. The largest banks in the Baltic States are Scandinavian owned banks Swedbank, SEB bank and DNB bank. Based on BankScope data, in 2012 Swedbank market share in Latvia, Lithuania and Estonia was 16%, 22% and 47% (!), respectively.

Correlation analysis based on data provided by BankScope in regards to individual countries did not revealed any statistically significant correlation between the analyzed measures. However, quite small number of banks in some countries

discredits the reliability of the received results. In turn, the analysis of the whole sample data provides the strong support to the stated hypothesis. Correlation coefficients between DEA scores and competition measures are significant. With the increase of HHI the situation in the market tends to monopoly, degree of competition decreases and, consequently, efficiency grows up. Similar conclusions can be made in regards to relationship between CR5 and DEA efficiency. In highly concentrated banking sector strong market players press out small banks and efficiency increases. Thus, the present study provided evidence on strong positive relationship between concentration and efficiency and, hence, a strong negative relationship between competition and efficiency in the banking sector of NMS.

However, it should be emphasized that the results could differ widely depending on methodological approach. In the present study simplified DEA model with only two variables was applied. Some researchers investigated the relationship between competition and cost and profit efficiency separately and received quite different results. The topic for further investigation could be testing the research hypothesis with application of different specification of DEA model. Besides, considering increased debate on the relationship between competition and concentration, it would be interesting to use other measures of competition in the analysis.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] S. Heffernan, **Modern Banking**, Chichester: John Wiley & Sons, Ltd., 2005.

[2] L. Drake, M. Hall, R. Simper, "The Impact of Macroeconomic and Regulatory Factors on Bank Efficiency: a Non-Parametric Analysis of Hong Kong's Banking System", **Journal of Banking and Finance,** Vol. 30, 2006, pp. 1443-1466.

[3] S. Ch. Karray, J. Chichti, "Bank Size and Efficiency in Developing Countries: Intermediation Approach versus Value Added Approach and Impact of Non-traditional Activities", **Asian Economic and Financial Review**, Vol. 3, No. 5, 2013, pp. 593-613.

[4] M. Zreika, N. Elkanj, "Banking Efficiency in Lebanon: An Empirical Investigation", **Journal of Social Sciences**, Vol. 7, No. 2, 2011, pp. 199-208.

[5] D. Pancurova, S. Lyocsa, "Determinants of Commercial Banks' Efficiency: Evidence from 11 CEE Countries", **Czech Journal of Economics and Finance,** Vol. 63, No. 2, 2013, pp. 152-179.

[6] S. Fries, A. Taci, **Banking Reform and Development in Transition Economies**, Working Paper No. 71, London: EBRD, 2002.

[7] J.P. Bonin, I. Hasan, P. Wachtel, Paul, "Bank Performance, Efficiency and Ownership in Transition Countries", **Journal of Banking & Finance**, Vol. 29, No. 1, 2005, pp. 31-53.

[8] A. Kasman, C. Yildirim, Cost and Profit Efficiencies in Transition Banking: The Case of New EU Members, **Applied Economics**, Vol. 38, 2006, pp. 1079-1090.

[9] Y. Fang, K. Marton, **Bank Efficiency in Transition Economies: Recent Evidence from South-Eastern Europe**, Research Discussion Papers 5/2011, Helsinki: Bank of Finland, 2011.

[10] V. Z. Toci, **Efficiency of Banks in South-East Europe: With Special Reference to Kosovo**, WP No 4, Prishtina: Central Bank of the Republic of Kosovo, 2009.

[11] G. Anayiotos, H. Toroyan, A. Vamvakidis, "The Efficiency of Emerging Europe's Banking Sector before and after the Recent Economic Crisis", **Financial Theory and Practice**, Vol. 34, No. 3, 2010, pp. 247-267.

[12] A. M. Andries, B. Capraru, "Competition and Efficiency in EU27 Banking Systems", **Baltic Journal of Economics**, Vol. 12, No.1, 2012, pp. 41-60.

[13] X. Vives, "Competition in the Changing World of Banking", **Oxford Review of Economic Policy**, Vol. 17, No. 4, 2001, pp. 535-547.

[14] J. A. Bikker, "Measuring Performance of Banks: An Assessment", **Journal of Applied Business and Economics**, Vol. 11, No. 4, 2010, pp. 141-159.

[15] J. A. Bikker, W. B. Bos, **Bank Performance: A Theoretical and Empirical Framework for the Analysis of Profitability, Competition and Efficiency,** New York: Routledge, 2008.

[16] L. Weill, "On the Relationship between Competition and Efficiency in the EU Banking Sector", Kredit und Kapital, Vol. 37, No. 3, 2004, pp. 329-352.

[17] L. Goldberg, A. Rai, "The Structure-Performance Relationship in European Banking", **Journal of Banking and Finance**, Vol. 20, 1996, pp. 745-771.

[18] A. Berger, T. H. Hannan, "Using Efficiency Measures to Distinguish among Alternative Explanations of the Structure-Performance Relationship in Banking" **Managerial Finance**, Vol. 1, 1997, pp. 6-31.

[19] L. Punt, M. Van Rooij, "The Profit-Structure Relationship and Mergers in the European Banking Industry: An Empirical Assessment", **Kredit und Kapital,** Vol. 36, 2009, pp. 1-29.

[20] J. Maudos, J. F. Guevara, "The Cost of Market Power in Banking: Social Welfare Loss vs. Cost Inefficiency", **Journal of Banking and Finance,** 2007, Vol. 31, 2007, pp. 2103 -2125.

[21] J. Hicks, "Annual Survey of Economic Theory: The Theory of Monopoly", **Econometrica**, Vol. 3, 1935, pp. 1-20.

[22] S. Rhoades, R. Rutz, "Market Power and □Firm risk - a Test of the "Quiet Life" Hypothesis", **Journal of Monetary Economics**, Vol. 9, 1982, pp. 73-85.

[23] J. Bain, **Barriers to New Competition,** Cambridge, Mass.: Harvard University Press, 1956.

[24] H. Demsetz, "Information and Efficiency: Another Viewpoint", **Journal of Law and Economics,** Vol. 10, 1973, pp. 1–22.

[25] S. G. Castellanos, J. G. Garza-García, **Competition and Efficiency in the Mexican Banking Sector**, Working Paper No. 13/29, Madrid: BBVA, 2013.

[26] Z. Fungáčová, P. Pessarossi, L. Weill, **Is bank competition detrimental to efficiency? Evidence from China**, BOFIT Discussion Papers No. 31. Helsinki: BOFIT, 2012.

[27] C. Ferreira, **Bank Market Concentration and Efficiency in the European Union: A Panel Granger Causality Approach**, WP 03/2012/DE/UECE. Lisbon: School of Economics and Management, 2012.

[28] B. M. Tabak, D. M. Fazio, D. O. Cajueiro, **Profit**, **Cost and Scale Efficiency for Latin American Banks: Concentration-Performance Relationship,** WPS 244, Brasilia: Banco Central do Brasil, 2011.

[29] J. Guillen, E. W. Rengifo, E. Ozsoz, "Relative Power and Efficiency as a Main Determinant of Banks' Profitability in Latin America", B**orsa Istanbul Review,** 2014, pp. 1-7.

[30] O. F. Abbasoğlu, A. F. Aysan, A. Gunes, **Concentration, Competition, Efficiency and Profitability of the Turkish Banking Sector in the Post-Crises Period.** MPRA Paper No. 5494. Istanbul: Boğaziçi University, 2007.

[31] G. E. Chortareas, J. G. Garza-Garcia, C. Girardone, **Banking Sector Performance in Some Latin American Countries: Market Power versus Efficiency**, WP 2010-20, Banco de Mexico, 2010.

[32] B. Casu, C. Girardone, "Bank Competition, Concentration and Efficiency in the Single European Market", **The Manchester School**, Vol. 74, No. 4, 2006, pp. 441-468.

[33] L. Weill, "Banking Efficiency in Transition Economies: The Role of Foreign Ownership", **Economics of Transition** Vol. 11, 2003, pp. 569-592.

[34] M. Kosak, P. Zajc, J. Zoric, "Bank Efficiency Differences in the New EU Member States", **Baltic Journal of Economics**, Vol. 9, No. 2, 2009, pp. 67-90.

[35] H. Yildirim, G. Philippatos, "Efficiency of Banks: Recent Evidence from the Transition Economies of Europe", **European Journal of Finance**, Vol. 13, 2003, pp. 123-143.

[36] R. Hogue, I. Rayan, "Data Envelopment Analysis of Banking Sector in Bangladesh", **Russian Journal of Agricultural and Socio-Economic Sciences**, Vol. 5, No. 5, 2012, pp. 17-22.

[37] K. Kosmidou, C. Zopounidis, "Measurement of Bank Performance in Greece", **South-Eastern Journal of Economics**, Vol. 1, 2008, pp. 79-95.

[38] K. S. Thagunna, S. Poudel, "Measuring Bank Performance of Nepali Banks: A Data Envelopment Analysis (DEA) Perspective", **International Journal of Economics and Financial Issues**, Vol. 3, No. 1, 2013, pp. 54-65.

[39] A. Nigmonov, "Bank Performance and Efficiency in Uzbekistan", **Eurasian Journal of Business and Economics**, Vol. 3, No. 5, 2010, pp. 1-25.

[40] E. Fiorentino, A. Karmann, M. Koetter, **The Cost Efficiency of German Banks: A Comparison of SFA and DEA**, Discussion Paper Series 2/ Banking and Financial Studies No 10/2006. Frankfurt am Main: Deutsche Bundesbank, 2006.

[41] H. Tetsushi, T. Yoshiro, U. Hirofumi, **Firm Growth and Efficiency in the Banking Industry: A New Test of the Efficient Structure Hypothesis,** RIETI Discussion Paper Series 12-E-060. Tokyo: RIETI, 2012.

[42] B. Rettab, H. Kashani, L. Obay, A. Rao, "Impact of Market Power and Efficiency on Performance of Banks in the Gulf Cooperation Council Countries", **International Research Journal of Finance and Economics**, Vol. 50, 2010, pp. 190-203.

[43] M. Koetter, O. Vins. **The Quiet Life Hypothesis in Banking: Evidence from German Savings Banks**, WPS No. 190, Frankfurt am Main: Fachbereich Wirtschaftswissenschaften Finance & accounting, 2008.

[44] J. C. Panzar, J. N. Rosse, „Testing for Monopoly Equilibrium", Journal of Industrial Economics, Vol. 35, 1987, pp. 443-456.

[45] J.A. Bikker, K. Haaf, "Competition, Concentration and Their Relationship: An empirical Analysis of the Banking Industry", **Journal of Banking and Finance**, Vol. 26, 2006, pp. 2191-2214.

[46] R. Griffith, J. Boone, Jan and R. Harrison, **Measuring Competition**, Research Paper No. 022, Advanced Institute of Management, 2005.

[47] European Central Bank. Consolidated Banking Data 2006-2012. Available from: http://sdw.ecb.europa.eu

[48] European Central Bank. Structural Financial Indicators, 2006-2012. Available from: sdw.ecb.europa.eu

[49] A. Charnes, W. W. Cooper, E. Rhodes, "Measuring the Efficiency of Decision-Making Units", **European Journal of Operational Research**, Vol. 2, 1978, pp. 429-444.

[50] R. Banker, A. Charnes, W. W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", **Management Science**, Vol. 30, 1984, pp. 1078-1092.

[51] J. Titko, D. Jureviciene, "DEA Application at Cross-Country Benchmarking: Latvian vs. Lithuanian Banking Sector", **Procedia – Social and Behavioral Sciences**, Vol. 110, 2014, pp. 1124-1135.

[52] M. D. Fethi, F. Pasiouras, "Assessing Bank Efficiency and Performance with Operational Research and Artificial Intelligence Techniques: A Survey", **European Journal of Operational Research**, Vol. 204, No. 2, 2010, pp. 189-198.

Message Dissemination Algorithm for Unreliable Broadcast Networks Guaranteeing Causal Order and Deadline Constraints

Jinho Ahn

Department of Computer Science, College of Natural Sciences
Kyonggi University
94-6 Iuidong, Yeongtonggu, Suwon Gyeonggi 443-760, Republic of Korea
e-mail: jhahn@kgu.ac.kr

Abstract—**Most of existing deadline constrained causal order broadcast algorithms force any group member to drop late messages received before the expiration of their deadlines, but not respecting causal order condition. However, their users want to see as many messages as possible in their cause-effect order within the earliest deadline among them. In this paper, we propose a highly efficient real-time constrained causal order broadcast algorithm to highly improve responsiveness and minimize the number of late messages discarded.**

Keywords-distributed system; realtime constraint; group communication; broadcast; message delivery order

I. INTRODUCTION

Causal order delivery to a broadcast group is a very important issue in the fields of sensor networks, video conferencing, stock trading, auction sales and so on [6, 7]. This message ordering condition can be satisfied if any two message sending events have cause-effect relation and the same destination, their corresponding delivery events should occur on the destination in their sending order. In order to ensure this ordering constraint, two approaches may generally be used as follows. First, if a group member receives a message capable of violating the constraint, the message delivery to the application is forced to wait for releasing the restriction caused by its predecessors [1, 4, 5]. Second, if deadline-constrained causal order requirement should be guaranteed, late messages, whose deadlines have passed or whose successors already received have exceeded their deadlines, are discarded. In the latter case, their users want to see as many messages as possible in their cause-effect order within the earliest deadline among them. However, the previous deadline-constrained causal order delivery algorithms [2, 3, 8] may not satisfy this important requirement. In this paper, we propose a highly efficient real-time constrained causal order broadcast algorithm to highly improve responsiveness and minimize the number of late messages discarded.

II. THE PROPOSED ALGORITHM

In figure 2, there is a broadcast group consisting of 4 processes, p1, p2, p3 and p4, sending 3 messages, m1, m2 and m3, to all members in order (by executing **Module** B-SEND(m)), whose deadlines are $deadline_{m1}$, $deadline_{m2}$ and $deadline_{m3}$ respectively. In the previous deadline constrained algorithms[2, 3, 8], p3 cannot receive m1 and m2 except for

delivering m3 in this example. In order to receive as many messages as possible before their earliest deadline like $deadline_{m3}$, our proposed algorithm allows each member like p1 and p2 to buffer received messages in its memory, $DLVD\_Q_{rcvr}$ (by executing **Module** B-RECV(m, $deadline_m$, $MVector_{sndr}$)). If a member, p3, receives a message like m3, from p2 in this figure, it requests m3's sender, p2, give m3's predecessors, m1 and m2, to itself by sending a solicitation message with m3's dependency vector, $MVector_{rcvr}$, (by executing **Module** SOLICIT-RECV($MVector_{rcvr}$)). After having obtained m1 and m2 from p2, p3 can deliver all three messages to their corresponding application (by executing **Module** RPY-RECV(*MSG_Q*)). In order to keep the deadline-constrained causal order requirement, our algorithm makes each member check deadline violation every time interval (by executing **Module** CHECK-MSGS()).

Module B-SEND(m) OF P_{sndr}
    $MVector_{sndr}[sndr] \leftarrow$ current time value of P_{sndr} ;
    broadcast (m, $deadline_m$, $MVector_{sndr}$) **to**
       all the other members ;

Module B-RECV(m, $deadline_m$, $MVector_{sndr}$) OF P_{rcvr}
    if(($deadline_m <$ current time value of P_{rcvr}) $\vee$
       ($MVector_{sndr}[sndr] \leq MVector_{rcvr}[sndr]$)) **then**
       discard message m **from** P_{rcvr} ;
    else if(($MVector_{sndr}[sndr] > MVector_{rcvr}[sndr]$) $\wedge$
       ($\forall i \neq sndr: MVector_{sndr}[i] \leq MVector_{rcvr}[i]$)) **then**
       $\forall i: MVector_{rcvr}[i] \leftarrow$
         $\max(MVector_{rcvr}[i], MVector_{sndr}[i])$;
       deliver m **to** its corresponding application ;
       insert (m, $deadline_m$, $MVector_{sndr}$) **into** $DLVD\_Q_{rcvr}$
         in m's sending time order *;*
       call CHECK-MSGS() ;
    else
       insert (m, $deadline_m$, $MVector_{sndr}$) **into** $RMSG\_Q_{rcvr}$
         in m's sending time order ;
       send solicitation($MVector_{rcvr}$, $MVector_{sndr}$) **to** P_{sndr} ;

// Every time interval, the procedure is executed.
Module CHECK-MSGS() OF PROCESS P_p
    for all e $\in RMSG\_Q_p$ in FIFO order **do**
       if($e.MVector_j[j] > MVector_p[j] \wedge \forall i \neq j: e.MVector_j[i] \leq$
       $MVector_p[i]$) **then**

$\forall$i:MVector$_p$[i] $\leftarrow$ **max**(MVector$_p$[i], e.MVector$_j$[i]) ;
deliver e.m **to** its corresponding application ;
insert e **into** DLVD_Q$_p$ in e.m's sending time order ;
remove e **from** RMSG_Q$_p$;
else if(e.deadline$_m$ = current time value of P$_p$) **then**
  for all c $\in$ RMSG_Q$_p$ in FIFO order st
  (c.MVector$_k$ $\leq$ e.MVector$_j$) **do**
    $\forall$i: MVector$_p$[i] $\leftarrow$
      max(MVector$_p$[i], c.MVector$_k$[i]) ;
  deliver c.m **to** its corresponding application ;
  insert c **into** DLVD_Q$_p$ in c.m's sending time
    order ;
  remove c **from** RMSG_Q$_p$;

Module SOLICIT-RECV(MVector$_{rcvr}$, MVector$_{upper}$) OF P$_{sndr}$
  MSG_Q $\leftarrow \Phi$;
  for all e $\in$ DLVD_Q$_{sndr}$ in FIFO order **st**
  (($\forall$i: e.MVector[i] < MVector$_{upper}$[i]) $\wedge$
  not($\forall$j: e.MVector[i] $\leq$ MVector$_{rcvr}$[i])) **do**
    insert e **into** MSG_Q in e.m's sending time order ;
  send reply(MSG_Q) **to** P$_{rcvr}$;

Module RPY-RECV(*MSG_Q*) OF P$_{rcvr}$
  for all e $\in$ *MSG_Q* in FIFO order **do**
    $\forall$i: MVector$_{rcvr}$[i] $\leftarrow$
      max(MVector$_{rcvr}$[i], e.MVector$_j$[i]) ;
    deliver e.m **to** its corresponding application ;
    insert e **into** DLVD_Q$_p$ in e.m's sending time order ;
    remove e **from** RMSG_Q$_p$;

Figure 1.   Procedures for each broadcast group member.

REFERENCES

[1] R Baldoni, "A Positive Acknowledgment Protocol for Causal Broadcasting", IEEE Transactions on Computers, vol. 47, no. 12, pp. 1341-1350, 1998.

[2] R. Baldoni, A. Mostefaoui and M. Raynal, "Causal Delivery of Messages with Real-Time Data in unreliable Networks", Real-Time Systems Journal, vol. 10, no. 3, pp. 245 -262, 1996.

[3] R. Baldoni, R. Prakash, M. Raynal and M. Singhal, "Efficient Δ-Causal Broadcasting", Journal of Computer Systems Science and Engineering, vol. 13, no. 5, pp. 263-270, 1998.

[4] K. Birman and T. Joseph, "Reliable Communication in the Presence of Failures", ACM Transactions on Computer Systems, vol. 5, no. 1, pp. 47-76, 1987.

[5] C. Kim and J. Ahn, "Causal order multicast protocol using minimal message history information", the 12th international conference on Algorithms and Architectures for Parallel Processing, vol. 1, pp. 546-559, 2012.

[6] J. Fanchon, K. Drira, S. P. Hernandez, "Abstract channels as connectors for software components in group communication services", Computer Science, ENC 2004. Proceedings of the Fifth Mexican International Conference in, pp. 88-95, 2004.

[7] C. Plesca, R. Grigoras, P. Queinnec, G. Padiou, and J. Fanchon, "A coordination-level middleware for supporting flexible consistency in CSCW", 14th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, 2006.

[8] L. Rodrigues, R. Baldoni, E. Anceaume, and M. Raynal, "Deadline-constrained causal order", Third IEEE International Symposium on Object-Oriented Real-Time Distributed Computing, pp. 234-241, 2000.

Figure 2.   An example of execution of our proposed algorithm supporting its high responsiveness.

Effect of Metrological Parameters Variation on the Radio Refractivity in Quebec-Canada

Yamina Bettouche, Basile L. Agba*, Ammar B. Kouki.

École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

* Institut de Recherche d'Hydro-Québec, 1800 Lionel-Boulet, Varennes, Québec, Canada J3X 1S1

Abstract – **In this paper, the influence of the variations of metrological parameters on the tropospheric radio refractivity during dry and wet month is analyzed. The used metrological data are provided by the "Environnement Canada" from four stations in Quebec for 2013. The daily variations of metrological parameters obtained from these stations are used to estimate the daily averages values of radio refractivity. Due to the difference in the variations of temperature, humidity and atmospheric pressure in the lower troposphere, the obtained values of radio refractivity during wet month were greater than the results in dry month. The analysis showed that, the stations located in land have higher difference wet to dry of radio refractivity, compared to the stations located coastal areas.**

Keywords—Troposphere; Radio refractivity; Meteorological parameter.

I. INTRODUCTION

The Troposphere is the lower part of the earth, which extends from the earth surface to an altitude of about 10 km at the earth poles and 17 km at the equator [1]. The state of the troposphere determined by atmospheric parameters such as temperature, pressure and relative humidity has a great impact on the propagation of electromagnetic waves in this layer of the atmosphere [2].

The changes of the atmospheric parameters result in the variation of the radio refractivity of the air in the troposphere from one point to the other. This variation of the radio refractivity is the source of the wave propagation's phenomena like ducting and scintillation [3].

The refractive properties of the troposphere are considered in the planning and the design of the terrestrial communication systems since they lead to multi-path fading and interference due to trans- horizon propagation [2].

In some seasons, the state of the troposphere constantly varies and affects the radio wave propagation. Various phenomena in the radio wave propagation such as ducting, scintillation, refraction and fading of electromagnetic waves are due to refractivity variations in the troposphere [2, 4].

The refractivity variations in the troposphere are characterized primarily by the state of pressure, temperature and relative humidity [5, 6].

The value of the radio refractivity index gradient which depends on the metrological parameters is used to estimate the quality of the radio wave propagation and its failure probability [5, 6].

In this paper, we estimate from meteorological parameters (temperature, humidity and atmospheric pressure) the variation of the radio refractivity in dry month as well as in wet month for four stations located in Quebec, (Canada) for 2013.

II. THEORY & METHODOLOGY

A. Radio Refractivity Theory

In normal atmospheric conditions near the Earth's surface, the radio refractive index is approximately 1.0003 [5, 6]. Since the value of the refractivity index is very close to unit, therefore the radio refractive index in the troposphere is often measured by a parameter called the radio refractivity N. This parameter is related to the radio refractive index, n by the following relationship:

$$N = (n-1) \times 10^6 \tag{1}$$

The radio refractivity is calculated using;

$$N = 77.6 \frac{P}{T} + 3.37 \times \frac{10^5 e}{T^2} = N_{dry} + N_{wet} \tag{2}$$

with the dry term, $N_{dry,}$ of radio refractivity given by:

$$N_{dry} = 77.6 \frac{P}{T} \tag{3}$$

and the wet, $N_{wet,}$ by:

$$N_{wet} = 3.73 \times 10^5 \frac{e}{T} \tag{4}$$

Where

P = Atmospheric pressure (hPa)

e = Water vapour pressure (hPa)

T = Absolute temperature (K)

The dry term contributes about 70 % to the total value of the refractivity but the wet term contributes to the major variation [8]. The equation 2 may be used for the propagation of radio frequencies up to 100 GHz [8]. The error associated with the application of the above formula is less than 0.5 % [9].

The water vapour pressure e, can be calculated from the relative humidity H and the saturation water vapour pressure e_s, given by:

$$e = e_s H \qquad (5)$$

With

$$e_S = 6.11 \exp\left[\frac{17.26(T - 273.16)}{T - 35.87}\right] \qquad (6)$$

Where:

H = Relative humidity (%)

e_s = Water vapour pressure (hPa)

B. Analysis Methodology

1) The precipitation is used as the meteorological parameter to determine the dry month and the wet month of the year for four stations (station 1, station 2, station 3) in Quebec province for 2013 (Table I).

2) The radio refractivity index is calculated based on the measured pressure, temperature and relative humidity collected at these stations provided by "Environnement Canada Agency".

3) The analysis was performed for all the months of 2013 to identify the dry and wet months. The results are summarized on Table II.

The analysis is performed based on the recorded data at the four stations. The geographical locations of these stations are shown in Fig.1 and in Table I.

Table II shows the driest and wettest month of each station for 2013.

TABLE I.   LOCATION OF METEOROLOGIQUE STATIONS

Stations	latitude	longitude	Altitude
1	48.42	-71.15	128.0
2	48.52	-72.27	178.6
3	48.84	-72.55	137.2
4	48.33	-71.00	159.0
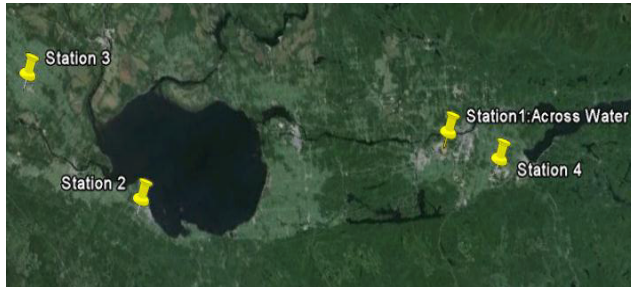
Fig. 1.   Geographical locations of the four stations

TABLE II.   DRY AND WET MONTHS

Stations	2013	
	Dry month	Wet month
1	Februry	August
2	Marsh	May
3	Marsh	September
4	Februry	August

III.   RESULTS AND DISCUSSION

The mean daily variations of N for each day are estimated from the data collected for dry and wet month in 2013, for the four stations. The obtained averages values are used to determine the radio refractivity for each day. Fig. 1 shows the comparison between the curve of daily variations of radio refractivity for dry month and the curve of daily variations of radio refractivity for wet month for station 1. Fig. 2, Fig.3 and Fig.4 show the comparison between the dry and the wet month for station 2, station 3 and station 4 respectively.
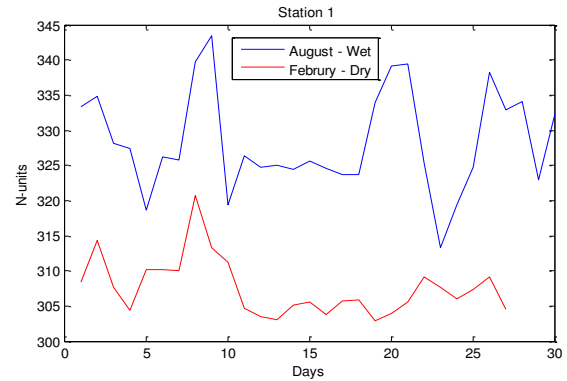
Fig. 2.   Daily variations of radio refractivity for both the dry and wet month.

For station 1 located across small lake, following remarks can be made:

- In wet month (August): the radio refractivity (N) is ranging from 313.3 to 343.5 and its average value ($Nwet_{mean}$) is 328.4.

- In dry month (Ferbruy): the radio refractivity (N) is ranging from 302.9 to 320.7 and its average value ($Ndry_{mean}$) is 307.6.

The difference between $Nwet_{mean}$ and $Ndry_{mean}$ ($Nwet_{mean}$-$Ndry_{mean}$) is 20.80

For station 2 located near water (about 4 km from water), following remarks can be made:

- In wet month (May): the radio refractivity (N) is ranging from 149.6 to 316.2 and its average value ($Nwet_{mean}$) is 204.6.

- In dry month (Marsh): the radio refractivity (N) is ranging from 96.75 to 183.4 and its average value ($Ndry_{mean}$) is 138.8.

The difference between $Nwet_{mean}$ and $Ndry_{mean}$ ($Nwet_{mean}$-$Ndry_{mean}$) is 65.80
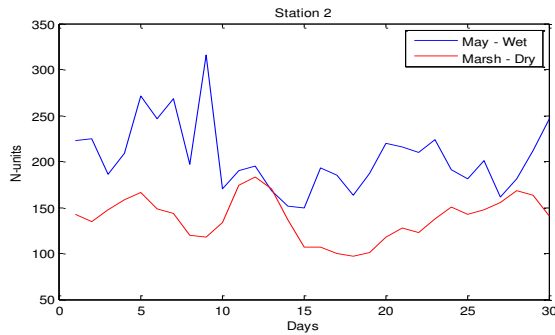
Fig. 3.   Daily variations of radio refractivity for both the dry and wet month.
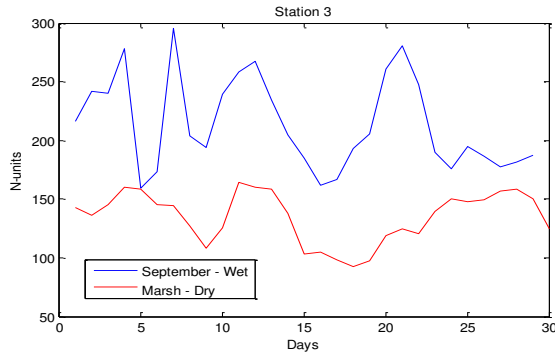
Fig. 4.   Daily variations of radio refractivity for both the dry and wet month.

For station 3 located in land (about 23 km from water), following remarks can be made:

- In wet month (September): the radio refractivity (N) is ranging from 159.3 to 295.4 and its average value ($Nwet_{mean}$) is 213.8

- In dry month (Marsh): the radio refractivity (N) is ranging from 92.34 to 134 and its   average value (Ndry_{mean}) is 135.1

The difference between $Nwet_{mean}$ and $Ndry_{mean}$ ($Nwet_{mean}$-$Ndry_{mean}$) is 78.7

For station 4 located in land (about 62 km from water), following remarks can be made:

- In wet month (August): the radio refractivity (N) is ranging from 201.4 to 420.1 and its average value ($Nwet_{mean}$) is 269.2.

- In dry month (Ferbruy): the radio refractivity (N) is ranging from 84.11 to 143.7 and its average value ($Ndry_{mean}$) is 112.

The difference between $Nwet_{mean}$ and $Ndry_{mean}$ ($Nwet_{mean}$-$Ndry_{mean}$) is 157.2.
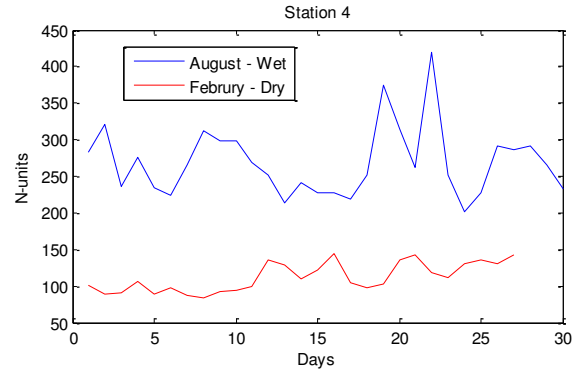
Fig. 5.   Daily variations of radio refractivity for both the dry and wet month.

We concluded that, dry to wet difference ($Nwet_{mean}$-$Ndry_{mean}$) of station located in land is highest compared to the two coastal stations and the across water station.

The difference decreases from its highest value (station 4) to the lowest (station 1) following the distance of the station to the water.

Some general conclusions can be made:

- The driest month occurs from January to April, and the wettest month is from May to September.

- The N_{dry} and N_{wet} curves have similar tendency. The curve of the radio refractivity in wet month is above the curves of the radio refractivity in dry month. This observation confirms the result in [8].

- The radio refractivity in wet month for stations located across water has higher N_{wet} value compared to the stations located in near water and in land.

The observations can be explained by some meteorological phenomena affecting coastal and in land regions. The water exists in the troposphere in three forms: liquid form (fog, clouds, rain), solid form (snow, hail, ice), or in vapour form. These three forms influence mainly the propagation of electromagnetic waves in the mentioned layer of the atmosphere. The rain and snowfall, as well the large amount of energy released in condensation process have a great influence on the value of the radio refractivity. In the wet month, the humidity is increased due to the presence the moisture of water vapour. This results in increase of the radio

refractivity. Contrary, in the dry month, there is a decrease in radio refractivity, observed during the period when there is sunshine (as shown on fig. 2 to 5).

IV. CONCLUSION

In this paper, the effect of the meteorological parameters variation on radio refractivity estimated from the data recorded at four stations in Quebec is analysed. The observations made in this paper confirmed that the variation of the radio refractivity in the troposphere is due to the changes of metrological parameters (temperature, humidity, atmospheric pressure and water vapour pressure). In general, the wave propagation is more affected during wet season than in dry season. In wet season, the humidity is high and consequently increases the value of the radio refractivity.

REFERENCES

[1]   References M. P.M Hall, "Effects Of The Troposphere On Radio Communication ," Numéro 8 de IEE electromagnetic waves series, Institution of Electrical Engineers. 206 pages, 1979-01-01.

[2]   Martin Grabner, Vaclav kvicera, "Refractive Index Measurement at TV Tower Prague," J. Radio Radioengineering, Vol. 12, No. 1, April 2003.

[3]   O. Agunlejika, and T.I. Raji, " Empirical Evaluation Of Wet -Term Of Refractivity In Nigeria," International Journal of Engineering and Applied Sciences (IJEAS), Vol.2, Issue 2, pp.63-68, 2010.

[4]   Famoriji J.Oluwole, "Variation of Metrological Parameters as They Affect the Tropospheric Radio Refractivity for Akure South-West Nigeria," World Academy of Science, Engineering and Technology, International Journal of Environmental, Earth Science and Engineering Vol.7 No.10, pp. 215- 217, 2013.

[5]   O. N. Okoro & G. A. Agbo, "The Effect of Variation of Meteorological Parameters on the Tropospheric Radio Refractivity for Minna, Global Journal of Science Frontier Research Physics & Space Science, Volume 12 Issue 2 Version 1.0, pp.37-41 February 2012.

[6]   Government of canada, historical climate: http://climat.meteo.gc.ca

[7]   A.T. Adediji, M.O. Ajewole,Falodun, S. E. and O.R. Olandosu, "Radio Refractivity Measurement at 150 m Altitude on TV Tower in Akure, South-West Nigeria," Journal of Engineering and Applied Sciences 2 (8),pp. 1308-1313, 2007.

[8]   IUT-Radio Communication Assembly: " The Radio Refractive Index: Its Formula and Refrsctivity Data, " Geneva, Switzerland, IUT-R P.453-9,2003.

Towards Energy-Efficient MAC Protocols

Tamer DAG
Computer Engineering Department, Kadir Has University
Istanbul, 34083, Turkey

and

Korhan CENGIZ
Electrical-Electronics Engineering Department, Trakya University
Edirne, 22030, Turkey

ABSTRACT

During the past two decades, the ICT (Information and Communication Technology) has become an integral part of our lives with exponentially increasing user interest and the vast number of applications being developed and used. The downside of the increase in these numbers is the continuous increase in the energy consumption. Various resources indicate that nowadays ICT is responsible for around 2% to 5% of the total global energy consumption and as new technologies are being developed, these numbers are very likely to increase in a proportional manner. In the past, networking protocols were developed for maximum efficiency, however without any energy constraints. Because of the analyses and expectations specified above, it has now become very important to develop new energy efficient networking protocols which will replace their corresponding ones. For all layers of the Internet protocol stack, studies have shown that significant energy savings can be achievable. This paper classifies the MAC protocols and describes some of the recent energy efficiency studies in the MAC sublayer. The reduction of energy consumption will not only have economical benefits but also environmental benefits that will have a positive impact for our lives in the future.

Keywords: Energy Efficient Protocols, MAC Sublayer, Green Communications

1. INTRODUCTION

ICT (Information and Communication Technology) has become an integral lifestyle commodity for most of the people around the world as it has played a central role in bringing people of the globe together. People and the companies have become dependent on the ICT for different kind of reasons such as social networks, e-commerce, online banking, home shopping, instant access to music, books and videos. With the wide usage of Internet of Things, Smart Cities and the invention of new technologies and applications, ICT's impact on human life will continue to increase rapidly.

It is estimated that ICT now uses approximately 1500 TWh of power per year which is equivalent to the combined power production of Germany and Japan. The power consumption of ICT is doubling every five years at an alarming rate. Thus, significant measures need to be taken to reduce the ICT power consumption. A significant number of studies are going on for each layer of the Internet protocol stack. While new energy saver applications are being developed, most of the energy hungry protocols are being modified or rewritten.

In this paper, the latest research on energy efficient MAC (Medium Access Control) protocols are described. The rest of the paper is organized as follows: Section 2 describes and classifies traditional MAC protocols into asynchronous MAC, syncronous MAC, frame slotted MAC and multi-channel MAC. Section 3 gives a summary of recent energy efficient MAC protocols with their properties and the paper concludes with the conclusions made in Section 4.

2. MAC PROTOCOLS

Medium Access Control data communication protocol is considered as a sublayer of the data link layer. A MAC protocol defines the rules of how a frame is transmitted on a communication link. When multiple nodes share a single link, the MAC protocol coordinates the channel access. This channel may provide unicast, multicast or broadcast communication services. In order to have high throughput in an energy-efficient way, designing an efficient MAC protocol is very important because MAC layer coordinates the access of the nodes to the shared wireless medium.

Traditionally, the MAC protocols are classified into four categories: synchronous MAC in which the neighboring nodes are synchronized to wake up at the same time, asynchronous MAC which aims to establish communication between two nodes which have different active/sleep schedules efficiently, frame-slotted MAC that is used for assigning time slots to nodes for receiving on behalf of transmission and multichannel MAC which provides parallel transmission by assigning nodes different channels.

2.1 Asynchronous MAC

In asynchronous MAC protocols, each node selects its active schedule autonomously without paying the price for synchronizing the neighbors' schedules. Asynchronous MAC protocols provide ultra-low duty cycle however they have to search efficient ways to constitute communication between two nodes.

Switching off lightly loaded devices is one of the extensively studied energy efficiency methods. Switching off network devices can result in significant energy savings provided that it doesn't have a significant impact on the network traffic. In preamble sampling (PS) a node turns on its radio when it has data to send. It aims to allow a node periodically wake up for a short time, the node goes back to sleep when the channel is idle.

However it brings overhearing problem because the node continues listening until the subsequent data frame is received or a timeout happens. This causes energy dissipation therefore this idea is not energy efficient. The PS is introduced together with the Mica wireless platform in [1]. Also, the performance of Aloha with PS and carrier sense multiple access (CSMA) are analyzed in [2,3]. To avoid collisions a transmitter performs clear channel assessment (CCA) in CSMA with PS. B-MAC [4] provides more accurate CCA and names its PS as low power listening (LPL). By the help of adaptive preamble sampling it reduces duty cycle and minimize idle listening thus obtain low power operation.

In PS, there is an unawareness of neighboring nodes' activities thus a node's preamble transmission may collide with the ongoing data transmission between neighboring nodes. For this reason STEM [5] uses two radios to distinguish the data transmission channel from wake-up channel.

To improve energy-efficiency the long preamble can be divided into a series of short packets which take some useful information. The non target nodes which have the destination address can instantly go back to sleep when they receive a short preamble packet thus more energy savings can be possible. Many protocols such as ENBMAC [6], MFP [7], B-MAC+ [8], SpeckMAC [9], DSP-MAC [10] and SyncWUF [11] use this idea for different problems. MH-MAC [12] and DPS-MAC [13] also contain timing information for broadcast messages, letting receivers to turn back to sleep and become active at the beginning of the data transmission.

The idea of shifting communication initiation from the sender side to the receiver side are presented in RICER [14] and PTIP [15]. Koala [16], AS-MAC [17], RI-MAC [18] and A-MAC [19] are also proposed for sensor networks. RI-MAC employs receiver created beacon packet to initiate data transmission instead of preamble packet. The beacon messages which are broadcasted by the nodes to announce that they are ready for receiving. RI-MAC unties collisions by reusing the beacon messages when a node broadcasts a beacon. According to [20], RI-MAC provides higher throughput, packet delivery ratio and power efficiency under a wide range of traffic loads compared with X-MAC [21]. However, only the receiver takes advantage of the receiver-initiated design. The sender has to stay awake until the data packets delivered. This is not energy-efficient thus PW-MAC [22] allows senders to predict target receiver's wake-up time. It also provides an efficient prediction-based retransmission scheme to obtain high energy efficiency even though wireless collisions happen and packets have to be retransmitted. The drawback of the PW-MAC is that the constant calculation of the neighbors' schedules brings necessary computational overhead which induces additional energy wastage.

2.2 Synchronous MAC

Synchronizing active time of neighboring nodes causes additional synchronization overhead. In prediction based synchronous MAC protocols only senders wake up at the target receiver's probing time while a cluster of nodes awake at the same time in synchronous MAC protocols. S-MAC [23] is specifically designed for Ad hoc wireless sensor networks and its primary goal is energy efficiency. In sleep period, nodes that are not involved in communication turn back to sleep. It also has collision and overhearing avoidance. However, it brings

latency and adaptive listening mechanism causes overhearing and idle listening brings inefficient battery usage. Instead of fixing the length of the active period, in T-MAC [24], the non sleep and sleep periods are fixed. It can easily manage variable load due to dynamic sleeping schedule. Nevertheless, there is an early sleep problem in T-MAC that nodes may sleep as per their activation time and data may get lost especially for long messages. Also, overhearing is still introduced because a node has to stay awake while it is not involved in data transmission thus this causes additional energy dissipation.

Shifting data transmission to the sleep period is discussed in RMAC [25] where instead of exchanging data, a control frame is forwarded by multiple hops and DW-MAC [26] which is same as RMAC but also introduces a one-to-one mapping function to procure collision-free data transmission in sleep period. Both RMAC and DW-MAC are not energy efficient because all downstream nodes are going to wake up pointlessly to receive the expected data packet that will not arrive due to the false alarm.

DMAC [27] is a proper approach for data gathering applications in wireless sensor networks where data are delivered from multiple sources to a sink. It gradates the active/sleep schedules of nodes thereby packets can flow continuously to toward the sink. Also Q-MAC [28] is very similar to DMAC. The energy efficiency of these protocols are questionable when the route length is difficult to determine.

2.3 Frame Slotted MAC

TDMA procure high throughput with maximized channel utilization under intensive contention. It is generally defined for global time synchronization. It has low channel utilization when slight number of nodes have data to send because a node can transmit only in its assigned time slot. To develop channel utilization of TDMA under low contention the one of the frame slotted MAC protocols Z-MAC which combines CSMA into TDMA is proposed in [29]. In every time slot a sender has to wait for a certain amount of time to ensure that the slot is abandoned by the owner. Also, each receiver has to be awake to control whether it is the target receiver. This brings additional energy consumption. TDMA-ASAP [30] is designed to improve the energy-efficiency of this idea. When a node abandons a slot, the data included nodes for the same parent can steal the slot, so other nodes do not need to wake up to check whether they have data to receive.

PMAC [31] reduces duty cycle by switching sending slots to receiving slots. It considers traffic load. A node is emboldened to increase its sleep time exponentially until the upper bound is reached if it does not have data to send. A node can quickly respond a new flow however it also increases energy consumption if only several packets will be delivered.

2.4 Multi-channel MAC Protocols

The radio bandwidth in WSNs is restricted and thus it is desirable to devise multichannel MAC protocols to conduct bursty traffic or provide multi-task support. By considering the energy efficiency and the cost, the channel allocation and cross channel communication problems can be achieved in multichannel MAC protocol designs.

For obtaining energy-efficient multichannel MAC schemes, authors in [32] aim to group nodes which communicate frequently into the same channel and separate nodes that communicate rarely into another channels. TMCP [33] similarly disintegrate a sensor network to several vertex-disjoint trees.

A TDMA based dynamic hybrid channel selection design Y-MAC is proposed in [34]. Time slots are assigned to nodes for receiving and less nodes wake up at each slot compared with scheduling senders thus it is energy-efficient but multiple senders in contention for sending. Also MuchMAC [35] is another hybrid design of TDMA and FDMA. It provides independent receiving channel selection for each slot for the nodes. It produces extra subslots, therefore the efficient way of assigning channel/slot to nodes is still an open issue.

3. RECENT ENERGY EFFICIENY STUDIES FOR MAC

Forcing the nodes to sleep adaptively while ensuring the continuation of communication is the purpose of [36]. The Base Station Controlled MAC (BSC-MAC) conserves energy by switching off the idle nodes. The source nodes generate the data and forward it to the base station through the root nodes. In BSC-MAC, the list of all nodes reaches the base station by flooding. The nodes with subsets are detected and they are designated as the root nodes. The seize of a path causes the neighbor nodes to go to sleep mode. This mechanism provides power savings, reduces packet collisions, inhibits packet losses and also prolongs the network lifetime. NS-2 simulator is used for simulations. BSC-MAC is compared with P-MAC, T-MAC and S-MAC. BSC-MAC performs better than S-MAC. It has worse results than P-MAC when the total number of generated packets increases. Also, BSC-MAC is compared with Adaptive AEEMAC which provides more energy efficiency than MAC by using additional improvements: adaptive sleep, reuse of the channel and combined control packets. BSC-MAC can result in energy savings compared with the AEEMAC.

[37] compares the performance of RI-MAC and S-MAC protocol for two different scenarios: contending flow and data gathering. In the first scenario, the RI-MAC outperforms S-MAC in terms of energy efficiency. For the second scenario, when the number of node increases average energy consumption is quite increased in RI-MAC even as in S-MAC it slightly decreases. Consequently, RI-MAC is more energy-efficient than S-MAC but it provides preferable packet reception ratio and throughput.

In [38], an Energy Efficient MAC (EE-MAC) protocol for distributed wireless sensor networks is presented. Protocol aims to reduce energy consumption and delay performance improvement. The goal is achieved by designating the optimal value of the sleep interval based on dominant conditions. A weighted linear combination of energy and delay is proposed for the objective function. To obtain the optimal value of the sleep times the objection function is minimized. In simulations, 700 nodes are scattered over a square area and the EE-MAC is compared with S-MAC in terms of energy consumption and delay. According to results, the energy saving of EE-MAC is little more than S-MAC for the small number of nodes, however when the number of nodes increase the energy saving of EE-MAC also increases accordingly.

Energy Efficient Sensor-MAC (ES-MAC) which focuses on optimizing energy efficiency of the practical applications of wireless sensor nodes is proposed in [39]. The protocol uses two schemes: Dynamic Data Cycle (DDC) and Selective Data Transmission (SDT) in other words it is substantially derived from S-MAC and T-MAC. ES-MAC aims to reduce the number of transmitted packets and also to let the mote to sleep for a short time when it is idle. However, the proposed protocol does not regard node discovery, multiple schedules of the nodes and network allocation vectors. For the simulations, TinyOS environment is used. A single sender in combination with a single receiver is considered. The number of transmitted packets and acknowledged packets are monitored. It is obtained that significant reduction in number of transmitted packets occurs when the DDC and SDT is used together. When only DDC is used the number of transmitted packets is increased. It is clearly shown that the proposed protocol provides significant energy savings and extends the battery life.

In [40], earliest deadline first (EDF) which is the dynamic priority algorithm based energy efficient earliest deadline first (EEDF-MAC) protocol is presented. The included scheduling algorithm increases the energy efficiency of the network and provides less latency. EEDF-MAC is proposed for WSN with both event-driven and clock driven nodes. Turning off the radios of nodes which are not involved in data transmission or data reception provides energy efficiency. Ns-3 is used for the simulation analyses of the EEDF-MAC. A single-hop network with an augmenting number of clock-driven nodes is designed. The nodes are separated into groups based on the period of data generation. The proposed protocol is compared with CSMA/CA which is widely used as a MAC protocol in WSNs in terms of latency, collision and energy efficiency. It is obtained that significant amount of energy consumption on listening can be saved by using EEDF-MAC. Also all the energy spent on channel sensing in CSMA/CA is saved by the use of EEDF-MAC.

T-MAC fulfills well in predefined stationary networks. Recently in some studies the performance of T-MAC is evaluated with mobile nodes. Authors in [41] present the performance of T-MAC protocol by considering node mobility. An energy-efficient solution is proposed to solve the mobility problem and it is called MT-MAC. The proposed protocol commences the scheduling process similar to T-MAC but it develops T-MAC by dividing the nodes into three different types: cluster-head, stationary-node and border-node. The improvements provide mobility detection and efficiently handling the movement of nodes between clusters. For simulations OMNET++ platform is used. The node mobility models are random walk and random waypoint. Average packet delivery ratio and average power consumption are the performance measurement parameters. The results show that MT-MAC provides slightly higher power consumption than T-MAC but on the other hand it outperforms T-MAC in terms of packet delivery ratio for all scenarios. As a result, MT-MAC does not show a marked improvement on energy-efficiency of T-MAC.

IEEE 802.15.4 standard defines the physical layer and MAC sublayer specifications. It defines 16 channels in the 2.4 GHz band. Especially it is proposed for low-rate wireless personal area networks (PANs). Low cost, low power consumption and low data rate are the key features of this standard. In [42], a novel wireless MAC mechanism which allows the IEEE 802.15.4 standard to procure an energy-efficient, reliable, and delay bounded data transmission for hybrid monitoring WSNs

is presented. The proposed protocol improves the IEEE 802.15.4 MAC standard by introducing a TDMA schedule to transmit neat long-term periodic monitoring (LTPM) traffic at the same time utilizing the standard CSMA/CA scheme to transmit seldom and randomly produced event detection (ED) traffic. The performance of proposed hybrid protocol is analyzed using ns-2 simulation based experiments. For all scenarios, a beacon enabled star topology network which includes several sensor nodes and a common coordinator is used. Standard IEEE 802.15.4 MAC is used for comparison. According to results, for all network sizes the proposed hybrid protocol provides less power consumption compared with the standard protocol. In addition, the energy stability and consuming constant amount of power unconcerned of the network size are also achieved by using proposed protocol.

4.   CONCLUSIONS

Power conservation has become a very crucial research area to support the enormous growth of the ICT industry. This paper describes recent research completed at the MAC sublayer of the Internet protocol stack which has addressed energy efficiency or in other words green communications. As the ICT sector growth rate is expected to accelerate with the invention of new technologies and applications, the traditional MAC protocols will not be able to satisfy the needs for green communications. However, the development of new MAC protocols is promising to reduce energy consumption as significant power savings can be achievable.

5.   REFERENCES

[1] J.L.Hilland, D.E.Culler, "Mica: A wireless platform for deeply embedded networks", **IEEE Micro,** Vol. 22, No. 6, 2002, pp. 12–24.

[2] A. El-Hoiydi, "Aloha with preamble sampling for sporadic traffic in adhoc wireless sensor networks," **in Proc. ICC**, 2002.

[3] El-Hoiydi, Amre, "Spatial TDMA and CSMA with preamble sampling for low power ad hoc wireless sensor networks", **7th International Symposium on Computers and Communications (ISCC),** 2002, pp. 685–692.

[4] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks", **2nd international conference on Embedded networked sensor systems ACM,** 2004.

[5] C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. Srivastava, "Optimizing sensor networks in the energy-latency-density design space", **IEEE Trans. Mobile Computing,** Vol. 1, No. 1, 2002, pp. 70–80.

[6] S. Lim, S. Kim, J. Cho, and S. An, "Medium access control with an energy-efficient algorithm for wireless sensor networks", **in Personal Wireless Communications, ser. Lecture Notes in Computer Science,** Vol. 4217, 2006, pp. 334–343.

[7] A. Bachir, D. Barthel, M. Heusse, and A. Duda, "Micro-frame preamble MAC for multihop wireless sensor networks", **IEEE International Conference on Communications (ICC),** Vol. 7, 2006, pp. 3365 –3370.

[8] M. Awenuti, P. Corsini, P. Masci, and A. Vecchio, "Increasing the efficiency of preamble sampling protocols for wireless sensor networks", **First Mobile Computing and Wireless Communication International Conference (MCWC),** 2006, pp. 117 –122.

[9] K.-J. Wong and D. K. Arvind, "SpeckMAC: low-power decentralized MAC protocols for low data rate transmissions in specknets", **2nd International Workshop on Multi-hop Ad Hoc Networks: from Theory to Reality,** 2006, pp. 71–78.

[10] S. Lim, Y. Ji, J. Cho, and S. An, "An ultra low power medium access control protocol with the divided preamble sampling", **Ubiquitous Computing Systems, ser. Lecture Notes in Computer Science,** Vol. 4239, 2006, pp. 210–224.

[11] X. Shi and G. Stromberg, "SyncWUF: An ultra low-power MAC protocol for wireless sensor networks", **IEEE Trans. Mobile Computing,** Vol. 6, No. 1, 2007, pp. 115 –125.

[12] L. Bernardo, R. Oliveira, M. Pereira, M. Macedo, and P. Pinto, "A wireless sensor MAC protocol for bursty data traffic", **IEEE 18th International Symposium on Personal, Indoor and Mobile RadioCommunications (PIMRC),** 2007, pp. 1–5.

[13] H. Wang, X. Zhang, F. Nait-Abdesselam, and A. A. Khokhar, "DPSMAC: An asynchronous MAC protocol for wireless sensor networks", **14th International Conference on High Performance Computing (HiPC),** 2007, pp. 393–404.

[14] E.-Y. Lin, J. Rabaey, and A. Wolisz, "Power-efficient rendez-vous schemes for dense wireless sensor networks", **IEEE International Conference on Communications (ICC),** Vol. 7, 2004, pp. 3769 – 3776.

[15] El-Hoiydi, Amre, and Jean-Dominique Decotignie, "Low power downlink MAC protocols for infrastructure wireless sensor networks", **Mobile Networks and Applications,** Vol. 10, 2005, pp. 675–690.

[16] R. Musaloiu-E, C.-J. Liang, and A. Terzis, "Koala: Ultra-low power data retrieval in wireless sensor networks", **7th international conference on Information processing in sensor networks. IEEE Computer Society,** 2008, pp. 421–432.

[17] B. Jang, J. B. Lim, and M. L. Sichitiu, "AS-MAC: An asynchronous scheduled MAC protocol for wireless sensor networks", **Mobile Ad Hoc and Sensor Systems, 25th IEEE International Conference,** 2008, pp. 434-441.

[18] Y. Sun, O. Gurewitz, and D. B. Johnson, "RI-MAC: A receiver-initiated asynchronous duty cycle MAC protocol for dynamic traffic loads in wireless sensor networks", **6th ACM conference on Embedded network sensor systems,** 2008, pp. 1–14.

[19] P. Dutta, S. Dawson-Haggerty, Y. Chen, C.-J. M. Liang, and A. Terzis, "Design and evaluation of a versatile and efficient receiver-initiated link layer for low-power wireless", **8th ACM Conference on Embedded Networked Sensor Systems,** 2010, pp. 1–14.

[20] Lakshmisudha, K., and C. Arun, "Research on power optimization in physical and MAC layer of wireless sensor networks-A survey", **IEEE International Conference on Intelligent Systems and Signal Processing (ISSP),** 2013, pp. 262-267.

[21] Buettner, Michael, et al. "X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks", **4th international conference on Embedded networked sensor systems,** 2006, pp. 307-320.

[22] Tang, Lei, et al. "PW-MAC: An energy-efficient predictive-wakeup MAC protocol for wireless sensor networks", **IEEE INFOCOM,** 2011, pp. 1305-1313.

[23] Ye, Wei, John Heidemann, and Deborah Estrin. "An energy-efficient MAC protocol for wireless sensor networks", **IEEE INFOCOM,** Vol. 3, 2002, pp. 1567-1576.

[24] Van Dam, Tijs, and Koen Langendoen. "An adaptive energy-efficient MAC protocol for wireless sensor networks", **1st international conference on Embedded networked sensor systems,** 2003, pp. 171-180.

[25] Du, Shu, Amit Kumar Saha, and David B. Johnson, "RMAC: A routing-enhanced duty-cycle MAC protocol for wireless sensor networks", **IEEE INFOCOM,** 2007, pp. 1478-1486.

[26] Sun, Yanjun, et al. "DW-MAC: a low latency, energy efficient demand-wakeup MAC protocol for wireless sensor networks", **9th international symposium on Mobile ad hoc networking and computing. ACM,** 2008, pp. 53-62.

[27] Lu, Gang, Bhaskar Krishnamachari, and Cauligi S. Raghavendra, "An adaptive energy-efficient and low-latency MAC for data gathering in wireless sensor networks", **18th International IEEE Parallel and Distributed Processing Symposium,** 2004, p. 224.

[28] Vasanthi, N. A., and S. Annadurai. "Energy efficient sleep schedule for achieving minimum latency in query based sensor networks", **IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing,** Vol. 2, 2006, pp. 214-219.

[29] Rhee, Injong, et al. "Z-MAC: a hybrid MAC for wireless sensor networks", **IEEE/ACM Transactions on Networking (TON),** 2008, pp. 511-524.

[30] Gobriel, Sameh, Daniel Mosse, and Robert Cleric, "TDMA-ASAP: Sensor network TDMA scheduling with adaptive slot-stealing and parallelism", **29th IEEE International Conference on Distributed Computing Systems,** 2009, pp. 458-465.

[31] Zheng, Tao, Sridhar Radhakrishnan, and Venkatesh Sarangan, "PMAC: an adaptive energy-efficient MAC protocol for wireless sensor networks", **19th IEEE International Parallel and Distributed Processing Symposium,** 2005, p. 8.

[32] Le, Hieu Khac, Dan Henriksson, and Tarek Abdelzaher, "A practical multi-channel media access control protocol for wireless sensor networks", **7th IEEE international conference on Information processing in sensor networks,** 2008, pp. 70-81.

[33] Wu, Yafeng, et al. "Realistic and efficient multi-channel communications in wireless sensor networks", **IEEE INFOCOM,** 2008.

[34] Kim, Youngmin, Hyojeong Shin, and Hojung Cha, "Y-mac: An energy-efficient multi-channel mac protocol for dense wireless sensor networks", **7th IEEE international conference on Information processing in sensor networks,** 2008, pp. 53-63.

[35] Borms, Joris, Kris Steenhaut, and Bart Lemmens, "Low-overhead dynamic multi-channel mac for wireless sensor networks", **Wireless Sensor Networks. Springer Berlin Heidelberg,** 2010, pp. 81-96.

[36] Toklu, Sinan, and O. Ayhan Erdem, "BSC-MAC: Energy efficiency in wireless sensor networks with base station control", **Computer Networks 59,** 2014, pp. 91-100.

[37] Simaiya, Deepak, Utkarsh Sharma, and Abhishek N. Tripathi, "Simulation and performance evaluation of energy efficient MAC protocols for wireless sensor networks", **10th IEEE International Conference on Wireless and Optical Communications Networks (WOCN),** 2013, pp. 1-5.

[38] Attiah, Afraa, et al. "EE-MAC: Energy efficient sensor MAC layer protocol", **38th IEEE Conference on Local Computer Networks Workshops (LCN Workshops),** 2013, pp. 116-119.

[39] Siddiqui, Shama, and S. Ghani, "ES-MAC: Energy Efficient Sensor-MAC protocol for Wireless Sensor networks", **10th IEEE International Conference on Networking, Sensing and Control (ICNSC),** 2013, pp. 28-33.

[40] Mohiuddin, Maaz M., I. Adithyan, and P. Rajalakshmi, "EEDF-MAC: An energy efficient MAC protocol for wireless sensor networks", **IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI),** 2013, pp. 1323-1329.

[41] Zareei, M. ; Islam, A.K.M.M. ; Baharun, S. ; Komaki, S. "Energy-efficient and mobility-aware MAC protocol for wireless sensor networks ", **19th IEEE Asia-Pacific Conference on Communications,** 2013, pp. 110-114.

[42] Wijetunge, S. ; Gunawardana, U. ; Liyanapathirana, R. "IEEE 802.15.4 based hybrid MAC protocol for hybrid monitoring WSNs ", **3**8th IEEE Local Computer Networks (LCN),** 2013, pp. 707-710.

Implementation of a SDN (Software Defined Network)

Mithat Sinan ERGEN[1], Sarp KUTLAN[1], Tamer DAG[2], Hasan DAG[1]
Management Information Systems Department[1], Computer Engineering Department[2], Kadir Has University
Istanbul, 34083, Turkey

ABSTRACT

Improving resource efficiency, enhancing network security and achieving simpler network management have become the main goals for networking in the previous years. To accomplish these ideas, an efficient routing, traffic monitoring, access control and server load balancing systems need to be designed. However, these objectives make the optimization and management of networks rather difficult. In this paper, Software Defined Networks (SDN) an alternative way for creating an optimized network by taking into account the difficulties met today is introduced. Software Defined Networks provide the separation of control and data planes for the switches which allows programming for a customized control plane. With simplified network management through SDN, it can be possible to dynamically adjust the behavior of the network equipment independently from equipment manufacturers. New mechanisms with new potential benefits can easily be explored and used. Quality of Service and network security problems can be solved rapidly. In this paper, it is described how to build a virtualized SDN in order to show the benefits of SDN over a distributed network. Virtualization can provide deployment and delivery flexibility, cost savings and improved user experience.

Keywords: Software Defined Networks, Network Security, Quality of Service, Data Plane, Control Plane

1. INTRODUCTION

The basic idea of Software Defined Networking (SDN) [1,2] lies under the principle of separating network intelligence from its hardware. The Internet protocols that we are currently using have started to be designed in the 1970s, and although more than 40 years have passed, an important part of the initial design is still protected and in use. Nowadays, extremely unique and different applications for different fields are being developed and are being adopted by users at a great pace. Quality of Service (QoS) and network security concepts have become critical issues based on current trends and unfortunately Internet lacks to solve these issues in an efficient manner as they were not even defined in the original Internet design. Thus, Software Defined Networking has become an extremely important step to solve these questions as it will allow developing new applications that cannot be supported by the current Internet standards.

The concept of Software Defined Networking and its underlying mechanisms have emerged about 5-6 years ago at Stanford. It is currently used extensively for data centers. A short time ago, Google has transformed the backbone forming its internal computer network into Software Defined Network. Many experts in the field foresee that in the next 15 to 20 years, the network technology covering the entire planet will turn out to be software defined.

In mid-90s or the early years of Internet, Internet was used mainly for transferring files. Sending different sized photos, video files or e-mail from point A to point B was the main objective. And the main problem was the speed of the file transfers. As a result, increasing the bandwidth of the infrastructure had become one of the most important priorities in those days.

Designed and developed in the 2000s, YouTube, Skype, VoIP, digital surveillance and its applications, has brought the issue of Real-Time Communications (RTC) to the agenda. With the widespread usage of these applications not only the speed, but also other parameters such as delay and jitter (the variance of delay) have started to become extremely important. With the differentiation of the used applications, the Quality of Service concept has began to be studied and different applications have started to be assigned different priorities. Each application's request from the network was different from the others. For example, VoIP packets has higher priority compared to FTP packets because their sensitiveness to delay. The routers started to process different application packets based on their priorities.

Starting with 2010s, different devices started to get connected to the network, turning the computer networks into more and more complex structures. The number of mobile devices connected to the Internet has exceeded the number of non-mobile devices for the first time. Other than individuals, smart devices began to connect to the Internet. With the diversification of applications and devices, the concept of priorities also started to change in a dynamic way. The priorities do not need to be protocol or application specific. For example, it might be needed for one application to assign a higher priority to the SIP protocol than to the FTP protocol, while another application might require exactly the opposite. The Quality of Service concept has also become dynamic. It is very possible that the Quality of Service requirements for an application might vary during the lifetime of the application. We might need to assign priorities packet based, not application based.

At this stage, Software Defined Networks has become very important and significant. Because right now, with the current network infrastructure fulfilling requests in a dynamic manner not possible. In a traditional router, data and control planes cooperate and work together. For these routers, forwarding decisions can be made with the help of protocols embedded into the network infrastructure. Because of this reason, it is very difficult to perform dynamic changes. However, with Software Defined Networking, the network traffic will be shaped and modeled depending on the immediate needs of the network or the applications.

Software Defined Networks [3] make this idea possible through the separation of the control plane and the data plane. As a result, the decision-making process which determines the

routing of the packets in the network is physically separated from the devices which make the routing decision. The data plane contains the network equipment such as routers and switches which basically transfer packets from point A to point B. On the other hand, the control plane contains a set of management servers which decide how the network traffic is going to move instantaneously by communicating and managing various network equipment in the data plane. Through the control plane, for example, the SIP traffic can be assigned higher priority than the SMTP traffic, and after a few minutes, depending on the changing needs, SMTP traffic can be assigned a higher priority than the SIP traffic. All network traffic can be controlled through a control console. The benefits for such a mechanism will lead into cheaper and simpler network equipment, In addition, simplified network management, the ability to dynamically adjust the behavior of network equipment independent from equipment manufacturers might be possible. New mechanisms with potential benefits [4] can easily be explored and used. Quality of Service, network security problems can be solved rapidly.

This paper describes the Software Defined Networking and how to implement SDN through OpenFlow and Mininet through a demonstration. Section 2, briefly describes SDN, OpenFlow and Mininet. Section 3, presents a demonstration on how to implement a SDN and the paper ends with the conclusions made in Section 4.

2. SOFTWARE DEFINED NETWORKS (SDN)

Software Defined Network is a dynamic, manageable and adoptable architecture that decouples the network control and data planes, forwarding functions enabling the network control, moving the control plane to an application called a controller. SDN is directly programmable, agile, centrally manageable, programmatically configured. SDN habitat uses open application programmatic interfaces to support applications. It also provides a centralized network operating system to communicate with a network of switches. Figure 1 illustrates the structure of a Software Defined Network.
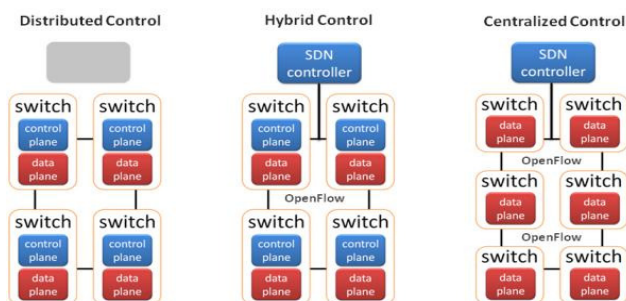
Figure 1: Structure of Software Defined Network

Why use OpenFlow ?

OpenFlow [5] is an architecture to run tentative and observational protocols in the network environment. There are two types of protocols we should indicate : A wire protocol and a configuration and management protocol. OpenFlow provides

an external controller to control a switch from the switch itself. OpenFlow provides an authorization to switches and controllers to communicate with each other. Decoupling of the control and data planes, using a protocol between networks and provide it programmability are the key components for OpenFlow. SDN supports the disintegration of data and control of packets and circuit networks. To create alternative routing and switching protocols in networks, we can benefit from OpenFlow.

Mininet

The software emulator which is used to produce a sample of a big network on a single computer is called mininet [6, 7]. A fast build of a rational virtual network is possible by utilizing mininet. Briefly, to make mininets virtual hosts switches, links and controllers, software is needed instead of hardware.

Creating a Simple Network with Two Hosts

In order to understand how Mininet works, it is necessary to start with a simple network which consists of two hosts connected to a single switch which is controlled by a reference controller. The reference controller is like a predefined controller which implements a bridge/learning switch like ovs-controller. It is used when a remote controller like POX is not created and allows the manage the communication in the network.

To create a network with 2 hosts using the command line interface, the command mn is used. If more details should be given, it is possible to specify them in the command line. For example mn –controller ref creates a virtual network with 2 hosts connected to a switch, controlled by a reference controller. This is the case in the first demo so after gaining administrator access to the system, we type in

"mn –controller ref –switch ovsk –topo single,2 --mac"

Figure 2: Creating a virtual network with two hosts

which creates a network with 2 hosts in a single topology, connected to an ovsk (openvSwitch-based) switch and sets the MAC address of the hosts equal to their IP as shown in Figure 2. The network points to a remote controller which is by default the localhost with IP address 127.0.0.1 on port 6633 as illustrated in Figure 3.
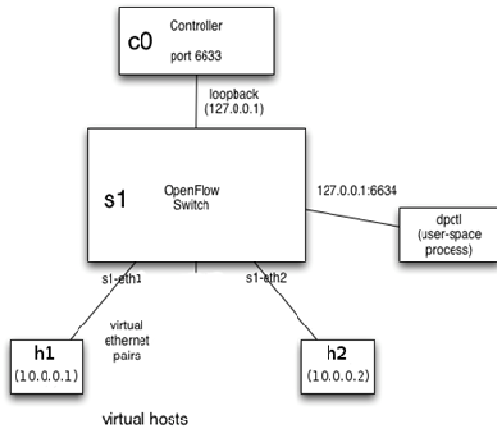
Figure 3: A virtual network with two hosts and a reference controller

3. IMPLEMENTING A SOFTWARE DEFINED NETWORK

Custom Topology with Multiple Controllers

For this demonstration, we used Python to create a custom topology which has seven hosts, three switches and two controllers. One of the controllers is a default controller which is running on port 6634. The other one is a remote controller which is defined in the class POXCont that we have programmed. We used MiniEdit to visualize our topology. It is a useful tool for visualization for networks as can be seen in Figure 4.
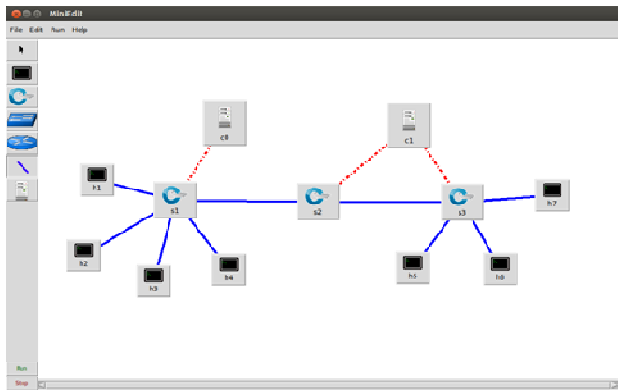
Figure 4: Custom topology visualization on MiniEdit.

In order to create this virtual network, a Python code is executed. In the code the hosts, their names, switches, controllers and their links are stated.

```
#!/usr/bin/python

from mininet.topo import Topo
from mininet.net import Mininet
from mininet.node import OVSSwitch, Controller, RemoteController
from mininet.log import setLogLevel
from mininet.cli import CLI

import os

class POXCont( Controller ):
```

```
    def start( self ):
            self.pox = '%s/pox/pox.py' % os.environ[ 'HOME' ]
        os.system( './pox/pox.py log.level forwarding.l2_learning
misc.pox_firewallson &' )
    def stop( self ):
            self.cmd( 'kill %' + self.pox )

controllers = { 'poxcont': POXCont }

class MyTopo( Topo ):

    def __init__( self, *args, **kwargs ):

        Topo.__init__( self, *args, **kwargs )

        h0 = self.addHost( 'h0' , mac='00:00:00:00:00:01')
        h1 = self.addHost( 'h1' , mac='00:00:00:00:00:02')
        h2 = self.addHost( 'h2' , mac='00:00:00:00:00:03')
        h3 = self.addHost( 'h3' , mac='00:00:00:00:00:04')
        h4 = self.addHost( 'h4' )
        h5 = self.addHost( 'h5' )
        h6 = self.addHost( 'h6' )
        leftSwitch = self.addSwitch( 's0' )
        rightSwitch = self.addSwitch( 's1' )
        midSwitch = self.addSwitch( 's2' )

        self.addLink( h0, leftSwitch )
        self.addLink( h1, leftSwitch )
        self.addLink( h2, leftSwitch )
        self.addLink( h3, leftSwitch )
        self.addLink( leftSwitch, midSwitch )
        self.addLink( midSwitch, rightSwitch )
        self.addLink( h4, rightSwitch )
        self.addLink( h5, rightSwitch )
        self.addLink( h6, rightSwitch )

topos = { 'mytopo': ( lambda: MyTopo() ) }

setLogLevel( 'info' )

c0 = Controller( 'Controller 1', port=6634 )

controller = POXCont ( 'Controller 2' )

controllerAssign = { 's0': controller, 's1': c0, 's2': c0 }

class MultiSwitch( OVSSwitch ):
    def start( self, controllers ):
        return OVSSwitch.start( self, [ controllerAssign[ self.name ] ] )

topo=MyTopo()
net = Mininet( topo=topo, switch=MultiSwitch, build=False
,controller=POXCont)
for c in [ c0,controller]:
    net.addController(c)
net.build()
net.start()
CLI( net )
net.stop()
```

First of all, we import all the necessary libraries from Mininet libraries. Topo class is the superclass which contains default topologies such as linear topology and tree topology and in this case it is used to create a custom topology. Mininet class is imported to create a virtual network using Mininet. OVSSwitch class is for creating an OVSSwitch. Controller and RemoteController classes are for defining controllers with given IP addresses on given ports. SetLogLevel class is for having logs of the work done in the code and CLI class and os class allow the execution of written strings in the command line interface.

It is possible to define a remote controller in the code and then start it from a separate terminal but the it can be started in the code as well. In the class POXCont we defined start and stop functions. Start function assigns the location of the pox file to the pox variable and "os.system" line types the command to the command line which we used to type by hand on a separate terminal in the other demos. After the Mininet is shut down, the stop functions is executed to kill the controller running on the reserved port.

After creating the remote controller's initiation class, we define our custom topology in the class MyTopo. As indicated, the network contains seven hosts and three switches. Hosts are created in the constructor by addHost function. AddHost function can get parameters such as name, IP address, MAC address. In this case we defined the MAC address of the first four hosts because we would like to specify communication rules for the first controller, which is the remote controller, to block traffic between h0 and h1 and allow traffic to anywhere else. After creating seven hosts, we defined three switches with the addSwitch function by giving their name as parameter. The connections between network elements are added with addLink function by giving the linked elements as parameters.

controllerAssign variable associates the switches to controllers and this variable is used in MultiSwitch class when the switches are started. The variable topo holds the properties of the custom topology we created and the network is created by using the Mininet class with the given topology, switch and controller definition. Build is given as false because it is done after the controllers are added manually in the for loop in the next line. After the controllers are added, the network is built and started. We execute the Python file on command line to create this network.

Between h0 and h1, the communication is not allowed because there is a firewall setting that is defined which doesn't allow communication between MAC addresses 00:00:00:00:00:01 and 00:00:00:00:00:02. h0 and h1 can communicate with any other host on the network.

Besides, if the communication time of the hosts is compared we can see that the hosts that are connected to the switch with the POX controller communicate with a higher average round trip time than the hosts connected on switches controlled by OVS controller.

4. CONCLUSIONS

SDN is considered as a new approach for designing, building and managing networks. By separating network's control and data planes from each other, optimizing and managing each plane can be handled in a much easier way. The most common protocol that SDN uses for communication between the controller and the switches is OpenFlow.

This paper describes SDN, the benefits of the separation of data and control planes and how one can implement SDN through a simple demonstration. It is shown how we can block communication between specific addresses, while those addresses can still communicate with other addresses.

Thus, with SDN organizations can rapidly deploy new applications that can answer their needs and changing business

goals and objectives. SDN can enable innovation with a centralized, programmable network. With SDN network engineers and administrators can respond quickly to changing requirements as traffic will be able to be shaped from a centralized control console as the switch rules could be easily changed whenever needed by prioritizing, de-prioritizing, blocking specific types of packets in a granular level of control.

5. REFERENCES

[1] Chung-Sheng Li, Wanjiun Liao, **Software Defined Networks**, IEEE Communications Magazine, February 2013, pp 113.

[2] Myung-Ki Shin, Ki-Hyuk Nam, Hyoung-Jun Kim, **Software-Defined Networking (SDN): A reference Architecture and Open APIs**, ICTC 2012, October 2012, Jeju, South Korea, pp 360-361.

[3] Flavio de Oliveira Silva, Joao Henrique de Souza Pereira, Pedro Frosi Rosa, Sergio Takeo Kofuji, **Enabling Future Internet Architecture Research and Experimentation by Using Software Defined Networking**, 2012 European Workshop on Software Defined Networking, October 2012, Darmstadt, Germany, pp 73-78.

[4] Sakir Sezer, Sandra Scott-Hayward, Pushpinder Kaur Chouhan et al., **Are We Ready for SDN? Implementation Challenges for Software-Defined Networks**, IEEE Communications Magazine, July 2013, pp 36-43.

[5] Thomas D. Nadeau and Ken Gray, **SDN - Software Defined Networks: An Authoritative Review of Network Programmability Technologies**, 2013.

[6] Mininet:
https://github.com/mininet/mininet/wiki/Introduction-to-Mininet#openflow

[7] Introduction to Mininet
https://github.com/mininet/mininet/wiki/Introduction-to-Mininet#openflow

Extension of Transportation Electronic Information System by Services for Visually Impaired Persons

Michal JERABEK, Jan KRCAL, Tomas ZELINKA
Faculty of Transportation Sciences, Czech Technical University in Prague
Konviktska 293/20, 110 00 Prague 1, Czech Republic
Email: jerabem1@fd.cvut.cz, krcaljan@fd.cvut.cz, zelinka@fd.cvut.cz

Abstract— Electronic information systems are an integral part of transportation arrangements in every city. They are built with the aim to offer to its constituents various services and to facilitate their orientation in a complex transportation network. These systems, however, do not necessarily take into consideration the needs of the handicapped, for example the visually impaired, although precisely this group needs correct and aptly transmitted information more than any other segment of the population. This article highlights the basic concept of the expansion of the current transportation information system in the Czech Republic by employing the Near Field Communication technology for the visually impaired passengers.

Keywords— Visually Impaired, Electronic Information System, Near Field Communication.

1. INTRODUCTION

Today, modern information systems in transportation are widespread and constitute an indispensible support of transportation processes for both the carriers and passengers. However, it is necessary to bear in mind already in the design phase that the transportation system should be designed so as to serve the highest possible number of users, not leaving any group of citizens behind. In case of people with limited movement and orientation capability, this is not an easy task, since each person faces special circumstances and has own special needs. What is suitable for one group may not fit another. When designing transportation information systems, it is therefore necessary to understand the requirements of the different groups and offer complex solutions suitable for all. The process should include continuous consultations with the representatives of the handicapped. When choosing such a consultation partner, it is necessary to exercise prudence, since "Diagnosis is not a qualification." (Viktor Dudr, Czech Blind United).

A whole array of adjustments and expansion of information systems was created for persons with visual impairments to ease their independent movement and reinforce their safety. The available solutions are far from perfect and there is room for improvement. This is the reason why the authors of this paper want to make a contribution in this area.

2. COMPENSATION DEVICE FOR VISUALLY IMPAIRED PERSONS

Compensation device is a tool, apparatus or mechanism specially designed or adjusted so that its use would at least partially compensate the limitations caused by the impairment.

A. White Cane

The white cane is a vital equipment of the visually impaired, since it enables independent orientation in space. According to their function, they can be categorized as signalization, protective, orientation, or support canes. Most of the white canes can combine up to three of these functions. No universal white cane that would offer all of these functions, however, exists, since some of these

functions, such as orientation and support, are mutually exclusive. White canes can also be further divided into folding, non-folding, telescopic, and combinational, according to their construction method. Each has its advantages and disadvantages.

B. Remote Controller

Acoustic orientation and information mechanism for the visually impaired on the territory of the Czech Republic are controlled by the VPN-type transmitters for the blind. There are six different encoded commands transmitted by the radio signal on the frequency 86.79 MHz. Two types of transmitters with identical functions are available:

- Type VPN 01 – pocket-size box with six buttons. The buttons are arranged in two columns with a tactile mark between button 1 and 2. The numbers of the buttons define the numbers of respective commands.

- Type VPN 03 – white cane component. The transmitter is built into the handle of the white cane (the orientation five-part compact type) with three buttons, which can, however, give six commands, such as the box type.

Fig. 1 VPN 01 and 03 transmitters

3. ACOUSTIC SOLUTIONS FOR VISUALLY IMPAIRED PERSONS

Hearing and touch are crucial senses for the visually impaired. Acoustic information can guide the blind and otherwise visually impaired to walk in the right direction or to safely cross the street.

A. Acoustic Signalization on Pedestrian Crossings

Acoustic signalization on pedestrian crossings indicates whether or not a street can be crossed. The emitted sound has a form of sound beats and their frequency allows to detect the instruction. The sound

beat with repeating frequency 1.5 Hz signals "red – stop" and repeating frequency 8 Hz (faster) signals "green – go."

Fig. 2 Acoustic Orientation Beacons

The SZN 01 device, which is so far most widely used in the Czech Republic, makes the characteristic hammer-like sound of an electromagnetic relay. The more modern electronic mechanisms utilize an interrupted sound at the frequency of approximately 400-500 Hz, if it is used at the pedestrian crossing. An analogue device, used where pedestrian crossings and railroads intersect, utilizes an interrupted sound at the frequency of approximately 800-1000 Hz. Thus, the sounds cannot be mixed. None of these devices, however, allows adjusting the volume of the acoustic signal. [2], [3]

B. Acoustic Orientation Beacons

Acoustic orientation beacon identifies a specific significant orientation point, such as:

- entrance to buildings, typically public buildings, such as administrative offices, shops, banks, post offices, health care centers, or social services providers;
- entrance onto escalators and moving pedestrian belts;
- underpasses, bus and train stations, or hospitals.

Fig. 3 Acoustic Orientation Beacons

Acoustic orientation beacons feature either a choice of a suitable trill, or a combination of the trill with a subsequent voice message and orientation tune. They can sound without an interruption in daily or weekly cycles or after activation by a command transmitter. The Czech Telecommunication Office has allocated for the entire territory of the Czech Republic a unique frequency of 86.79 MHz for the control by radio commands. They are typically placed in the height of 2.2-4 m and their sounds differ according to the purpose of their use.

Expansive underpasses, bus terminals or train stations, hospital complexes etc. can be equipped with a set of remote controlled acoustic beacons so that the blind or visually impaired person can be guided by their sound. Location of these beacons must be carefully chosen and their information message carefully formulated. It is important to take into consideration that beacons located close to one another would be activated by the transmitter simultaneously and thus disorient the user. [2], [3]

C. Command Systems in Public Transport Vehicles

Clearance and information system of the public transportation vehicles (trams, trolleys, buses) and some trains is typically supplemented by a command system. Once activated by the visually impaired, a voice message containing information about the connection (vehicle number and direction) is sounded. By pushing an additional button, the visually impaired can signal to the driver the intention to board. This device can also announce the current and upcoming stops by activating the respective command. [2], [3]

D. Electronic Information Systems with Output for the Blind

This category includes for example the so-called "intelligent stops," which reproduce, based on the signal from the transmitter, a voice message containing for example information about arrival of the next vehicle. Another type of such a device are information systems in terminal halls, such as airports, where the system can give information about arrivals and departures. The highest level constitute information electronic stands with voice output, which provide information about timetables, cultural events, accommodation, restaurants etc. [2], [3]

4. REQUIREMENTS ON THE DESIGN OF EXTENSION OF INFORMATION SYSTEM

A. Key parameters of Transportation Information System

1) Response Duration

The response duration is the ability of the information system to serve request for navigation to a certain specified maximum duration that can be defined as the probability

$$P((t_{R,i} - T_R) < \varepsilon_{RD}) > \gamma_{RD} \qquad (1)$$

that the difference between the measured duration of *i*-th response $t_{R,i}$ and the specified maximum duration T_R will not exceed the value ε_{RD} on the probability level γ_{RD}.

A research study conducted by the (now defunct) Institute for Rehabilitation of the Visually Impaired of the Charles University in Prague revealed that three seconds is the maximum response time of the system that the blind person can comfortably accept. A longer response time is less acceptable, because the blind person cannot scan the environment to understand the delay. An acoustic signalization indicating that the system is processing the response would be a fitting solution in such a case.

The total time of response is generally broken down into sub-periods:

- the period of generating the request (p_greq);
- the period of sending the request (p_sreq);
- the period of request processing (p_reqp);
- the period of generating the response (p_gres);
- the period of sending the response (p_sres);
- and the period of response processing (p_resp).

$$response\ duration = p_{greq} + p_{sreq}$$

$$+ p_{reqp} + p_{gres} + p_{sres} + p_{resp} \qquad (2)$$

2) Intelligibility

Intelligibility is the ability of an information system to provide transportation-related information in a comprehensible form, so that the passenger would not need to repeat his/her request for navigation for the final bus destination or connection. Intelligibility can be defined as probability

$$P\left(\left(\frac{1}{c_{req,i}}\right) \geq \varepsilon_c\right) > \gamma_c \qquad (3)$$

meaning that the converted value of the number of requests for navigation at *i*-th electronic information stand $c_{req,i}$ will not be lower than the requested value ε_c on the probability level γ_c. Since the electronic information stands are in a majority of cases located in open space, the sound can be drown out by other passengers or passing vehicles. However, NFC technologies allow for a very short distance between the passenger and electronic information stands.

B. System's parameters of passengers

From the viewpoint of system analysis every passenger can be characterized by the following parameters, which can be interconnected and they limited the passenger in the selection of paths:

- average speed of movement - it is determined not only by medical conditions but also by other limiting factors. The average walking speed is reported 5 kilometers per hour, but in the case of a person with a walking frame, crutches, mechanical wheelchairs or small child (and thus also his accompaniment) this speed is lower.
- spatial requirements - ordinary passenger needs about 500-700 millimeters for passage. The table below [4] gives the required width in millimeters for people with various disabilities.

person in a wheelchair	800-1200
person on crutches	900-1200
person with a walking frame	700-900
blind	900-1200
person with a stroller	700-900

- movement impairment - this parameter expresses the degree of the locomotor disability and therefore also the higher requirements for barrier-free routes in terms of the locomotor.
- visual impairment - this parameter expresses the degree of of visual impairment and therefore also the higher requirements for barrier-free path from the orientation point of view.
- hearing impairment - this parameter indicates the degree of hearing impairment and therefore also the higher requirements for barrier-free path from orientation point of view.

The disability of passenger is expressed by this scale:

1	full disability
2	partial disability
3	without disability

The passenger can thus be characterized by a vector **p**, whose components are the parameters mentioned.

$$\boldsymbol{p} = (p_1, p_2, p_3, p_4, p_5)$$

C. System's parameters of path

The path can be expressed as a graph that consists of individual nodes and edges. Each edge is characterized by the following parameters:

- length of the route - the total distance of the route, respectively edge.
- width of the route - if the width of a route is not the same everywhere, gives the value of narrowest part
- capacity of the route - on the route can accumulate passengers who are waiting for a means of transport. The lower capacity routes means less passage the passenger route.
- barrier-free route for people with movement impairment
- barrier-free route for people with visual impairment
- barrier-free route for people with hearing impairment

Barrier-free path is expressed by this scale:

1	accessible route
2	partially accessible route
3	route is inaccessible

The route can thus be characterized by a vector **r**, whose components are the parameters mentioned.

$$\boldsymbol{r} = (r_1, r_2, r_3, r_4, r_5, r_6)$$

5. EXTENSION OF TRANSPORTATION ELECTRONIC INFORMATION SYSTEM FOR VISUALLY IMPAIRED PERSONS

Acoustic orientation beacons are very useful for the visually impaired. They are activated by a command from the VPN-type transmitter and this command is transmitted by a radio signal – it is all-directional. Their main disadvantage is that the signal is not transmitted to the selected beacon only, but throughout the entire area that the signal covers (tens of meters in the open space), thus activating all the beacons in the area. The information from all the activated beacons can interfere with one another and instead of providing a guidance creates chaos. Visual sense has the ability to filter the unnecessary information, for example by closing the eyes, focusing on the desired information etc., but the sense of hearing does not provide for any such possibility. Another disadvantage of the acoustic orientation beacons is their ability to react only by pre-formulated reaction (trill, trill + voice message). Their reaction is therefore basically static and their adjustment for a new situation cause a delay in providing the right information. Information and communication technology is so advanced today that it is possible to remove these disadvantages. It would be a mistake not to promote the acoustic orientation beacons despite their limitations, since they require only low maintenance (they only require an electronic connection) and can prove to be very effective and practical.

At the Faculty of Transportation Sciences at the Czech Technical University in Prague, we are working on a proposal to expand the information system for bus terminals, which would help the visually impaired in their independent and safe movement in the bus terminal. The information electronic stands with the guidance trill and voice

message provide information about arrivals and departures of buses, but they do not provide any guidance on how to reach a particular boarding point. The key feature of the proposed expansion of the information system is a simple electronic information stand, proposed as a thin client, when the main logistics is programmed in the transportation information system of the bus terminal. This information stand has a NFC reader and voice input only. It provides two types of information:

- information about the required connection (arrival time or delay);
- information for independent and safe relocation to the required bus boarding point.

Should the path towards the boarding point be complicated and difficult to remember for the visually impaired person, the electronic stand would provide additional guidance to the next nearest stand in the desired direction. The path between the electronic stands reflects the need of the visually impaired persons and is dotted with tactile elements detectable with the white cane or distinguished by special surface stepping area. Since the bus terminals tend to vary in size and shape, they would have to be equipped with several electronic information stands, while each stand would need to have its clearly distinguishable identifier (S_ID), which would help to identify the position of the passenger who could thus receive the relevant information for further relocation to the desired bus arrival point.

For easy identification of the required connection, technology NFC is used. For the purposes of the passenger, it is sufficient if it is equipped with the NFC tag. For the information stand to provide the required information, the information system needs to know the connection which the passenger wants to take. This process of data upload into the central database of the transportation information system can be done in two ways at the minimum:

- when purchasing the ticket;
- through a web application.

Using one of these two methods, a unique identification number is read from the NFC tag and deposited by the chip producer, along with the data about the required connection into the database. The passenger is not identified and can own several NFC tags. Given the consent by the transportation providers (and the passengers), the ticket could be uploaded directly onto the NFC tag or a ticket medium equipped with the NFC tag.

The request sent from the electronic information stand contains not only the identification number for the NFC chip of the passenger, but also identification number of the stand in order to determine the position of the passenger in the bus terminal.

Where possible, the connection between the information electronic stand and the transportation information system is carried out through Ethernet, but the method of choice is more typically WiFi due to the existence of the infrastructure and minimalization of costs. The transfer of data is being carried out by the protocol TCP/IP. No sensitive data is being transferred and therefore the communication does not need to be encoded.

6. ALGORITHMS FOR DETERMINING THE NUMBER AND LOCATION OF ELECTRONIC INFORMATION STANDS

In the previous chapter has been introduced the use of electronic information stand on the example of the bus station. Its use, however, can be successfully used on railway stations or other public transport interchanges.

For the successful use of electronic information stands is important their location. This chapter describes algorithm which is used to determine their location and at the same time it is determined their number needed to cover the terminal, the interchange node. Application of the algorithm will be shown at the bus station Florenc in Prague, Czech Republic.

The algorithm is based on a gradual breadth first search. First, it is necessary to express bus terminal using a graph, where nodes are bus stations and crossing roads (on the example the distinction between empty and full circles) and edges represent the route between them.

The starting node from which the algorithm begins and where it is also located electronic information stand, a stand at the entrance to the terminal. For algorithm it is necessary to define this input parameters:

- maximum distance (maxDistance) of electronic information stand from the farthest station
- maximum complexity (maxComplexity) of routes - the complexity of the route is determined by the amount turn of the last electronic information stand.

The algorithm works with a vector **p**, that characterizes the passenger, in our case, the blind, and each edge with the vector **r**, that characterizes the route between two nodes. Both vectors were explained in Chapter 4.
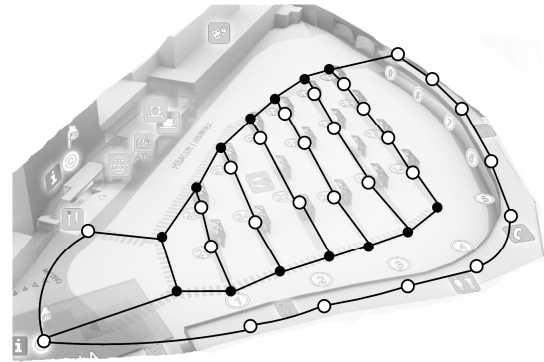
Fig. 4 Bus terminal Florenc, graph of trace
http://florenc.cz/images/mapy/stanoviste/map_stanoviste.jpg

```
procedure LocateStand(maxDistanceFromStand; maxComplexity)
numberStands :=1;
FOR every node n ∈ N
    LOOP
        d[n] := ∞;              -- distance from start node
        pred[n] := NULL;        -- predecessor of node n
        c[n] := ∞;              -- distance from stand
        complex[n] := 0;        -- complexity from stand
END LOOP;

FOR every node n ∈ N
    LOOP
        IF (r₅ = 1 AND r₂ >= 1200 THEN
            ENQUEUE(Q, n);
        END IF;
END LOOP;

IF doesn't exist SpanningTree(Q) THEN
    EXIT("It is impossible to find a way to all stations.");
END IF;

WHILE not EMPTY(Q) LOOP
    n := EXTRACT-MIN(Q);
    S := S ∪ {n};
    FOR every node v ∈ Successor[u] LOOP
        IF d[v] > d[n] + r₁ THEN
            d[v] := d[n] + r₁;
            pred[v] := u;
            c[u] := c[n] + r₁;
            complex[u] := complex[n] + 1
```

```
            END IF;
            IF d[n] + r₁ > maxDistanceFromStand  THEN
                numberStands := numberStands + 1;
                c[n] := 0;
                complex[n] := 0
            END IF;
        END LOOP;
    END LOOP;
```

7. CONCLUSION

The visually impaired persons do not want to live in seclusion from the world, but want to live a full and meaningful life to the extent their disability allows them. At minimum, they want to be self-reliant. When designing an information system (or other applications) it is therefore necessary to respect this and adjust the design to the specific needs and abilities of the visually impaired. The aim of this paper was to introduce a proposal for expansion of transportation information system of a bus terminal, which by utilizing modern information and communication technologies would afford the visually impaired an independent and safe movement. Technical proposal is supplemented by performance indicators, which help to determine if the solution is sufficiently fast and comprehensible for the passengers.

REFERENCES

[1] Vyhláška Ministerstva pro místní rozvoj ČR č. 398/2009 Sb. o obecných technických požadavcích zabezpečujících bezbariérové užívání staveb. Online [2013-07-10] <http://www.mmr.cz/getmedia/f015224c-ff91-4cad-a37b-dc0dc1072946/Vyhlaska-MMR-398_2009

[2] J. Matuška, Bezbariérová doprava, Pardubice: Institut Jana Pernera, 2009. ISBN 978-80-86530-62-8

[3] V. Dudr, Opatření na podporu samostatného a bezpečného pohybu pro zrakově znevýhodněné na ulici a v dopravě. Online [2013-06-26] http://www.okamzik.cz/view/okamzik/Kurzy_zp/Studijni_material y_osvetovy_pracovnik/V.Dudr_Opatreni_na_podporu_samostatneh o_a_bezpecneho_pohybu_ZP__na_ulici_a_v_doprave.pdf

[4] I. Šestáková, P. Lupač, Budovy bez bariér, Praha: Grada, 2010. ISBN 978-80-247-3225-1

[5] L. Krcalova, Influence of Age and Medicaments on Biometric Verification. In Recent Researches in Information Science & Applications. Athens: WSEAS Press, 2013, pp. 229-232. ISBN 978-1-61804-150-0

[6] Regulation (EC) No 552/2004 of the European Parliament and of the Council of 10 March 2004 on the interoperability of the European Air Traffic Management network

SCOUT: Efficient and Mobility Based Approach to Detect Clone Attacks in Wireless Sensor Networks

Mohammad Hasan Ansari
School of Electrical Engineering
Iran University of Science and Technology (IUST)
Tehran,Iran

Vahid Tabataba Vakili
School of Electrical Engineering
Iran University of Science and Technology (IUST)
Tehran,Iran

Abstract—**Regarding accelerating development and increasing utilization of mobile sensor nodes technology, this paper proposes a new approach for detecting the replica node attack in static wireless sensor networks by using mobile sensor nodes as node detector, which is distributed and independent of location information. In this approach, one general framework and two protocols for detection of mobile and static node are proposed. For efficiency and performance analysis of the proposed approach, security and cost are analyzed and the result of this new approach is compared with classical approaches based on only static nodes. Finally, for accurate performance evaluation of proposed approach to multiple networks configuration with actual protocol layers, vast variety of simulations results are discussed.**

Keywords—Scout nodes, Mobility,Wireless sensor network, Clone attack

I. INTRODUCTION

Utilizing WSNs in different environments, such as medical and military areas, because of the low cost, self-organized and no need to continues supervision of sensor nodes is increasing. Due to lack of physical shield layer on these nodes and utilize them in enemy environment without protection, usually these networks expose to different internal and external attacks [1]-[4]. The limitations of energy and memory sources in these sensor nodes, causes more security challenges in these networks in compare with to another mobile telecommunication networks. These complexities are significantly increasing, if the sensor nodes have mobility. Regarding the structure and architecture of WSNs among different attacks introduced in papers [1]- [7], because of the clone node attacks capability in passing through encrypting layer and authentication and also proper conditions for other attacks, these attacks are considered as the most serious security threat for the WSNs. In clone node attacks, first the attacker compromises the network sensor node and then by using side channel attacking techniques exploits the information on the node in a certain amount of time and finally uploads the information on any number of nodes. Now the attacker can make any type of attack on the network by taking the control of these nodes. Because after compromising the node, the attacker exploits the confidential information, including secret keys and uploads it on other clone nodes. From point of view of other nodes, the clone nodes seem to be valid. To avoid these types of attacks, new defensive solutions are required as a second security layer. Regarding the importance of this subject, many researchers have focused their efforts on it and a lot of procedures have been recommended for this attack. By analyzing various procedures, it can be observed that different criteria and variant methods and sometimes impractical hypothesis are considered, in which, simulation by real conditions such as the life of the battery, lower layer protocol, and real mobility models can clearly prove this point. Furthermore, choosing criteria, hypothesis, and as the result, proper detection procedure in the networks is essential. Regarding the most of proposed approach to detect replica node attacks in static sensor networks exposed high communication and memory overhead [2]. In this by using the idea of moving policeman in the real society scenario where he inspects the streets to find the bad person which is more efficient than checking of all the citizens and report their neighbors, a replica node attacks detection approach is proposed. The proposed approach assumes that network sensor nodes are static and otherwise mobile nodes are taken into account detector. Thus, mobile nodes with movement in the network and inspection; detect malicious node and broadcast revocation messages on the network. Access to location information of network nodes is a strict and costly assumption; therefor, detection in the proposed approach is performed independently of location information. Furthermore, proposed approach can be use distributed or centralized form in network with implementation of BS. In the following, first in Section II, the network and attacker models, capabilities and limitations of them is explained. In Section III, general framework of proposed approach and two protocols for detection of static and mobile sensor nodes is detailed. In Section IV, performance and security capability of proposed approach is evaluated, also communication and storage cost is calculated for it. In Section V, vast simulation of the proposed approach is done with using real protocol layers and the results are analyzed. Finally, in Section VI, conclusion of theoretical and simulation analysis is presented.

II. NETWORK AND ATTACKER MODEL

A. network model

The WSN usually contains hundreds and sometimes thousands of cheap and small size wireless sensor nodes, which
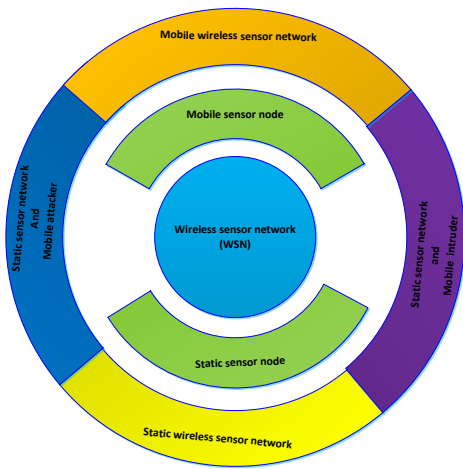
Fig. 1. Schematic diagram of all types of wireless sensor networks in terms of nodes mobility

are accidentally or in a pre-designed way distributed in a vast geographic area. In WSNs it is assumed that every moment there is the possibility that a number of sensor nodes get lost or be added to the network [1]. While in old networks, all the sensor nodes are static and have a base station for gathering data, today, respecting the robotics technology advancement and the emergence of mobile wireless sensor nodes, the structure of sensor networks has been changed. Different types of sensor networks with the combination of mobile and static are illustrated in Fig. 1. Therefore, here it is assumed that scout nodes are mobile and regarding this criterion, a hybrid network is considered. Also it is possible that, mobile nodes in compare with static nodes may have greater resources and capabilities [15]. In considered network, mobile nodes are considered as node detector and static nodes as normal nodes of WSN. Also, because of the severe security challenges in homogeneous network, general architecture of networks is considered homogeneous. Furthermore, possibility of base station in the network is considered.

B. attacker model

With referring to the definition of clone node attack [7]-[9], in this attack, the attacker compromises between one or some network nodes and exploits their stored information and replicates a preferred number of nodes from a certain node with specific identity and places them in appropriate locations in network, so that respecting the desired goals, it will be able to make different attacks including, eavesdropping, DoS, inject fake data and attacking the network protocols. Also it is assumed that the attacker is not able to allocate a new identity to the clone nodes and mostly the attacker can only compromise a small part of the network nodes. Otherwise, with referring to clone attacks cost, there is no need for node replica. In addition, it is assumed the attacking nodes can communicate and even collude with each other. It is also possible that the attacker nodes have more capability and flexibility compared to the valid nodes of the network which is a reasonable hypothesis. Finally, it is supposed that the attacker nodes can use various mobility models for moving in the network. But, due to generality and ideality of RWM model, in this paper it

is assumed that the attacker utilizes this model. Indeed, from a general point of view, the attacker node can be either mobile or static in the network.

C. Related work

After Parno et al. [7] pointed out the concept of replica attack, different detection methods were proposed, such as centralized detection, local detection, and distributed detection to deal with these kinds of attacks. In general, centralized methods such as SET [30] will bring out the problem of single point failure, and many communications are converged in the neighborhood of the central node, that cause sharp decline in energy BS neighboring nodes compared to other nodes in the network. Distributed approaches such as LSM [7], RED [8], RAWL [10] and SDC [9], which are used Birthday Paradox and witness nodes to detection replica, the communication and storage cost is high. Furthermore, it can be seen that most of the approaches proposed so far, are non-efficient against smart attacker [10]. Thus, it can be concluding that all of the approaches proposed based on static nodes have a lot of weakness. Regarding that, in [18] for the first time a mobile wireless sensor network node replica detection approach based on nodes encounter each other is proposed. Also in [15]-[16], several approaches based on define the maximum speed threshold for the mobile sensor nodes are presented. With respect to this social principle that the community policemen inspect the various regions and encounter with malicious people and with modeling this principle, a new idea for node replica detection is proposed in this paper. Also it is observed that, almost all previous node replica detection approaches for WSN are based on location information nodes in each moment [5]- [16] that it needs to spend high budget. Regarding resource and budget limitation in WSN, this assumption is not executable. Thus, we proposed an approach that is efficient, distributed and independent from location information based on mobile scout nodes.

III. PROPOSED APPROACH

Generally, if two or more nodes in WSN have the same identity, then, they will be regarded as compromised node or its replicas. We assume that WSN implementation is random and mobile node movement is modeled as RWM; then if a mobile node in a given time interval can meet one static node for several times that it exceeds the threshold, the mobile scout node is considered as compromised node or its replica.

A. Basic Scout approach

The mobile scout nodes move in the WSN and broadcast their claims to static nodes. The static nodes first authenticate validity of received claims and respond to it. Note that, if the node does not respond to the claims scout node, static node will be isolated from the network in the next round. Whereas, in all previous papers in this area it has been assumed always that any attack in the beginning stages of implementation cannot be done by an attacker [5]- [27], so this assumption is considered in this paper. Firstly, after implementation the WSN, each scout nodes tracking set are assigned randomly, and at least two mobile nodes are inspect each static node. Also, when for the first time, static nodes take the scout claim and respond to it, they consider scout nodes as anchor nodes. Then, static

nodes send list of its neighbors to the scout node. Scouts after receiving all claims check freshness and validity of received claims. If their message is a valid claim; then scout saves received claim with their ID in the table.

After the implementation phase, each detection round is divided into several intervals with same length. In order to node replica detection, scout node in each time slot moves randomly within a region and broadcast his claims. Static nodes take the scout claims and respond to them with using the approach framework. It can be observed from algorithm

---

Algorithm 1 Detection protocol framework.

1: Mobile nodes(scout) send claim$M \to n$:
   $C_M = \{ID_M \| T \| sig_M\}$
2: Static nodes executes the algorithm$2n \to M$:
   $C_n = \{ID_n \| neighbor - list \| T \| sig_n\}$
3: Mobile nodes(scout) executes the algorithm:
   $M \to all network nodes : Revoke the replica with ID_i$
4: Static nodes executes the algorithm3 and detect scout replica

---

(1), when scout nodes arrive into a new area, at first, they broadcast a claim that contains ID and time of claim to all nodes within their communication range. To avoiding high power communication attack, we can use a random seed or a waiting time to receive an ack from the mobile node. Indeed, we have $C_M = \{ID_M \| T \| sig_M\}$, and signature of the claims base on public key cryptography is generated as below:

$$sig_{M_j} = \{ID_M \| T\}_{K_s(M_j)} \tag{1}$$

Where, $K_s(M_j)$ is scout private key. Also, scout public key can scout ID node and it is known to all nodes in the network. Immediately, after reception of scout claim by the nodes in the network area, the first nodes check validity of claim sign, if the signature is valid. The node to avoid repeat attacks, check claim freshness using the following equation:

$$|T' - T| \geq \delta + \varepsilon \tag{2}$$

Where, T' is receiving claim time, δ is maximum delay in the communication range of the nodes and ε is maximum acceptable error rate. But, if the claim freshness is not approved, the node will broadcast the following error message in the network:

$$e_{n_i} = \{ID_{n_i} \| ID_{M_j} \| sig_{n_i} \| sig_{M_j}\} \tag{3}$$

Where sig_{n_i} and sig_{M_j} are static and mobile nodes claim signature, respectively. If time difference of received claim is acceptable and claim is fresh, static nodes will send their claims $C_{n_i} = \{ID_{n_i} \| neighbor - list \| T \| sig_{n_i}\}$ to scout M_j. The scout M_j, at first, check the validity and freshness of received claim , by using equations (1) and (2). Then, if claim is valid, it will be stored in scouts table with its ID. Scout node checks its table, after collecting the claim message, and if there is two or more same ID, it compares neighbors list with each other. Finally, if there is a conflict and some nodes with same ID are there, scout detects node replica and it save replica node ID in its blacklist and furthermore, scout broadcast revocation message. In next time interval, scout nodeM_j moves to other areas randomly and repeats this process.

B. Clone node detection

According to the network model, both static and mobile sensor nodes have been considered here. Therefore, an efficient node replica detection approach should be able to identify static and mobile attacker.

1) Static clone node detection: In static WSN, each node has a unique identity and it is immovable. For each node list of neighbors is fixed and subscription of these lists is non-empty. Using this principle, if scout nodes see some node with same ID and different neighbor list, the node is considered as a replica. Note that, if attacker implements replica nodes in an area, so that an scout collect and check their claims, then the scout can detect them after receiving the second claim, immediately. Otherwise, if different scouts collect claim of same replica, when they met each other, they exchange tracking set and detection is occurred. In general, static clone node detection by mobile scout nodes algorithm is as follows:

Algorithm 2 Algorithm 2: Static node replica detection

1: Receive static node claim message $C_{n_i} = \{ID_{n_i} \| neighbor - list \| T \| sig_{n_i}\}$
2: If receive more than on claim for a one ID ; go to stage5 and else go to 10
3: If ID_{n_i} is in blacklist; go to stage10
4: If ID_{n_i} isnt in blacklist; continue
5: check the authenticity of C_{n_i} ;if authenticity is true, continue; else go to stage10
6: check the freshness of C_{n_i} ;if freshness is true, continue; else go to stage10
7: search for ID_{n_i} in whitelist and if there are two ID_{n_i}
8: compare neighbor- list of two nodes
9: If different between two neighbor- lists, go to stage10
10: Revoke the ID and broadcast revoke message
11: End protocol

---

2) Mobile clone node detection: Sometimes, it is possible that attacker compromises a mobile scout node and replicates it in the WSN. In this scenario, assumption of static replica node detection cannot be used. Indeed, all mobile scout nodes are identified as a replica, because their neighbor list is changed permanently.

Fortunately, mobility of scout nodes and their use of random mobility model (RTM), provides a good criterion for detection of replica nodes. Regarding the network is implemented with uniform random distribution and random mobility models; it can be concluded that, in a normal state of the network for a specific area, number of meetings with static nodes is uniform and always lower than a certain threshold. Therefore, this criterion is used in approach to detect mobile replica nodes.

According to the selected criterion, mobile replica scout node detection approach is a two-stage approach. Firstly, in an offline stage, the meet threshold of static nodes with mobile scout nodes (λ) in time interval T_1 is calculated. Secondly, in an online stage, each static node for time interval T_1counters the number of meetings with certain scout and records it. If number of meetings between static and scout mobile nodes exceeds the threshold level, it considered as a mobile scout node and revocation messages are broadcast by the static

nodes in the network. Generally, mobile scout replica nodes algorithm is as follows:

Algorithm 3 Algorithm 3: Scout node replica detection

1: Receive scout node claim message $C_{M_j} = \{ID_{M_j}\|T\|sig_{M_j}\}$
2: If receive more than on claim for a one ID ; go to stage12 and else go to3
3: If ID_{M_j} is in blacklist; go to stage12
4: If ID_{M_j} isnt in blacklist; continue
5: check the authenticity of C_{M_j} ;if authenticity is true, continue; else go to stage12
6: check the freshness of C_{M_j} ;if freshness is true, continue; else go to stage12
7: add one to C_{M_j} in third column of table
8: compare$C_{M_j}$with threshold λ
9: If$C_{M_j} > \lambda$, go to stage12; else
10: check the interval time t; if $t_{timer} < T_1$ go to stage 12; else
11: if $t_{timer} > T_1$, then $C_{M_j} = 0$ and $t_{timer} = 0$, go to stage 12
12: Revoke the ID and broadcast revoke message
13: End protocol

IV. PROPOSED APPROACH PERFORMANCE ANALYSIS

A. Security analysis

An efficient replica node detection protocol should perform replica detection in a secure manner. To evaluate the security of proposed approach, we investigate the attacks and confronting strategies that are proposed.

In the classic mode, following attacks are significant: attacker node may try to defame an scout or static nodes with some forged claim. However,sig_{n_i} and sig_{M_j} digital signatures are used to counter with this attack for mobile and static nodes, respectively. Also, to prevent replay attack and to keep the fresh claim, the time is placed in the message and then it is signed. Therefore, malicious node cannot forge time T, because the time T is encrypted by the private key. Finally, message integrity is provided by using the SHA1 hash function. Thus, if mobile nodes try to revoke static nodes, they should forge the legitimate claim, that is impossible, because claim is protected with digital signature and integrity.

Moreover, the adversary cannot gain more benefit from collusion of malicious static and mobile nodes. In these conditions, the attacker will deploy many replicas in the area of a malicious scout. Regarding the malicious node cannot use a new ID to the replica nodes and the area will be inspected by another scout node in next round, however, the replica will not be revoked in a round, but they can be detected quickly, because they have high density nodes in an area and they be inspected by another legitimate scout nodes in next round. Furthermore, if several replicas of a single node are physically close to each other and respond to all requests with a same claim, they are able to escape from detection by detection approach. But this strategy substantially limits the region affected by the replicas and thus the attacker will not gain much benefit from using the replicas in the limited region. For example, in a false data injection attack, it would be easy to ensure that only one of the replicas data values at a time is accepted by the data

aggregators, that the existing mechanisms at the application layer can thwart it, effectively.

B. Performance analysis

To evaluate the performance of the proposed approach, network with n static nodes and m mobile nodes is considered and the deployment field divides into k claim areas. It also assumes that, nodes are distributed randomly and uniformly in the network and each round divides to t time interval.

1) Detection probability: For discuss about the detection probabilities, a node with r replicas $\{n_r^1, n_r^2, \ldots, n_r^r\}$ is considered. Each replica is inspected by scout nodes at an interval with probability of $\dfrac{m}{k}$, and there is $t.\dfrac{m}{k}$ chances to visit by mobile nodes. Thus, according to the standard derivation of the birthday paradox, the probabilityP_1that$t.\dfrac{m}{k}$ scout nodes inspect the area located byn_r^1 does not patrol thenn_r^2s area is given by:

$$P_1 = (1 - \frac{m.t}{k^2})^{\frac{m.t}{k}} \tag{4}$$

Similarly, the probability P_i that $(i.\frac{t.m}{k})$ scout nodes inspect the areas located by one replica of $\{n_r^1, n_r^2, \ldots, n_r^i\}$ does not inspect the n_r^{i+1}s zone is given by:

$$P_i = (1 - \frac{i.m.t}{k^2})^{\frac{m.t}{k}} \tag{5}$$

Thus, the probabilityP_{none} that neither two areas with any nodes in $\{n_r^1, n_r^2, \ldots, n_r^r\}$ are inspected by scout nodes is:

$$P_{none} = \prod_{i=1}^{r-1}(1 - \frac{i.m.t}{k^2})^{\frac{m.t}{k}} \leq \prod_{i=1}^{r-1} e^{-\frac{i.m^2.t^2}{k^3}} \tag{6}$$

$$= e^{-\sum_{i=1}^{r-1}\frac{i.m^2.t^2}{k^3}} = e^{-\frac{m^2.t^2.r.(r-1)}{2k^3}} \tag{7}$$

Thus, the detection probability is:

$$P_{detection} \geq 1 - e^{-\frac{m^2.t^2.r.(r-1)}{2k^3}} \tag{8}$$

2) Communication overhead: The proposed approach can be assessed in both centralized (with BS) and distributed scenario. If the network has a base station, then scout nodes send stored claims to BS, after each round and BS decides about replica nodes. If there is no base station, then the scouts should contact to exchange the detected information and scout node with smaller node ID make decision about replica nodes. For a better understanding of the issue, at first, the scenario that the network has a base station is considered. In centralized architecture, BS is considered as a trusted center and it can arrange the mobile nodes to scout the nodes. Thus, with each round division to $\lfloor \frac{k}{m} - 1 \rfloor + 1$ interval time, we can be sure that, at least each area is inspected by one node in each round. So, the nodes of k areas receive and respond claim at each interval. Therefore, total communication overhead is $(\lfloor \frac{k}{m} - 1 \rfloor + 1) \times (\frac{n}{k} \times m) \simeq n$ and communication overhead of each node is equal to $O(1)$. According to pervious centralized approach, we have communication overhead in order of $O(n\sqrt{n})$, thus it observed that, proposed approach has much smaller overhead than pervious approach[7]. Proposed

approach significantly reduces the communication overhead compared with the static node-based approach, because three-tier sensor network architecture including BS, static and mobile scout nodes is used. Also, proposed approach use routing based on mobility, that, it reduces the communication overhead, significantly. Otherwise, it can be seen that, the communication overhead of the proposed approach in compare with the hierarchical approach is also lower, because the communication overhead in hierarchical approach are in order of $O(m)$ that in compare with the proposed approach is higher.

In distributed scenario without a base station in the network, communication costs are contained local communication overhead in each area and communication overhead due to exchange information between two or more scouts.

Local detection of communication overhead, with assume uniform distribution of sensor nodes in network and m scout nodes with RTM mobility model is equal to $(\lfloor \frac{k}{m} - 1 \rfloor + 1) \times (\frac{n}{k} \times m) \simeq n$. In addition, the communication cost of nodes exchange of claims of tracking set between two scout nodes is equal to $O(1)$, because at first, each scout node listen to request, then check them and if there are common nodes, it responses them. Thus, communication overhead per node is equal to $O(1)$, even if piggybacking is used to broadcast messages, it is possible this cost tend towards zero. Therefore, the total communication cost in a distributed approach is $O(2n)$. It can be seen that, communication cost of the proposed approach is also much lower than the static nodes based on distributed approach [7]-[12].

3) Computational and memory overhead: In the proposed approach, the meeting threshold level between scout and static nodes calculation is time-consuming work, but this calculation is offline and it is performed before the running of approach. Thus, this cost will not be considered in computational overhead of detection approach. Generation and verification of digital signature cost is as another cost. In the proposed approach, ECDSA signature due to the computational speed is considered as signing claims, and signature generation and verification is performed with TinyEcc library [28], that it is very fast and takes about milliseconds, computation complexity is acceptable. Finally, proposed approach memory overhead is in order of $O(\frac{n}{k} + m) \simeq O(\frac{n}{k})$.

V. SIMULATION

A. The principles of simulation

1) Simulation environment : Regarding the special characteristics of the WSNs, many papers have focused on analysis and classification of simulation software's of these networks [33]-[36]. Regarding criteria like, development capability, and being free, by investigating the common network simulator packages [33]-[35], the network simulator software "OMNeT4.2.2" has been selected for the purpose of performing the clone node detection procedures in WSNs with mobile node and comparative study. In simulation, to gain the highest similarity of real results, the protocol layers are in accordance with both IEEE802.15.4 [44] and IEEE82.11 standards and are separately simulated. For radio specifications, a real model "CC2420" the real radios of the name by Texas Instruments is utilized and unit disk mode is considered as propagation model. Beside, in simulation, the energy, memory and calculation power limitations have been considered. It is also supposed that the extent of the area, in which the WSN is performed, is area.

2) Sensor nodes mobility model: Till now lots of different models have been proposed for the mobility of wireless sensor nodes [38]. RTM and RWM could be mentioned as the most famous ones that are used in the WSNs with mobile nodes. Therefore, to simulate the mobility of sensor nodes, two models of RWM and RTM are considered and the simulation is performed by both of them. But we believe that RWM is much perfect. In RWM the mobility of a sensor node are independent from each other and after each node reaches its location, it remains there for a randomly chosen time from $[T_{min}, T_{max}]$ and then it randomly chooses a location in the network, and with randomly chosen speed from $[V_{min}, V_{max}]$ moves towards the chosen destination. Of course, it must be noted that by using RWM the average speed of the network decreases during the life of the network. And if the least speed of the nodes is supposed to be zero, the average speed will converge at zero. To stop the occurrence of that, the least permitted speed should be regulated with more than zero. The details of RTM are similar to RWM. The only difference is the time of the presence of the mobile node in the destination which is the same for all the nodes and equals a fixed time. Then the minimum and the maximum permitted speed for the sensor nodes are supposed $1m/s$ and $20m/s$ respectively. Also the waiting interval for RTM model in the destination consider equals $20s$.

3) Cryptography mechanisms: In order to establish security against the cryptography common attacks as the first security layer, confidentiality and authentication mechanisms based on TinyECC [45] software package is utilized. Regarding the cost analysis carried out [46], in simulation, the hash function SHA-1 is used to provide the integrity of the nodes claims. Also digital signature algorithm ECDSA-160 is used for claims signing of each node. Which in accordance with [47], the cost of consumed energy used in ECDSA-160 for signing and confirming equals $22.83mj$ and $45.09mj$ respectively. Also the cost of calculating the checksum by SHA-1 is supposed to equal to $0.0059mj$.

B. Results of simulation

To compare the performance and capability replica node detection approaches in WSNs, criteria presented in Table (I) is used. In our simulations, we randomly deploy 1000 nodes within a $1000m1000m$ square. The transmission range is set to $25m$. So that, proposed approach is simulate and we compare its performance with the classical approach based on static nodes. Also, geographic routing (GPSR [31]), that is accepted in most of the papers [7]-[11] has been implemented to forward the claim in simulation of classic approaches. Vast simulations have been performed using the variable parameters, to investigate the impact of number of mobiles, time interval number in each round and size of areas. Note that, we assume all packets can successfully reach next hop and packet loss can be compensated by another protocols layer that use retransmission mechanism. We randomly generate a topology and execute the protocols for a random selected node 1000 times to get the average values, then the simulation results are calculated. The performance metrics used in the simulations is as follows:

TABLE I.   PERFORMANCE METRICS FOR SIMULATION

Performance Metrics	Description
Total consumed energy(E_d)	The sum of communication and computation costs consumed by all nodes to detect a clone
Clone detection probability(P_d)	The probability of successful detection
False alarm error (e_{FA})	Major detection error measured in mobile WSNs, which may deteriorate the quality of clone detection schemes due to costly false detection with regard to the high clone detection ratio

Fig. 2.   Detection probability versus different node number
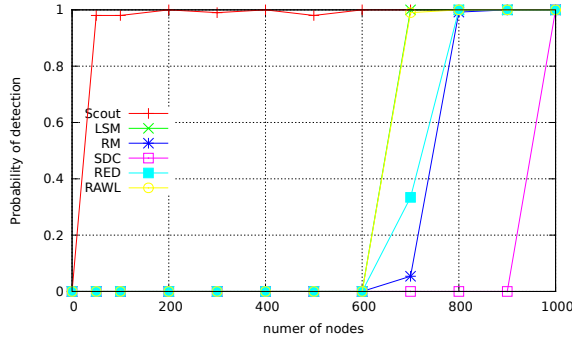
Fig. 3.   Detection probability versus different mobile scout nodes number

Detection probability(P_s): Similar to the simulation in [7] as the main reference for all papers in this field, this article focuses on the detection of a single node replica. Detection approach repeats for a given times (1000 in our simulation) and in start a replica is inserted into the network randomly and then start the detection protocol. Then, detection probability(P_s) is calculated $\frac{successful\ detection\ times}{repeat\ times}$.

Communication overhead: The average number of messages per node is considered as communication overhead.

Memory overhead: the average number of bytes that each node stores is considered as memory overhead.

False alarm: the average number of legitimate nodes is announced as a replica.

Fig. 4.   Detection probability versus different cell size

C. Detection Probability results

According to proposed approach, it can be seen that detection probability depends on the number of cells(k), the number of scout nodes(m) and the number of time intervals for each round(t). Also, the approach performance dependency to node number of network is analyzed. Fig.2 depicts the detection probability of the proposed approach in compare with other classical approaches, for different number of nodes in the network, while network size is same and number of scout nodes equals the witness node in classic approaches. It can be seen that, detection probability of proposed approach in a wide range of network is better than classic approaches. Because of moving of the mobile scout node in the network to everywhere that causes the better detection of the replicas. Also, it shows that, detection probability of classical approach based on static nodes in sparse WSNs due to failure of claims achievement to witness nodes is weak. Therefore, for tactical applications in serious environments, that node is distributed randomly by UAV or airplane. The performance of classical approach is unacceptable but proposed approachs performance is convenient. According to the power limitation in wireless sensor networks, it is obvious that with increasing the number of scouts and number of time intervals, energy consumption is
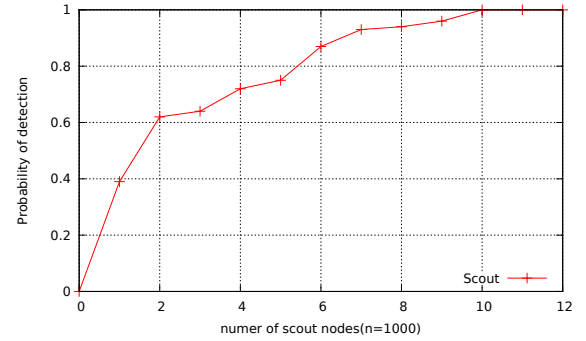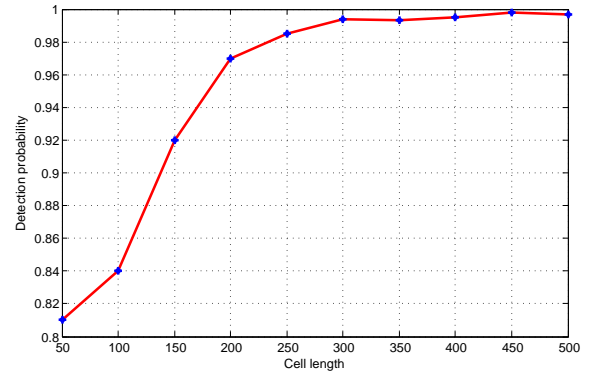
raised up. Therefore, evaluation of the impact of these factors on the probability of detection is important. Fig.3 illustrates the change of detection probability for a fixed number of time intervals and variant number of mobile scout nodes. From Fig.3 it can be known that, the detection probability is increased with increasing of the numbers of scout nodes, that, it is an obvious result. Also, it can be seen that, performance of proposed approach for low numbers of mobile scout is very good. Fig.4 shows the change of detection probability versus different cell size. It can be seen that, the probability of detection increases with increasing of the cell size, because scouts number in each cell increases and a replica node is inspected with more scout nodes. Fig.5 depicts variation of probability of detection for different time interval numbers in each round. It can be seen that, by increasing the number of time interval, probability of detection is increased. But, according to that the increasing number of time intervals the network, communication overhead increases, so optimal values of scout number, cell size and number of time interval for an efficient approach should be determined. In future work we will focus on it. Finally, regarding proposed approach has ability to detect mobile scout node replica, for given time simulation equal 2000 seconds, simulation is done for a mobile node replica. Fig.6 shows the change of detection probability versus different node numbers.
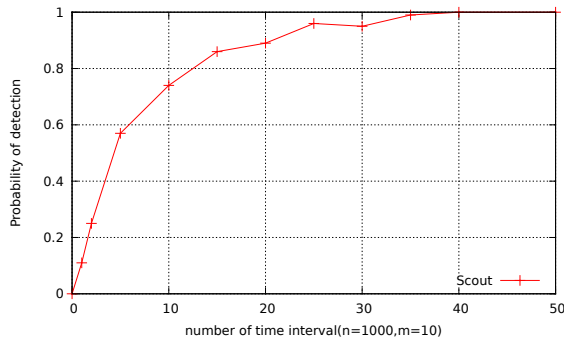
Fig. 5.   Detection probability versus different time interval number in each round
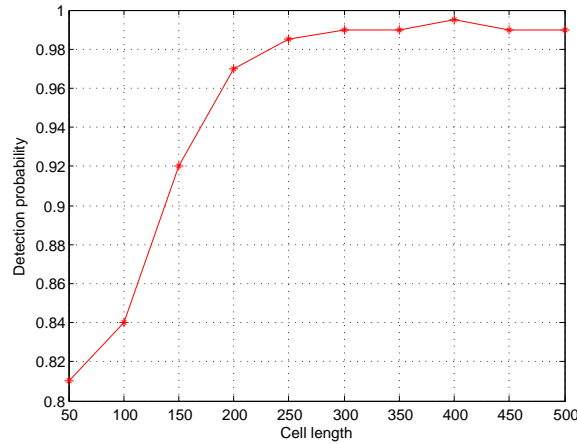
Fig. 6.   Detection probability of mobile nodes versus different scout node numbers (n=1000)

It is observed that the proposed approach always detects a mobile replica node with a probability higher than %95.

D. Communication overhead

To assess the performance of the proposed approach in compare with the classical approach, such as LSM [7], RED [8], RAWL [10], SDC [9] and RM [7], we simulate communication overhead of them with a profile (n=1000, m=G = 20) and %100 detection probability is simulated. Note that, in the proposed approach, number of mobile scout node to witness node in classical approach is considered equal. Fig.7
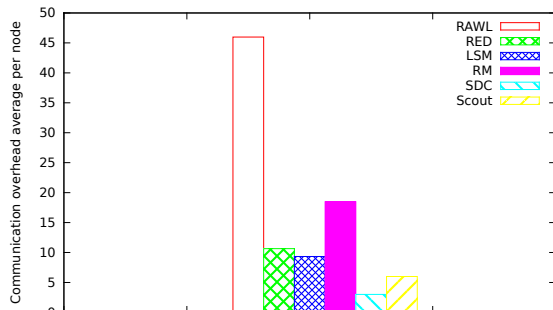
Fig. 7.   Request message average per node of different approaches
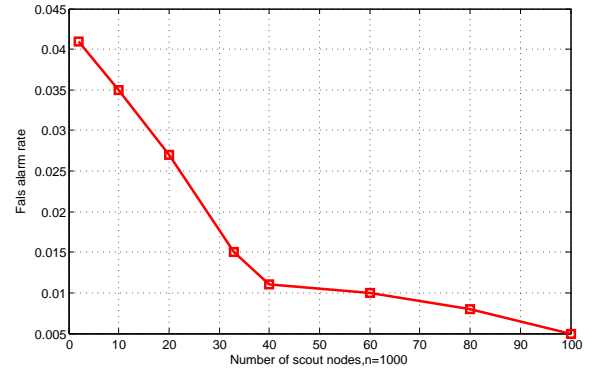
Fig. 8.   False alarm rate versus different scout node numbers (n=1000)

shows the communication overhead of different approaches for detection of a node replica with one probability. It can be seen that, the proposed approach has less communication overhead in compare with other classical approaches; except SDC. Regarding transmit claim and routing in proposed approach that is based on mobility, the results are rational. Indeed, in the proposed approach, number of claims that is transmitted in the WSN is lower than the classical approach. But we note that, the SDC approach against smart attacker is vulnerable [10], however its communication overhead is good, but it cannot establish the necessary security.

E. False alarm rate

In mobile-WSN, it can be obvious that, replica node detection approaches has false alarm, because of different factors such as memory and receiving error. Thus, according to proposed approach, detection is based on motion of scout nodes. This parameter has been examined with various network configurations. Fig.8 shows the probability of a legally node that is identified as a replica. It can be seen that, its probability is always less than%5.

VI. CONCLUSION

Replica node attack is one of the most serious security challenges in wireless sensor networks. Classical detection approaches to deal with these attacks are based on static nodes and they have some drawbacks. Indeed, due to the resource constraints of the sensor nodes, using of them in large-scale networks is very difficult and sometimes impossible. Therefore, regarding the progress made in mobile sensor nodes, in this paper, a new efficient approach based on mobile sensor node is presented. This approach is able to use in fully distributed or centralized network and mobile nodes have role of scout in the network. The important point here is that, while previous classical approaches are based on access to location information of node, this assumption is a very strict, but proposed approach is completely independent of location information. The proposed approach has better detection probability and less communication overhead than classic approaches. Also, memory overhead of the proposed approach with utilizing tracking set is significantly reduced and thus the memory overhead in compare with the approaches based on witness nodes is low. Furthermore, simulation results

demonstrate that the proposed approach has good detection rates and it is resistant to smart attackers.

REFERENCES

[1] Chen et al, Sensor Network Security: A Survey, IEEE Communications Survey & Tutorials. Vol 11, No2, Second, Quarter2009

[2] W.T. Zhu et al, Detecting node replication attacks in wireless sensor networks: A survey, Journal of Network and Computer Applications 35 (2012) 10221034

[3] Michael Riecker et al, A Survey on Intrusion Detection in Wireless Sensor Networks, Technical Report, SEEMOO-TR-2011

[4] V.Manjula et al, Replication attack mitigations for static and mobileWSN, International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.2, March 2011

[5] S.K.Das et al, A synopsis on node compromise detection in wireless sensor networks using sequential analysis, Computer Communications 34 (2011) 20032012

[6] Roberto Di Pietro et al, Securing Mobile Unattended WSNs against a Mobile Adversary, IEEE 2010

[7] Bryan Parno et al, Distributed Detection of Node Replication Attacks in Sensor Networks, IEEE 2005

[8] Mauro Conti, Distributed Detection of Clone Attacks in Wireless Sensor Networks, IEEE Transactions on Dependable and Secure Computing, VOL. 8, NO. 5, September/October 2011

[9] Bo Zhu et al, Localized Multicast: Efficient And Distributed Replica Detection in Large-Scale Sensor Networks, IEEE Transactions on Mobile Computing, Vol. 9, No. 7, July 2010

[10] Yingpei Zeng et al, Random-Walk Based Approach to Detect Clone Attacks in Wireless Sensor Networks, IEEE Journal on Selected Areas in Communications, Vol. 28, No. 5, June 2010

[11] Carlos Aguilar Melchor et al, Active Detection of Node Replication Attacks, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.2, February 2009

[12] Bo Zhu et al, Efficient Distributed Detection of Node Replication Attacks in Sensor Networks, 23rd Annual Computer Security Applications Conference,2007

[13] Chia-Mu Yu et al, CSI: Compressed Sensing-Based Clone Identification in Sensor Networks, 8th IEEE International Workshop on Sensor Networks and Systems for Pervasive Computing 2012, Lugarno (19 March 2012)

[14] Wen Tao Zhu, Node Replication Attacks in Wireless Sensor Networks: Bypassing the Neighbor-Based Detection Scheme, 2011 International Conference on Network Computing and Information Security

[15] Ho et al, Fast Detection of Mobile Replica Node Attacks in Wireless Sensor Networks Using Sequential Hypothesis Testing, IEEE Transactions on Mobile Computing, Vol. 10, No. 6, June 2011

[16] J.-W. Ho et al,Distributed detection of mobile malicious node attacks in wireless sensor networks, Ad Hoc Networks 10 (2012) 512523

[17] YanYxainaxniga and et al, Single Hop Detection of Node Clone Attacks in Mobile Wireless Sensor Networks, 2012 International Workshop on Information and Electronics Engineering (IWIEE)

[18] Chia-Mu Yu et al, Mobile Sensor Network Resilient against Node Replication Attacks, 2008 IEEE

[19] Mauro Conti et al,Emergent Properties: Detection of the Node-capture Attack in Mobile Wireless Sensor Networks, WiSec08, March 31April 2, 2008, Alexandria, Virginia, USA.

[20] J.-W. Ho et al.,Distributed detection of replica node attacks with group deployment knowledge in wireless sensor networks, Ad Hoc Networks 7 (2009) 14761488

[21] XiaoAmutihnogr et al, A Replication Detection scheme for Sensor Networks, rPinrogc,(2012) 21 26

[22] Tamara Bonaci et al, Distributed Clone Detection in Wireless Sensor Networks: An Optimization Approach,2011 IEEE

[23] Abu Saleh Md. Tayeen et al, Mobility Assisted Solutions for Well-known Attacks in Mobile Wireless Sensor Network, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 5, May 2011

[24] Tamara Bonaci et al, Node Capture Attacks in Wireless Sensor Networks: A System Theoretic Approach, IEEE2010

[25] Roberto Di Pietro et al,Intrusion-Resilience in Mobile Unattended WSNs, IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings

[26] Mauro Conti, The Quest for Mobility Models to Analyses Security in Mobile Ad Hoc Networks, WWIC 2009, LNCS 5546, pp. 8596, 2009.

[27] Patrick Tague et al,Modeling Node Capture Attacks in Wireless Sensor Networks, IEEE 2008

[28] A. Liu and P. Ning, TinyECC: A configurable library for elliptic curve cryptography in wireless sensor networks, in Proc. IPSN, Apr. 2008,pp. 245256.

[29] A. S. Wander, et al,Energy analysis of public-key cryptography for wireless sensor networks, in Proc. IEEE Int. Conf. Pervasive Comput. Commun., Mar. 2005, pp.324328.

[30] H. Choi, S. Zhu, and T. F. La Porta, SET: Detecting node clones in sensor networks, in Proc. Third Int. Conf. Security and Privacy in Communications Networks and the Workshops (SecureComm 07), Sept.2007, pp. 341350

[31] B. Karp and H. T. Kung, GPSR: greedy perimeter stateless routing for wireless networks, in Proc. ACM MobiCom, 2000, pp. 243254.

[32] Bo Zhu et al, Localized Multicast: Efficient And Distributed Replica Detection in Large-Scale Sensor Networks. *IEEE Transactions on Mobile Computing, Vol. 9, No. 7, July* 2010

[33] Muhammad Imran et al, A Survey of Simulators, Emulators and Testbeds for Wireless Sensor Networks.*IEEE*2010

[34] Ansgar Kellner et al, Simulation Environments for Wireless Sensor Networks.*Technical Reports of the Institute of Computer Science at the Georg-August-University at Gottingen*, June 2010

[35] Xuedong Xian et al, Comparison of OMNET++ and Other Simulator for WSN Simulation. *IEEE* 2008

[36] Klaus Wehrle et al,*Modeling and Tools for Network Simulation*, Springer-Verlag Berlin Heidelberg 2010

[37] Roberto Di Pietro et al, Intrusion-Resilience in Mobile Unattended WSNs. *IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM* 2010*proceedings*

[38] Mauro Conti, The Quest for Mobility Models to Analyses Security in Mobile Ad Hoc Networks. *WWIC 2009, LNCS 5546, pp. 85-96,* 2009

[39] Marjan Kuchaki Rafsanjani et al, Identifying Monitoring Nodes with Selection of Authorized Nodes in Mobile Ad Hoc Networks. *World Applied Sciences Journal 4* , 2008 444-449

[40] Patrick Tague et al, Modeling Node Capture Attacks in Wireless Sensor Networks. *IEEE* 2008

[41] T. C. Lam et al, A Mobile Agent Clone Detection System with Itinerary Privacy. *Proceedings of the Eleventh IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE?02)* 2002

[42] Yingpei Zeng et al, Random-Walk Based Approach to Detect Clone Attacks in Wireless Sensor Networks. *IEEE Journal on Selected Areas in Communications, Vol. 28, No. 5, June* 2010

[43] Chia-Mu Yu et al, CSI: Compressed Sensing-Based Clone Identification in Sensor Networks.*8th IEEE International Workshop on Sensor Networks and Systems for Pervasive Computing* 2012, Lugarno

[44] The IEEE 802.15.4 standard (ver. 2006) http://standards.ieee.org/getieee802/download/802.15.4-2006.pdf

[45] A. Liu and P. Ning, TinyECC: A configurable library for elliptic curve cryptography in wireless sensor networks.*in Proc. IPSN, Apr.*2008,pp. 245?256.

[46] A. S. Wander, N. Gura, H. Eberle, V. Gupta, and S. C. Shantz, Energy analysis of public-key cryptography for wireless sensor networks.*in Proc. IEEE Int. Conf. Pervasive Comput. Commun., Mar* 2005, pp.324-328.

[47] P. Ganesan, R. Venugopalan, P. Peddabachagari, A. Dean, F. Mueller,and M. Sichitiu, Analyzing and modeling encryption overhead for sensor network nodes. *in Proc. 2nd ACM Int. Conf. Wirel. Sensor-Netw. Applicat.,* 2003, pp. 151-159.

Research and Implementation of a Distributed Performance Testing Execution System Based on TTCN-3

Yongpo Liu

School of Computer Science and Engineering
BUAA
Beijing, China

Shuangmei Liu

College of Science and Information,
Qingdao Agriculture University
Qingdao, China

Abstract—**A distributed testing system is designed in this paper, which provides a mechanism of node communication, test script deployment, test scheduling, execution-driving and test result collection in distributed environment. A workload model is established, by which testers can describe the performance testing requirement. A performance testing framework is given, which simulates user behaviors in real environment based on virtual users so as to generate workload from the system under test (SUT). It can control the execution of virtual users by TTCN-3 standard interface. After executing the performance testing, test report is generated by extracting log. A method of generating performance test-case is studied by reusing functional test scripts. By executing performance testing on an online bookstore, this paper demonstrates the availability of the method of reusing TTCN-3 functional test scripts and the capability of distributed performance testing system established.**

Keywords-TTCN-3; Performance Testing; Distributed Testing System; Test Suite Reuse; System Under Test (SUT)

I.   INTRODUCTION

TTCN-3(Testing and Test Control Notation-3) provides a set of unified language standard for different users and tool developers, which can be used as a description language to test the response system of various communication ports[1]. Research on performance testing based on TTCN-3 is very important to broaden the application field of TTCN-3. At present, the performance testing tools based on TTCN-3 are very rarely, so as to simulate the load, it is usually done by adding the network traffic generator on the basic TTCN-3 test system[2]. But it is a network testing tool, which can only test the network delay between two nodes and didn't consider the business logic in a real scenario and other load indexes [3].

In this paper, the distributed performance testing system has been designed, which expands TTCN-3 by using typical performance testing methods. The transparent communication method between nodes is also designed, which can solve many communication problems such as the node addressing, message management, timeout retransmitting, message distribution, event registration, monitoring etc[4]. This paper also provides the scheduling mechanism to manage the test process and the test-driven technology to control the execution of TTCN-3 code. The two load indexes, which the existing study did not involve, have been introduced, which can provide the possibility to establish the association relationship between various indexes. The establishment of the association management makes the load model can be correctly controlled. Through parsing and automatic conversion of TTCN-3 function test-cases, the performance test-cases can be generated. The transformation rules were proposed in this paper, which realized the target of reusing the test suite and reducing the workload of performance testing code.

II.   TTCN-3 AND THE TESTING SYSTEM

TTCN-3 is a testing language based on text. The syntax of its core language is similar with the traditional programming language, but it has special test propagation characteristics. TTCN-3 is a flexible and powerful test description language, which is applicable to a variety of interactive system test. Its typical application are protocol test (including mobile and Internet protocol), service test, module test, platform test based on CORBA, API test etc. [5-6].

The top element TTCN-3 is module which is the basic building block. When the code in a module needs the elements defined in another module, the elements can be exported from the definition by importing mechanism. TTCN-3 is used to describe testing[7]. The test requirement is represented by test-cases and test data. Test-case is the most important behavior of TTCN-3, which describes the test requirement, depicts the test behavior logic and specifies the test criteria. In a TTCN-3 testing system, a test-case is allowed to run at the same time [8]. Namely, TTCN-3 doesn't support the concurrency execution between test-cases [9]. The TTCN-3 testing system can be viewed as the entity sets which are interactive [10]. Each entity is associated with a specific function in the testing system. These entities manage many operations, such as executing test, interpreting or executing the compiled TTCN-3 code, realizing the correct SUT, executing the external function, processing the clock and so on [11]. TTCN-3 solves the problem of describing the testing system from two levels, namely, platform-independent and platform-dependent. In the platform-independent level, it focuses on describing the test logic (including test-case, test data, test components), and cannot solve the test execution. In the platform-dependent level, it focuses on describing how to interact with SUT and cannot describe the test logic [12].

TTCN-3 looks like an ordinary programming language with expanded test, including dynamic testing configuration, matching mechanism, timer support, processing test decision and communication mechanism [13]. The testing system provides the mechanism to execute TTCN-3 from two levels of the execution logic and test adapter. The TTCN-3 test-case can be independent to the SUT and the test platform through a standard interface [14].

III.   THE DISTRIBUTED TEST EXECUTION SYSTEM

The wide application of distributed technology also gradually affects the design of testing system. On one hand, in order to test the distributed software system, the testing system has to support the distributed testing [15]; On the other hand, the testing system which adopts distributed architecture not only make full use of limited hardware resources and human resources, improve the efficiency of executing test, reduce the cost of time and resources, but also can simulate the actual operating environment, support some special testing scenarios and produce the required test load [16].

The distributed execution framework defines the system structure of testing execution platform, realizes the effective management of all testing execution nodes, and provides the basic services needed to execute the distributed testing. For the choice of the distribute architecture, this paper adopts the distributed strategy which is centralized [17]. Namely, a central computer controls the execution of a plurality of controlled computers, the whole testing process and resource management is accomplished by a center, which grasps the distributed testing environment condition and sends out the control command. The centralized structure is applicable to the distributed system for testing. Because the testing process is a kind of activity with high quality of process control, the testing system needs to timely obtain the global state to guide the process correctly [18]. At the same time the structure can be convenient for testers to monitor and operate the testing. In this paper, the distributed system is shown in Fig.1.
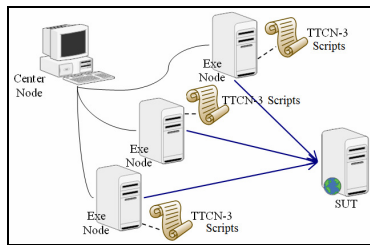
Fig.1 The Distributed System Environment

It includes two kinds of nodes, that is Center Node and Execution Node. The TTCN-3 testing script runs on the execution node, it describes a use-case how to test the system, including test input, test results and test execution conditions. It completely describes the process of testing the system in a given test target.The distributed testing system designed in this paper is shown in Fig.2.
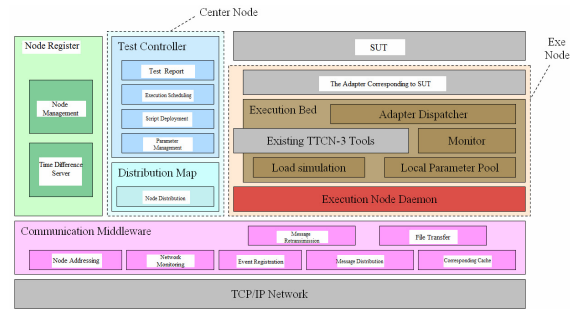
Fig.2 The Architecture of Distributed Testing System

It includes three kinds of nodes, which are Center Node, Execution Node and Nodes Register. The interaction among them is realized by the communication middleware which transfers the asynchronous communication information based on message and is realized by the socket based on TCP. The execution node communicated with the SUT through the test adapter, as to meet the needs from the SUT which needs different codec rules and interactive communication modes.

Based on the transaction as a unit, the communication resources are organized and managed. In the process of running transaction, the thread to deal with the communication process needs to put the resources into the recovery device, the resources are not able to determine whether will be used in the future and are recycled when not in use.

The distributed task can be completed by collaborating which involves information interaction. Therefore, in order to make the different nodes exchange data and cooperate effectively according to the requirements, we must formulate corresponding rules for interactive tasks and synchronous timing. This paper adopts the hierarchical division way to organize the testing task execution. Tasks are independent between layers, each layer doesn't need to know the mechanism to implement the next layer of it, and only need to know the interfaces between the layers. When the mechanism to implement a layer is changed, the upper and lower layers of it are not affected.

The test task of each layer involves the interaction between two nodes. In the test layer, this paper uses the interactive C/S structure. The execution node is a server which passively receives the request from the center node as a client. The center node actively sets up the test. Once the execution node receives the transaction request to start a test, a thread will be established immediately. Then the execution node users process all the corresponding behaviors under the test. Once the execution node receives the transaction request to end the test, the corresponding thread will be suspended and the related resources will be cleaned. A complete testing process is as follows: Firstly, the center node sends a command to notify the execution node to create a test. This command contains the TestID of the test. After receiving the notification, the execution node starts a thread to handle all the interactive activities corresponding to TestID. All the interaction will be carried out via the event registration and monitoring mechanism of the communication system. Once the test began, all the events and messages are registered and sent by the execution node. The center node will be coupled with TestID through the communication system. They both start the

deployment session to prepare for executing the test. Then, according to the scheduling scripts, the center node will start the scheduling session sequentially or concurrently and interact with the execution node. After executing the scheduling scripts, the center node will start the transfer result session to collect the test log from the execution node, so as to analysis the performance.

The distributed testing system can solve the problems in the distributed testing environment, such as node management, process scheduling, scripts deployment, results collection, task organization The communication system can support asynchronous message and file transfer. It can also resist abnormal communication in the distributed environment and guarantee to run the testing process stably in the risk communication environment. It provides the basic platform support for performance testing.

IV. DESCRIBE AND CONTROL THE PERFORMANCE TESTING

The performance testing model studied in this paper combined with the advantages of load simulation based on network characteristics and user behaviors, the performance testing scenarios based on virtual users and load indexes are presented. They provide the basis to describe and control the performance test-cases.

Virtual users represent the users to use a function of the SUT, so as to simulate user behaviors in accordance with the business process in an actual operation. The testing process based on virtual users is an iterative process of concurrent behaviors. To design the performance testing needs to consider how to organize virtual users. So the virtual user organization hierarchy is proposed and shown in Fig.3, which can concurrent virtual users between the execution nodes and in the execution nodes at the same time.
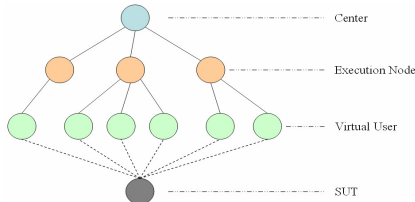
Fig.3 The Virtual User Organization Hierarchy

The node mappings allocate the tasks between the execution nodes to make use of the execution resources in the distributed environment. Within a node, a test bed is responsible for creating and destroying the virtual users, which maintains the number of concurrent virtual users required at a level, so as to simulate the actual load.

The various load indexes do not exist in isolation, which have related and restricted relations. This paper introduces two new indexes and a performance index.

- The Number of Concurrent Users. It is the number of users running at the same time and represents the use intensity from the user group to use the SUT.

- The Request Interval in a session. It is the interval between the request sequences from the same session.

- The Response Time. It is the interval between receiving request and returning the last response byte from the SUT.

The load model is based on above three indexes and shown in Fig.4.
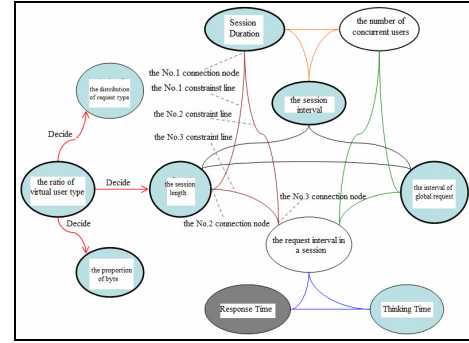
Fig.4 The Load Model

In the Fig.4, each index is connected with the curve which is called as the constraint line. If each index is regarded as a circle, the two ends of each connecting line were located separately in a point on different circles. The point is called the connection point.

For the set V composed of three indexes, there are three connecting points according with the indexes located in the V. If there are three constraint lines respectively to the three connection points for the end points to form a circle, the three indexes from V constitute a constraint group.

A complete process of performance testing is to execute a performance testing scenario. A performance testing scenario is composed of a Phrase sequence. In a performance testing, testers can choose a model to generate the load. Each Phrase contains the suite of next load indexes configuration in the present model. The composition of performance testing scenario is shown in Fig.5.
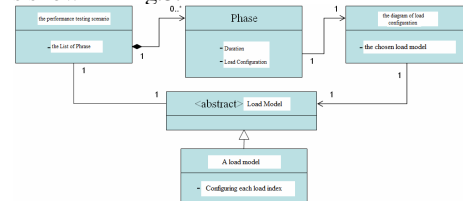
Fig.5 The Structure of Performance Testing Scenario

Through the above performance testing scenario, testers not only can configure every load index by Phrase, but also can change the indexes in the process of testing by the sequence composed of different Phrase.

V. THE PERFORMANCE TESTING BY TTCN-3

Through TTCN-3 language and TTCN-3 system, we can construct the performance testing framework and realize the virtual user management, load controlling, dynamic parameter management, communication connecting management, execution monitoring and track record.

The performance testing framework implements the execution bed of the execution node in a distributed testing environment. The execution bed is the driver to execute TTCN-3. Its purpose is to control the execution of TTCN-3, and then implement the virtual user management, load generation, test monitoring and resource management. The relationship among the function nodes from the framework is shown in Fig.6.
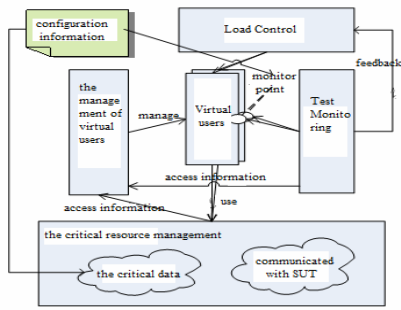
Fig.6 The Relationship of Performance Testing Framework

In order to reduce the development workload and shorten the cycle with greater efficiency, an automatic generator of TTCN-3 performance testing code is implemented in this paper. In a test cycle, the function testing precedes the performance testing, which means that a lot of available function test suite have already existed before the beginning of the performance testing, including TTCN-3 testing code, test adapter and codec. On the other hand, after 7 years of developments, TTCN-3 already has a plenty of function test suite that are developed, which are the basis of performance testing.

In this paper, the TTCN-3 function test-cases are parsed by the TRex open source tool (the TTCN-3 Refactoring and Metrics Tool), then an abstract syntax tree is generated. According to the position information of the source code corresponding to the node in the syntax, the replacement node is labeled in the source code. By analyzing the syntax tree, the new code is generated for the replacement corresponding to the nodes, and at last the original code is read. The characters outside the replacement are output as original. The characters inside the replacement are output as new codes to generate the performance test-cases. The conversion process of code is shown in Fig.7.
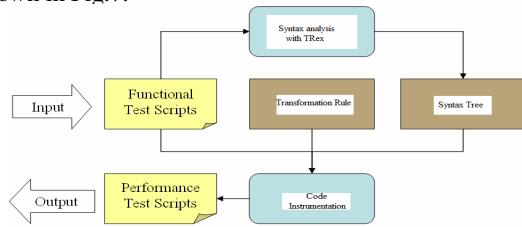
Fig.7 The Conversion Process of Code

VI. ANALYZE AN EXAMPLE

In this paper, the performance test-cases are generated on the basis of existing TTCN-3 function test-cases which describe the process of using an online bookstore and inspecting and deciding the behaviors of SUT.

The test deployment scripts are mainly used to design the distribution scheme of test scripts in different test points and the communication between different test points and the test control center, and the path of the TTCN-3 script file deployed in each test execution node. The test scheduling scripts are mainly used describe the execution sequence and the dependencies between the execution nodes during a test, so as to design the specific test logic corresponding to the particular test needs. The load configuration describes the performance testing scenario which is composed of Phrases with differ load

levels to describe the input intensity submitted by different users, including the command, task and data submitted. In this example, the test load configuration script contains 11 Phrases of the static load index to describe the ratio of different types of virtual users. Every Phrase has a different dynamic load index. The performance of SUT is focused on when the number of concurrent users is different, so the number of concurrent users and the thinking time are chosen to be controlled. The thinking time is constant in each phrase.

After the test execution, the center node receives the log returned from the execution node, which records the execution path and time and other information form every virtual user. After the performance measuring the test log, it will obtain a performance index file, which describes the test results, namely, the performance level of SUT. The file can be output automatically by the center nod after the end of test scenario. The test results are extracted from the performance measurement. The number of concurrent users is a variable. The response time and throughout and the ratio of success are dependable variables, it will get the metric map shown in Fig.8.
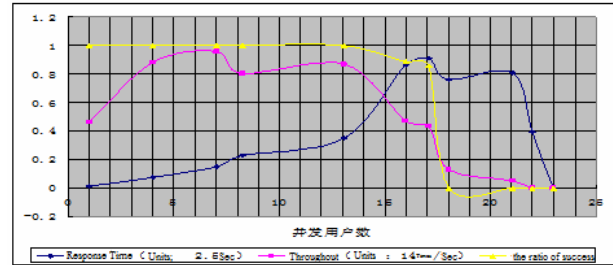
Fig.8 The Performance Metric Map

In Fig.8, as the number of concurrent users began to gradually increase from 1, the response time of SUT increased linearly, and the ratio of the success session retained 100%. The throughout increased at the beginning, the SUT worked normally and can respond to the load smoothly. When the number of concurrent users increased to 13, the response time significantly increased, the throughout began to decrease, the ratio of the success session also gradually decreased, then the SUT emerged the performance bottleneck. When the number of concurrent users exceeded 17, the ratio of the success session reduced to 0, the throughout reduced to near 0, then the SUT failed and cannot respond to the user requests. We get the following conclusion from this example. The users access to the online bookstore with 0.1sec thinking time, the online bookstore can support a maximum of 13 concurrent users.

VII. CONCLUSION

In this paper, the distributed execution strategy and framework are designed to support the performance testing. The method of describing and controlling the performance testing load is also proposed. And the technology to implement the load description controlling is studied, which is based on TTCN-3 system. This paper also analyzes and compares the function test-cases and the performance test-cases based on TTCN-3. Then it puts forward a performance testing method based on the test suite reuse, which can automatically reconstruct the TTCN-3 function test-case and then transfer the function test-case into the performance test-case. It will reduce the development workload of testers. Finally, through the performance testing of an online bookstore, the method and

process of TTCN-3 performance testing are demonstrated in a distributed environment by use of the system designed in this paper. Through the analysis of the results, the performance bottleneck of SUT is found.

REFERENCES

[1] Gábor Ziegler, György Réthy. Performance testing with TTCN-3 [R], TTCN-3 User Conference, 2010

[2] Mahnaz Shams, Diwakar Krishnamurthy, Behrouz Far. A model-based approach for testing the performance of Web Applications [J]. Proceedings of the 3rd international workshop on Software quality assurance: 54 - 61, 2011

[3] Katerina Goseva-Popstojanova, Fengbin Li, Xuan Wang, and Amit Sangle.A Contribution Towards Solving the Web Workload Puzzle [J]. Proceedings of the 2006 International Conference on Dependable Systems and Networks, 2011

[4] Raúl PeñaOrtiz Julio Sahuquillo, Ana Pont, José A. Gil. Modeling continuous changes of the user's dynamic behavior in the WWW [J]. Proceedings of the 5th international workshop on Software and performance: 175 - 180 2012

[5] Barford, P. and Crovella, M. Generating Representative Web Workloads for Network and Server Performance Evaluation [J]. Proceedings of the 1998 ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems:151-160, 1998

[6] Diwakar Krishnamurthy, Jerome A. Rolia, Shikharesh Majumdar. A Synthetic Workload Generation Technique for Stress Testing Session-Based Systems [J]. IEEE Transactions on Software Engineering. 32(11):868-882, 2006

[7] Helmut Neukirchen, Benjamin Zeiss, Jens Grabowski. An approach to quality engineering of TTCN-3 test specifications [J]. International Journal on Software Tools for Technology Transfer, 10(4):309-326 , 2010

[8] Ina Schieferdecker, Bernard Stepien, Axel Rennoch. PerfTTCN, a TTCN language extension for performance testing [J]. Testing of Communicating Systems, 2012

[9] Zhen Ru Dai, Jens Grabowski, Helmut Neukirchen. Timed TTCN-3 - a real-time extention for TTCN-3 [J]. Proceedings of the IFIP 14th International Conference on Testing Communicating Systems XIV: 407-424. 2012

[10] Giovanni Denaro, Andrea Polini, Wolfgang Emmerich. Early performance testing of distributed software applications [J]. ACM SIGSOFT Software Engineering Notes. 29(1):94-103, 2012

[11] Dorina Petriu. Architecture-Based Performance Analysis Applied to a Telecommunication System [J]. IEEE Transactions on Software Engineering, 26(11): 1049 - 1065, 2011

[12] Matthew J. Rutherford. A Case for Test-Code Generation in Model-Driven Systems [J]. Generative Programming And Component Engineering, Vol.48:377-396. 2012

[13] Dipl.-Ing. Martin Steppler. Performance Analysis of Communication Systems Formally Specified in SDL [J]. Workshop on Software and Performance :49-62,1998

[14] Paul Baker, Dominic Evans. TRex – The Refactoring and Metrics Tool for TTCN-3 Test Specifications [J]. IEEE Computer Society: TAIC-PART: 90-94, 2012

[15] Steve Counsell, Rob M. Hierons. Refactoring test suites versus test behaviour - a TTCN-3 perspective [J], Fourth international workshop on Software quality assurance: 31 - 38, 2010

[16] Robert L. Probert, Bernard Stepien, Pulei Xiong. Formal Testing of Web Content using TTCN-3 [R], TTCN-3 User Conference, 2012

[17] Ina Schieferdecker, Jens Grabowski. Introduction to the special section on advances in test automation: the evolution of TTCN-3 [J]. International Journal on Software Tools for Technology Transfer, 10(4): 281-283, 2012

[18] SIEMENS. Synchronized Distributed Testing using TTCN-3 [R], TTCN-3 User Conference 2007

Qualitative Analysis Using Rough Set in the Tourism and Hospitality Sectors

Chun-Che Huang
Information Management
Department, College of Man-
agement, National Chi Nan
University
Puli Township, Nantou
County 545, Taiwan (R.O.C.)

Tzu-Liang (Bill) Tseng
Department of Industrial,
Manufacturing and Systems
Engineering
The University of Texas at El
Paso
El Paso, Texas, USA

Meng-Ying Shen
Information Management
Department, College of Man-
agement, National Chi Nan
University
Puli Township, Nantou
County 545, Taiwan (R.O.C.)

Yu-Hsien Wang
Information Management
Department, College of Man-
agement, National Chi Nan
University
Puli Township, Nantou
County 545, Taiwan (R.O.C.)

ABSTRACT

Nowadays, many of tourism and hospitality studies are con-
ducted through survey mechanism and analyzed with a particu-
lar statistical analysis technique for a particular objective. Alt-
hough several mathematical techniques such as statistical tech-
niques are used for qualitative studying, but the questionnaire
contains both quantitative and qualitative data which are diffi-
cult for people to integrally explore the nature of entire data by
standard statistical techniques. To demonstrate nature of the
data set, the rough sets approach is a promising tool, particular-
ly suitable for uncovering patterns in a database with qualitative
and quantitative data in the questionnaire. In this paper, to deal
with the multiple- value problem in outcome features, the tradi-
tional rough set is modified. A new extended approach ─ The
Multi-Value Rule Generation (MVRG) Algorithm is developed
to bring a new perspective on data interpretation and knowledge
discovery. A case of tourism and hospitality is investigated to
perform comparison between the proposed approach and the
traditional analytical technique. Such comparison reveals the
proposed solution approach provides additional implication in
tourism and hospitality application.

Keywords: Questionnaire, Rough sets, Rule induction, Data
mining, Tourism

1. INTROSUCTION

There are so many issues discussed in the domain of tourism
and hospitality over decades. Oh et al., (2004) specified com-
ments and suggestions: There still are several fundamental is-
sues for hospitality and tourism research that were detected
during their review. For example, numerous studies give an
impression that the data were gathered to serve for a particular
application by a particular chosen statistical analysis technique.
Such method-driven studies typically lacked both compelling
problem statements and strong conceptual footholds (Oh, et al.,
2004). Data and methods are just selected in such ways or
means to achieve the (conceptual) goals of a study (Oh, et al.,
2004). This situation may occur often in the use of some meth-
odologies which questionnaire is collected and analyzed.

Such a methodology using questionnaire is broadly applied
in the domain of tourism and hospitality area, e.g., Researchers
used descriptive data analysis methods including most fre-
quently, for instance, content analysis (Bowen, 1998), correla-
tion (Baloglu, et al., 1996), t-test (Lam, et al., 1996), frequency
and cross-tabulation (Baloglu, et al., 1999), and im-
portance-performance analysis (Chu et al., 2000). Since the data
of questionnaire may also be both quantitative and qualitative, it
may be difficult for people to look inside the nature of data if
only by standard statistical techniques. In addition, a survey is

usually investigated with multiple dimensional perspectives.
Each dimension uses multiple similar question items. For ex-
ample, there may be five non-absolute responded values which
are identified by different levels of satisfaction. The question-
naire could be seen as a vague information system. Therefore,
the purpose of the system is to extract significant features that
have impacts on the results. Noticed by Au & Law (2000), the
Rough Set (RS) based approach provides a tool that is particu-
larly suitable for uncovering patterns in a database with hybrid
data. In the RS theory, a questionnaire may be viewed as an
object and question items may be viewed as attributes. It is
expected to apply the rough set theory to such a questionnaire
information system.

The rough set theory proposed by Pawlak in 1982 (Pawlak,
1982) has been used in reasoning and knowledge acquisition
(Grzymala-Busse, 1988). However, the traditional RS approach
assumes an unique value for an attribute corresponding to a
particular object. Thus, in this study, traditional rough set is
modified and extended to deal with the multiple value problems
since in the questionnaire, there may be multiple choices in a
question. To discover decision rules from a set of question-
naires with multiple choice items, a new extended approach ─
The Multi-Value Rule Generation Algorithm (MVRG) is pro-
posed.

To compare with the traditional analytical technology, "A
study of customer-satisfaction attribute matrix in the hotel sec-
tor including different constituencies like Japanese, American,
and Chinese travelers" is used. The goal of this paper is to apply
the proposed algorithm not only to induct rules and discover
hidden knowledge from questionnaires with the perspective of
complementary to traditional statistical analysis, but also pro-
vide different suggestions in accordance to the generated rules.
It is anticipated that the proposed research methodology for
mining data from such a vague system could establish a new
paradigm on data interpretation and knowledge discovery,
which is the core contribution from our research study.

2. METHODOLOGY

The main idea of the proposed algorithm is to generate reducts,
where condition attributes in a decision table may contain mul-
tiple values and multiple outcomes. The solution searching
process is as follows: According to the interaction sets
$\{[A_i,1]\} \cap \{[A_i,2]\} \cap \{[A_i,3]\} \cap \dots \cap \{[A_i, j]\}$ where $[A_i, j]$
refers to Attribute i corresponding to object j, check all subsets
of value sets of outcomes, if the value set of attributes is a sub-
set of value set of outcomes, and then output the result into
reduct table. Repeat the aforementioned steps until all objects
are implemented, then terminate the multi-value rules with mul-
ti-outcome reduct generation procedure. A multiple outcome

table (Table 1) represents the relationship between condition attributes and outcomes. In Table 1, the element ($A_{.cond\ i\ m}$) denotes the value of attribute ($A_{.cond\ m}$) that an object (tuple)(X_i) contains, while $A_{.out\ n}$ depicts the different outcome set of the corresponding tuple. In the proposed solution approach, Step 1 generates an initial decision table (Table 1).

Table 1 Fundamental structure of a multiple outcome decision table

Object (Xi)	Attributes ($A_{.cond\ m}$)				Outcomes ($A_{.out\ n}$)			
	$A_{.cond\ 1}$	$A_{.cond\ 2}$	...	$A_{.cond\ m}$	$A_{.out\ 1}$	$A_{.out\ 2}$	...	$A_{.out\ n}$
1	$A_{.cond\ 11}$	$A_{.cond\ 12}$	...	$A_{.cond\ 1m}$	$A_{.out\ 11}$	$A_{.out\ 12}$	...	$A_{.out\ 1n}$
2	$A_{.cond\ 21}$	$A_{.cond\ 22}$	...	$A_{.cond\ 2m}$	$A_{.out\ 21}$	$A_{.out\ 22}$	...	$A_{.out\ 2n}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x	$A_{.cond\ i1}$	$A_{.cond\ i2}$	...	$A_{.cond\ im}$	$A_{.out\ i1}$	$A_{.out\ i2}$	...	$A_{.out\ in}$
weight	w1	w2	...	wm				

Next, a pre-process procedure (Steps 2.1-2.4) is used to resolve the multiple value problem and reduce the comparisons in each iteration. For example to compare object 1 with other objects from 2 to x, there is a loop runs for $x*(x-1)/2$ times of comparison. In each comparison, the value of condition attributes and outcomes transfer into an array and do the XNOR operation, $\{\{[A_{.cond\ 1,1}] \oplus [A_{.cond\ 2,1}]\}, \{[A_{.cond\ 1,2}] \oplus [A_{.cond\ 2,2}]\},...\{[A_{.cond\ 1,m}] \oplus [A_{.cond\ 2,m}]\}, \{[A_{.out\ 1,1}] \oplus [A_{.out\ 2,\ 1}]\}, \{[A_{.out\ 1,\ 2}] \oplus [A_{.out\ 2,\ 2}]\}...\{[A_{.out\ 1,n}] \oplus [A_{.out\ 2,\ n}]\}\}$. The result of the operation leads to "True or False." If all of the value set results in "True" (T), which means the two objects have the same characteristics for each condition attribute and outcome attribute, then next combine them into an individual object given by a new object number and remove the original object number from the new table. On the other hand, if it results in "False" (F), which means this object is an individual one and will be added to the reduct table. The XNOR operation reduce the comparison significantly.

In Step 3, number of outcomes is used to divide the process into single outcome solution and multiple outcome solution. From Step 4 to the end of algorithm (Step 14), it is under the assumption that the number of object cardinality is greater than N_{min} to make sure that each object type is significant large, rather than a particular small case. In addition, if the merged reduct or individual reduct has the highest SI (strength index), then it refers the reducts can be qualified to decision rule. However, if the reduct with highest SI is not greater than N_{min}, then the reduct is excluded in the decision rule set. Repeat the aforementioned steps (1-14) until all objects are evaluated, then terminate the Multi-Value Rules Generation Algorithm procedure.

Note that the desired reducts (rules) are often comprised of the maximum of Strength Index (SI). In this study, SI is obtained from the work of Tseng (2008). The strength index of reduct f is computed as:

$$SI(f) = \sum_{j=1}^{m} v_j W_j * n_f, \text{ where:}$$

f is the reduct number, $f = 1, ..., n$;
$v_j = 1$ if condition attribute j is selected, 0 otherwise (A_j = "x");
W_j is the weight of condition attribute j;
n_f is the number of identical reducts f.

In this study, alternative rules are defined since there may be more than two reducts with the same maximum of SI. Assume the first desired reduct is selected as a decision rule. The remaining reducts with the same SI are defined as alternative rules. In general, the proposed search process adopts the breadth first search strategy. Figure 1 shows the process of the proposed Multi-Value Rules Generation Algorithm. The difference between this approach and the traditional rule generation algorithm is in Step 3. Most traditional rule generation algorithm only deals with single value for each attribute. The algorithm is presented next and the flow chart is presented in Figure 1.

Notations:
C: the condition domain;
S_v: the set of value for each condition attribute;
$A_{.cond}$: the condition attribute (feature);
$A_{.out}$: the outcome attribute
p: multiple value index ;
q: the number of values in the condition attribute (multiple value) for each object;
i: object index;
j: condition attribute index;
k: the outcome index;
D_{set}: the data set.
D_{reduct}: the ad hoc data set used for store the results.
O_{total}: the total number of outcomes.
m: the total number of condition attributes.
n: the total number of outcome attributes
x: the total number of objects;
t: reduct index;
X_i: object number;
Ψ: t-th reduct for object X_i;
Ψ'_{it}: t-th merged reduct for object X_i;
R_m: set of merged reducts;
Q: set of valid merged reducts from R_m;
N_{min}: minimum number of objects for the reduct required to be a qualified rule;
$S_{reducts}$: set of rules (reducts) selected from object X_i;
$A_{cond\ ip,j}$: the objects where each $A_{.cond,j}$ attribute contains $A_{.cond\ ip,j}$.
$A_{.outi,k}$: the objects where each $A_{.out,k}$ outcome contains $A_{.outi,k}$.
$A_{in}()$: a generation attributes interaction set function.
$A_{in}(i) = \{[A_{i,1}]\} \cap \{[A_{i,2}]\} \cap \{[A_{i,3}]\} \cap ... \cap \{[A_{i,final}]\} = [A_i] \cap D_{set}$.
$A_{comb}()$: a generation combination set function. For example, $A_{comb}(3, 2)$ can generate a set of two possible from a set $\{1, 2, 3\}$, and the result is $\{12, 13, 23\}$, i.e., $A_{comb}(3, 2)$ is equal to C_2^3.
$S_{.a}$: set of alternative rules selected from object X_i
$S_{.q}$: set of qualified alternative rules selected from $(S_{.a} \cup S_{.q})$
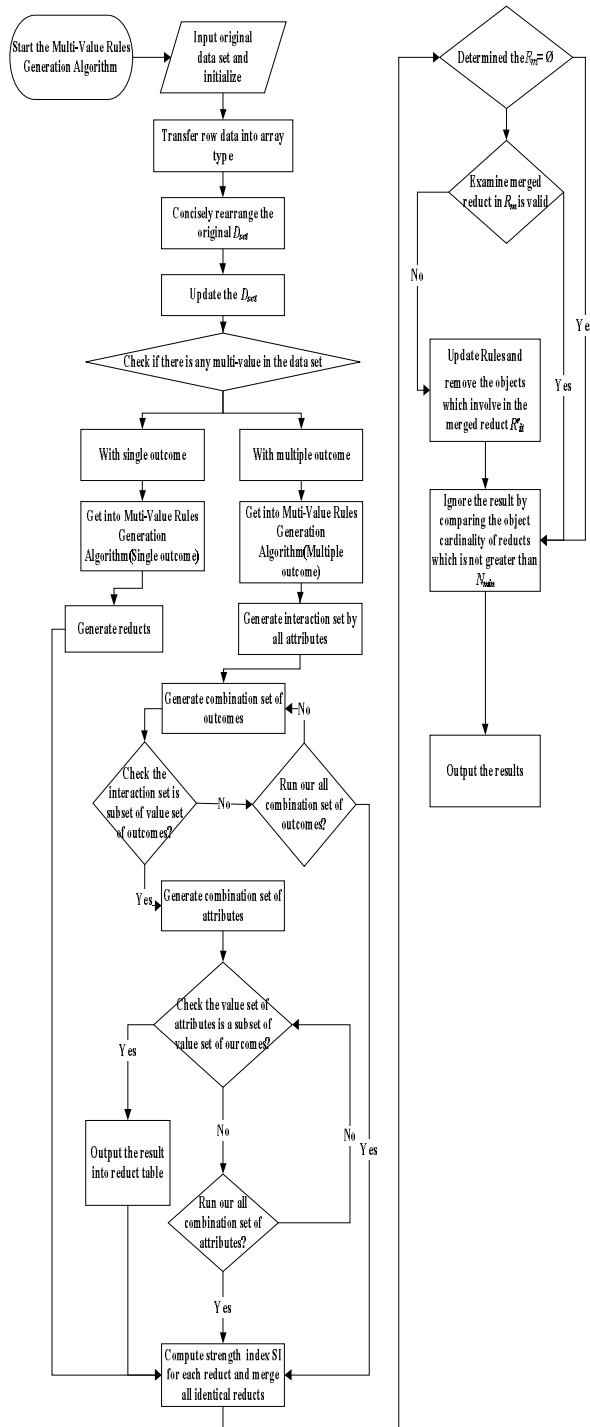
Figure 1. The flowchart of the MVRG algorithm

 The MVRG Algorithm
Input: $D_{set.}$
Output: The set of decision rules and alternative rules.
Step 1. Initialization: List all objects in D_{set}.

Step 2. Get into Pre-procedure.
Step 2.2. For $i = 1$ to x; Next i
Step 2.3 For $i = 1$ to x, $j = 1$ to m, $k = 1$ to n; Next j; Next k; Next i
Step 2.4. If $([V]A.cond\ i,m$ XNOR $[V]A.cond\ i{+}1,m)$ &&
          $([V]A._{.out\ i,n}$ XNOR $[V]A._{.out\ i{+}1,\ n})$= True
          Combine $X_i\{[V]A._{.condj}, [V]A._{.out\ k}\}$ with
           $X_i\{[V]A._{.condj+1}, [V]A._{.out\ k+1}\}$ into new_X_i
              Do new_X_i _cardinal number+1,count+1,
              Delete $X_i\{[V]A._{.condj}, [V]A._{.out\ k}\}$ and
               $X_i\{[V]A._{.condj+1}, [V]A._{.out\ k+1}\}$ from D_{set}
                 Add new_X_i into D_{set}
              Else
              Go to Step 3.
              End for
          Go to Step 3.
          End for
Step 3. Generate the reducts for each object.
Step 3.1. For $i = 1$ to x, $p = 1$ to q (multi-value)
Step 3.2. Set new_Out_Set $= \emptyset$
Step 3.3. Set mini_Set $= A_{in}$ (i)
Step 3.4. For out $= n$ to 1
Step 3.5.   Set comb_Out_Set $= A_{comb}$ (n, out)
Step 3.6.   For each k in comb_ Out_Set
Step 3.7.   If mini_Set $\subset \{\{[V]A._{.out,k} \cap [V]A._{.out\ ip,k}\} \cup \{[V]A._{.out,k} \cap [V]A._{.out\ iq,k}\}\}$ then add k into new_Out_Set
              End for
Step 3.8.   If new_Out_Set $= \emptyset$
Step 3.9.   Go to Step3.4
Step 3.10.   For Attr $= 1$ to m $- 1$
Step 3.11.   Set A_{comb} _Set $= A_{comb}$ (m , Attr)
Step 3.12.   For each k in new_Out_Set
Step 3.13.   For each j in A_{comb} _Set
Step 3.14.   If $\{[V]A._{.cond,j} \cap [V]A._{.cond\ ip,j}\} \subset \{[V]A._{.out,k} \cap [V]A._{.out\ ip,k}\}$
Step 3.15.   The reducts for X_i is formed
Step 3.16.   Else
Step 3.17   The reducts for X_i is not formed, go to Step 3.4 until that X_i formed.
Step 3.19.   The reducts for X_i is not formed
                 End for
              End for
              If any reducts for X_i is formed
                    Break Loop
              Else
                 The reducts for X_i is not formed
           End for
      End for
Step 4. Compute strength index SI for each reduct.
Step 5. Select one reduct at a time and compare it with all reducts in the remaining objects.
Step 6. Merge all identical reducts determined by this comparison and store them in R_m and go to Step 7. Else if $R_m = \emptyset$, then go to Step10.

Step 7. Examine if each merged reduct in R_m is valid. If the SI of each merged reduct in R_m is greater than the SI of each individual reduct from merged objects (except itself), then include the qualified merged reducts in Q, go to Step 8; otherwise, select one Ψ out of the merged objects (X_i) containing the highest SI and go to Step 9.

Step 8. In the set Q, select the reducts containing the max value of SI, and update

$S_{reducts} = S_{reducts} \cup \{\Psi'_{it}\}$. If the rest of merged reduct Ψ'_{it} from Q has more than one highest SI reduct, update $S_{.a} = S_{.a} \cup \{\Psi'_{it}\}$. Remove the objects which are involved in the merged reduct Ψ'_{it}.

Step 9. Select the Ψ with highest SI for X_i. Update $S_{reducts} = S_{reducts} \cup \{\Psi\}$. If there is more than one reduct with the highest SI for X_i, store the rest of the reducts Ψ in $S_{.a}$. Select the reducts in $S_{reducts}$ and $S_{.a}$ whose object cardinality is greater than N_{min} into $S_{reducts}$ as qualified alternative rules from $(S_{.a} \cup S_{.q})$. If not greater than N_{min}, then ignore this reduct, and go to Step 10.

Step10. Do while $i < n$, $i = i + 1$ then go to Step 9 until end up the loop.

3. CASE STUDY

Huang & Chen (2004) proposed a research to explore the concept of customer-satisfaction attribute matrix in the hotel industry. The customer-satisfaction matrix grids illustrated that the room quality, service quality, and overall efficiency factors fell into the competitive strength quadrant; public facilities, name/reputation, and reception service in the relative indifference quadrant. Focusing on Attribute importance, original case study adopted the concept "direct ratings of importance" suggested by Jaccard et al. (1984) which is an outstand attribute and dominate the decision of buying behavior about to happened.

The survey questionnaire in this case was designed in three kinds of language, English, Japanese and simplified Chinese. There are 613 valid questionnaires, respectively are 229 sheets for American travelers, 212 sheets for Japanese travelers and 172 sheets for China travelers. Questionnaire had been separated into two parts. First part is used to find out the most considerable element when choosing hotel and satisfaction of actual experience after using hotel. The second part is basic information which separated into two sub parts, respectively is tourism property and social background. Besides, there are six dimensions for data analysis, including (1) room quality factor, (2) public facilities factor, (3) service quality factor, (4) reputation factor, (5) greeting service factor and (6) entire service efficiency factor. Moreover, the data analysis method mainly includes Descriptive Statistics, Factor Analysis, Paired-Samples T-Test and One Way ANOVA. Finally build up a satisfaction attribute matrix.

This study also classify all samples of questionnaires with 3 different nationalities of questionnaire respectively and use the same 613 objects of data. All of question items were classified to the same 6 dimensions. And 15 question items are extracted as the survey subjects. The value of each item coded from 1 to 5, respectively are "extremely not important/satisfied", "not important/satisfied", "normal", "important/satisfied "and "extremely important/satisfied". Questionnaire attributes are presented in Table 2.

The proposed algorithm is applied to the data. The comparison between of the original findings and new findings are presented as follows:

Table 2 Questionnaire attributes (6 dimensions and second part of items)

Questionnaire	Dimension	Questionnaire attributes	Value
First part	1.Room quality	A1(importance of clean appearance) A2(satisfaction of clean appearance) A3(importance of reasonable room price) A4(satisfaction of reasonable room price) A5(importance of service quality) A6(satisfaction of service quality)	1~5
	3.Service quality	A7(importance of room cleanliness) A8(satisfaction of room cleanliness)	1~5
	6.Entire efficiency	A9(importance of entire efficiency/fluently service) A10(satisfaction of entire efficiency/fluently service)	1~5
Second part	Basic information	A11 (nationality) A12 (purpose of traveling to Taiwan) A13 (information resource: 1. Travel agency, 2. Family member or friends, 3. Guide book or magazines, 4. Media or advertisement, 5. Airline, 6. Government organization) A14 (travel type)	1~5

<Note> A13 is multiple choice

The similar part

Both results are discussed from perspective of nationality. For example, in the perspective on American travelers, the information from the previous study shows that American travelers are more care about dimension of room quality, and those sub factors are clean room, comfortable room condition, quiet room, safe room condition and clean appearance. The results by applying the proposed approach sows that attributes "clean room" and "entire efficient service" which are detaild elements in service quality dimension and entire efficiency dimension would be important reasons of satisfaction to affect American travelers' consumer loyalty.

The dis-similar part and additional discovery

In the part of strategy suggestion, the previous study said that consumer loyalty is complexly related with satisfaction but they are not symmetrical, which means that it cannot tell the exact direct relationship between consumer loyalty and satisfaction in the previous study. The re-visit willing and four elements, cleanliness of appearance, price, inner cleanliness and service are extracted as emphasis of this study implying the relationship between consumer loyalty and their satisfaction.

The data expressing of this study is different from statistical method. The new results are in format of a rule, "If…then…".

- If those who are American and were satisfied with clean appearance and extremely satisfied with clean room, then they would be willing to visit again.
  (Rule3. If A2 = 4 and A7 =5 and A11=A then O1 = Y)

- If those who are American and were satisfied with clean appearance, extremely satisfied with service quality satisfied with clean room, then they would be willing to visit again.
  (Rule18. If A2 = 4 and A6=5 and A7= 4 and A11= A then O1 = Y)

- If those who are Japanese and were satisfied with clean appearance and satisfied with service quality and satisfied with clean room, then they would be willing to visit again.
  (Rule 9. If A2 = 4 and A6=4 and A8=4 and A10=4 and A1tose who are Japanese and consider clean appearance, fair price, service quality, clean room and entire service efficiency are extremely important, then they would be willing to visit again.
  (Rule 37. If A1= 5 and A3=5 and A5=5 andA7=5 and A9=5 and A11 = J then O2 = Y)

- If those who are Chinese and consider clean appearance is extremely important and clean room is important and extremely satisfied with clean appearance and service quality, then they would be willing to visit again.
  (Rule 62. If A1=5 and A2 =5 and A6=5 and A7=4 and A11 = C then O1 = Y)

- If those who are Chinese and consider that service quality is important and also satisfied with service quality and clean room but feel normal about affair price, then they would be willing to visit again.
  (Rule 50. If A4=3 and A5 =5 and A6=4 and A8=4 and A11 = C then O1 = Y)

From the six dimension perspectives, additional discovery from traditional statistical method are presented;

1. According to Rules 3 and Rule 18, from the perspective of America travelers, "Tourists who are American and were satisfied with clean appearance, extremely satisfied with service quality satisfied with clean room, then they would be willing to visit again." That is, the attributes "clean room", "clean appearance" and "Service quality" which are detailed elements in room and service quality phase would be important reasons to affect their consumer loyalty.

2. According to Rules 9 and 37 from the perspective of Japan travelers, "Tourists who are Japanese and consider clean appearance, fair price, service quality, clean room and entire service efficiency as extremely important. They were satisfied with clean appearance and satisfied with service quality and satisfied with clean room, which would be important reasons to affect their consumer loyalty." Specifically, the attributes "clean appearance", "service quality" and "entire service efficiency (not mentioned in impervious study)" which are detailed elements in service quality phase, room quality phase and entire efficiency phase would be important reasons to affect their consumer loyalty.

3. According to Rules 62 and 50, "Tourists who are Chinese and consider that service quality is important and also satisfied with service quality and clean room but feel normal about affair price, then they would be willing to visit again." Specifically, the attribute "clean appearance", recognition of "importance of service quality" and "fair price (not mentioned in previous study) "which are detailed elements in room quality phase and service quality phase would be considerable elements to affect their consumer loyalty.

The brief findings are summarized: (1) The previous study shows the more important dimension to American is room quality, to Japanese are room quality and service quality, as well as to Chinese. By using proposed approach, the rules are able to show the cause and effects behind the questionnaire data set. (2) Both offer different perspectives from three nation travelers. But this study not only finds out the more specific reasons why those people who would be willing to visit again, but also finds out the more specific reasons why those people who would be not willing to visit again. (3) This study only focus on three parts of all dimensions for data analysis. If researchers want to extend to extract different dimensions to analyze in compliance with their requirement, it is possible to obtain the customized results by using the proposed methodology.

4. Conclusions

This study introduced the problem that methodology uses questionnaire and proposed useful rule induction approach that could mine the questionnaire data. The multi-value rule generation algorithm was developed such that all different data types and multi-values could be handled in a uniform manner. The proposed solution approach aims people to easily find out the nature of questionnaire data. Furthermore, this study is believed to have contribution on that it can result in specific rules and integrally reveal knowledge hidden behind large data. This research has significant implications not only in relation to proposing a different DM method for questionnaire survey, but also expect to help managers analyze market strategies in tourism and hospitality sectors.

In the future, it is critical issue to extend the study by text mining and association mining techniques to mine questionnaire data, including both open and closed questions. In addition, in this study, one case is re-studied only by applying the approach in tourism and hospitality related questionnaire, it is necessary to compare addition cases in other research area with different types of questionnaires. Besides, with the aid of advanced information technology, business intelligence is now a trend to transform raw data into meaningful and useful information for some business purposes. Therefore, how to conduct the decision rules inducted by the proposed algorithm to data will also be an interesting research topic in future.

5. References

[1] Au, N., & Law, R. (2000). The application of rough sets to sightseeing expenditures. Journal of Travel Research, 39(1), pp.70-77.

[2] Grzymala-Busse, J. W. (1988). "Knowledge acquisition under uncertainty: A rough set approach." Journal of Intelligent Robotic Systems. Vol. 1, pp.3–16.

[3] Huang, C. T. & Chen, F. Y. (2004). A Study of Customer-Satisfaction Attribute Matrix in the Hotel Industry－A

Comparison of Japanese, American, and Mainland Chinese Travelers. Journal of Tourism Studies, 10(2),

[4] Oh, H., Kim, B. Y., & Shin, J. H. (2004). Hospitality and tourism marketing: recent developments in research and future directions. International Journal of Hospitality Management, 23(5), 425-447.

[5] Pawlak, Z. (1982). "Rough set", International Journal of Computer and Information Sciences, pp. 341–356.

[6] Tseng, T.L. (Bill), Huang, C.C. and Ho, C. (Johnny). (2008)."Autonomous Decision Making in Customer Relationship Management: A Data Mining Approach," Proceeding of the Industrial Engineering Research 2008 Conference, Vancouver, British Columbia, Canada.

[7] Jaccard, J., & Sheng, D. (1984). A comparison of six methods for assessing the importance of perceived consequences in behavioral decisions: Applications from attitude research. Journal of Experimental Social Psychology, 20(1), 1-28.

[8] Bowen, J.T., Sparks, B.A. (1998). Hospitality marketing research: a content analysis and implications for future research. *International Journal of Hospitality Management 17, pp.125–144.*

[9] Baloglu, S., & Uysal, M. (1996). Market segments of push and pull motivations: a canonical correlation approach. *International Journal of Contemporary Hospitality Management*, *8*(3), 32-38.

[10] Baloglu, S., & Assante, L. M. (1999). A content analysis of subject areas and research methods used in five hospitality management journals. *Journal of Hospitality & Tourism Research*, *23*(1), 53-70.

[11] Chu, R. K., & Choi, T. (2000). An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: a comparison of business and leisure travellers. *Tourism management*, *21*(4), 363-377.

[12] Lam, T., & Zhang, H. Q. (1999). Service quality of travel agents: the case of travel agents in Hong Kong. *Tourism Management*, *20*(3), 341-349.

Using Cycles to Detect Strengths and Weaknesses in the Software Development Lifecycle

David L. Ivan and James P. Buckley
Computer Science Department
University of Dayton
300 College Park
Dayton, Ohio 45469-2160

Abstract—**This paper proposes data mining utilization techniques that can identify strengths and weaknesses in software projects and sets of projects that can be used by business leaders, managers, and engineers to strengthen the software development project with greater efficiency. Strengths and weaknesses in the software development lifecycle are identified using the meta-pattern extraction of relationship cycles[13] in mined data.**

1. Introduction

We have the capability to store extremely large amounts of data in organized and automated systems. The preponderance of data warehouses and datamarts [5] [9] are concrete evidence that this is not only possible, but of great interest to researchers, government agencies, and large corporations.

Data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [7]. We are looking for *patterns* within the data that reveal knowledge previously unknown. One of the most common applications of data mining is to generate all significant association rules between items in a data set. We can employ an efficient algorithm to mine a large collection of basket data type transactions for association rules between sets of items with some minimum specified confidence [2]. There may be chains of rules forming patterns of patterns, or *meta-patterns*, where the head of one rule is the body of another rule. In particular, the chain of rules may form a *cycle*.

In this paper, we propose detecting cycles in the software development lifecycle (SWDLC) to reveal opportunities for strengthening positive cycles and weakening negative cycles. A general software development lifecycle is introduced with sources of data that can be mined. The concept of situational factors[6] and sub factors in software development brings possible relations of interest in the SWDLC that may or may not be available in the data mined. The relationships, meta-patterns and cycles are mined from the data using existing algorithms [13][3][4].

2. Cycle Mining Formalism

Cycles exist commonly in everyday life. As would be expected, the evidence of these cycles can be found in our data. It is important to identify such patterns because this identification assists us in a better understanding of the data itself. A powerful aspect of a cycle is its inherent implication of continuity. Extraction of a cyclic pattern alerts a system that targeting any of the cycle participant activities, the rules, assures continuous attainment of the goal (at least until the cycle is broken). Likewise, an undesirable cycle can be broken by failing to fire any of the constituent rules.

In previous work [13], [4], the authors developed a methodology for discovering cycles and a methodology that uses the individual rule supports and confidences to detect and categorize different types of cycles. Using metric τ computed from constituent rule support and confidence factors, we are able to characterize the strength of the encompassing cycle. The following definitions were used and presented in that work.

Definition 2.1 (support and confidence) *The* support *for a rule, $C \leftarrow B$, is the percentage of data tuples that satisfy $C \land B$; the* confidence *for rule $C \leftarrow B$ is the percentage of tuples that satisfy C given all tuples that satisfy B [12]. The* support *of a cycle is the minimum support value of any of the constituent rules forming the cycle; the* confidence *of a cycle is the minimum confidence of any of the constituent rules forming the cycle.*

We use τ, the average of the above two metrics, to be our threshold measurement for any specific rule. No rule with τ less than a user-specified threshold U will be considered meaningful enough to be placed in the system knowledge base. Hence, it will not be detected as part of any cycle.

Definition 2.2 (τ)
$$\tau = (support + confidence) / 2$$

Cycles are composed of n individual rules, so we define the strength metric τ applied to cycles as

Definition 2.3 (T)
> $T = min(\tau_1, \ldots, \tau_n)$, where τ_i is the strength measurement of rule i.

Our methodology and cycle detection algorithms do not consider rules with τ less than user-specified threshold U. Thus, any detected cycle has T at least U.

3. Software Development Lifecycle and Data Mining

Data mining applications have been used to detect patterns in specific aspects of the software development lifecycle in the past. The software development lifecycle (SWDLC), or software development process, is the methodology imposed on the development of a software product and has a number of implementations, methods and representations. The SWDLC can be based on the general, waterfall, V-model, object oriented, spiral, agile, or a number of other methods. The basic components of the SWDLC that occur in the general model include the requirements, design, implementation/coding, verification, and optional maintenance phases. The implementation and repetition of the phases vary per lifecycle method. The basic waterfall model completes each phase before proceeding to the next (Fig. 1).
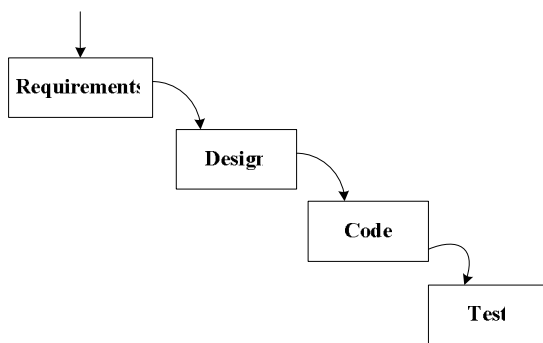
Fig. 1 Waterfall Model

The V-model extends the waterfall model raising the bottom of the model back up after the code phase and breaks the test phase into a unit test, integration test, system test, and acceptance test phases [8]. The spiral model uses multiple prototypes to refine objectives and resolve identified risks prior to following the basic waterfall phases. The object oriented model uses an object oriented analysis to tie system and user requirements to the project domain, and an object oriented design to develop an architectural model of the project domain in objects. It then follows with the object oriented coding and testing phases. The agile model iterates over the entire lifecycle rapidly creating prototypes after each lifecycle. The common model referenced for the purposes of this paper will be a lightly modified waterfall model with additional timeline concurrency and feedback flexibility features to reflect many modern SWDLC implementations. (Fig 2).
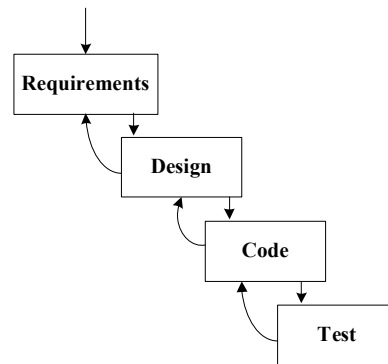
Fig 2. Lightly Modified Waterfall Model

The lightly modified waterfall model allows for a late phase concurrency of SWDLC phases. Allowing for the start of next timeline phase prior to the completion of the previous, helps prevent entire projects from getting frozen pending the resolution of a troublesome requirement, design detail, or code implementation. Complete and mostly complete modules with various levels of independence can continue to the next SWDLC phase. The lightly modified waterfall model also allows for immediate feedback reaction of defects found at the next phase. A design implementation that reveals a conflicting requirement defect can be acted upon immediately using the available defect system the feedback path presents in the model for each phase. Defects can be introduced at any phase of the SWDLC. A defect that crosses back into a previous phase or phases has greater scope. The cost of defects generally increases with increased scope, especially when a rigid documentation formalism (or safety nets) exists.

Mining for strengths and weaknesses in the SWDLC that can prevent or lead to defects respectively requires an abundance of project data. A list of data sources that can be used to mine the project, corporation, and software development lifecycle for patterns and relationships can include:

1) Historical records, version control systems.
   - Includes a lines of code(LOC) count, authors of sections/lines, and other software metrics.
2) Software problem report and tracking systems.
   - Data includes problem reports (PR) per project, user, or module. Patterns on PR priorities, detection phase, effect/scope, author, owner, etc.
3) Electronic communications.
   - Communication logs including email, meeting minutes, instant messages can be mined to detect: requirement, design, implementation, test, and bug correcting decision dialogs; team cohesion, leadership, and other issues.
4) Software development documentation and processes [14]
   - To detect conformance to standards and project development processes.
   - To gather the volume of changes and or rework in the planning aspects, requirements, and design.
   - Can be mined for short comings and or repeated bottlenecks in SDLC.
5) Organizational, employee, management records and additional project information.

   Additional source of information to mine for the software development lifecycle includes:
   - Organizational information like size, maturity, stability.
   - Employee records/profiles that include information like: seniority, overall experience, experience on project types and projects, experience at different software lifecycle stages.
   - Management parameters like experience and accomplishments.
   - Project properties like team size, overall team experience, experience with the project focus, experience with current tool set and others.

   List A: SWDCL Mined Data Sources

Data mining has been used in the SWDLC for management rule generation, risk assessment, and software component testing. Data mining methods have also been used for maintenance, component discovery, and automation of component reuse using predictive modeling, link analysis, and text mining analysis[10]. One link analysis of mined data was used to discover associations among software problem report characteristics using the Apriori algorithm[11].

Example 1: Rules generated using link analysis of mined data[10]

Rule 1: When a problem report was low priority and the time spent to fix it was about a ½ day, then the problem has a high probability (87.5%) of being a document related bug.

Rule 2: When a problem report was a low priority and the time spent to fix it was about 1 week then the problem had a high probability (83.3%) of being a software bug related bug.

Other methods of data mining usage in the SWDLC include the use of association rules and frequent patterns, bug classification, error source clustering, and text mining [14].

4. Software Development Lifecycle Relationships

Identified factors that can be used to mine the SWDCL for associations and relationships must be accessible, accurate, and meaningful. The accessibility of a software development factor requires the data representing that factor can be mined from one or more sources similar to List A. The accuracy of data mined enforces proper management of duplicate data and proper matching of similar but not exactly the same text data representing the same information. How meaningful the data is to the user can be based on the confidence and support of a relation, but must also be determined by the consumer as relevant before action will be taken on it. The useful detection of relationships or rules in the SWDLC requires meaningful factors to be mined. Using a subset of the situational factors and sub-factor concepts for the SWDLC presented in [6], meaningful relations can be harvested if the data is accessible. (Note that this is not an exhaustive study of the 8 factor classifications, 44 factors, and 170 sub factors presented in [6], and based on the source of the data to be mined and the level at which an association is made, relationships may differ from that model).

Some sample classification factors and sub factors follow:

- Personnel: turnover, skill, productivity, cohesion, culture, experience.
- Application: quality, application size, complexity, performance, reuse.
- Management: expertise, continuity, knowledge.
- Requirements: changeability, scope creep, feasibility, rigidity.
- Organization: maturity, facility remoteness, stability, size, commitment.
- Business: time to market.

Associations or rules mined from the SWDLC data can be specific to a single project or multiple projects depending on the granularity of the data. Rules are represented in the form of *if-then* patterns and have an associated τ value. Only rules that meet the user specified valid threshold are considered relevant. Some example mined rules and their threshold follow.

1) If Employee.ProgExperience = 'High' then Employee.Program.Reuse = 'Above Average' ($\tau = 0.88$).

2) If Employee.Program.Reuse = 'Above Average' then Employee.Productivity = 'High' ($\tau = 0.86$).

3) If Employee.Productivity = 'High' then Employee.Program.Cost = 'Below Average' ($\tau = 0.87$).

4) If Employee.Program.Cost = 'Below Average' then Employee.Review = 'Above Average' ($\tau = 0.85$).

5) If Employee.Review = 'Above Average' then Employee.Turnover = 'Below Average' ($\tau = 0.88$).

Coupling of mined relationships can be used to develop cycles in the SWDLC using the Cycle Mining Algorithm 2.1 in [4].

5. Cycle Detection
There may be chains of rules forming patterns of patterns, or *meta-patterns*, where the head of one rule is the body of another rule. We now represent our discovered rules using a Causal Graph. This allows us to depict the causal links and hypergraphs formed by interconnecting the consequent of one rule to the antecedent of another. Storing the knowledgebase as a hypergraph enables us to find meta-pattern by applying simple graph algorithms. We are primarily interested in chains of rules that form *cycles*.

An example of a detected cycle follows.

- The example rules above are present in a dependency hypergraph showing an informal relationship in the system.
- The user specified valid threshold is 0.83.
- A new rule is discovered:

  -If Employee.Turnover = 'Below Average' then Employee.ProgExp = 'High' ($\tau = 0.84$).

- The Cycle Mining Algorithm detects the following cycle with T = 0.84.
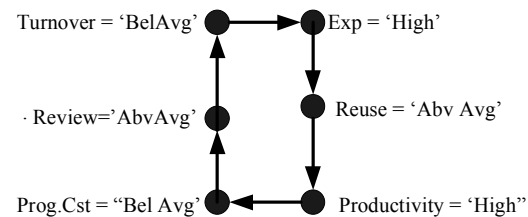
Fig. 3 Detected Cycle

This cycle represents an alpha cycle[3] since it is positive cycle. The corporate management might decide to reinforce it by emphasizing software reuse in the SWDLC.

6. Conclusion and Future Work
In this paper, we have shown how utilizing data mining to identify cycles[13] in the SWDLC can be used to identify strengths and weaknesses in a corporation's software projects and sets of projects. We propose that mining these cycles may reveal previously unknown efficiencies and inefficiencies in the SWDLC that would provide new insights as to possible options for increasing positive (alpha) cycles and decreasing negative (beta) cycles. The granularity of the data mined provides for cycle information at many levels from a very specific SWDLC procedure to handling project level resources and personnel. This granularity could

provide useful cycles for many of a program's business leaders, managers, and engineers.

In future work, we are considering new algorithms to mine cycles using an exhaustive traversal tree. We also will examine certain properties that hold on a particular cycle and show that it also holds for all sub-cycles.

Bibliography

[1] Adriaans, P., and Zantinge, D., *Data Mining*, Addison-Wesley, 1996.

[2] Agrawal, R., Imielinski, T., and Swami, A., "Mining association rules between sets of items in large databases," *SIGMOD Bulletin*, May 1993, pp. 207-216.

[3] Buckley, J. P., Seitzer, J., M., "A Framework for. Detecting Vulnerable, Cascaded Fuzzy Cycles in the. Carbon Chain", Proceedings of the the 16th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2012), Vol 1.

[4] Buckley, J. P., and Seitzer, J. M., "A Paradigm for Detecting Cycles in Large Data Sets via Fuzzy Mining," *Proceedings of the 1999 IEEE Knowledge and Data Engineering Exchange Workshop*, Chicago, Illinois, November 8, 1999. Pages 49-55.

[5] Chaudhuri, S. and Dayal, U., "An overview of data warehousing and OLAP technology," *SIGMOD Record*, Vol.26, Num. 1, 1997, pp. 65-74.

[6] Clark P. and O'Connor R.V. 'The situational factors that affect the software development process: Towards a comprehensive reference framework', Journal of Information Software and Technology, Vol. 54, Issue 5, May 2012, pp. 433-447.

[7] Frawley and Piatetsky-Shapiro, editors, *Knowledge Discovery in Databases*, chapter Knowledge Discovery in Databases: An Overview, AAAI Press/The MIT Press. 1991.

[8] Joan Rigat, Sup. Dr Trienekens, J., Samalikova, J., 'Data mining analysis of defect data in software development process', Eindhoven University of Technology, (February 2009)

[9] Kimball, R., *The Data Warehouse Toolkit*, John Wiley and Sons, 1996.

[10] Nayak, Richi & Qiu, Tian (2006) Use of Data Mining in System Development Life Cycle. In Williams, Graham J. & Simoff, Simeon J. (Eds.) Data Mining - Theory, Methodology, Techniques and Applications. Springer-Verlag, Germany, pp. 105-117.

[11] Rakesh Agrawal. Ramakrishnan Srikant. Fast Algorithms for Mining Association Rules. IBM Almaden Research Center. 650 Harry Road, San Jose, CA 95120.

[12] Ramakrishnan, R., *Database Management Systems*, McGraw-Hill, 1998.

[13] Seitzer, J., Buckley, J. P., Monge, A., "Meta-Pattern Extraction: Mining Cycles", *Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS-99)*, Orlando, FL; pp. 466-470.

[14] Taylor, Q and Giraud-Carrier, C. (2010) 'Applications of data mining in software engineering', Int. J. Data Analysis Techniques and Strategies, Vol. 2, No. 3 pp.243-257

Computational Currency Swapping

John JENQ

Montclair State University

Montclair, NJ, 07043, USA

and

Hubert JOHNSON

Montclair State University

Montclair, NJ, 07043, USA

ABSTRACT

In the world's financial market today, there are many exchange rates. Banks have their own exchange rates posted and these rates are dynamic in nature. There are arbitrage opportunities when there exist price differences among different currencies in the international exchange market. The major process of the arbitrage strategy is to convert one currency to another, then convert it again to a third currency…etc., and then eventually convert it back to the original currency within a short time span. The simplest form of this process is the triangular arbitrage which involves only three currencies. It is possible to process more than three currencies in order to apply the arbitrage process. Although it may seem as if arbitrage is a risk-free process, in reality, there are risks involved. One of the risk factors is time; i.e., how fast one can process and how much capital one has in order to take advantage of these price discrepancies. It is worth noting that the arbitrage opportunities only exist when a bank's quoted exchange rate is not equal to the market's implicit cross exchange rate; these all happen in the range of seconds. The high frequency market fluctuation makes this process a challenge and calls for the utilization of high-performance computation. In this paper, we investigate the arbitrage problem by utilizing GPU to identify the opportunities and speed up the process.

1. INTRODUCTION

Currency swaps may occur because of the interest rate discrepancies that companies are facing due to loan regulations of doing business in other countries [3]. In order to buy goods, companies need to reserve foreign currency, such as US dollars, in order to do international trading because US dollars are trusted by most nations.

In today's worldwide financial market, there are many exchange rates. Banks have their own exchange rates posted and these rates are dynamic in nature. It is possible that these exchange rates are not consistent during a short period of time and that they will go back to a so-called *equilibrium state*. This, therefore, raises a question of currency swapping arbitrage. Traditionally, arbitrage of currency swaps occurs when the interest rates of two parties in two countries are different. By swapping the currency from one to the other and then saving the money to other country, this may give opportunities to profit more. This practice usually comes with a forward contract to safe guard the outcome. It is sometimes used to hedge assets as well. Currency swaps are over-the-counter derivatives that serve two main purposes. First of all, they can be used to minimize foreign borrowing costs. Secondly, they can be used as tools to hedge exposure to exchange rate risk. In [5], the authors proposed that currency swaps can help global firms achieve long-term financing and financial risk management objectives. Additionally, in [6], they show the growth of currency swapping due to firm's economic exposure.

There are arbitrage opportunities when there exist price differences among different currencies in the international exchange market over a very short period of time, such as seconds. The major process of this kind of arbitrage strategy is to convert one currency to another, then convert it

again to a third currency…etc., and then eventually convert it back to the original currency within a short time span. The simplest form of this process is the triangular arbitrage which involves only three currencies. It is possible to process more than three currencies in order to apply the arbitrage process. Although it may seem as if arbitrage is a risk-free process, in reality there are risks involved and therefore the gain may be limited [1]. In [4], the authors investigate the role of execution risk in high-frequency trading through arbitrage strategies. They show that if rational agents face uncertainty about completing their arbitrage portfolios, then arbitrage is limited. Note that arbitrage not only occurred in currency exchange market, but also only equity market and stock market as well[12].

For high frequency trading, the risk factor is time; i.e., how fast one can process compared with your competitors and how much capital one has in order to take advantage of these price discrepancies. It is worth noting that the arbitrage opportunities only exist when a bank's quoted exchange rate is not equal to the market's implicit cross exchange rate—all these happen in the range of seconds.

The high frequency market fluctuation makes this process a challenge and calls for the utilization of high-performance computation. In this paper, we accomplish this by using GPU to identify the opportunities and speed up the process in our investigation of the arbitrage.

In the past, CPU clusters have been used to achieve high performance computation. GPU computing uses GPU as a co-processor to accelerate CPUs for general purpose scientific and engineering computing. It shifts computation- intensive program segments into GPU while keeping the rest of the program segments, which are serial in nature, on the CPU. This kind of hybrid computing improves the performance of many computer applications.

The GPU computation can be used on financial computations as well. Researchers in the financial world find the benefits of using GPU in financial computation. Examples include computing option pricing, financial derivative modeling, moving averages, random number generation and the Monte-Carlo simulation to predict future stock prices [7][8][9][10][11].

In this report, we focus on the computational aspect of rate discrepancies of exchange rates among different currencies. Specifically, we focus on rate changes within the range of seconds. Although we don't expect there to be a large amount of discrepancies among exchange rates, we would like see if we can find out that in the time range of seconds, exchange rate inconsistency exists.

We organize our report in the following manner: section two discusses the organization of forex data and programs; section three compares the experimental results; section four concludes the report with some remarks.

2. SYSTEM IMPLEMENTATION

We collect ForEx tick data from Internet. There are total of ten currencies that are under consideration: AUD, CAD, CHF, EUR, GBP, JPY, NZD, TRY, GSD, and USD. These data were preprocessed and extracted to form a single data file, which contains date and time, bid and ask prices, and base and quote currency IDs. We assign each currency an integer ID to facilitate the program processing.

There are two programs involved. One is serial and the other parallel counterpart.

2.1 Serial implementation

The serial program takes each individual tick data and updates the data structure. Each tick only involves two pieces of data, one for base currency and the other for quote currency. As soon as the data was processed, a signal will be raised if there is an arbitrage opportunity that occurred when the data structure was updated. The information will then be output onto a data structure that can be output to a file at the end of this simulation. Note that for each tick data process, the pair of data only affects a portion of the data structure.

We represent the exchange rates as a multiplier, as shown in Figure 1. For example, to compute the result of converting currency A to B, B to C, then from C back to A, we will compute M_{AB} * M_{BC} * M_{CA}, as shown in Figure 1. Note that we ignore all the transaction costs which were involved in the currency exchange processes.
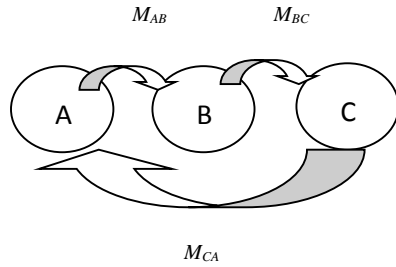
Figure 1 Compute the possibility of currency swapping arbitrage

2.2 Parallel arbitrage Algorithm

Because currency swapping has data dependency due to the fact that tick data are in time series, both the time sequence nature and the current tick data will affect and change the data structure. Therefore, in order to speed up the process, we spread the work which is supposed to be performed in each tick to as many threads as possible. Based on this observation, we assign each thread to process one operation which will check possibility of arbitrage. The next tick data can be accessed by all devices and will be put into shared memory by a particular device with a mapped device ID.

Because only ten currencies were considered and since we only concerned about triangular currency swapping, it can be proved that there are at most twenty operations needed to be considered each time there is a tick. We don't have to consider all 1000 operations because many of them remain the same.

Assuming a tick data contains a base currency b and quote currency q, with bid and ask prices given, we only need to consider the operation of

M_{bi} * M_{iq} * M_{qb}, for $0 < i < 10$. Similarly, we have to perform M_{qi} * M_{ib} * M_{bq}., for $0 < i < 10$. Since there are 1000 possible permutations, we will allocate 1000 threads for our application. A thread with *tid* of b_{tq} will process M_{bt} * M_{tq} * M_{qb} each time base currency b, quote currency q (or reversed when base currency q and quote currency b tick data were received). For example, if the *tid* of a thread is 983, then when base currency $b = 9$, quote currency $q = 3$ were received, it will compute the value of M_{98} * M_{83} * M_{39}. Because this assignment is unique, there is no conflict among threads. This assignment will reduce the 20 fold host operations to one operation in each thread.

3. EXPERIMENTAL RESULTS

Without considering transaction overhead of currency swapping, we might get 2% incidents of possible arbitrage. But these transactions must be carried out within seconds. It was observed that when the overhead fee of transaction rose to 1%, for example, then the percentage of arbitrage opportunity by swapping currencies was reduced to 0, i.e., it is impossible to have financial gain by swapping currency in very short periods time.

Consider 3091 ticks; there are 89 possible arbitrages when zero overhead is considered. When there is 1% overhead, the opportunity reduces to 0. The running time to process the tick data in the host computer took an average of 22 ms, while the parallel counterpart only achieves 17 ms with speed up less than 2.

The speedup factor of using GPU in this application is therefore minimal. There is no significant gain of using parallel implementation to compute currency swapping arbitrage. One of the reason is the syncthreads() operation which is used to avoid the so call read-modify-write race condition constraint. The other reason is due to the processing speed factor of CPU and GPU.

We expect that if more currencies are involved in the computation instead of just the triangular swapping as discussed in this report, the parallel implementation will give us better speed up results. Also note that in this report, we only used ten currencies; if more currencies are involved in

the parallel computation, our implementation shall receive better performance compared with the serial implementation.

4. CONCLUSION AND REMARKS

The experiments were conducted on an Intel i7 with Nvidia GeForce 550M. We focus mainly on computational aspect of short term currency swapping. It is possible to come up with arbitrage if there is no transaction fee involved and spreads are not considered. We found that it is unlikely to have financial gain when there are transaction fees and other fees, such as a regulatory fee, required. In our experiment, we found that even with a 1% overhead, it will wipe out all opportunities of high frequency arbitrage. One thing we did not do is predict the trend of various exchange rates among all these currencies. It would be interesting to find out if more than these currencies are involved. However, we think it is somewhat unlikely to gain when more than three currencies are involved. It would also be interesting to use the historical data to predict the trends and find out the future arbitrage opportunities.

5. REFERENCES

[1] Andrei Shleifer and Robert W. Vishny, "The limit of Arbitrage", The Journal of Finance, vol. LII, No 1, pp 35-55

[2] Nilufer Usmen , "Currency Swaps, Financial Arbitrage, and Default Risk ", Financial Management, Vol. 23, No. 2, Summer 1994, pp 43-57

[3] Y. S. Park, "Currency Swaps as a Long-Term International Financing Technique", Journal of International Business Studies, Vol. 15, No. 3, Winter, 1984, pp. 47-54

[4] Roman Kozhan, Wing Wah Tham, "Execution Risk in High-Frequency Arbitrage", Management Science, Vol. 58, No. 11 (November 2012), pp. 2131-2149

[5] Gautam Goswami, , Jouahn Nam, Milind M Shrikhande, "Why do global firms use currency swaps?: Theory and evidence",

Journal of Multinational Financial Management, Vol 14, Issues 4–5, October–December 2004, pp 315–334

[6] Gautam Goswami and Milind M. Shrikhande, "Economic exposure and currency swaps", Journal of Applied Finance 2007, Vol. 17, no. 2, pp 1-9

[7] Myungho Lee, Chin Hong Chun, and Sugwon Hong, " Financial Derivatives Modeling Using GPU's", International Conference on Scalable Computing and Communications; The Eighth International Conference on Embedded Computing, pp 440 - 445

[8] Myungho Lee, Jin-hong Jeon*, Joonsuk Kim, and Joonhyun Song, " Scalable and Parallel Implementation of a Financial Application on a GPU: with focus on out-of-core case", 2010 10th IEEE International Conference on Computer and Information Technology, pp 1323 - 1327

[9] Ying Peng, Bin Gong, Hui Liu, and Bin Dai, "Option Pricing on the GPU with Backward Stochastic Differential Equation", 2011 Fourth International Symposium on Parallel Architectures, Algorithms and Programming, pp 19 - 23

[10] Steven Solomon, Ruppa K. Thulasiram and Parimala Thulasiraman, " Option Pricing on the GPU", 2010 12th IEEE International Conference on High Performance Computing and Communications, pp 289 - 296

[11] John Jenq, "Parallel Implementation of Moving Averages and Stock Market Prediction". *The Proceedings of the 2012 International Conference on Parallel and Distributed Processing Techniques and Applications*, pp 833-837.

[12] John A. Doukas, Chansog (Francis) Kim, Christos Pantzalis, "Arbitrage Risk and Stock Mispricing", The Journal of Financial and Quantitative Analysis, Vol. 45, No. 4 (AUGUST 2010), pp. 907-934

New CA based image encryption-compression scheme using wavelet transform

Bala Suyambu Jeyaram

Department of Mathematics, IIT Madras
Chennai, Tamil Nadu-600036, India

and

Rama Raghavan

Department of Mathematics, IIT Madras
Chennai, Tamil Nadu-600036, India

ABSTRACT

The wide use of digital images leads to the necessity of securing them when they enter into an insecure channel. Image cryptography plays a vital role in the modern communication. In this paper we propose a new image encryption compression scheme, which will do both, lossy image compression and encryption. Cellular automata is used for key generation and wavelet transformation is used for image compression. Encryption has been done in two steps: one before compression and another one after the compression. Performance evaluation results clearly show that the proposed method is better in all aspects.

Keywords:

Elementary cellular automata (ECA), wavelet transformation, image compression, image encryption and decryption.

1. Introduction

Nowadays images are widely used in the communication networks. Even though the computer network has been largely developed, this development is not achieved fully without passive and active attacks. Secure ways of storing and transmitting the images have become the need of the hour. Image cryptography helps to achieve these requirements. The randomness in the encryption method and the size of the encrypted image can be the important parameters as far as the strength of the algorithm and the speed of transmission are concerned. The method of generating pure random numbers is always an open problem. Because of this, researchers are getting motivated to develop new methods to generate good random numbers. Cellular automata(CA) with its simple implementation nature, unpredictability, parallelism and homogeneity are found as very good tools for generating random numbers. Starting from Wolframe [1, 2] many researchers have worked in CA based cryptography. Due to high data redundancy and capacity, images have to be compressed before transmission to increase the speed and reduce the storage place. So the study on image compression

has been increased tremendously. There are two types of image compression: lossy and lossless. In lossless compression, we can recover the original image, whereas in the lossy compression, we can recover only similar image of the original image, in which some data would be lost. Each compression type has its own field of applications. In [3, 4, 5, 6, 7], some techniques on image compression have been proposed with good performance on compression. But these methods cannot take care of secrecy and integrity of the travelling images in the communication channel. In the recent years, it is essential to carry out both encryption and compression of the images for secure and fast transmission. A number of image encryption and compression schemes have been proposed [8, 9, 10]. Wavelets [11, 12, 13] are used in signal and image processing as well as in its analysis applications. Several researchers have worked on applications of wavelets in image cryptography[14, 15]. In [16], we have proposed an image encryption scheme for RGB color images using ECA for key generation and $GF(2^8)$ for encryption and decryption. In [16], we have not dealt with image compression and the pixel intensity distribution of the encrypted images are not following uniform distribution. In this paper we have proposed a new encryption step along with the encryption step in [16], which fairly produces the uniform distribution in the pixel intensities of the encrypted images. Lossy Image compression is carried out in this paper using wavelet transformations. This paper is organized as follows. Section 2 describes the basics of Cellular Automata; section 3 presents some basics of wavelet transformations; Section 4 describes the proposed encryption scheme; In section 5 we present some experimental results and security analyses are given in section 6; Concluding remarks are given in section 7.

2. Cellular Automata

A Cellular Automaton is a mathematical model of a system with discrete inputs and outputs. A cellular automaton (CA) is a finite state machine with infinite, regular lattices that change the states synchronously, according to a local rule. Binary state automaton takes only the states 0 and 1. Elementary Cellular Automata is a

one dimensional, binary state CA that uses the nearest neighbors to determine their next state. If a neighborhood has 3 states then there are 2^8=256 elementary CA. The cell i is denoted by (i) and the state of the cell (i) at time t is denoted as S_i^t. The neighborhood of radius r is defined for each cell (i) is defined as $N(i) = ((i - r), .., (i-1), (i), (i+1), .., (i+r))$. The state S_i^{t+1} of the cell i at time $t + 1$ depends only on states of its neighborhood at time t i.e., $S_i^{t+1} = f(N(i))$ where f is the transition function, called a rule. When there are n number of states in a neighborhood, the number of rules can be expressed as 2^{2^n}. CA's are classified into two ways in terms of the number of rules used to update the cells. If the same rule is used to update the cells, then the CA is called uniform, in contrast if different rules are used to update the cells, then the CA is called non-uniform. An evolution of rule 30 is given in Figure 1.

1D Elementary Cellular Automata

Neighbourhood radius r=1, rule $(00011110)_2 = (30)_{10}$

Neighbourhood

| 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | t = 0 |

| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | t = 1 |
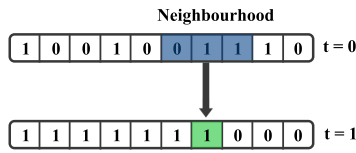
Number	7	6	5	4	3	2	1	0
Neighbourhood	111	110	101	100	011	010	001	000
Rule Value	0	0	0	1	1	1	1	0

Fig. 1.   Example of an evolution of Rule 30

3. Wavelet Transformation

The wavelet transformation W(s,t) of a signal f(x) in terms of an orthonormal basis formed by a mother wavelet family is defined as, $W(s,t) = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{s}} \psi^*(\frac{x-t}{s}) dx; s > 0$, where t corresponds to the translation of the mother wavelet function and s corresponds to the scaling of the mother wavelet. When digital images are to be viewed or processed at multiple resolutions, the discrete wavelet transformation (DWT) is the mathematical tool of choice. Discrete Fourier Transform defined by two straightforward equations that resolve around a single pair of transformation kernals, where as the Discrete Wavelet Transform refers to a class of transformations that differ not only by transformation kernels but also by the nature of these functions and in the way they are applied. We cannot write a single equation for a DWT that completely describes all since it encompasses a variety of unique but related transformations. A transform kernel pair as set of parameters that defines the pair can characterize DWT. The various transforms are related by the fact that their expression functions are small waves of varying frequency and limited duration. The kernel can be represented as these separable 2-D wavelets.
$\psi^H(x,y) = \psi(x)\phi(y)$
$\psi^V(x,y) = \phi(x)\psi(y)$
$\psi^D(x,y) = \psi(x)\psi(y)$, where $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$

are called horizontal, vertical and diagonal wavelets, respectively and one separable 2-D scalling function $\phi(x,y) = \phi(x)\phi(y)$, Each of these 2-D functions is the product of two 1-D real, square integrable scaling and wavelet functions $\phi_{i,j}(x) = 2^{j/2}\phi(2^j x - k)$ and $\psi_{i,j}(x) = 2^{j/2}\psi(2^j x - k)$, where the translation parameter k determines the position of these 1-D functions along the X-axis, the scaling parameter j determines their width-how broad or narrow they are along X and $2^{j/2}$ controls their height or amplitude. The associated expression functions are binary scaling and integer translates of the mother wavelet $\psi(x) = \psi_{0,0}(x)$ and scaling function $\phi(x) = \phi_{0,0}(x)$. The wavelet decomposition of an image is used to analyze the image in low and high frequencies with different resolutions. This information can be used to compress the images.Figure 2 shows the first level approximation,horizontal,vertical and diagonal details of lena image.
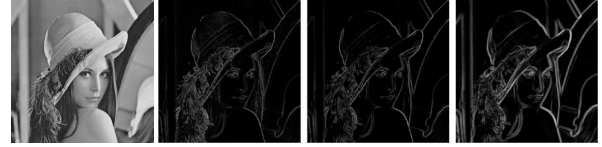
Fig. 2.   First level decomposition of lena image

4. Proposed Image Encryption Scheme

In [16] we proposed an encryption method for RGB color images as well as gray scale images. Algorithm 1 given below is proposed in [16].Let P be a gray scale image of size MxN, and Q be the encrypted image encrypted by Algorithm 1 [16].
Algorithm 1:
Input:
1. Plain Image P of size MxN.
2. Key Image K of size MxN.
Output: Encrypted Image Q of size MxN.

Step 1: Input the plain image P to the algorithm.
Step 2: Input the Key Image K.
Step 3: Change the values of the pixels in P to 1 wherever the value is 0.
Step 4: Consider each pixel values of K and P as elements in $GF(2^8)$. Perform the multiplication operation in $GF(2^8)$ between the corresponding elements in P and in the key image K to get the encrypted image Q.
$Q(i,j) = P(i,j) \odot K(i,j)$, where $\odot$ is the element wise multiplication in $GF(2^8)$.
Construct the matrix K_{in} from the matrix K by replacing the elements of K with its multiplicative inverse in $GF(2^8)$.This matrix K_{in} is used as the inverse key image for the decryption in the receiver side. So the original image can be found using the following operations.
$P(i,j) = C(i,j) \odot K_{in}(i,j)$, where $\odot$ is the element wise multiplication in $GF(2^8)$.
Now this Q is considered as the input of the proposed method. Algorithm 2 describes the over all encryption scheme of the proposed method.
Algorithm 2:
Input:
1. Encrypted image Q of size MxN (Output from Algorithm 1)
2. Key image K2 of size M/2 x N/2.

Output: Encrypted image C of size M/2 x N/2.

Step 1: Input the image Q of size MxN.
Step 2: Fix the wavelet for wavelet transformation.
Step 3: Decompose Q using wavelet transformation and get the approximation coefficient matrix QA of Q.
Step 4: Generate the key image K2 of size M/2 x N/2 using ECA as in [16]
Step 5: Do element wise Exclusive OR operation between QA and K2.
Step 6: The resultant M/2xN/2 matrix is the encrypted image C.
Encrypt the plain image P using the key K1 of size MxN as in [16] and get the intermediate encrypted image Q. Decompose Q using wavelet transformation and get C and S. Extract the approximation coefficient matrix QA of Q. Get the encrypted image C using the key K2 as follows:
$C(i,j) = QA(i,j) \oplus K2(i,j) \ \forall \ 1 \leq i \leq M/2$ and $\forall \ 1 \leq i \leq N/2$, In the receiver side the decrypted image can be got as follows:
$QA(i,j) = C(i,j) \oplus K2(i,j) \ \forall \ 1 \leq i \leq M/2$ and $\forall \ 1 \leq i \leq N/2$, where $\oplus$ is the element wise Exclusive OR operation.
After getting QA, using C and S by applying wavelet reconstruction method we will get back Q. From Q by applying the decryption process described in [16], we will get back the original image P.

5. Simulation Results

Different Gray scale images are used to test the performance of the proposed method. In this paper we have given the test results on Lena images. Figures in 3(a),(b) and (c) show the original, compressed and encrypted as well as decrypted images of the proposed method. This clearly shows the definitive difference between the original and the compressed encrypted images. Figures in 3 (d) and (e) show the encrypted images of the Lena by [16] and by the proposed method. When compare to the existing method, the newly proposed method gives not only a perfectly encrypted image but in addition causes compression of the image which helps in increasing the transmission speed.
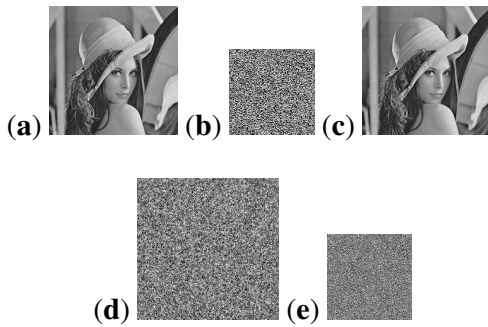
Fig. 3. (a) Original Lena image (b) Encrypted lena image (c)Decrypted lena image (d)Encrypted lena image by existing method (e)Encrypted lena image by proposed method

6. Security Analysis

In this section we carry out the performance analysis of the proposed scheme. We provide: key space analysis, histogram analysis, correlation analysis, key sensitivity analysis. All the experiments are performed on a personal computer with an 3.1 GHz Intel Core Quad Core i5 Processor, 4 GB RAM and 1 TB hard disc with apple i mac operating system.

Key Space Analysis

The key space of a strong encryption algorithm should be very large to make the brute force attack ineffective. Since the proposed method is the extension of the existing method [16], the key space includes the space for both the methods. The key space size of the existing method for a 256x256 gray scale image is 2^{2066} when we fix the third parameter as 1024. The seed to generate M/2xN/2 key image for the proposed method can take 2^{8N} possible values and the wavelet decomposition vector C as well as the corresponding book keeping matrix S. Excluding C and S, the over all key space size of the newly proposed method for a 256x256 gray scale image is $2^{2066} \mathrm{x} 2^{2048} = 2^{4114}$, which is considerably very large to make the brute force attack very hard. This is very larger than the existing system's key space size.

Histogram Analysis

Histogram is the gray value distribution graph, which shows the pixel distribution of an image. If each gray value of the encrypted image has equal probability, then the encryption method is more robust against statistical attack and differential attack. This can be easily checked by the histogram of the encrypted image. That means the histogram of an encrypted image should be uniform for a well encrypted image. Figures 4(a) represent the original Lena image and its histogram chart respy. Figures in 4(b) represent the encrypted Lena of the existing system and its histogram chart respectively. Figures in 4(c) represent the encrypted Lena of the proposed method and its histogram chart respy. It is clear from the figures that the gray value of the encryption image is fairly uniform and significantly different from the gray value distribution of the plain image. And they show that the gray value distribution of the newly proposed method is more fairly uniform than the gray value distribution of the existing method. So this will not provide any hint to perform statistical and differential attacks on the proposed image encryption and compression scheme.

Key Sensitivity Analysis

This section shows the key sensitivity of both encryption key as well as the decryption key. Figure 5 (b) is the encrypted lena image with key K1 and (c) is the encrypted Lena image with key K2. Figure 5(d) shows that the difference between (b) and (c). This clearly shows that the wrong encryption key leads to entirely different image with good difference. Next we encrypted the original Lena image with the encryption key K1 and decrypted with another key K2, which fails to get back the original image. Figures 6(a),(b),(c) show this analysis. From this we can conclude that the proposed scheme is sensitive to both encryption as well as decryption keys.

Correlation Analysis

Plain image and imperfectly encrypted images have correlation between the adjacent pixels, which makes statistical attack possible. We have tested the correlation between horizontally adjacent pixels, vertically adjacent pixels and diagonally adjacent pixels, to show the diffusion effect of our scheme. We have randomly selected 2000 pairs of adjacent pixels from the plain image and the encrypted image. We have plotted the distribution graph of two horizontally, vertically and diagonally adjacent pixels of the plain image and encrypted image in figure 4. This figure

(a)

(b)

(c)
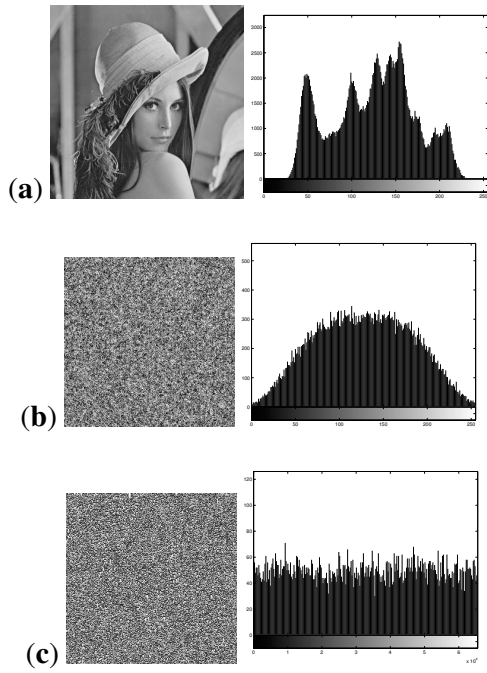
Fig. 4. (a) Original Lena image and its Histogram. (b) Encrypted Lena image and its Histogram in existing system. (c) Encrypted Lena image and its Histogram in proposed system.

(a) **(b)**

(c) **(d)**
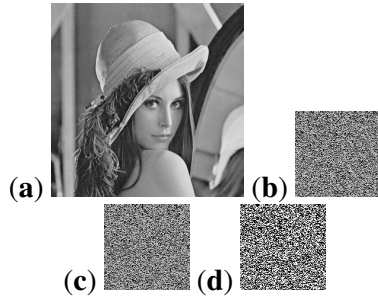
Fig. 5. (a) Original Lena Image. (b) Encrypted image of (a) with key K1. (c) Encrypted image of (a) with key K2. (d) Image difference between (b) and (c).

(a) **(b)**

(c)

Fig. 6. (a) Original Lena image. (b) Encrypted image of (a) with key K. (c) Decrypted image of (b) with key K3.

Original   **Encrypted**

(a)

(b)

(c)
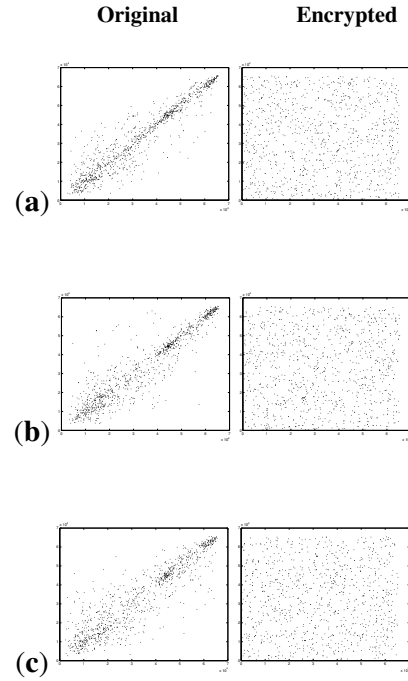
Fig. 7. Correlation Distribution of the pairs of adjacent pixels: (a)Horizontal. (b) Vertical. (c) Diagonal

clearly shows that the pixels in the encrypted image are highly uncorrelated. We have also calculated the correlation coefficient of adjacent pixels of the plain image and the encrypted image by the Eqs. (1),(2),(3) and (4). Table 1 shows that the correlation coefficients of two horizontally adjacent pixels, vertically adjacent pixels and diagonally adjacent pixels of original lena image and encrypted image.This confirms that the adjacent pixels in the plain images are strongly correlated where as the adjacent pixels in the encrypted images are weakly correlated. Figure 7 illustrates the correlation distribution of the horizontal adjacent pixels, vertical adjacent pixels and diagonal adjacent pixels of the plain and the corresponding encrypted images using the proposed method. This shows that the encrypted images are very weakly correlated.

$$E(x) = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad (1)$$

$$D(x) = \frac{1}{N} \sum_{i=1}^{N} ((x_i - E(x))^2 \qquad (2)$$

$$cov(x,y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - E(x))(y_i - E(y)) \qquad (3)$$

$$\gamma_{xy} = \frac{cov(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \qquad (4)$$

Table 1: Correlation coefficients for adjacent pixels between original and encrypted images.

	Horizontal	Vertical	Diagonal
Original Lena Image	0.9385	0.9313	0.8965
Encrypted Lena Image	-0.0216	0.0273	0.0154

7. Conclusion

In this paper, we have proposed a new image encryption compression scheme with Cellular Automata and Wavelet transformation. In this scheme we have used ECA to generate good random key image and wavelet transformation to carry out the compression. These two tools together have made a strong encryption-compression scheme. We have analyzed the proposed scheme by carry out different performance analyses tests to demonstrate the security and strength of the proposed method. The compression we have performed here using wavelet transformation is a lossy compression. Our next work is to design a new image cryptosystem for lossless image compression with better randomness in the encryption.

1. REFERENCES

[1] S. Wolfram, Cryptography with Cellular Automata, in advances in cryptology: Crypto '85 proceedings, Lecture notes in Computer Science, vol. 218. Springer; 1986 p.429-32.

[2] S. Wolfram, Theory and Applications of Cellular Automata, Advanced series on complex systems-Volume 1.

[3] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, 2nd Edition, Pearson Education 2004.

[4] Bryan Usevitch, "A Tutorial on Modern Lossy Wavelet Image Compression: Foundation of JPEG 2000", IEEE Signal Processing Magazine, 2001.

[5] DONOHO D. Compressed sensing [J], IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.

[6] S. Singh, V. Kumar, H. K. Verma, " DWT-DCT hybrid scheme for medical image compression", Journal of Medical Engineering and Technology, Volume 31, Issue 2, pp.109-122, March 2007.

[7] Macarena Boix, Begona Canto, Wavelet Transform application to the compression of images, Elsevier, Mathematical and Computer Modelling 52 (2010), 1265-1270.

[8] Debayan Goswami, Naushad Rahman, Jayanta Biswas, Anshu Koul, Rigya Lama Tamang, Dr. A. K. Bhattacharjee, "A Discrete Wavelet Transform based Cryptographic Algorithm", International Journal of Computer Science and Network Security, Volume 11, Number 4, April 2011.

[9] Ch.Samson, V. U. K. Sastry, A Novel Image Encryption Supported by Compression Using Multilevel Wavelet Transform, International Journal of Advansed Computer Science and Applications, Volume 3, Number 9, 2012.

[10] Xiping He, Qionghua Zhang, Image Encryption Based on Chaotic Modulation of Wavelet Coefficients, 2008 Congress on Image and Signal Processing, May 2008, Sanya, Hainan, China.

[11] Robi Polikar, The Wavelet Tutorial, "http://users.rowan.edu/ polikar/WAVELETS/WTtutorial.html"

[12] K. P. Soman, K. I. Ramachandran, Insight into Wavelets from theory to practice, Second Edition, PHI, 2006.

[13] Jatan K. Modi, Sachin P. Nanavati, Amit S. Phadke, Prasanta K. Panigrahi, "Wavelet Transforms- Application to Data Analysis-I", Resonance, Nov.2004.

[14] Zhu Yu Zhou Zhe Yang Haibing Pan Wenjie Zhang Yunpeng, A Chaos-Based Image Encryption Algorithm Using Wavelet Transform, ICACC 2010 IEEE,pp.217-222.

[15] Juan M Vilardy, J. Useche, C. O. Torres and L. Mattos, Image encryption using the fractional wavelet transform, Journal of Physics: Conference Series 274 (2011) 012047.

[16] Bala Suyambu Jeyaram, Rama Raghavan, Krishna Shankara Narayanan, New CA based key generation for a robust RGB color image encryption scheme, International Journal of Computer Applications (0975-8887), Volume 80, Number 7, October-2013.

Neurofuzzy Decision Making Support Model for
Stock Trading based on Strategic Trading Signals and Technical Analysis Approach

M.L. Kulthon KASEMSAN

and

Monruthai RADEEROM
Faculty of Information Technology, Rangsit University,
Pathumthani 12000, Thailand

ABSTRACT

In this paper, a Neurofuzzy (NFs) adopted on a Takagi-Sugeno-Kang (TSK) type model was created, and its parameters were tuned by neural network methods. The proposed model uses a unique combination of technical analysis (TA) and rule-based expertise to study the behavior of a stock price's movement, and volume patterns such as indicators and oscillators as the input variables. The model involves an initial prediction of next day close price which adapts to up and down fluctuations, and also applies a strategic trading signals approach to support decision making in stock trading. The combination of all of this information input into this unique model results in better trading decision support, and in the end is more profitable for the investor. The capability of the proposed approach was tested by the Bangchak Petroleum Public Co., LTD (BCP) stock price from the Stock Exchange of Thailand (SET). The input variables group showed possible profits reaching 84.21% in May 2010 and 94.44% in July 2010. These results are very encouraging, and they can be implemented and provide investors with highly accurate decision support to increase the efficiency of stock trading.

Keywords: Intelligence System, Fuzzy Logic, Neuro-Fuzzy System, Stock Index, Decision Making System.

1. INTRODUCTION

The prediction of financial market indicators is a topic of considerably practical interest and, if successful, may involve substantial pecuniary rewards. People tend to invest in equity because of its high returns over time. Considerable efforts have been put into the investigation of stock markets. The main objective of the researchers is to create a tool, which could be used for the prediction of stock markets fluctuations; the main motivation for this is financial gain. In the financial marketplace, traders have to be fast and hence the need for powerful tools in order to efficiently and profitably make important decision.

The use of Artificial Intelligence (AI) had a big influence on the forecasting and investment decision-making technologies. There are a number of examples using neural networks in equity market applications, which include forecasting the value of a stock index [4,5], recognition of patterns in trading charts[6] rating of corporate bonds[8], estimation of the market price of options[7], and the indication of trading signals of selling and buying[3,8].

Even though most people agree on the complex and nonlinear nature of economic systems, there is skepticism as to whether new approaches to nonlinear modeling, such as neural networks, can improve economic and financial forecasts. Some researchers claim that neural networks may not offer any major improvement over conventional linear forecasting approaches [7]. In addition, there is a great variety of neural computing paradigms, involving various architectures, learning rates, etc., and hence, precise and informative comparisons may be difficult to make. In recent years, an increasing amount of research in the emerging and promising field of financial engineering has been incorporating Neurofuzzy approaches [9]. Almost all models are focused on the prediction of stock prices. The difference of our proposed model is that we are focusing on decision-making in stock markets, but not on forecasting in stock markets.

In contrast to our previous work [9], we are not making a direct prediction of stock markets, but we are working on a one-day forward decision-making tool for buying/selling stocks. We are developing a decision-making model which, besides the application of Fuzzy Logic and Neurofuzzy system (NFs), At first, Our proposed Trading Strategy based on fuzzy logic captured knowledge from experts who are making decision buy, hold and sell decisions based on technical analysis as well as input for our proposed trading system based on NFs. Moreover, optimization algorithms are based on the rate of the return profit of each stock index constructed from our NFs Model. In this paper, we present a decision-making model that combines a technical model and a NFs model. The technical analysis model evaluated knowledge about buy, hold and sell strategy from each technique. Our proposed model used result from technical analysis model to input of our NFs. The NFs trading system decides the buy, sell and hold strategy for each stock index. The objective of this model is to analyze the daily stock and to make one day forward decisions related to the purchase of stocks.

The paper is organized as follows: Section 2 presents the background about the neural network and the Neurofuzzy system. Section 3 presents the NFs decision-making model; Sections 4 is devoted to experimental investigations and the evaluation of the decision-making model. This section provides the basis for the selection of different variables used in the model, and models the structure. The main conclusions of the work are presented in Section 5, with remarks on future.

2. NEURAL NETWORK AND NEUROFUZZY APPROACHES FOR THE INTELLIGENCE PORTFOLIO MANAGEMENT SYSTEM S

Both neural networks and the fuzzy system imitate human reasoning process. In fuzzy systems, relationships are represented explicitly in forms of if-then rules. In neural networks, the relations are not explicitly given, but

are coded in designed networks and parameters. Neurofuzzy systems combine the semantic transparency of rule-based fuzzy systems with the learning capability of neural networks. Depending on the structure of if-then rules, two main types of fuzzy models are distinguished as mamdani (or linguistic) and takagi-sugeno models [1]. The mamdani model is typically used in knowledge-based (expert) systems, while the takagi-sugeno model is used in data-driven systems

In this paper, we consider only the Takagi - Sugeno-Kang (TSK) model. Takagi, Sugeno and Kang [1] formalized a systematic approach for generating fuzzy rules from an input-output data pairs. The fuzzy if-then rules, for the pure fuzzy inference system, are of the following form:

$$if\ x_1\ is\ A_1\ and\ x_2\ is\ A_2\ and\ x_N\ is\ A_N\ then\ y = f(x) \quad (1)$$

Where $x = [x_1, x_2, ..., x_N]^T$, $A_1, A_2, ..., A_N$ fuzzy sets are in the antecedent, while y is a crisp function in the consequent part. The function is a polynomial function of input variables $x_1, x_2, x_3, ..., x_N$. The aggregated values of the membership function for the vector are assumed either in a form of the MIN operator or in the product form. The M fuzzy rules in the form Eq. (4) are N membership functions $\mu_1, \mu_2, \mu_3, ..., \mu_N$. Each antecedent is followed by the consequent:

$$y_i = p_{i0} + \sum_{j=1}^{N} p_{ij} x_j \quad (2)$$

Where p_{ij} are the adjustable coefficients, for

$i = 1,2,3,...,M$ and $j = 1,2,3,...,N$.

The first-order TSK fuzzy model could be expressed in a similar fashion. Consider an example with two rules:

$$if\ x_1\ is\ A_{11}\ and\ x_2\ is\ A_{21}\ and\ then\ y_1 = p_{11}x_1 + p_{12}x_2 + p_{10}$$

$$if\ x_1\ is\ A_{12}\ and\ x_2\ is\ A_{22}\ and\ then\ y_2 = p_{21}x_1 + p_{22}x_2 + p_{20}$$

Figure 2 shows a network representation of those two rules. The nodes in the first layer compute the membership degree of the inputs in the antecedent fuzzy sets. The product node $\prod$ in the second layer represent the antecedent connective (here the "and" operator). The normalization node N and the summation node $\sum$ realize the fuzzy-mean operator for which the corresponding network is given in Figure 1 Applying fuzzy singleton, a generalized bell function such as membership function and algebraic product aggregation of input variables, at the existence of M rules the Neurofuzzy TSK system output signal upon excitation by the vector, are described by
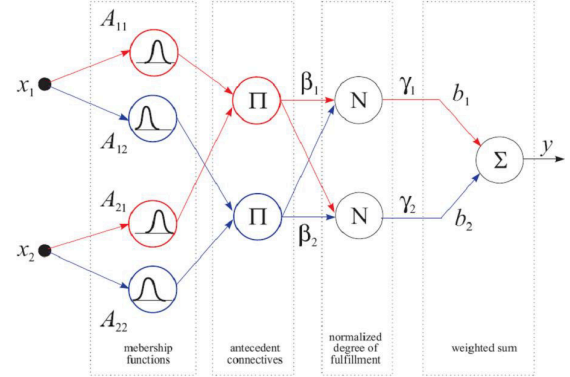
Fig.1 An example of a first-order TSK fuzzy model with two rules systems [1]

$$y(x) = \frac{1}{\sum_{r=1}^{M} [\prod_{j=1}^{N} \mu_r(x_j)]} \times$$
$$\sum_{k=1}^{M} \left(\left[\prod_{j=1}^{N} \mu_r(x_j) \right] \left[p_{k0} + \sum_{j=1}^{N} p_{kj} x_j \right] \right) \quad (3)$$

The adjusted parameters of the system are nonlinear parameters of bell function ($c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$), the fuzzier functions and linear parameters (weight) of the TSK function for every $j = 1,2,...,N$ and $k = 1,2,...,M$. In contrast to the mamdani fuzzy inference system, the TSK model generates crisp output values instead of fuzzy ones. This network is simplified. Thus, the defuzzifier is not necessary. So, the learning of Neurofuzzy network, which adapts parameters of the bell shape membership functions ($c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$) and consequent coefficients,

p_{ij} can be done either in supervised or self-organizing modes. In this study, we apply a hybrid method which is one-shot least-squares estimation of consequent parameters with iterative gradient-based optimization of membership functions. The important problem in the TSK network is to determine the number of rules that should be used in modeling data. More rules mean better representation of data processing, but increased complexity of the network and a high cost of data processing. Therefore, a procedure for automatically determining the number of rules is required. In our solution, each rule should be associated with one cluster of data. Fuzzy c-means is a supervised algorithm, because it is necessary to indicate how many clusters C looks for. If C is not known beforehand, it is necessary to apply an unsupervised algorithm. Subtractive clustering is based on a measure of the density of data points in the feature space [1]. The idea is to find regions in the feature space with high densities of data points. The point with the highest number of neighbors is selected as the center for a cluster. The data points within a prespecified data, fuzzy radius are then removed (subtracted), and the algorithm looks for a new point having the highest

number of neighbors. This process continues until all data points are examined.

In conclusion, figure 2 summarizes the Neurofuzzy Networks System (NFs). Construction process data called "training data sets," can be used to construct Neurofuzzy systems. We do not need prior knowledge called "knowledge-based (expert) systems". In this way, the membership functions of input variables are designed by the subtractive clustering method. Fuzzy rules (including the associated parameters) are constructed from scratch by using numerical data. And the parameters of this model (the membership functions, consequent parameters) are then fine-tuned by process data.
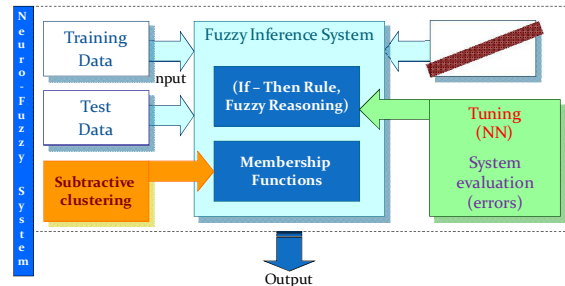
Fig. 2 Constructing Neurofuzzy Networks

3. METHODOLOGY FOR THE INTELLIGENCE PORTFOLIO MANAGEMENT SYSTEM

Table. 1 Description of technical index used in input NFs

Technical Index	Description
Exponential Moving Avarege (EMA)	EMA used to emphasize the direction of a trend and smooth out price
Relative Strength index (RSI)	RSI compares the magnitude of recent gains to recent losses in attempt to determinate overbougth and oversold conditions of an asset
Moving avarege convergence and divergence (MACD)	MACD shows the difference between a fast and slow exponential moving average(EMA) of closing price.
Williams %R (W%R)	W%R is the ratio of the number of rising periods over the total number of periods. It reflects the buying power in relation to the selling power
On-Balance Volume (OBV)	OBV calculates from the stock closing price and volume traded data.
Price rate of change	PROC is rate of Close Price in 12 Time difference
Price and Volume Trend (PVT)	PVT calculated from the stock closing price and volume traded data.
Typical price(TPRC)	TPRC is the average of the high, low, and closing prices for each period.
Volume rate of change (VROC)	VROC calculates from the volume between the current volume and the volume 12 Times ago.

Technical indexes are calculated from the variation of stock price, trading volumes and time according to a set of formulas to reflect the current tendency of the stock price fluctuations. These indexes can be applied for decision making in evaluating the phenomena of oversold or overbought stock. In soft computing, many researches use as inputs technical analysis factors that are sometimes combined with daily or previous stock index price[7]. The

technical analysis factors range from 2 to 25 indicators. For input data, several technical indexes are selected in this paper, as shown in Table 1.

Such as, The Relative Strength Index (RSI) is a momentum oscillator used to compare the magnitude of a stock's recent gains to the magnitude of its recent losses, in order to determine the overbought or oversold conditions. The calculation formula used is

$$RSI = 100 - \frac{100}{1 + \frac{\sum(positivechange)}{\sum(negativechange)}} \qquad (4)$$

Where RS=Average gains/Average losses

Figure 3 shown a RSI index in 4 period's time. After that, we can be evaluating buy, hold and sell signal by experience trader. Sell zone was RSI more than 70 and buy zone was RSI below 30. Thus, Hold zone was between 30 and 70. Sell, buy and hold zone shown Fig 1.
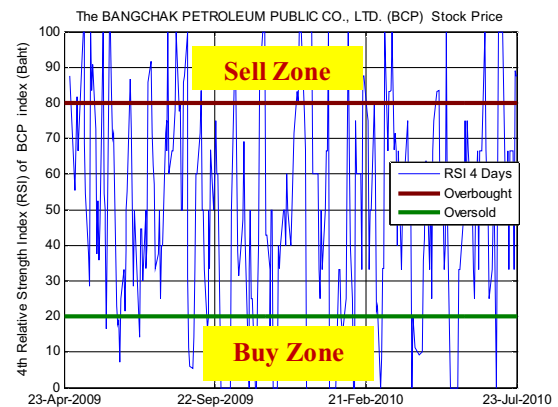
Fig.3 The 4 periods RSI index (RSI4 (t)), calculated close price (t)

This paper attempts to build a collaborative trading model to detect daily trading signal. The system of collaborative trading model developed from technical analysis and computational intelligent approach to learning overall trading signal knowledge (all trading signal include buy-sell and holding point) and dynamic trading threshold knowledge (only buy – sell points). We selected a Neurofuzzy system (NFs), which are now studied and incorporated into the emerging and promising field of financial engineering.

We proposed NFs for the decision making system, called intelligence trading system. See the subsequent steps, shown in figure 4, where the system can be trained and combined with Decision Making System based on Technical Analysis and Finally Decision Making System based on NFs model. The model scenario represents one time calculations made in order to get decisions concerning the purchase of stocks. For this paper, historical data of daily stock returns was used for time

interval. At first step of the model realization, technical analysis techniques are used to the decision strategy recommendation. The recommendations (R) represent the relative rank of investment attraction to each stock in the interval $[-1, 1]$. The values -1, 0, and 1 represent recommendations: Sell, Hold and Buy, respectively.
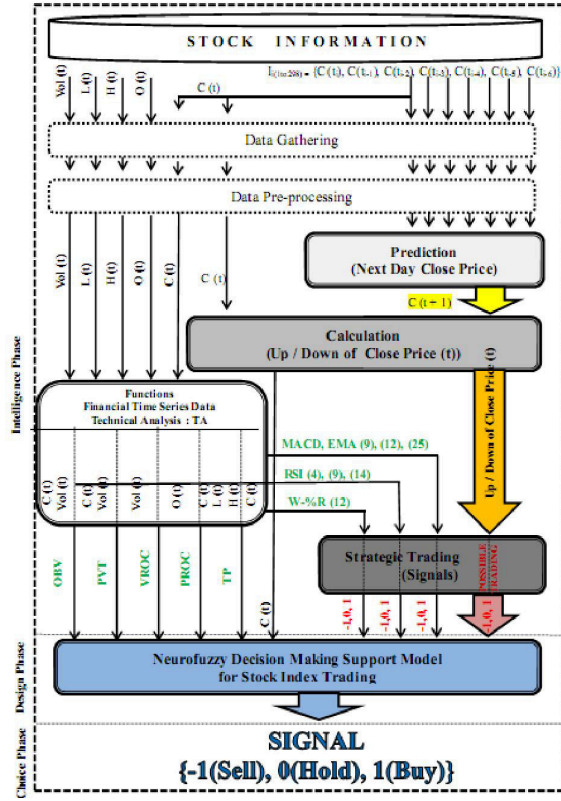
Fig. 4 The scenario of Intelligence Trading System

After that, the recommendations are included in the input of a proposed intelligence system. Output intelligence system is the evaluating recommendation based on possible trading signal, which are defined, as following as

$$if \ cp_t < cp_{t-1} \ and \ cp_t < cp_{t-1} then \ 1 \ (Buy \ signal),$$
$$if \ cp_t > cp_{t-1} \ and \ cp_t > cp_{t-1} then -1 \ (Sell \ signal),$$
$$otherwise, 0 \ (Hold \ signal) \qquad (5)$$

When cp_t is close price in time (t)
           cp_{t-1} is close price in previous time (t-1)
           cp_{t+1} is close price in next time (t+1)

After normalization, input scale is between -1 and 1. The normalization and scaling formula is

$$y = \frac{2x - (\max + \min)}{(\max - \min)}, \qquad (6)$$

Where x the data is before normalizing, y is the data after normalizing.

There are several kinds of error function used in evaluating of approximating method, namely, Mean absolute Deviation (MAD), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). In this paper, like a neural network model, we used two error functions for our NFs system; the Percentile Variance Accounted For (VAF) [1] is selected for evaluating the NFs model. The VAF of two equal signals is 100%. If the signals differ, the VAF is lower. When y1 and y2 are matrices, VAF is calculated for each column. The VAF index is often used to assess the quality of a model by comparing the true output and the output of the model. The VAF between two signals is defined as follows:

$$VAF = 100\% * [1 - \frac{var(y1 - y2)}{var(y1)}] \qquad (7)$$

For NFs trading system, we interested in profitability of the model[5]. The expected returns are calculated for considering profit of stock trading market. That are, the value obtained on the last investigation day is considered the profit. The trader's profit is calculated as

$$Profit(n) = Stock \ Value(n) - Investment \ value \qquad (8)$$

Where n is the number of trading days.
And, the Rate of Return Profit (RoRP) is

$$RoRP = \frac{Profit(n)}{Investment \ value} \times 100 \qquad (9)$$

4. RESULTS AND DISCUSSION
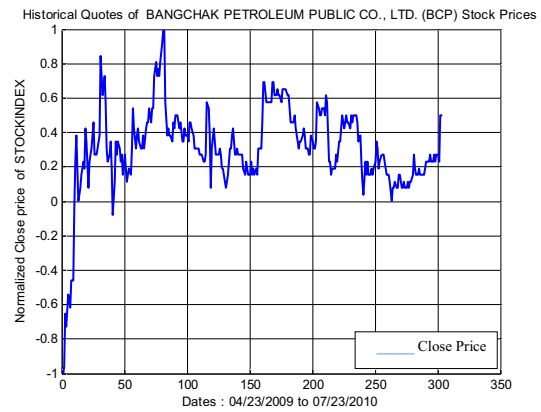
Fig. 5 Historical Quotes of Bangchak Petroleum public Co., Ltd. (BCP) Stock Prices

At the beginning of each realization, the start investment is assumed to be 1,000,000 Baht (Approximately USD 29,412). The data set, including the Stock Exchange of Thailand (SET) index, Historical Quotes of Bangchak Petroleum public Co., Ltd. (BCP) Stock Prices, Siam Commercial Bank (SCB) and Petroleum Authority of Thailand (PTT) stock index, has been divided into two different sets: the training data and test data. The stock index data is from April 23, 2009 to July 23, 2010 totaling 304 records. The first 266 records are training data, and the rest of the data, i.e., 38 records, will be test data. Moreover, the data for stock prices includes the

136

buy-sell strategy, closing price and its technical data. Consequently, max-min normalization can be used to reduce the range of the data set to appropriate values for inputs and output used in the training and testing method.

For input data, several technical indexes are proposed 3 input models as shown in table 2(left), correspondingly. Our 3 models are difference number of inputs and 1 output which is trading signal (sell (-1), hold (0) and buy (1). Model 1 was totally 7 inputs. Model 2 was totally12 inputs. Model 3 was totally12 inputs.

And, Table 3 (right) has shown 3 proposed scenario models, respectively. Model 1 used close price and decision values from technical analysis as input to NFs. Model 2 was difference from Model1 which are including some volume indicators as inputs of NFs. Model 3 was difference from Model 2 which is using NFs for predicting close price on next day as input of Decision Trading NFs. In NFs for predicting Close price, we use our NFs model from previous work[8] to making a direct prediction of stock markets on next day.

After setting up the parameters of the experiments, we compared the proposed NFs decision system with the Buy and Hold, and Neural Network. The step-by-step procedure of the evaluating our proposed system is explained in detail in the following

Step 1 Calculating technical indicator
Step 2 Evaluating buy and sell of some technical
      indicator depend on any model
Step 3 Normalized variables input to any models for NFs
      trading system
Step 4 Training NFs trading system, for one day, we
      are selecting 100 days for sliding window.
Step 5 Evaluating Trained system by simulating buy and
      sell
Step 6 Calculating Rate of Return and Profit
Step 7 Repeating 5 to 6 for Buy & Hold and 4 to 6 for
      Neural Network
Step 8 Repeating 4 – 7 for testing data that means
      training process, which is repeated for one testing
      data

In step 5 – 8, after developing the intelligence trading system, we were given 1,000,000 baht for investment at the beginning of the testing period. The decision for to buy and sell stocks is given by proposed intelligence output. We translated the produced RoRP results that verify the effectiveness of the trading system. Table 3 is show a Financial Simulation Model for calculating profit in our trading strategy. For example of experimental result, results of Model 3 within training days and testing day are shown in figure 6 and 7, respectively.

Moreover, our proposed decision-making NFs model compared RoRP performance with Buy & Hold Strategy and NNs. The antithesis of buy and hold is the concept of day trading in which money can be made in the short term if an individual tries to short on the peaks, and buy on the lows with greater money coming with greater volatility.

Table. 2 Input of NFs portfolio management

INPUTS
Close Price – C (t)
Buy & Sell from Up / Down of Close Price (t)
Buy & Sell from EMA 10 Days (t) with EMA 25 Days (t)
Buy & Sell from MACD (t) with EMA 9 Days (t)
Buy & Sell from RSI 4 Days (t)
Buy & Sell from RSI 9 Days (t)
Buy & Sell from RSI 14 Days (t)
Buy & Sell from W-%R 10 Days (t)
On Balance Volume – OBV (t)
Price and Volume Trend – PVT (t)
Volume Rate of Change – VROC (t)
Price Rate of Change – PROC (t)
Typical Price – TP (t)

Table. 3 Example of Financial Simulation Model in trading strategy

Stock Index	Possible Buy & Sell				NFs Buy & Sell			
		Action	# of Shared	Cash(Baht)		Action	# of Shared	Cash(Baht)
12.00	1	STAY	-	1,000,000	1	STAY	-	1,000,000
12.00	0	HOLD	-	1,000,000	1	HOLD	-	1,000,000
12.00	1	BUY	83,333		0	HOLD	-	1,000,000
12.50	0	HOLD	83,333		1	BUY	80,000	-
13.00	#	SELL	-	1,083,329	0	HOLD	80,000	
12.70	0	HOLD	-	1,083,329	0	HOLD	80,000	
12.60	0	HOLD	-	1,083,329	0	HOLD	80,000	
12.55	0	HOLD	-	1,083,329	#	SELL	-	1,004,000
12.45	0	HOLD	-	1,083,329	0	HOLD	-	1,004,000
12.40	1	BUY	87,365		0	HOLD	-	1,004,000
12.45	#	SELL	-	1,087,694	#	SELL	-	1,004,000
.....	...	....	....	......	...	....	....	......
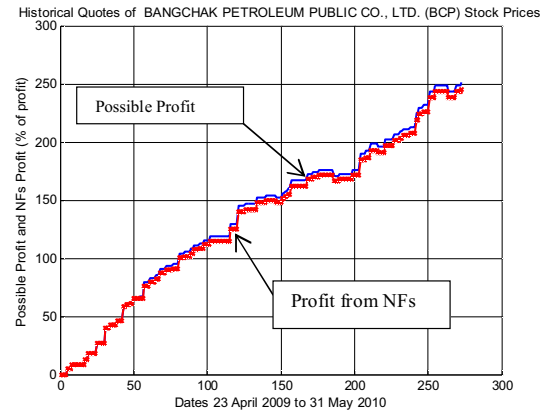
Fig. 6 Comparison profit between Possible Rate of Return Profit (Possible RoRP) and Profit from our proposed NFs Trading System in Training Days

The performance by each stock index is illustrated on table 4. It reflects the performances of investment strategies in Buy and Hold, NN and NFs model, respectively. Each line implies the performance of the NFs system in terms of cumulative profit rate of return gained from each stock index.

Historical Quotes of BANGCHAK PETROLEUM PUBLIC CO., LTD. (BCP) Stock Prices
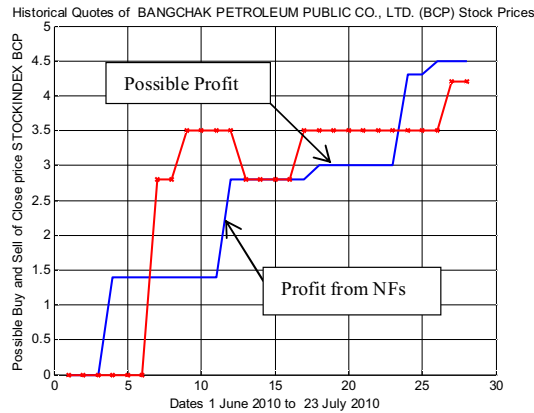
Fig. 7 Comparison profit between Possible Rate of Return Profit (Possible RoRP) and Profit from our proposed NFs Trading System in Testing Days.

Table 4   Rate of Return Profit (RoRP) gained from each trading stock index

	Training Set (23 April 2009 – 23 April 2010)						
	POSSIBLE	Model 1		Model 2		Model 3	
Index	RoRP	VAF	RoRP	VAF	RoRP	VAF	RoRP
BCP	207.4	36.68	94.20	52.97	176.50	97.20	204.40
SCB	457.7	21.63	174.24	80.64	279.72	98.20	455.34
	Testing Set I (4 May 2010 – 18 June 2010)						
Index	RoRP	VAF	RoRP	VAF	RoRP	VAF	RoRP
BCP	3.80	13.30	1.50	9.45	2.20	66.90	3.20
SCB	16.11	28.93	2.94	39.27	6.49	8.52	11.21
	Testing Set II (1 June 2010 – 23 July 2010)						
Index	RoRP	VAF	RoRP	VAF	RoRP	VAF	RoRP
BCP	4.50	46.88	1.40	79.95	3.50	54.32	4.43
SCB	28.75	30.89	5.50	19.98	8.00	29.93	14.73

In the case of experimental results, NFs display a greater rate of return than the "buy, sell and hold" model and NN model. Moreover, when comparing of 3 Models, Profit from Model 1 is lower than Model 2 and 3. Differencing Model 1 from Model 2 and 3 was not volume indicator. Model 3 was profit than Model 2 because it was including close price next day into its model. The results of difference in the stock index results are small. It is more valuable to calculate the loss and gains in terms of profitability in practice.

From table 4, On testing day, Profit from Decision-Making System based on Neurofuzzy Model for any stock index on 4 May 2010 to 18 June 2010 are 3.2% and

11.21% for BCP and SCB respectively. And, the profit on 1 June 2010 to 23 July 2010 are 4.43% and 14.73% for BCP and SCB respectively. We were simulating on investment within 1,000,000 baht for any stock index. If we separated investment for any stock index, such as 300,000 baht for BCP, 400,000 baht for PTT and 300,000 baht for SCB. Total profit was increased to 14.1%. Thus, we suggested multi trading from Decision-Making System based on Neurofuzzy Model more than invest only one stock index.

5. CONCLUSION

This paper presented the decision-making model based on the application of NFs. The model was applied in order to make a one-step forward decision, considering from historical data of daily stock returns. The experimental investigation has shown NFs trading system to make a trading strategy, achieving more stable results and higher profits when compared with Buy and Hold strategy. For future work, several issues could be considered. Other techniques, such as support vector machines, genetic algorithms, etc. can be applied for further comparisons. And other stock index groups, another stock exchange or other industries in addition to electronics one is further considered for comparisons.

6. REFERENCES

[1] Babuska, A. R., **Neuro-fuzzy methods for modeling and identification, Recent Advances in intelligent Paradigms and Application**, 2002, pp. 161–186.

[2] Cardon, O., Herrera, F. and Villar, P., **Analysis and Guidelines to Obtain A Good Uniform Fuzzy rule Based System Using simulated Annealing**, , Int'l J. of Approximated Reason 25, No. 3,2000, pp. 187–215.

[3] Chapman, A. J., **Stock market reading systems through neural networks: developing a model**, Int'l J. of Apply Expert Systems, Vol. 2, No. 2, 1994, pp. 88-100

[4] Chen A.S., Leuny, M.T., and Daoun, H., **Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading The Taiwan Stock Index**, Computers and Operations Research, Vol. 30, 2003, pp. 901 – 902.

[5] Conner, N.O. and Madden, M., **A Neural Network Approach to Pre-diction Stock Exchange Movements Using External Factor**', Knowledge Based System, Vol. 19, 2006, pp. 371 – 378.

[6] James N.K. Liu and Raymond W.M. KWong, **Automatic Extraction and Identification of chart Patterns Towards Financial Forecast**, Applied soft Computing, Vol. 1, 2006, pp. 1 – 12.

[7] Doeksen, B., Abraham, A., Thomas, J. and Paprzycki, M. **Real Stock Trading Using Soft Computing Models**, IEEE Int'l Conf. on Information Technology: Coding and Computing, Las Vegas, Nevada, USA, 2005, pp. 123-129.

[8] Dutta, S. and Shekhar, S., **Bond rating: A non-conservative application of neural networks**, IEEE Int'l Conf. on Neural Networks, San Diego, CA , USA, 1990, pp. 124 – 130.

[9] Radeerom, M., Srisa-an, C., K., Kasemsan M. L.K, **Prediction Method for Real Thai Stock Index Based on Neurofuzzy Approach**, Lecture Notes in Electrical Engineering : Trends in Intelligent Systems and Computer Engineering, Vol. 6, 2008, pp. 327–347.

[10] Radeerom, M., and M L. Kasemsan, K., **Neurofuzzy Decision-Making Approach for the Next Day Portfolio Thai Stock Index Management Trading Strategies**, Advances in Intelligent Information and Database Systems 2010, pp. 371-381.

[11] Radeerom, M., and M L. Kasemsan, K., **Intelligence Decision Trading Systems for Stock Index.** ACIIDS, Vol. 3, 2012, pp. 366-375

Object-Oriented Software Engineering and Anti-Patterns

Lisa MACLEAN
Wentworth Institute of Technology
Department of Computer Science and Networking
550 Huntington Avenue
Boston MA 02115 USA

ABSTRACT

Industry estimates of software project failures run at over 80%. Points of failure include the systems architecture, the people on the project, management issues, and within the software coding itself. Object-oriented (OO) software is a common software development framework that encompasses many programming languages. OO software languages produce code that is highly flexible and reusable and promotes valuable design patterns, falling into the categories of creational, structural, and behavioral. They are also susceptible to problems called anti-patterns. Being aware of these anti-patterns is the first step to eradicating them. This is necessary to push more software projects to completion and successful functioning.

Keywords: object-oriented software, anti-patterns, software engineering, programming

INTRODUCTION

The use of object-oriented (OO) programming techniques has become the de facto standard for new software development [1]. In procedural programming, the program is conceived as a series of steps which need to be taken in sequence in order to process data [2]. The steps are broken into blocks of statements that are used to manipulate data.

In OO programming, code is created in reusable modules called objects. OO programs focus on combining data and functions into these objects.

The two main constructs of OO are classes and objects. Creating a new class creates what is called a "type," whereas creating a new instance of that class is an "object." These objects can then store data using ordinary variables. Variables can belong to an object or class, and are called as fields [3].

These objects can also store functions that belong to a class. These functions are called "methods" of the class. This terminology helps to differentiate between functions and variables which exist by themselves and those which belong to a class or object [3].

These fields and methods are referred to as the attributes of that class. Fields can be of two types. They either belong to an instance/object of the class, or they can belong to the class itself. If they belong to the object, they are called instance variables. If they belong to the class itself they are called class variables [3].

"Instantiation" is the creation of a class of objects or a computer process. To instantiate is to create such an instance by, for example, defining a variation of object within a class, naming it, and saving it to a physical location. In OO programming, a class is instantiated to create an object, a concrete instance of the class. The object is the executable file that is run on a computer [4].

These terms and concepts are necessary to the continued discussion of anti-patterns and software engineering.

OBJECT-ORIENTED SOFTWARE ENGINEERING

There is a classic design textbook first released in 1995 concerning flexible and reusable code for object-oriented systems: Design Patterns: Elements of Reusable Object-Oriented Software, published by Addison-Wesley. The authors of the book are Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides, and are commonly called the "Gang of Four."

This is a valuable web site address to the source code for this book: http://hillside.net/component/content/article/51-design-patterns/131-elements-of-reusable-object-oriented-software-book. However, design patterns have an equal and opposite force: anti-patterns. This is also spelled as AntiPatterns or antipatterns.

WHY STUDY ANTI-PATTERNS?

It is estimated that five out of six software projects still fail or do not produce the desired result. The failures are sometimes architectural, sometimes within the software, and sometimes managerial. The failure of projects repeatedly due to the same causal factors provoke the need for a common vocabulary to discuss these issues and give designers and developers the means to address chronic problems. One of these known problems is anti-patterns [5]. There are times when established information theory can become its own hindrance.

By being aware of design patterns and anti-patterns, it is hoped that systems analysts and developers can avoid the inherent traps and save projects in crisis.

DESCRIPTIONS OF PATTERNS AND ANTI-PATTERNS

What are design patterns? Design patterns are a relatively recent software engineering design phenomenon that grew from the object-oriented coding community. They must be repeatedly observed in practice in order to be considered a valid Pattern [5]. There are three main patterns: Creational, Structural, and Behavioral [6].

Creational patterns have to do with class instantiation. These are divided into class-creation patterns and object-creational patterns. While class-creation patterns use inheritance effectively in the instantiation process, object-creation patterns use delegation instead [7].

Structural patterns concern class and object composition. They use inheritance to compose interfaces and define the composition of objects to obtain new functionality [7].

Behavioral patterns are about a class' objects' communication. They are specifically for communication between objects [7]. This grounding in the theory of patterns leads into a discussion of architectural anti-patterns.

What are architectural anti-patterns? Anti-patterns are closely related to design patterns, which document recurrent solutions. A design pattern becomes an antipattern when it causes more problems than it solves [8].

"The essence of an Anti pattern is two solutions, instead of a problem and solution for an ordinary design pattern. The first solution is problematic. It is a commonly occuring [sic] solution that generates overwhelmingly negatively consequences. The second solution is called the refactored solution. The refactored solution is a commonly occuring [sic] method in which [an] antipattern can be resolved and reengineered into a more beneficial form [8]."

Kärpijoki makes an elegant distinction in saying that the fundamental difference between patterns and anti-patterns is that a design pattern provides a solution, while the AntiPattern starts from that solution [5].

Srinivasan takes the stance that anti-patterns can actually be helpful, in that they, when recognized, will allow the avoidance of making

the same, or the same type, of mistakes seen in a previous instance of the code:

"Anti-patterns are more difficult to draft, but rapidly lead to sharing of experiences. In practice it is much more difficult to identify the situations that should be avoided, particularly so in software development and that is the challenge tackled carefully through Anti-patterns in many phases of software life cycle development. Through Anti-patterns it is possible to clearly identify bad concepts followed in design [8]."

But how shall they be recognized? While some characterize anti-patterns into groups, such as user, managerial, developer, etc. this section will focus upon those anti-patterns found in and defined exclusively by OO code objects.

Brown, Malveau, McCormick, Mowbray, and Thomas present the following set:

The Cover your Assets anti-pattern is where a developer writing software processes lists all possible alternatives but makes no decisions [8].

Intellectual Violence: Where one developer insists upon sticking to obscure theories, esoteric standards, and untested theories for short-term gain [8].

Jumble: The user interfaces are so badly defined or designed that the objects written for them are not reusable [8].

Reinventing the Wheel: Legacy systems with overlapping but interoperable functionality are used as extensible. Each object is written in isolation [8].

Stovepipe Systems: Ad hoc code objects with lack of integration or abstraction [8].

The Swiss Army Knife: Over-design leads to objects with multiple functions and methods that try to anticipate every possible scenario; leads to overly complicated and less flexible code. Also known as the Avarice Anti-pattern [8].

Brown et al. also mention other such coded Anti-patterns such as:

The Blob: This anti-pattern uses one class to monopolize all the processing; there is a lack of OO standards and design [10].

The Functional Decomposition anti-pattern: Classes that have names such as "Calculate_Interest" or "Display_table" may indicate this pattern. This antipattern contains no way to clearly document or explain how the system works. This is typically a disaster anti-pattern [10].

The Golden Hammer: One developer uses a favorite software concept obsessively. A developer may learn one or two of the "Gang of Four" patterns and apply them to everything [10].

Spaghetti code: The classic anti-pattern. The flow of execution is dictated by object implementation, not by clients of the objects. Efforts to maintain the code only exacerbates the problem. There may be procedural code embedded in the objects. This code is unmaintainable and non-reusable [10].

Cut and paste: The same bug reoccurs due to the fact that the intent behind an object is propagated but not preserved in the code [10].

Lava flow: the software has code whose purpose is unseen or unclear. It tends to be OO code that originally worked and was added to too many projects. Its intended use is probably lost. Useless code is maintained and executed again and again, in vain, hence its name. Lava flow consumes resources, both human and computational, in its path [10].

Kärpijoki expands upon anti-patterns by examining their root causes. Software is written, designed and maintained by fallible human beings. They fall prey to such things as:

Haste: Decisions are made too quickly. Software is of poor quality and can cause catastrophic

flaws in the code. Haste leads to panic: as deadlines loom, precious testing is left out [5].

Apathy: This leads to common problems being overlooked, or discovered and ignored, due to breakdown in the group or group-management dynamic [5].

Sloth: it is easier to write bad or simple code and not follow established procedures [5].

Avarice: this is one cause of the "Swiss Army Knife" anti-pattern. This can occur when a module is written exclusively by one developer, but feels the module is so important that it must do more: more features, more designs, even if not required by the systems analysis. A developer can become very "identified" with the code and over engineer out of pride in his work [5].

Ignorance: This is intellectual sloth. A developer may be lazy and seek the easiest, fastest solution he can code with no thought to standards, reusability, flexibility, complexity or even functionality. This is a dangerous root cause because ignorance is hard to discern; however, as the project goes over schedule as necessary modifications appear again and again, the problem can become severe [10].

Some Anti-patterns may even be a valid instance of a Pattern, but used in the wrong context. Anti-patterns may also be valid and good instances of a known Pattern that drifted from its original use as code is tacked on here and here in the name of reusability.

Knowing what patterns are and why they are valuable is an important skill in systems design and the development of object-oriented software. Conversely, being able to recognize anti-patterns can be just as important. Project failure can be minimized by following well-established project and personnel management practices. Jobs should be clearly defined and well-monitored. Communication within project members is crucial. People not using, or misusing, time-honored techniques should be removed from a

project. Code should be peer-reviewed and well-tested.

With a library of standard objects within an object-oriented system that can be used on varying operating systems, projects can be completed more quickly with less testing and a common look-and-feel. A goal of high cohesion in tandem with low coupling increases the reusability of code objects [11].

Spaghetti code anti-Patterns can be avoided by using standard, strictly-defined templates for objects from the same source; use mentoring of new programmers and implement code reviews. Maintain proper code documentation. Use critical performance measurement standards to identify poorly-performing code and establish timely code cleaning techniques to fix them after identification [12].

Make sure programmers are trained in OO technologies and that the code is not inherently procedural and therefore inappropriate for OO solutions. Modules with a high number of attributes or operations can lead to "the blob" anti-pattern and the modules need to be restructured using proper OO techniques. Lava code, also known in procedural programming as "kruft," occurs when old modules of code of unknown origin or benefit are left in the system. These poorly-designed modules clog systems and must be removed [12].

Functional decomposition can be identified by looking for code written in a procedural style that has all private classes with no interaction outside the object. There is no inheritance or polymorphism, and no hope of reusability. This code must also be rewritten using OO techniques. This same holds true for the "Golden Hammer," which occurs when programmers work in isolation and use proprietary technology [12].

When programmers attempt to reuse bad code instead of designing fresh modules using proper OO techniques, the cut-and-paste anti-pattern appears. This can be avoided by watching for

root modules which are constantly causing errors and slowdowns in the code and applying proper code cleaning techniques. Assigning lead developers to this task is advantageous [12].

The Intellectual Violence anti-pattern and Reinventing the Wheel also need to be rewritten using proper form and the latest OO training. These occur when development takes place in a vacuum and developers are working in isolation [12]. The Swiss Army Knife modules must be divided into smaller modules with proper cohesion and coupling; one module cannot do all the work in a proper OO environment [13]. Stovepipe anti-patterns occur when pre-existing software is simply migrated into a new environment. The time and cost to rewrite new code should be undertaken instead, especially when moving to a new architecture [13].

The Cover Your Assets anti-pattern is the fault of end users who cannot decide what the system is actually supposed to do. Indecision and attempting to take into account any and all possible occurrences leaves developers with voluminous specifications but no outcomes, or pages upon pages of requirements all declared mandatory and of equal priority. Bad systems analysis and design leads to bad code. Architecture blueprints can be used to generate dialog across organizational units to bring agreement on systems specifications [12].

The Jumble anti-pattern occurs when horizontal and vertical design elements are mixed. Horizontal design elements, those that are common across applications, and vertical design elements, those dependent upon specific software implementations, are put together by systems architects and software developers. This intermingling makes all software less stable and less reusable. All horizontal design elements must be separated out and consigned to a single layer. This horizontal layer can then be used to identify common interoperability functionality within the architecture and allow for metadata to be incorporated while still using vertical elements for specialized functionality [12].

These anti-patterns occur when programmers are insufficiently trained, time is tight, and oversight is minimal. Rewriting poorly designed code is time-consuming and management can often be unsupportive of the investment when systems are running and appear to be functioning, if not functioning well. The problems will only increase when systems are scaled to larger applications. Communication between developers should be encouraged, and developers should be open to peer review and feedback instead of being heavily invested in their code.

The proper use of judicious OO techniques must be insisted upon and followed through to avoid these anti-patterns from forming. Once the code is written and deployed, or when a legacy system is approached, the system must be carefully watched and analyzed to ensure that anti-patterns are not overlooked. When new features are required, code should be properly restructured and not simply modified. Time spent in this manner will save time after deployment.

CONCLUSION

Object-oriented programming languages are valuable design tools offering flexibility, reusability and rapid development when properly used. With a high failure rate of software projects in the age of information, it is necessary to move more projects to not only completion, but success. These advantages of OO code must be balanced against the tendency for it to produce anti-patterns. When wisely used and implemented, OO anti-patterns can be easily avoided.

REFERENCES

[1]. Scott W. Ambler (2009). **Introduction to Object-Orientation and the UML.** Available at http://www.agiledata.org/essays/objectOrientati on101.html

[2]. WiseGeek.com. (2012). "What Is Procedural Programming?" Available at

http://www.wisegeek.com/what-is-procedural-programming.htm

[3]. A Byte of Python. Available at http://www.ibiblio.org/g2swap/byteofpython/read/oops.html, n.d.

[4]. TechTarget.com. "Instantiation," 2012, Available at http://searchcio-midmarket.techtarget.com/definition/instantiation

[5]. Vesa Kärpijoki, "AntiPatterns," Aalti University, Multimedia Laboratory, Seminar on Design Patterns, 2001. Available at http://users.tkk.fi/~vkarpijo/dp01/dp01_anti-patterns.pdf

[6]. Data and Object Factory. ".NET Design Patterns," Available at http://www.dofactory.com/Patterns/Patterns.aspx, 2014.

[7]. Erich Gamma, Richard Helm, Joseph Johnson, John Vlissides. **Design Patterns: Elements of Reusable Object-Oriented Software,** Boston: Pearson, 1995.

[8]. William J. Brown, Raphael C. Malveau, Hays McCormack III, Thomas Mowbray. "Anti-patterns: Refactoring Software, Architectures and Projects in Crisis," Available at http://silanian.tripod.com/Articles/antipattern.htm, 2005.

[9]. Srinivasan, R., Dr., CTO, iCMG. (2001, December 17). "Anti-patterns identify bad design concepts," Available at http://www.icmgworld.com/corp/news/Articles/RS/dec_0301.asp

[10]. Brown, William, Malveau, Raphael, McCormick, Hays, Mowbray, Thomas, Scott W. Thomas. (1999, December 13). "What are Architectural Anti-patterns?" Available at http://www.anti-patterns.com/arch_cat.htm

[11]. Jeffrey L. Whitten, and Lonnie D. Bentley, (2007). **Systems Analysis and Design Methods,** (7th ed.). Boston: McGraw and Hill:Irwin.

[12]. Alexander Shvets, Gerhard Frey, Marina Pavlova. **AntiPatterns: The Survival Guide,** SourceMaking.com

[13]. Valentin Corneliu Pau, Octavian Stanescu, Marius Iulian Mihailescu, Titu Maiorescu. "Antipatterns - Implementing Productive Solutions to Avoid Developing Problems for Web and Software Applications," Journal of Applied Computer Science and Mathematics, no. 7, July 11, 2012.

Efficient Work Team Scheduling: Using Psychological Models of Knowledge Retention to Improve Code Writing Efficiency

**Michael J Pelosi, Michael Scott Brown, and
Henry Dirska
University of Maryland University College
Largo, Maryland, United States**

and

**Mir Mohammed Assadullah
Career Opts
Springfield, VA, United States**

ABSTRACT

Development teams and programmers must retain critical information about their work during work intervals and gaps in order to improve future performance when work resumes. Despite time lapses, project managers want to maximize coding efficiency and effectiveness. By developing a mathematically justified, practically useful, and computationally tractable quantitative and cognitive model of learning and memory retention, this study establishes calculations designed to maximize scheduling payoff and optimize developer efficiency and effectiveness.

Keywords: WBS; scheduling; team; knowledge; retention.

1. INTRODUCTION

An interesting and classical quantitative law of cognitive psychology is that forgetting curves are well described by power functions (e.g., [1], [2], [3]). For example, Wixted and Ebbesen (1991) [3] and Wixted and Carpenter (2007) [4] show that diverse measures of forgetting various items such as words, faces, and nonsensical syllables can be well described as power functions of a retention interval.

In one seminal and classic article, Wickelgren (1974) [4] derived an equation that is robust in several respects, including in its ability to characterize the previously famous Ebbinghaus (1885, 1913) [5] memory savings function. With most typical conditions, the Wickelgren power law is reduced to:

$$m = \lambda(1 + \beta t)^{-\psi} \qquad (1)$$

where m is memory strength, and t is time (i.e., the retention interval). The equation has three parameters: λ is the state of long-term memory at $t = 0$ (i.e., the degree of learning), ψ is the rate of forgetting, and β is a scaling parameter.

Wixted (2004) [6] also substantiated that Equation 1 provides a very accurate description of forgetting data that have been averaged over many subjects. It therefore not only fits the data well in terms of the percentage of variance, a relatively weak test, but also accurately predicts where future points will fall as the retention interval increases, which is a relatively stronger test. As a result of these averaging of artifacts in the group data, a possible stronger tests would be to accurately predict the

degree of forgetfulness for individual learning subjects. One practical problem with averaging the testing artifacts is that such data are usually quite noisy. However, the eight measured data points of the classic Ebbinghaus (1885, 1913) [5] savings function have offered one possible rare and well-known exception. Previous research work has shown that the Ebbinghaus data can be reasonably well characterized by a two parameter power function of the form:

$$m = \theta t^{-\psi} \qquad (2)$$

This power function has been considered an approximation of Equation 1 (Anderson & Schooler, 1991 [1]; Wixted & Ebbesen, 1991 [3]). Although Equation 2 offers a much better fit of the savings function than many other two-parameter candidates, it is undefined at $t = 0$, which is theoretically unsatisfying and limits the equation's practical utility. As one example, it could not be used to estimate the degree of learning. Typical memory decay rates based on measured experimental results from Equation 2 are shown below in Figure 1.

2. RETENTION EFFECT
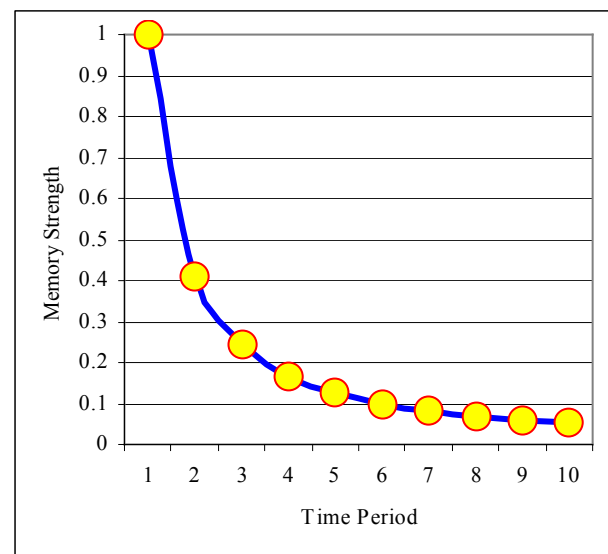
Figure 1. Typical retention decay rates ($\psi = -1.286$) from Donkin, 2012 [7].

Another study suggests that the probability that a previously learned memory item can be correctly recalled decreases at an exponential rate over time like radioactive decay. However, a so-called "spacing effect", defined as the repeated exposures to the same information, substantially boosts future retention probability and reduces the rate of memory loss. Individuals who have experienced two or more learning sessions of the same item will remember more when the sessions are separated by time. The degree of memory retention directly correlates to the size of the time gap (or spacing) between sessions—the greater the spacing, the more likely that a subject will retain the memory. For experimentally measured results and worthwhile discussions of these spacing effects, see Bahrick (1987 and 1993) [8], [9], Cepeda (2008 and 2009) [10], [11], Goverover (2009) [12], Pavlik (2003 and 2008) [13], [14], and Rohrer (2010) [15].

3. APPLICATION TO SOFTWARE ENGINEERING

A Work Breakdown Structure (WBS) is a hierarchical decomposition of the work activities in a software project. The lowest level activities in the WBS hierarchy are tasks. Each element of a WBS is named using a verb phrase to denote the process-oriented nature of a WBS. Another technique, the "architecture decomposition view" technique is normally used in close cooperation with the WBS technique to assemble WBS packages into larger groupings.

A WBS specifies work packages for tasks. Work packages for activities aggregate the work packages for subordinate activities and tasks. A WBS decomposes large work activities such as analysis, development, design, coding, and testing, into smaller tasks of 40 to 80 staff-hours each using the "augmented rolling wave" approach to planning. The WBS documents each task in a work package and each work package becomes a negotiated contract between the team leader and the teams or individuals assigned to that work package.

Each specific work package describes a task as follows: the order of precedence for the task. For example the magnitude of importance for that task in comparison to other activities and tasks, the planned duration of the task, a description of resources needed to accomplish the task, the work products to be produced, the risk factors associated with completing the task, and the acceptance criteria for the resulting product. Each work package must produce one or more tangible work products that satisfy some objective acceptance criterion.

The project manager estimates the time durations needed to complete each task and will negotiate resource allocation with each team member, and then assign work packages to teams or individuals subject to overall resource and scheduling constraints. Assignment of a work package from a team leader to a team member is analogous a contract for completion of one or more work products that meet an acceptance criteria within a specified time duration. By choosing the most qualified team or individual to complete a specific work package, a project manager can mitigate project risks, improve effectiveness, reduce defect levels, and speed completion.

4. EXAMPLES

How can knowledge of the Ebbinghaus [5] memory retention decay rate curve help to improve the effectiveness of project managers who allocate resources for work projects? This knowledge can help by allowing the project manager to schedule the work team with the greatest memory retention of prior programming work to specific work packages and maintenance following completion of the project. The project manager can predict the likely code retention of different teams of programmers, and make decisions based on these quantitative memory retention metrics. The conducted research provides two examples worthy of further exploration: using memory retention metrics for scheduling work packages for a large software project, and using the metrics for software maintenance by individual programmers.

A. Work Package Scheduling Example

Assume a large software development project is progressing as shown in Figure 2 below. Work packages are scheduled using the PERT analysis technique as shown in the chart, with two teams A and B performing the development of work packages 1 through 8, upon which work package number 9 depends.
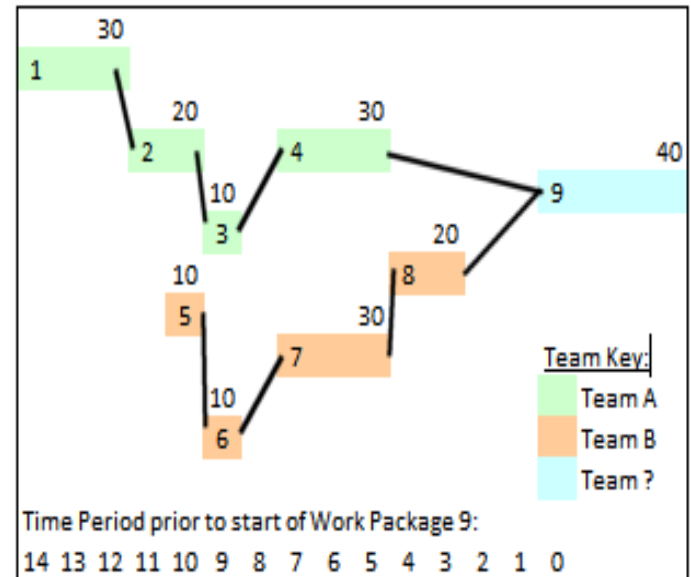
Figure 2. Sample work scheduling PERT chart.

Under our scenario, Team A has completed work packages 1 through 4, and Team B has completed work packages 5 through 8. The scheduling and completion of these various work packages is shown in the above PERT chart, as well as starting completion time in weeks prior to the scheduled start of work package number 9. The "thousands of lines of code" (KLOC) for each work package appears above the respective work packages.

A description of the work completed is shown below in Table I for work packages 1 through 8. Team A has completed 90,000 lines of code, and Team B has completed 70,000 lines of code for the various work packages. However, notice in Figure 2 that Team A's experience started at least four weeks prior to the start of Team B's work, and ended two weeks prior to the end of Team B's most recent effort on work package 8:

TABLE I.   TEAM WORK PACKAGE HISTORY

Work Package	KLOC	Team	Tot. KLOC
1	30	A	
2	20	A	
3	10	A	
4	30	A	90
5	10	B	
6	10	B	
7	30	B	
8	20	B	70

This poses the question—which of these two teams should the project manager schedule to begin work on work package number 9? Team A's greater experience in terms of lines of code (90,000 versus 70,000) suggests that Team A would have a greater knowledge of the existing code that will be used to help complete work package 9. However, according to the Ebbinghaus [5] decay rate equation, it turns out Team B's knowledge of the code through memory retention is expected to be nearly double that of Team A's.

Table II outlines how the likelihood of memory retention for both Team A and Team B could be calculated. The retention rate is calculated according to the decay rate function (y = -1.286, from Donkin, 2012 [7]), using as the time variable interval the number of periods a work package ended prior to the start of an impending work package (such as work package number 9).

This retention rate is then multiplied by the number of thousands of lines of code in the respective work package. The result is a weighted retention rate that takes into consideration the quantity of code written, as well as its recency. After adding the weighted retention rates together for each work package for both teams, Team B's weighted retention comes in at nearly double that of Team A, at 14.525 versus 8.290, respectively.

TABLE II.   TEAM WEIGHTED RETENTION CALCULATION

Work Package	KLOC	Team	Periods	Retention	W. Ret.	Total
1	30	A	11	0.0458	1.373	
2	20	A	9	0.0592	1.185	
3	10	A	8	0.0689	0.689	
4	30	A	4	0.1681	5.043	8.290
5	10	B	9	0.0592	0.592	
6	10	B	8	0.0689	0.689	
7	30	B	4	0.1681	5.043	
8	20	B	2	0.4100	8.200	14.525

Although the PERT chart seems to indicate that the teams have nearly identical experience and would assumedly maintain the same memory retention, the calculated memory decay rate suggests that Team B is a much better choice in terms of current code familiarity and should therefore be more efficient at completing work package number 9 than Team A.

This result suggests that a calculated quantitative metric contrasts sharply with typical intuitive and subjective judgments. Since Team B's familiarity is nearly double that of Team A's, this degree of memory retention could result in significantly higher productivity and more efficient code

development with less defects introduced due to lapses of knowledge.

B. Software Maintenance Scheduling Example

The proposed technique can be applied to software maintenance as well. Given a choice between two programmers, programmer X and programmer Y, which should be assigned to perform maintenance on a large software program? Assume programmer X was initially responsible for writing the program, although this development took place ending approximately one year, or 52 weeks, ago. Another programmer, programmer Y, was brought in relatively recently to modify 10,000 lines of code, and this project ended approximately 8 weeks ago. Also, programmer Y added 2,000 lines of new code, and this effort ended five weeks prior to the scheduled start of maintenance. At first glance, it would seem programmer X, having written 100,000 lines of code on the software, would be the better choice in terms of code familiarity. However, when applying the memory decay calculation, programmer Y actually has a better memory of the code.

The weighted retention of programmer X turns out to be 0.620, based on 52 time periods and 100,000 lines of code. The weighted retention of programmer Y is 0.942, based on the 10,000 lines of code ending eight weeks ago, and 2,000 lines of code ending five weeks ago. These results are shown below in Table III.

TABLE III.   PROGRAMMER EXPERIENCE

Work	KLOC	Prog	Periods	Retention	W. Ret.	Total
1	100	X	52	0.0062	0.620	0.620
2	10	Y	8	0.0689	0.689	
3	2	Y	5	0.1262	0.252	0.942

This analysis demonstrates that programmer Y is a better choice than programmer X for maintenance work—which once again contrasts with the likely human subjective judgment as to which of the two programmers would be more familiar with the code and would perform more efficient maintenance as a result.

5. CONCLUSION

A novel technique has been shown for scheduling more efficient work teams and programmers to perform software development and project maintenance. Calculating likely code retention is a relatively simple and straightforward technique using the Ebbinghaus [5] decay rate function.

The technique could be expanded to many different areas of software engineering, as well as to project management in general. Being able to calculate ahead of time how much people will remember can be useful in identifying the most efficient programmers.

This technique could possibly be combined with the aforementioned "spacing effect", which postulates that an increased length in time in between reinforcement sessions increases the likelihood of memory retention. For more on this "spacing effects" phenomenon, see the reference for the 2008 article by Cepeda [10].

6. REFERENCES

[1] Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, 2, 396–408.

[2] Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. Memory & Cognition, 2, 775–780.

[3] Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. Psychological Science, 2, 409–415.

[4] Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. Psychological Science, 18, 133–134.

[5] Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. New York: Teachers College, Columbia University. (Original work published 1885).

[6] Wixted, J.T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. Psychological Review, 111, 864–879.

[7] Donkin C., & Nosofsky M. (2012). A Power-Law Model of Psychological Memory Strength in Short- and Long-Term Recognition. Psychological Science, 23(6) 625–634.

[8] Bahrick, Harry P.; Phelphs, Elizabeth. Retention of Spanish Vocabulary Over 8 Years. (1987) Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 13(2), Apr 1987, 344-349

[9] Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. Psychological Science, 4, 316–321.

[10] Cepeda NJ, Vul E, Rohrer D, Wixted JT, Pashler H. (2008). Spacing effects in learning: a temporal ridgeline of optimal retention. Psychological Science. 2008;19(11):1095–102.

[11] Cepeda, N.J., Coburn, N., Rohrer, D., Wixted, J.T., Mozer, M.C., Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. Experimental Psychology.

[12] Goverover Y., Arango-Lasprilla J. C., Hillary F. G., Chiaravalloti N., Deluca J. (2009). Application of the spacing effect to improve learning and memory for functional tasks in traumatic brain injury: A pilot study. American Journal of Occupational Therapy 63:543–548.

[13] Pavlik, P. I., Anderson, J. R. (2003). An ACT-R model of the spacing effect. In F. Detje, D. Dorner, & H. Schaub (Eds.), Proceedings of the Fifth International Conference of Cognitive Modeling (pp. 177–182). Bamberg, Germany: Universitats-Verlag Bamberg.

[14] Pavlik Jr., P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. Journal of Experimental Psychology: Applied.

[15] Rohrer, D. and H. Pashler (2010). Recent Research on Human Learning Challenges Conventional Instructional Strategies. Educational Researcher, 39(5), 406

Measuring and minimizing the badness of exam timetables

Elmien Spoelstra and Jacob Spoelstra
School of Computer Science, Statistics and Mathematical Sciences
Potchefstroom Campus of the North West University
Potchefstroom, South Africa
Email: elmien.spoelstra@nwu.ac.za, japie.spoelstra@nwu.ac.za

Abstract—**Students and staff usually prefer exam timetables that are available before the start of the academic year, and in which adequate study time is available between exam papers. One method of constructing exam timetables before registration data of the students are available, is to divide the possible subjects into non-clashing subject groups and schedule the groups as units. As an example the exam timetable for all possible curricula and choices in the curricula of the North West University in South Africa had to be drawn up, giving students as fair an exam timetable as possible. As the lecture timetable is based on subject groups, the problem was to schedule all the groups of subjects, from first year up to fourth year students, in such a way over the available days of the exam, that students would have as many free days available between papers as can be achieved. The main focus of this paper was to define an objective function, taking into account the number of students and a measure of the 'badness' of each student's exam schedule. The objective function was minimized by a variation of the method of simulated annealing, over all possible permutations of the order of papers. The process resulted in a viable exam schedule, with virtually no student having a bad schedule, demonstrating that the method is a practical and effective method of construction fair exam schedules.**

I. Introduction

A brief background, based on the North West University experience, is first given on why 'subject grouping' is considered instead of the more recent flexible scheduling methods and commercial programs, followed by defining an objective function for optimizing the exam schedule. The optimizing method is briefly discussed in the following section, the results are presented and finally conclusions are drawn.

In construction and general scheduling of projects, the concept of dividing tasks into non-clashing operations that can each be done by a different subcontractor is the same as the scheduling discussed in this paper. One problem that often occurs is overrun of time by subcontractors, delaying all subsequent operations. This can lead to costly fees for exceeding the contract time. By the same strategy as discussed here for exams, all operations can be scheduled to allow for the maximum time between different contracts.

A. Historical background

The Potchefstroom Campus of the North West University (formerly called the Potchefstroom University) had for many decades used a system of dividing all subjects into 10 groups and scheduling the lecture timetable for these groups. By this method the complete timetable could fit onto one page [1, p. C123]. The only limitation on students' choices was that the subjects that they chose should not clash on the timetable.

In the year 2000, due to complaints concerning restrictions on possible subject choices caused by this system, the university introduced fixed, 'logical' curricula. These offered wider options for students to choose a curriculum with subject combinations that were more suited to modern society — in particular to allow for subjects from IT and Computer Sciences to be included in other combinations than the sciences and engineering.

No timetable restrictions were placed on lecturers planning these new curricula — the administrators and education experts claimed that they had the best computer program available to form a timetable and therefore that the academics can design curricula without constraint.

The timetable computer program could not form a timetable. A mathematician was seconded to the task to clean up curricula and scrap problematical subject combinations that blocked all options for a timetable, resulting in a timetable using shortened lectures from 7:40 until 19:30 and less lectures per subject. This was not satisfactory to staff and students.

A group based system for exams seemed no longer possible for all the possible combinations and it was decided to schedule the exam timetable base only on registration data of the enlisted students' choices, referred to as 'registration data'. The exam timetables were made available three to four weeks before the onset of the exams. Because the programs scheduled the subjects according to the number of students per combination, it happened that some curricula with less than 200 students would get a extremely bad timetable with 3 or more consecutive days of exams, while combinations of 1000 students would get five or more days between papers. This seeming unfairness, coupled with the late exam timetable, cause official complaints.

In 2008, due to a merger between three universities into one university with three campuses, all curricula were once again redesigned.

Academic staff saw an opening to ask for lecture timetable reform. It was decided to return to the grouping system. This was implemented successfully by the authors from 2009 until the present [2]. Figure 1 shows the classification of the subjects for the second year engineering students — which is

a small subset of the total table that contains approximately 650 subjects per semester.

1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B	10
EERI212 EERI213		CEMI211 FSKS211 INGM211				CHEN211 INGM212						
				TGWN211a TGWN212b	TGWN211b TGWN212a					WISN211b WISN212c	WISN211c WISN212h	WVTS211b

Fig. 1. Subject groups for second year engineering students

The lecture timetable simply places the groups in the time slots on the daily schedule, more or less evenly distributed over the week. Practical sessions fall mostly in the afternoons after a lunch break. One subject can be taken in a large number of curricula over the different faculties. The structure of the timetable is shown in Figure 2.

	07:30 08:15	08:15 09:00	9:10 10:00	10:10 11:00	11:10 12:00	12:10 13:00	14:00 14:50	14:50 15:40	15:50 16:40	16:40 17:30
Mo	6A	6A	5B	4B	3B	2B	1A	1A	1B	1B
Tu	5A	5A	6B	3B	2B	1B	4A	4A	4B	4B
We	4A	4A	6A	6A	6B	6B	2A	2A	2B	2B
Th	3A	3A	4B	1B	10	10	5A	5A	5B	5B
Fr	2A	2A	6B	5B	1A	1A	3A	3A	3B	3B

Fig. 2. The timetable structure

The success in switching back to a system of timetable groups prompted staff to ask if we could also return to an exam timetable that is based on these groups. This was to be available before start of term to allow for long term planning of staff and students. The principals of the university asked me to investigate the possibilities. Our findings and work on this consultation work led us to write this paper and develop and improve the system.

B. Technical background

Even though the method presented here are applicable to different scheduling applications, we will focus on exam timetables.

All subjects must be divided into groups. The groups will never clash. This is based partly on available curricula, but this can be still be too lenient and some extra constraints have to be added (*e.g.*, Sociology is not allowed to be registered for at the same time as Electronic Engineering).

The optimal division into groups, even with accurate information, is not a simple matter, and one for which no viable algorithm existed at the time. We have since developed methods of optimizing the grouping. The structure of the placing has to allow for students repeating a subject.

The basic concept was that all subjects in a timetable group should write their exams at the same time. The *groups* are then scheduled rather than *individual subjects*.

Some students repeat subjects of a previous year. In a few cases they are allowed to do this without attending classes to avoid clashes. This implies that papers for different year levels have to be written in different time slots.

The following were the basic constraints on the exam timetable:

- Only one paper per subject is allowed.
  This implies that, *e.g.*, second year Linear Algebra (which

is repeated three times for 200 students each), had to have one time slot for all students, which may come from Engineering, Science, Economy and the Humanities.
- Subjects that were repeated for different groups had to have one time slot for all.
  This implies that, *e.g.*, Business Management, with 1500 students and four lecturers spread over six class groups and four timetable groups, had to write one paper at the same time.
- Some subjects, with the same code on all three campuses, had to write the same paper at the same time.
- Students should have at least one day open between papers, if possible.
- If some students could not have a schedule with a day open between papers, they should not have two papers on the same day under any circumstance. In cases where students repeated subjects without attending lectures this restriction could be lifted.
- All options for electives allowed for in the lecture timetable should be free of clashes in the exam timetable.
- The exam had to fit into eighteen days.
- There was to be a second exam opportunity (for students who could not write due to medical reasons and who failed the first opportunity). This should accommodate all the papers of the first opportunity. All papers that were scheduled on a Saturday or at noon on Friday for the first exam must fall on different days for the second exam due to religious restrictions. Due to time constraints on grading and publishing results, they had to be in approximately the same order as for the first opportunity.

II. EXAM TIMETABLE GROUPS

Subjects that were presented in more than one group for classes were not to clash with possible electives in the exam. Therefore, more groups were needed for the exam, into which such subjects could be reclassified.

Different year level subjects were placed in different time slots. First and fourth year students wrote at the same time, *e.g.*, from 8:00 to 11:00. Second year papers were then scheduled for 12:00–15:00 and third year papers for 16:00–19:00. As it may not be advisable, for example, for third year students to write every paper at 16:00, this order could be scrambled for different days of the exam.

In the example we present, the duration of the exam was 18 days, with three time slots per day, we decided on 18×3 subject groups, one per year level (instead of the 13 groups for lectures). Subjects were placed in the same groups as for the lectures, except for subjects in which classes were repeated in different time slots. These subjects were placed in the 5×3 additional groups. In some cases, due to the structure of curricula and electives, it was possible to keep the subjects in the same group. In other cases, for subjects that are electives in virtually every curriculum (like Economics and Psychology), a subject may have to occupy a group by itself, with no other subjects in the group.

This — as was the original grouping — was done by hand initially, with the computer only used for checks. We have since developed and implemented an algorithm and a program to divide the subjects into groups.

The problem thus came down to the decision of when to schedule each group. The exam schedule then amounted to an ordering of the eighteen group numbers, for example,

7 11 5 15 2 1 17 10 3 8 14 6 9 12 4 18 16 13

which signifies that group seven is placed on day one, then group eleven, *etc*.

The order was chosen by minimizing a function describing the overall 'badness' of the schedule. In the next section we describe this function and the process of obtaining it. That is followed by a discussion of the minimization procedure.

III. A MEASURE OF THE BADNESS OF AN EXAM SCHEDULE

A. *Quantifying perceptions of students*

The perception of how bad or good an exam schedule is, is a very subjective topic. However, a measure that could be used in optimization had to be devised. To obtain such a measure, a group of students from different faculties and different year levels were selected and interviewed informally.

In the interviews they were presented with possible exam schedules and their opinions were asked. Based on their answers, the line of questioning differed from student to student: for some quantifying their feelings came easy and for others this was quite difficult.

The first step was to try and obtain an indication of what would be considered an ideal placement of 8 papers (a typical number) over the exam period of 18 days. The general consensus resulted in the scheme given in Figure 3. This shows open days as a 0, Sundays by a vertical line and days on which a paper is written as a 1:

| 1 0 | 1 0 0 1 0 1 | 0 1 0 0 1 0 | 1 0 0 1 |

Fig. 3.   An example ideal exam timetable, starting on Friday (1≡paper, 0≡open day). Sundays are indicated by a vertical line.

The original questions were of the following type:

- Do you find the following placements good or bad?

| 0 1 | 1 0 1 0 1 0 | 1 0 1 0 1 0 | 1 0 0 0 |
| 0 0 | 0 1 0 1 0 1 | 0 1 0 1 0 1 | 0 1 0 1 |

Fig. 4.   Two suboptimal exam timetables, starting on Friday (1≡paper, 0≡open day).

They were then asked to quantify how much worse a timetable as shown in Figure 4 would be. Even though the second option in this figure is basically the same as the first, except for allowing more study days at the beginning of the exam, the students preferred the first — stating as reason that their motivation for studying was exhausted near the end of such an exam. However, the opinions were not very firmly stated. Consequently the two schedules were decided to be of equal badness.

| 0 1 1 1 0 0 |
| 0 0 1 1 1 0 |
| 0 0 0 1 1 1 |
| 0 0 1 1 1 1 |
| 0 1 1 1 1 1 |
| 1 1 1 1 1 1 |

Fig. 5.   Quantifying the badness of a number of papers without open days.

- They were then asked to consider other options, as given in Figure 5. with each row in the table a possible schedule of exam papers on the days from Monday to Saturday. In each case we kept on trying to get an opinion as to how much worse the cases were.
- The next line of questioning was: As you consider the first case bad, let's give it a badness of 5. How much worse or better is the following alternatives: Double as bad? Half as bad? . . . .

It was clear from the answers that many papers crowded together was considered very bad by all. In fact, two papers in two days was not good, the placement of 3 papers in three days was very bad, 4 papers in four days was considered worse than the previous two combined. This line of thought immediately led to Fibonacci numbers, where each number is the sum of the previous two. We developed a scale for the badness based on the Fibonacci numbers and presented it to the student as a measure of how bad an exam schedule is.

The response from all was that we undervalued the badness — that the scale increases too slowly for cases of n papers in n days. Therefore we suggested the use every second Fibonacci number, to get a very rapidly increasing function of the number of days in a row on which papers are written.

This was still perceived as growing too slowly. However, by adding the badness of each subset of the series of papers — by which we mean that as a set $\boxed{1\ 1\ 1\ 1}$ contains $\boxed{1\ 1}$ three times and $\boxed{1\ 1\ 1}$ twice — we arrived at a measure that the students considered acceptable.

B. *Computing the badness of the schedule of an individual student*

Finally the following function for the badness of the timetable of an individual student was formed after repeated iterations of the process described above. Additionally, as described in the following section, the badness was applied to the student group as a whole, which caused repeated refinements and resulted in the following algorithm.

1)   a) An open day gives a badness of 0.
   b) Two open days give a badness of -1.
      (The negative sign implies an improvement.)
   c) Three open days give an additional badness of -1.
   d) More open days do not help and do not improve or worsen the badness.

   This means that a schedule $\boxed{0\ 1\ 0\ 1\ 0\ 1}$ gives a badness of 0.

2) the basic penalty for two papers on two days is 3, for three papers in three days 8, for four in four 21. *etc.*,

which is every second Fibonacci number starting at 3: $[3, 8, 21, 55, 144, 377, 987, \ldots]$.

But: as a combination of $\boxed{1\ 1\ 1}$ also contains the $\boxed{1\ 1}$ at the beginning and at the end, the penalty of 3 (for two papers in two days) is added twice, giving a total penalty of $8 + 2(3) = 11$.

The basic penalty for four papers in a row is 21, but it contains two strings of three in a row and three of two in a row, and therefore has a badness of $21 + 2(8) + 3(3) = 46$.

3) We therefore get the following rapidly increasing measure of badness:

   a) 0 1 1 0 gives a badness of 3.
   b) 0 1 1 1 0 gives a badness of 14
   c) 0 1 1 1 1 0 gives a badness of 46.
   d) 0 1 1 1 1 1 0 gives a badness of 133.
   e) 0 1 1 1 1 1 1 0 gives a badness of 364.
   f) 0 1 1 1 1 1 1 1 0 gives a badness of 972 — but this case can never occur, as there would be at least be a Sunday after 6 days of exams.

In the next section the individual badness is applied to get a measure of the badness for the schedule for all students combined.

IV. A MEASURE FOR THE BADNESS FOR THE EXAM AS A WHOLE

The fundamental idea is that the badness for every student had to be computed and added to obtain a measure for the badness as a whole.

The fundamental problem here was that if a projected schedule is good for 1000 students, but extremely bad for one student, it is still bad. On the other hand, if it is very good for 1000 and moderately bad for one student, it may be acceptable — especially as it is meant to be available before registration, and thus a student would know that his electives will give him a bad exam schedule. He can then still change his choices.

In essence the number of students in a curriculum with a certain badness was multiplied by the badness of their schedule, and all these numbers were added to obtain the overall badness. In an attempt to ensure that one or two students with a high badness number is not totally overwhelmed by the large group of students for whom the schedule is very good, a basic number was added to the number of students before multiplying. In the end it was found that this number can be as low as 10 (*i.e.*, 1 student in a curriculum was counted as 11 student-units while 100 was counted as 110 and 1000 as 1010.

This then resulted in function for the overall badness of the exam schedule, which was then minimized, as will be discussed subsequently. Because the badness is such a rapidly increasing function, the weighting of the numbers by 10 was sufficient to ensure that no really bad schedule was obtained for any student.

V. MINIMIZING THE BADNESS

The total badness function had to be minimized over all possible permutations of the placement of the groups, of the order of 10^{15} possibilities. We thought that it would be sufficient to play around with the options and that we would soon see a pattern from which we could determine the structure, based on our experience, but it proved to be a far more intractable problem than first anticipated. Simplistic methods of running through all possible permutations was clearly not feasible, as the badness function was quite expensive to compute in terms of time and the options too many. Such an exhaustive search for a minimum would have taken years on even the fastest computer.

We therefore implemented the simulated annealing method, which became famous as a solver of the travelling salesman problem [3], [4, p. 350]. This was implemented in Matlab®.

In this method there are three basic strategies:

- A randomly chosen part of the sequence is interchanged with another of the same length, starting at a randomly chosen place.
- A randomly chosen part of the sequence is reversed in order.
- If a new ordering gives an improvement, it is accepted. If it does not give an improvement but instead a higher value, it may still be accepted, if

$$e^{-I/T} > p$$

where I is the difference between the new value of the objective and the previous one, p is a random number between 0 and 1 and T is a given factor which is originally large and decreased as the optimization progresses. T corresponds to the temperature in a real annealing process.

The last — seemingly illogical — step is similar to the process in annealing in which a transfer to a higher energy state may occur at times. It avoids the program getting stranded in local minima. Removing it dramatically reduces the effectiveness of the method.

A few steps in the procedure is illustrated in Figure 6.

Option	Step	Badness	Order of groups							
Switch	13	-78210	9	5	8	14	7	15	3	$\cdots$
Rev:	14	-78675	9	5	8	14	7	15	11	$\cdots$
Rev:	15	-80161	9	5	8	14	7	15	11	$\cdots$
Rev:	16	-79677	9	5	8	14	7	11	15	$\cdots$
Rev:	17	-80173	9	11	7	14	8	5	15	$\cdots$

Fig. 6. The three basic strategies in simulated annealing

In the first row of Figure 6 the sequence from the first entry to the twelfth had been removed and inserted after the thirteenth entry.

In the third row the order of entries from number 12 to number 13 was inverted.

In the third column the fourth entry shows that a worse value was accepted and the procedure continued from this.

VI. Results

In Figure 7 the schedules of the students with the worst exam schedules, together with the number of students for which this is their schedule, are shown. The last column gives the value of the badness function for the specified exam schedule for every student. This was obtained for a student body of more than 15000 undergraduate students.

45 students:	1 0 0 1 1 1 0 0 0 0 0 0 1 0 0 1 0 0	4
8 students:	1 0 0 1 1 1 0 0 0 0 0 0 1 0 0 1 0 0	4
4 students:	1 1 1 0 0 0 0 1 1 0 0 0 1 1 0 0 0 1	2
93 students:	1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0	2
10 students:	0 0 0 0 0 0 1 1 0 0 0 1 0 1 0 0 0 0	2
1 students:	1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 0	2
9 students:	1 0 0 0 0 0 1 0 0 0 1 0 0 1 1 1 0 0	4
125 students:	0 0 1 0 1 0 1 0 0 1 1 0 0 0 0 1 1 0	5
98 students:	0 0 1 0 1 0 1 0 0 0 1 0 0 0 0 1 1 0	1

Fig. 7.   The students with the worst exam schedule, badness > 0

As can be seen in Figure 7, the worst schedule gives a badness of 5, but it affects 125 students of a specific curriculum.

Figure 8 gives a histogram of the number of curricula for values of the badness function in 24 equally spaced intervals. The most curricula have a negative badness, which implies a good schedule.
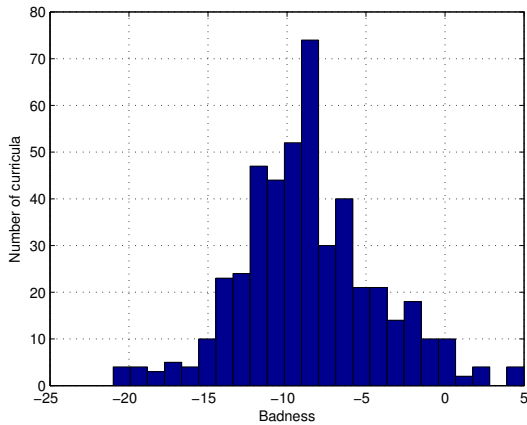
Fig. 8.   A histogram of the curricula versus badness

VII. Conclusion

As shown in Figure 7 and Figure 8, even the worst schedules are still acceptable. The method constructed good timetables for the vast majority of students. As the precise number of students was not critical in drawing up the exam timetable, it can be constructed before the start of an academic year, based on data from the previous year. The schedule satisfies all the requirements that were requested.

We therefore contend that the method of sorting subjects into non-clashing groups, and scheduling the groups according to the method describe here, is a viable and competitive way of scheduling exams in an as fair as possible way, even for curricula with just a few students.

References

[1] Potchefstroom University, *Yearbook 1999*, Potchefstroom, South Africa: PU for CHE, 1998.
[2] J. Spoelstra, (2011, August 26). Class timetables. [PDF]. Available: http://www.nwu.ac.za/export/sites/default/nwu/p-aka/Klasroosters.html
[3] Wikipedia, (2011 August 25). *Simulated annealing*. Available: http://en.wikipedia.org/wiki/Simulated_annealing
[4] Press, W.H., Flannery, B.P., Teutolsky, S.A. and Vetterling, W.H. *Numerical Recipes in Pascal: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1989.

The Formation of Dynamic Objects Trajectories in Conditions of Control Signals Saturation

Alexander V. LEBEDEV
Far-Eastern Federal University
8, Sukhanova Str., 690950, Vladivostok, Russia
avlebedev@mail.primorye.ru

ABSTRACT

The new approach to the synthesis of desirable trajectories of dynamic objects spatial motion is proposed in this paper. The method of formation of specifying influences for separate control channels of system on whole as parametrical equations of given spatial curves is developed. This method allows to provide automatically any required laws of velocity change during of the object motion on these trajectories and to decrease object velocity in the case of insufficient control resource. The analogue system for the formation of specifying influences realizing offered control laws is synthesized.

Keywords: Vehicles and transportation systems, robotics, control, nonlinear systems, spatial trajectories

1. INTRODUCTION

Numerous dynamic objects (such as underwater robots, flying and space vehicles) must move precisely on the given spatial trajectories with the various laws of velocity changes in different modes of its functioning [1-3].

The various approaches to the decision of a control problem for these objects are known. So, in the papers [4-6] the general task of the ensuring of the object motion on the given spatial trajectory is divided on two subtasks. First subtask is system stabilization rather this trajectory and second subtask is organization of a desirable internal motion on the trajectory. These subtasks are decided at the expense of application of relative and longitudinal control respectively. A weakness of described approach is necessity of construction of various complex algorithms for realization of the appropriate control influences.

Other way [7,8] consists in the ensuring of required spatial motion of objects with the help of universal robust and adaptive control laws on each degree of mobility. In this case specifying influences for separate control channels are formed as the parametrical equations of the given spatial curve. And then these signals are realized by control system with maximal precise. However there are open questions about realization of desirable trajectories

and desirable velocity modes in the conditions of power saturation of object thrusters.

As shown in the paper [9], the simultaneous ensuring of desirable trajectory and desirable velocity is a nontrivial task. The original method for this task decision is proposed in paper [9]. But this method do not takes into account control signals saturations which distort desirable trajectory if object velocity is too high.

The development of new method of dynamic objects spatial trajectories formation (i.e. the laws of program input signals changing) with the simultaneous ensuring of a given velocity modes is the main purpose of this paper. The trajectories formed by this method can be realized in the conditions of thrusters power saturations. The method proposed allows to provide the limited accelerations for each spatial coordinates of dynamic object. And this method allows to decrease automatically object velocity if desirable mode of motion requires too high accelerations.

The efficiency of synthesized system of trajectory formation is confirmed by numerical simulation.

2. THE TASK SETTING

Generally the desirable trajectory of motion of the object mass centre represents some spatial curve G (or planar curve in case of object motion in the given plane). As known, such curve is determined in absolute Cartesian coordinate system by two following equations:

$$y = g_y(x), \quad z = g_z(x), \qquad (1)$$

where x, y, z are Cartesian coordinates of the mass centre, $g_y(x)$, $g_z(x)$ are desirable functions.

Let some functions $x = x(t)$, $y = y(t)$, $z = z(t)$ satisfy to system (1), i.e. the equations (1) transform to the identities rather t at substitution of these functions. Then they are parametrical equations of curve G and they determine the required laws of changes of the mass centre coordinates.

The task of formation of the object desirable trajectory is reduced to definition of the specified functions within the framework of the considered approach. Further these functions are used as specifying signals (influences) for control subsystems of each degree of the object mobility.

It is obvious, the required functions $x = x(t)$, $y = y(t)$, $z = z(t)$ are determined in accordance with equations (1) by a not unique image if any additional requirements are absent. In particular, follow decision of this task is trivial:

$$x(t) = t, \quad y(t) = g_y(t), \quad z(t) = g_z(t). \qquad (2)$$

But the object velocity and acceleration is not controlled in this case during object motion along the trajectory which has formed above. Often required control signals may be higher than its saturated values. As a result thrusters can not be able to provide precise object motion along given trajectory with desirable velocity.

Now we shall formulate the additional requirements for the functions $x = x(t)$, $y = y(t)$, $z = z(t)$ in order to decide the problem mentioned above.

For this purpose we shall set some desirable law of changes of the absolute value v of object velocity vector during its motion on a trajectory G. Moreover we shall take into account the saturations of control force projections. These saturations can be easy transformed to inequalities for the object acceleration projections $\ddot{x}(t)$, $\ddot{y}(t)$, $\ddot{z}(t)$ on the axes of Cartesian coordinate system:

$$v = f_v(t), \qquad (3)$$
$$|\ddot{x}(t)| \leq a_{x\max}, \quad |\ddot{y}(t)| \leq a_{y\max}, \quad |\ddot{z}(t)| \leq a_{z\max}, \qquad (4)$$

where $f_v(t)$ is a desirable function, $a_{x\max}$, $a_{y\max}$, $a_{z\max}$ are maximal values of acceleration projections.

The performance of requirement (3), (4) is not guaranteed in case of a choice of specifying influences from a condition (1) only (as functions (2), for example).

Further we shall develop a new method of functions $x = x(t)$, $y = y(t)$, $z = z(t)$ definition. These functions must be the parametrical equations of a curve G and they must provide the desirable law of changes of the absolute value v of object velocity vector. Also they must provide the limitation of object acceleration during motion on given trajectory, i.e. these functions must satisfy simultaneously to equations (1) and to requirements (3), (4).

3. THE DEFINITION OF THE DESIRABLE LAWS OF COORDINATES CHANGES

First of all we shall describe shortly the original method of the formation of desirable input signals in accordance

with conditions (1), (3), which was proposed early in paper [9].

We shall use known expression for the module v of a velocity vector in absolute coordinates system for the decision of the put task. Taking into account the expression (3), let's write follow equation for the projections $\dot{x}(t)$, $\dot{y}(t)$, $\dot{z}(t)$ of object velocity vector on the axes of absolute coordinate system:

$$\sqrt{(\dot{x}(t))^2 + (\dot{y}(t))^2 + (\dot{z}(t))^2} = f_v(t). \qquad (5)$$

The derivatives of functions $x(t)$, $y(t)$, $z(t)$ must satisfy to the expression (5) in order to guarantee the desirable law (3) of velocity module changes. We shall receive the derivatives of two last functions by differentiating on time complex functions $g_y(x(t))$, $g_z(x(t))$ in the expressions (1):

$$\dot{y}(t) = g'_y(x)\dot{x}(t), \quad \dot{z}(t) = g'_z(x)\dot{x}(t). \qquad (6)$$

Here and after a symbol "'" is used for a designation derivative on coordinate (against derivative on time).

We shall substitute the expressions (6) in equation (5) and we shall have as a result:

$$\sqrt{(\dot{x}(t))^2 + (g'_y(x)\dot{x}(t))^2 + (g'_z(x)\dot{x}(t))^2} = f_v(t). \qquad (7)$$

By solving the equation (7) rather $\dot{x}(t)$, we shall receive the first order differential equation with divided variables. This equation contains required function $x(t)$ and its derivative:

$$\dot{x}(t) = \frac{f_v(t)}{\sqrt{\Phi(x)}}, \qquad (8)$$

where $\Phi(x) = 1 + (g'_y(x))^2 + (g'_z(x))^2$ is an auxiliary function.

The main idea of the new method of trajectory formation proposed in this paper is the following. The input signal $x(t)$ is generated not only from equations (1). The function $x(t)$ is found as the solution of equation (8). Note the integration of this equation is carried out with known functions $g'_y(x)$ and $g'_z(x)$, which are unequivocally determined from (1). It is always possible on intervals of a continuity of included functions. And it is easy to determine the stayed functions $y(t) = g_y(x(t))$ and $z(t) = g_z(x(t))$ by the substituting of received expression for $x(t)$ to equations (1). In accordance with this new method the desirable velocity mode (3) is provided automatically during the object motion along the trajectory formed.

4. THE METHOD OF CORRECTION OF GIVEN SIGNALS IN ACCORDANCE WITH THE CONTROL SATURATIONS

Now we shall develop the method of correction of motion laws $x = x(t)$, $y = y(t)$, $z = z(t)$ in case of desirable velocity mode requires too high acceleration, i.e. inequalities (4) are disturbed.

This method must provide the automatic variation of object velocity absolute value in order to exclude the inequalities (4) disturbance without desirable trajectory deformation.

Let's show that any given trajectory can be realized without the inequalities (4) disturbance by the reduction of velocity of control object motion only. For this we shall deduce the expressions for the first and second derivatives of the functions $x(t)$, $y(t)$, $z(t)$ formed above.

We shall substitute expression (8) to both equations (6) and we shall obtain the first derivatives $\dot{y}(t)$ and $\dot{z}(t)$:

$$\dot{y}(t) = g'_y(x)\frac{f_\upsilon(t)}{\sqrt{\Phi(x)}}, \qquad \dot{z}(t) = g'_z(x)\frac{f_\upsilon(t)}{\sqrt{\Phi(x)}} . \qquad (9)$$

Now we shall obtain the expressions for the second derivatives $\ddot{x}(t)$, $\ddot{y}(t)$ and $\ddot{z}(t)$ from (8) and (9), using the rules of differentiation of product, quotient and complex function:

$$\ddot{x}(t) = \frac{\dot{f}_\upsilon(t)}{\sqrt{\Phi(x)}} - \frac{f_\upsilon^2(t)\Psi(x)}{\Phi^2(x)}, \qquad (10)$$

$$\ddot{y}(t) = g''_y(x)\frac{f_\upsilon^2(t)}{\Phi(x)} + g'_y(x)\left(\frac{\dot{f}_\upsilon(t)}{\sqrt{\Phi(x)}} - \frac{f_\upsilon^2(t)\Psi(x)}{\Phi^2(x)}\right), \qquad (11)$$

$$\ddot{z}(t) = g''_z(x)\frac{f_\upsilon^2(t)}{\Phi(x)} + g'_z(x)\left(\frac{\dot{f}_\upsilon(t)}{\sqrt{\Phi(x)}} - \frac{f_\upsilon^2(t)\Psi(x)}{\Phi^2(x)}\right), \qquad (12)$$

where $\Psi(x) = g'_y(x)g''_y(x) + g'_z(x)g''_z(x)$ is an auxiliary function.

It is obvious, the values of functions $\Phi(x)$ and $\Psi(x)$ in the expressions (10)-(12) do not depend on function $f_\upsilon(t)$. They are determined by a kind of trajectory only. Therefore it is possible to reduce the absolute value of each addend in these expressions by a proper choice of function $f_\upsilon(t)$ (i.e. by decreasing of $\dot{f}_\upsilon(t)$ and $f_\upsilon^2(t)$ values). As a result the inequalities (4) will be fulfilled during the object motion along given trajectory.

The function $f_\upsilon(t)$ variation indicated above will not deform the desirable trajectory. It will adjust velocity and

acceleration only, because of the proposed method guarantees formation of $x(t)$, $y(t)$, $z(t)$ functions in accordance with equations (1) at any $f_\upsilon(t)$ function.

On the base of fulfilled analysis further we shall develop the new method of signal $f_\upsilon(t)$ correction, which provide the execution of given requirements.

Let first, second or third inequality (4) is disturbed at some time moment when the object move along the formed trajectory with given velocity mode $f_\upsilon(t)$. Then it is necessary to replace the function $f_\upsilon(t)$ in the equation (8) to another function $\tilde{f}_\upsilon(t)$ in order to find the current value of $x(t)$ function. This function $\tilde{f}_\upsilon(t)$ must be determined from the equations (10), (11) or (12) with the substitutions $\ddot{x}(t) = a_{x\max}\operatorname{sign}\ddot{x}$, $\ddot{y}(t) = a_{y\max}\operatorname{sign}\ddot{y}$ or $\ddot{z}(t) = a_{z\max}\operatorname{sign}\ddot{z}$ respectively.

Let's consider the formation of signal $\tilde{f}_\upsilon(t)$ in case of inequality $|\ddot{y}(t)| \le a_{y\max}$ disturbance, for example. We shall obtain the following equation after some transformations:

$$\frac{g'_y(x)}{\sqrt{\Phi(x)}}\dot{\tilde{f}}_\upsilon(t) + \frac{g''_y(x)\Phi(x) - g'_y(x)\Psi(x)}{\Phi^2(x)}\tilde{f}_\upsilon^2(t) =$$
$$= a_{y\max}\operatorname{sign}\ddot{y} . \qquad (13)$$

Then we shall write new law of function $\tilde{f}_\upsilon(t)$ formation in the final view on the base of expression (13):

$$\dot{\tilde{f}}_\upsilon(t) = \left(a_{y\max}\operatorname{sign}\ddot{y} - \frac{g''_y(x)\Phi(x) - g'_y(x)\Psi(x)}{\Phi^2(x)}\tilde{f}_\upsilon^2(t)\right)\frac{\sqrt{\Phi(x)}}{g'_y(x)} . \qquad (14)$$

It is easy to obtain the analogous expressions by using the equation (10) or (12).

Now we shall propose an alternative approach for the desirable trajectories formation. For simplicity we shall consider the case of planar motion of dynamic object. The object trajectory is given by equality $y = g_y(x)$ only.

Let's put the task of first and second inequalities (4) fulfillment (without the expression (3) satisfaction). Taking into account the first ratio (6) for $\dot{y}(t)$, we shall write the expression for $\ddot{y}(t)$ in the view:

$$\ddot{y}(t) = g''_y(x)(\dot{x}(t))^2 + g'_y(x)\ddot{x}(t) . \qquad (15)$$

We shall obtain the expression for x from the first equation (1):

$$x = g_x(y), \tag{16}$$

where $g_x(y)$ is a reverse function for the function $g_y(x)$.

Let's obtain expressions for $\dot{x}(t)$ and $\ddot{x}(t)$ from (16):

$$\dot{x}(t) = g'_x(y)\dot{y}(t), \tag{17}$$

$$\ddot{x}(t) = g''_x(y)(\dot{y}(t))^2 + g'_x(y)\ddot{y}(t). \tag{18}$$

We shall suppose that $u_x = \ddot{x}(t)$ and $u_y = \ddot{y}(t)$ are given signals in some system of desirable trajectories formation. Then we shall obtain searched functions $x(t)$, $y(t)$ after double integration of those signals.

In order to object always move along desirable trajectory with maximal possible accelerations (without this trajectory deformation and conditions (4) violation) it is sufficient to provide the following law of signals u_x, u_y formation:

$$u_x = g''_x(y)(\dot{y}(t))^2 + g'_x(y)u_y, \quad \text{if } |u_y| = a_{y\max},$$

$$u_x = a_{x\max}, \qquad\qquad\qquad \text{if } |u_y| < a_{y\max},$$

$$u_y = g''_y(x)(\dot{x}(t))^2 + g'_y(x)u_x, \quad \text{if } |u_x| = a_{x\max},$$

$$u_y = a_{y\max}, \qquad\qquad\qquad \text{if } |u_x| < a_{x\max}. \tag{19}$$

Obviously the maximal fast motion is provided in the case of law (19) application.

5. THE SYNTHESIS OF THE SYSTEM OF DESIRABLE TRAJECTORIES FORMATION

Now we shall develop the original structure of the system of desirable trajectories formation. This system generates specifying influences $x(t)$, $y(t)$, $z(t)$ for the dynamic object in accordance with the method proposed in this paper.

The structural scheme of such system is given in fig. 1. Here $g_y(x)$, $g_z(x)$, $g'_y(x)$ and $g'_z(x)$ are blocks realizing appropriate functional dependences, $\delta = 1$ and μ is a constant signal. Only one of three channels of signal $\tilde{f}_\upsilon(t)$ formation is shown in fig. 1.

This scheme contains four interconnected subsystems. The subsystem of trajectory formation generates signals $x(t)$, $y(t)$, $z(t)$ on the basis of the information about the desirable law $f_\upsilon(t)$ or $\tilde{f}_\upsilon(t)$ of velocity change. Also the information about the concrete kind of functions $g_y(x)$ and $g_z(x)$ is taken into account. It is necessary to create the special subsystem realizing an offered method of formation of a desirable trajectory, because of the features of the differential equation (8) do not allow to receive required function $x(t)$ in an analytical kind in

most cases (even for enough simple dependences $g_y(x)$ and $g_z(x)$).

The various approaches to such subsystem construction are possible. The equation (8) can be solved numerically with application of known ways of integration. In this case subsystem of trajectory formation can be realized on the basis of the specialized processor or computer. As the alternative approach, the subsystem of trajectory formation is synthesized as a special analogue device in the given paper.

The subsystem of function $f_\upsilon(t)$ calculation and the subsystem of auxiliary functions calculation provide the correction of desirable law $f_\upsilon(t)$ of velocity change in order to guarantee the conditions (4). These subsystems are also constructed as the analogue devices on the base of expression (14) or similar one.

The switching subsystem provides the replacement of signal $f_\upsilon(t)$ to another signal $\tilde{f}_\upsilon(t)$ on the input of dividing block in case of inequalities (4) disturbance. Signal μ takes the value 1 if inequalities (4) are fulfilled and it takes the value 0 if any of these inequalities are disturbed.

We shall note that the opportunity of readjustment must be stipulated in the given system for realization of different functions $g_y(x)$, $g_z(x)$ and their derivatives at formation of various trajectories.

6. THE NUMERICAL SIMULATION RESULTS

The results of numerical simulation of synthesized system in case of planar parabolic trajectory formation (when $y = x^2$, $g_y(x) = x^2$) are shown in fig. 2. The analogues results in case of planar sinusoidal trajectory formation (when $y = \sin x$, $g_y(x) = \sin x$) are shown in fig. 3. The constant value $\upsilon = 2$ m/sec is given as a desirable law of object velocity module variation, i.e. $f_\upsilon(t) = 2$. The values $a_{x\max} = a_{y\max} = 0.5$ m/sec^2 and $a_{x\max} = a_{y\max} = 1$ m/sec^2 are given as saturations of the object acceleration. The generated trajectories $y(x)$ of the dynamic object are submitted in fig. 2a and 3a. The time dependences of coordinates $x(t)$ (curve 1) and $y(t)$ (curve 2) are shown in fig. 2b and 3b. The time dependences of velocity module $\upsilon(t)$ are shown in fig. 2c and 3c. The time dependences of accelerations $\ddot{x}(t)$ (curve 1) and $\ddot{y}(t)$ (curve 2) are shown in fig. 2d and 3d. As numerical simulation results show, formation of desirable trajectory $y(x)$ and law $f_\upsilon(t)$ of changes of object velocity are provided automatically in both cases. Developed system also provides necessary correction of this velocity law and guarantees fulfilment of saturations (4) on the acceleration projections.
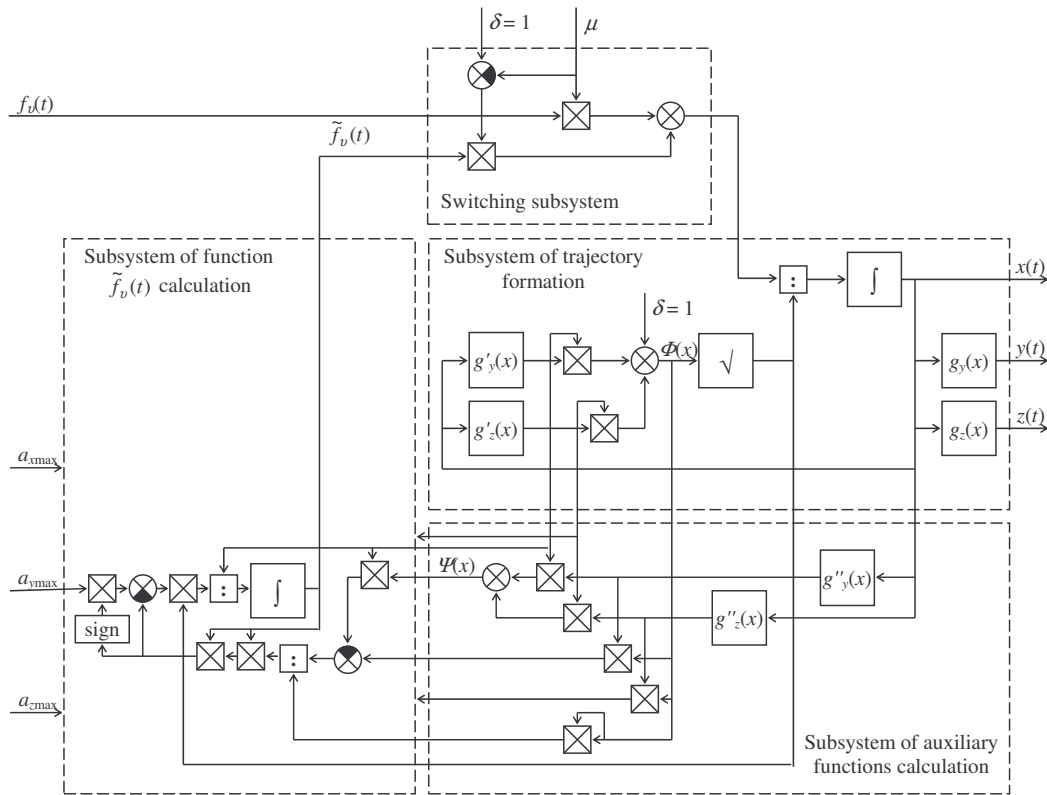
Fig. 1. The structural scheme of the synthesized system

7. CONCLUSIONS

New approach to the construction of desirable trajectories of dynamic objects spatial motion is proposed in this paper. The new method of formation of specifying influences for separate control channels as parametrical equations of the given spatial curves is developed. The originality of this method is the following. Any required laws (3) of velocity change are guaranteed automatically during the object motion on these trajectories by the using of equation (8) solution. Automatic variations of velocity mode of motion in accordance with nontrivial expression (14) are provided in case of insufficient control resource. As a result the conditions (4) are always fulfilled. The original scheme of analogue system of trajectory formation realizing offered control laws is synthesized. Efficiency of developed method is confirmed by numerical simulation results. The simplicity of created control laws allow to use them at designing of technical objects intended for complex and precise spatial motion.

REFERENCES

[1] Craig J.J. **Introduction to Robotics**, Addison-Wesley Publishing Company, 1989.

[2] Ramadorai, A.K., and Tarn, T.J. On Modelling and Adaptive Control of Underwater Robots, **Journal of Robotic Systems**, *5*(1), 1993, 47-60.

[3] Dubowsky S., Papadopoulos E. The Kinematics, Dynamics and Control of Free-Flying and Free-Floating Space Robotic Systems, **IEEE Tran. on Robotics and Automation**, 1993, Vol.9, No5, 531-543.

[4] Yoerger, D.R., and Slotine, J.-J.E. Adaptive Sliding Control of An Experimental Underwater Vehicle, **Proc. of IEEE International Conf. on Robotics and Automation**, Sacramento, USA, 1991, 2746-2751.

[5] Suzuki H., Yoshida K. Trajectory Tracking Control of A ROV for Lifting Objects, **Proc. of 1st International Offshore and Polar Engineering Conf.**, 1990, 545-552.

[6] Yoerger D.R., Slotine J.-J.E. Robust Trajectory Control of Underwater Vehicles, **IEEE Journal of Oceanic Engineering**, 1985, Vol.OE.-10, No4, 462-480.

[7] Lebedev A.V., Filaretov V.F. The Synthesis of Multi-Channel Adaptive Variable Structure System for the Control of AUV, **Proc. of The IEEE/RSJ International Conf. on Intelligent Robots and Systems**, Nice, France, 2008, 2834-2839.

[8] Sanyal A., Chyba M. Robust Feedback Tracking of Autonomous Underwater Vehicles with Disturbance Rejection, **Proc. of American Control Conference**, St. Louis, MO, USA, 2009, 3585-3590.

[9] Lebedev A.V. Synthesis of Desirable Trajectories of Dynamic Objects Spatial Movement, **Proc. of The 6th IASTED International Conf. Intelligent Systems and Control**, Honolulu, Hawaii, USA, 2004, 36-39.

a)

b)

c)

d)
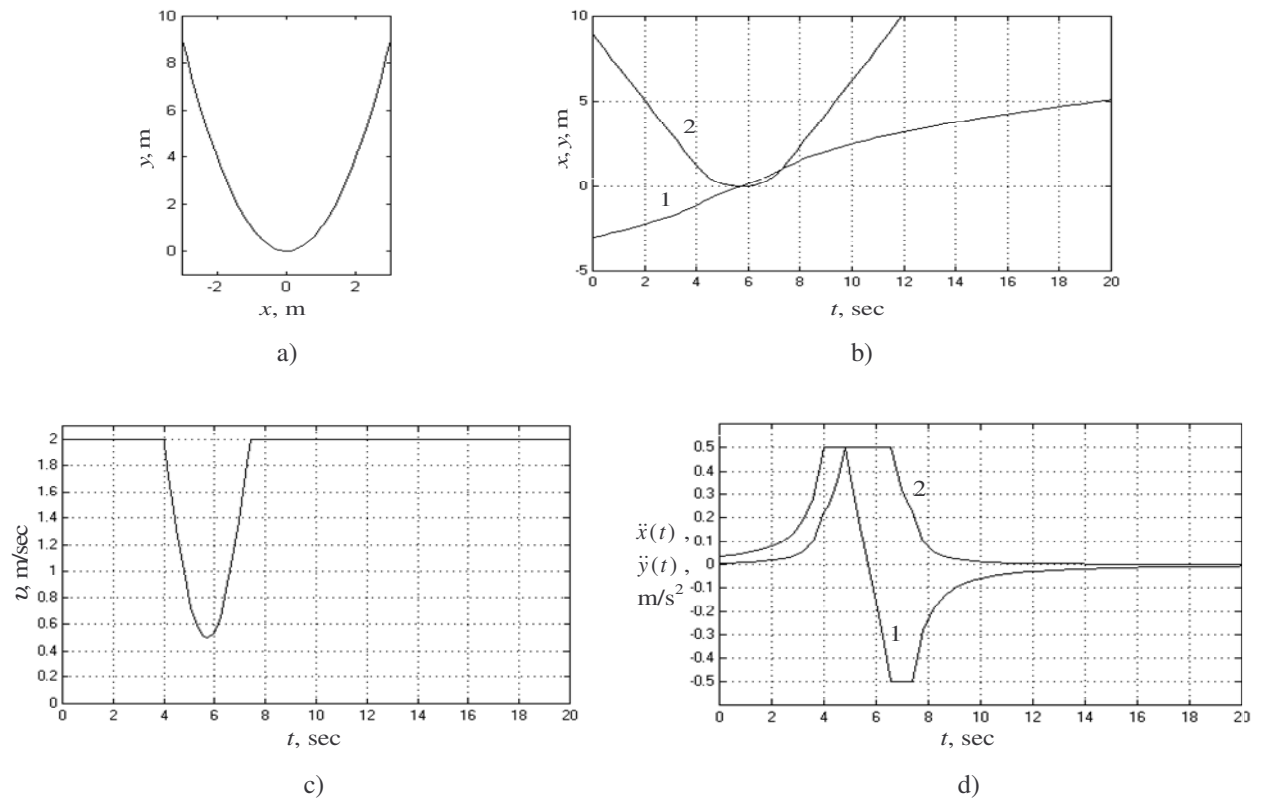
Fig. 2. The formation of parabolic desirable trajectory

a)

b)

c)

d)
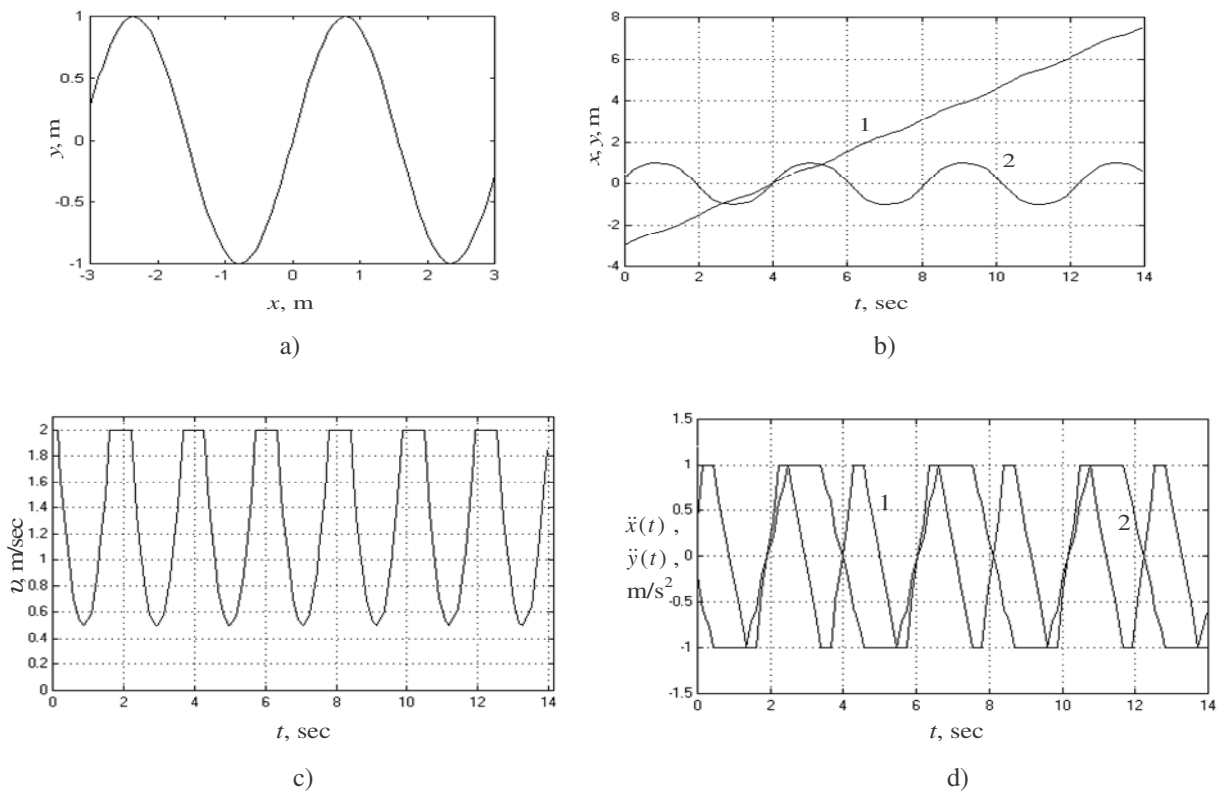
Fig. 3. The formation of sinusoidal desirable trajectory

Adaptive Control of a PPR Planar Manipulator with One Passive Joint

Licheng WU
School of Information Engineering, Minzu University of China
Beijing, 100081, China

and

Xiali LI
School of Information Engineering, Minzu University of China
Beijing, 100081, China

ABSTRACT

In this paper, a direct adaptive control scheme of a planar under-actuated manipulator with parameter uncertainty is presented. The dynamic equation of the manipulator is rewritten in operational space using augmented Jacobian matrix method. Based on this dynamic equation, a direct adaptive control scheme to stabilize the entire system to track a desired trajectory against parameter uncertainty is proposed by Lyapunov stabilization theory. A simulation example is included to demonstrate the results.

Keywords: Under-actuated, Planar manipulator, Adaptive control, Passive joint.

1. INTRODUCTION

The number of degrees of freedom of a conventional manipulator is equal to the number of joint actuator. Planning and control of the under-actuated manipulator with passive joints have attracted a lot of attention in the recent past [3-6][13] to reduce the weight, cost, and energy consumption of the manipulator. Moreover, a fully actuated manipulator systems become into under-actuated systems when one or more actuators fail, In this case, it is very meaningful to study the control for the under-actuated systems.

The under-actuated manipulator systems belong to a class of nonholonomic systems, in which constraints include kinematical constraints that geometrically restrict the direction of mobility or dynamic constraints due to dynamic balance at passive degree of freedom where no force or torque is applied. The dynamic constraints introduced by passive joints are also presented as a differential equation including the generalized coordinates, the generalized velocities, and the generalized accelerations, and called 2nd-order nonholonomic constraints. These constraints make it difficult to directly apply the mathematical methods developed so far for 1st-order nonholonomic systems.

A general theory for planning and control of planar under-actuated manipulators is not yet available, and most of the methods so far proposed rely on the specific case considered. The case of 2R planar robots with a single actuator in the presence of gravity has been considered, e.g., in [5], [7] and [8] another case that has been studied in details are three-link planar manipulators with a passive rotational third joint in the absence of gravity. Oriolo et al. [9] showed how to make the system fully linearizable and input-output decoupable by means of a nonlinear dynamic feedback. Arai et al. [10] showed how to plan rest-to-rest motions through a sequence of elementary maneuvers consisting of pure translations of the third link or pure rotations around its center percussion (CP). Imura et al. [11] proved that the system could be transformed into 2nd-order chained form via static state feedback. Kobayashi et al. [12] have showed the n-link planar manipulator where the first joint is actuated and only one of the other joints is passive is completely controllable, etc.

However, in most of their study, the dynamic parameters of equation of motion of the under-actuated manipulator are certainty. But we know, the modeling error and the external disturbance always exist in real plant. To deal with the dynamic parameters uncertainty, an adaptive control scheme is needed. Many researchers [1][2] developed the adaptive control for the fully actuated systems, but these schemes are not completely applicable to the under-actuated systems due to the nonlinear parameterization.

In this paper, we introduce a adaptive control to a PPR planar manipulator with one passive joint to track a desired trajectory against parameter uncertainty. The remainder of this paper is organized as follows: In section II the equation of motion with operational coordinates are presented. Direct adaptive control scheme is considered in section III. Simulation results are in section IV, Conclusion is considered in section V.

2. DYNAMIC MODEL

The PPR planar Under-actuated manipulator system is shown in Fig.1. It consists of two active joints and one passive joint. The first and second prismatic joints are orthogonal and the third is a revolute joint. The generalized coordinate vector of the manipulator is supposed as $q = \begin{bmatrix} r_x & r_y & \theta \end{bmatrix}^T$, i.e. r_x and r_y are the position, and θ is the orientation of the 3rd link. Let $u = (f_1, f_2)^T$ be the generalized torque in the active joint and the torque in passive joint be zero, $m_i (i = 1,2,3)$ be the mass of the link i and I_3 be the inertia of the third link, and d be the distance between the center of the mass of the 3rd link and the 3rd joint. Then the equation of motion of the planar manipulator under no friction in joint space can be written as

$$W(q)\ddot{q} + C(q,\dot{q})\dot{q} = \tau \tag{1}$$

where $W = \begin{pmatrix} m_1 + m_2 + m_3 & 0 & -m_3 d \sin\theta \\ 0 & m_2 + m_3 & m_3 d \cos\theta \\ -m_3 d \sin\theta & m_3 d \cos\theta & I_3 + m_3 d^2 \end{pmatrix}$,

$C = \begin{pmatrix} 0 & 0 & -m_3 d \dot{\theta} \cos\theta \\ 0 & 0 & -m_3 d \dot{\theta} \sin\theta \\ 0 & 0 & 0 \end{pmatrix}$, $\tau = \begin{pmatrix} u^T & 0 \end{pmatrix}^T$.
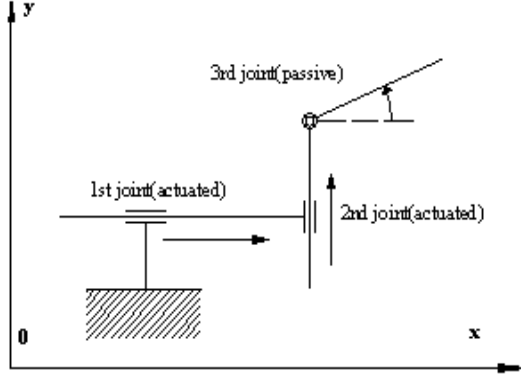
Fig.1 Scheme of the three link planar manipulator.

C has the following property:

$$z^T C z = \frac{1}{2} z^T \dot{W} z \tag{2}$$

for an arbitrary vector $z \in \Re^3$. Inertia matrix W can be portioned into four blocks,

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \tag{3}$$

where W_{11} is 2 by 2 symmetric sub-matrix, W_{22} is 1 by 1 symmetric block, and $W_{12} = W_{21}^T$ is 2 by 1 sub-matrix representing the interaction between active joints and passive joint.

Likewise, Joint displacement vector q can be portioned into two blocks,

$$q = \begin{bmatrix} q_a^T & q_p^T \end{bmatrix}^T \tag{4}$$

Where q_a is 2 degrees of freedom for active joints, q_p is 1 degree of freedom of passive joint.

The Jacobian Matrix

Since the operating task for the manipulator task is usually specified in term of operational coordinates as $p = h(q) \in \Re^2$, the manipulator tip position with respect to the fixed based is

$$p = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r_x + 2d \cos\theta \\ r_y + 2d \sin\theta \end{pmatrix} \tag{5}$$

Then, Jacobian matrix is determined by

$$J = \frac{\partial h}{\partial q} = \begin{pmatrix} J_1 & J_2 \end{pmatrix} \tag{6}$$

where $J_1 = \begin{pmatrix} \frac{\partial h}{\partial r_x} & \frac{\partial h}{\partial r_y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$,

$J_2 = \frac{\partial h}{\partial \theta} = \begin{pmatrix} -2d \sin\theta & 2d \cos\theta \end{pmatrix}^T$.

Equation of motion with operational Coordinates

The equation of motion of the manipulator is rewritten in terms of operational coordinates p. From eq. (6), J is a 2 by 3 matrix. Thus we define an augmented operational coordinates vector $y_e = \begin{pmatrix} p \\ q_p \end{pmatrix} = h_e(q)$, and its time-derivative

$$\dot{y}_e = \begin{pmatrix} \dot{p} \\ \dot{q}_p \end{pmatrix} = \begin{pmatrix} J_1 & J_2 \\ 0 & 1 \end{pmatrix} \dot{q} = J_e \dot{q} \tag{7}$$

where $J_e(q) \in \Re^{3 \times 3}$ is an augmented Jacobian matrix.

When eq. (7) is differentiated with respect to time, we obtain

$$\ddot{y}_e = J_e \ddot{q} + \dot{J}_e \dot{q} \tag{8}$$

If J_1 in eq. (6) is nonsingular, using eq. (8) and substituting $\ddot{q} = J_e^{-1}(\ddot{y}_e - \dot{J}_e \dot{q})$ into eq. (1), We obtain

$$W J_e^{-1} \ddot{y}_e - W J_e^{-1} \dot{J}_e \dot{q} + C\dot{q} = \begin{pmatrix} u \\ 0 \end{pmatrix} \tag{9}$$

Substituting $\dot{q} = J_e^{-1} \dot{y}_e$ into eq. (9), we obtain the dynamic equation of the manipulator in operational space as

$$M\ddot{y}_e + N\dot{y}_e = J_e^{-T} \begin{pmatrix} u \\ 0 \end{pmatrix} \tag{10}$$

where $M = J_e^{-T} W J_e^{-1}$, $N = J_e^{-T} C J_e^{-1} - M\dot{J}_e J_e^{-1}$.

3. DIRECT ADAPTIVE CONTROL SCHEME

We now show that if $q_p, \dot{q}_p, \ddot{q}_p$ are exactly measurable for this manipulator, then a direct adaptive control scheme can asymptotically stabilize the entire system to a desired trajectory against parameter uncertainty.

We define the extended output error function between the desired $(y_e)_d = \begin{pmatrix} y_d & q_p \end{pmatrix}^T$ and the actual $y_e = \begin{pmatrix} y & q_p \end{pmatrix}^T$ as $e_e = (y_e)_d - y_e = \begin{pmatrix} e \\ 0 \end{pmatrix}$, let

$$s = \dot{e}_e + k_p e_e = \begin{pmatrix} \dot{e} + k_p e \\ 0 \end{pmatrix} \tag{11}$$

where $e = y_d - y$ is the output error function and $k_p > 0$ are constant gains matrix.

We define reference output velocity η and reference output acceleration $\dot{\eta}$ as follow,

$$\eta = \begin{pmatrix} \dot{y}_d + k_p e \\ \dot{q}_p \end{pmatrix} \tag{12}$$

and

$$\dot{\eta} = \begin{pmatrix} \ddot{y}_d + k_p \dot{e} \\ \ddot{q}_p \end{pmatrix} \tag{13}$$

Comparing eq.s (12), (13) with eq. (11) yields

$$s = \eta - \dot{y}_e \tag{14}$$

and

$$\dot{s} = \begin{pmatrix} \ddot{e} + k_p \dot{e} \\ 0 \end{pmatrix} = \dot{\eta} - \ddot{y}_e \tag{15}$$

Now we define a control law as follows,

$$M_m \dot{\eta} + N_m \eta + \begin{pmatrix} K_v(\dot{e} + k_p e) \\ \sigma \end{pmatrix} = J_e^{-T} \begin{pmatrix} u \\ 0 \end{pmatrix} \tag{16}$$

where M_m and N_m represent the inertial matrix M and the matrix N in a model plant of the planar manipulator, respectively, and

$$M_m = J_e^{-T} W_m J_e^{-1} \tag{17}$$

$$N_m = J_e^{-T} C_m J_e^{-1} - M_m \dot{J}_e J_e^{-1} \tag{18}$$

In eq. (16), k_v is a 2 by 2 positive-definite and symmetric constant weighting matrix.

The equation eq. (17) can be decomposed into two parts

$$\begin{pmatrix} W_{21} & W_{22} \end{pmatrix}_m J_e^{-1} \dot{\eta} + \begin{pmatrix} J_2^T & 1 \end{pmatrix} N_m \eta + J_2^T K_v(\dot{e} + k_p e) + \sigma = 0 \tag{19}$$

and

$$\begin{pmatrix} W_{11} & W_{12} \end{pmatrix}_m J_e^{-1} \dot{\eta} + \begin{pmatrix} J_1^T & 0 \end{pmatrix} N_m \eta + J_1^T K_v(\dot{e} + k_p e) = u \tag{20}$$

From eq. (19), we can find σ can only be evaluated to ensure eq. (19) disappear. Let ψ be the parameter column vector which list all real physical objective parameter should be identified.

Let ψ_m be the corresponding parameter vector for the model plant of the planar manipulator.

To derive the direct adaptive, we define a Lyapunov function

$$V = \frac{1}{2} s^T M s + \frac{1}{2} \phi^T \Gamma \phi \tag{21}$$

Where $\phi = \psi - \psi_m$, Γ is a constant adaptive gain and is positive definite and symmetric.

Clearly, $V \geq 0$, and $V = 0$ if and only if $\begin{pmatrix} e & \dot{e} \end{pmatrix}^T = 0$ and $\phi = 0$. Taking time-derivative for V, we obtain

$$\dot{V} = s^T M \dot{s} + \frac{1}{2} s^T \dot{M} s + \dot{\phi}^T \Gamma \phi$$

$$= s^T M \dot{\eta} - s^T M \ddot{y}_e$$

$$- s^T M \dot{J}_e J_e^{-1} s + \frac{1}{2} s^T J_e^{-T} \dot{W} J_e^{-1} s + \dot{\phi}^T \Gamma \phi$$

$$= s^T M \dot{\eta} + s^T N \ddot{y}_e - s^T J_e^{-T} \begin{pmatrix} u \\ 0 \end{pmatrix} - s^T M \dot{J}_e J_e^{-1} s$$

$$+ s^T J_e^{-T} C J_e^{-1} s + \dot{\phi}^T \Gamma \phi$$

$$= s^T M \dot{\eta} - s^T J_e^{-T} \begin{pmatrix} u \\ 0 \end{pmatrix} + s^T N \eta + \dot{\phi}^T \Gamma \phi$$

$$= s^T \left[(M - M_m)\dot{\eta} + (N - N_m)\eta \right]$$

$$- s^T \begin{pmatrix} k_v(\dot{e} + k_p e) \\ \sigma \end{pmatrix} + \dot{\phi}^T \Gamma \phi$$

$$= s^T Y \phi - s^T \begin{pmatrix} k_v(\dot{e} + k_p e) \\ \sigma \end{pmatrix} + \dot{\phi}^T \Gamma \phi \tag{22}$$

where $Y\phi = \left[(M - M_m)\dot{\eta} + (N - N_m)\eta \right]$, and Y is a matrix function of $q, \dot{q}, \ddot{q}_p, y_d, \dot{y}_d$ and $\ddot{y}_d$, and is independent of the objective physical parameters.

From eq. (22), we can define an adaptive law for the planar manipulator as follow

$$\dot{\phi} = -\Gamma^{-1} Y^T s \tag{23}$$

Then we obtain

$$\dot{V} = -(\dot{e} + k_p e)^T K_v (\dot{e} + k_p e) \tag{24}$$

which is negative-definite and is zero when $\dot{e} + k_p e = 0$, that is to say $s = 0$.

Therefore, the control law eq. (20) and the adaptive law eq. (23) asymptotically stabilize the planar manipulator system to track a desired trajectory described in terms of $y_d, \dot{y}_d, \ddot{y}_d$.

4. SIMULATION RESULTS

We show a simulation result in order to verify the effectiveness of the adaptive control scheme. We consider a three link planar manipulator with two prismatic actuated joints and one revolute passive joint as shown in Fig.1.Total three parameters involved in W are to be identified. The parameter vector ψ can be defined as $\psi = (m_x, m_y, m_3)^T$, The real parameters are fixed to be $\psi = \begin{pmatrix} 10 & 5 & 5 \end{pmatrix}^T$, all the model plant parameters in ψ_m used for adaptive control simulation are simply defined to

be $\begin{pmatrix} 2 & 2 & 2 \end{pmatrix}^T$, and the control gains are set to be $k_p = 4$, $K_v = 10I$, The length of the third link is $2\sqrt{2}$.

The desired trajectory of the tip is supposed as $x_d = e^{-2t}$, $y_d = e^{-t}$.

As the simulation results, Fig 2 shows the position tracking errors versus time with adaptive control, Fig. 3 shows Parameter estimation result.
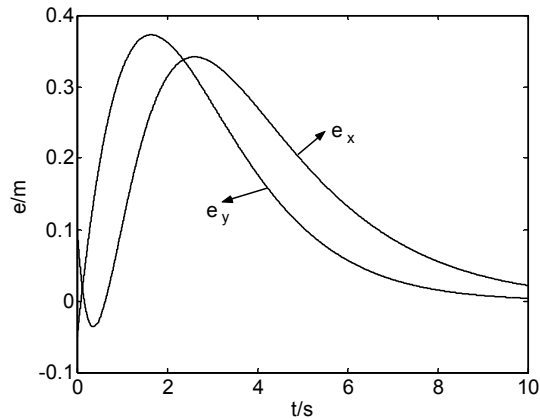
Fig. 2 Position tracking error.
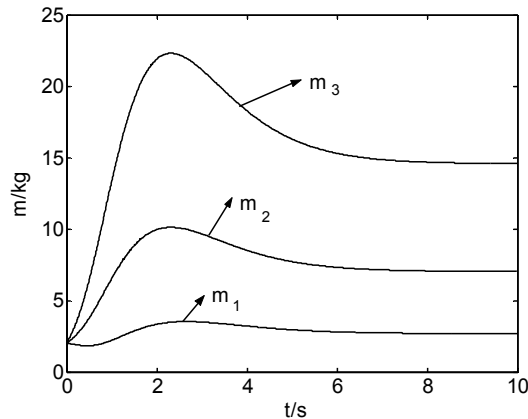
Fig.3 Parameter estimation result.

5. CONCLUSIONS

We have introduced an adaptive control law scheme for the PPR planar under-actuated manipulator with a passive joint to track a desired trajectory against the parameter uncertainty. We have shown the asymptotical stability of the scheme for the system. The stability is guaranteed if the square Jacobian matrix is invertible and all the passive joint positions, velocities and accelerations are measurable and bounded.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] S. Sastry and M. Bodson, **Adaptive Control: Stability, Convergence, and Robustness**, Prentice Hall, 1989.

[2] J. J. E. Slotine, W. P.Li., "On the adaptive control of robot manipulators. **Int. J. of Robotics,** vol. 6, no, 3, 1987, pp. 49-59.

[3] J.D. English, A.A. Maciejewski., "Fault tolerance for kinematically redundant manipulators: anticipating free-swinging joint failures," **IEEE Trans. on Robotics and Automation**, vol. 14, no. 4, 1998, pp. 566-575.

[4] G. Oriolo, Y. Nakamura, "Free-joint manipulators: motion control under second-order nonholonomic constraints," **Proc. IEEE/RSJ Int. Workshop on Intelligent Robots and Systems**, 1991, pp. 1248-1253.

[5] A. De Luca, R. Mattone, and G. Oriolo, "Control of underactuated mechanical systems: Application to the planar 2R robot," **Proc. 35th IEEE Int. Conf. Decision and Control**. 1996, pp. 1455-1460.

[6] H. Arai, K. Tanie and S. Tachi, "Dynamic control of a manipulator with passive joints in operational space," **IEEE Trans. Robotics and Automation**, Vol. 9, No. 1, 1993, pp. 85-93.

[7] A. De Luca. R. Mattone, and G. Oriolo, "Stabilization of an underactuated planar 2R manipulator," **Int. J. Robust and Nonlinear Control**, no.4, 2000, pp. 181-198.

[8] H. Arai and S. Tachi, "Position control system of a two degree of freedom manipulator with a passive joint," **IEEE Trans. on Industrial Electronics**, vol. 38, no. 1, 1991, pp.15-20.

[9] A De Luca and G. Oriolo, "Motion planning and trajectory control of an underactuated three-link robot via dynamic feedback linearization," **IEEE Int. Conf. on Robotics and Automation**, 2000, pp. 2789-2795.

[10] H. Arai, K.Tanie, and N. Shiroma, "Nonholonomic control of a three-dof planar underactuated manipulator," **IEEE Trans. on Robotics and Automation**. vol.14, no. 5, 1998, pp. 681-695.

[11] J. Imura, K. Kobayashi, and T. Yoshikawa, "Nonholonomic control of 3 link planar manipulator with a free joint," **35th IEEE Conf. on Decision and Control**, 1996, pp. 1435-1436.

[12] K. Kobayashi, T. Yoshikawa, "Controllability of under-actuated planar manipulators with one unactuated joint," **Proc. 2000 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems.** 2000, pp. 133-138.

[13] H. Arai and S. Tachi, "Position control of a manipulator with passive joints using dynamic coupling," **IEEE Trans. on Robotics Automation**, 1991, pp. 528-534.

Effectiveness and Utility of a Case-Based Model for Delivering Engineering Ethics Professional Development Units

LA-UR 14-20045

Heidi Ann HAHN

Engineering Sciences Directorate, Los Alamos National Laboratory
Los Alamos, NM, 87545, USA

ABSTRACT

This article describes an action research project conducted at Los Alamos National Laboratory (LANL) to resolve a problem with the ability of licensed and/or certified engineers to obtain the ethics-related professional development units or hours (PDUs or PDHs) needed to maintain their credentials. Because of the recurring requirement and the static nature of the information, an initial, in-depth training followed by annually updated refresher training was proposed. A case model approach, with online delivery, was selected as the optimal pedagogical model for the refresher training. In the first two years, the only data that was collected was throughput and information retention. Response rates indicated that the approach was effective in helping licensed professional engineers obtain the needed PDUs. The rates of correct responses suggested that knowledge transfer regarding ethical reasoning had occurred in the initial training and had been retained in the refresher. In FY13, after completing the refresher, learners received a survey asking their opinion of the effectiveness and utility of the course, as well as their impressions of the case study format vs. the typical presentation format. Results indicate that the courses have been favorably received and that the case study method supports most of the pedagogical needs of adult learners as well as, if not better than, presentation-based instruction. Future plans for improvement are focused on identifying and evaluating methods for enriching online delivery of the engineering ethics cases.

Keywords: Case Studies in Education, Ethics Case Studies, Engineering Case Studies

1. INTRODUCTION

As one of the premier research laboratories in the United States, Los Alamos National Laboratory (LANL) employs many engineers who have become credentialed within their discipline either by attaining a Professional Engineer (PE) license or through certification conferred by a professional society. Generally, maintaining such licenses and certifications requires participation in continuing education to obtain professional development units or hours (PDUs or PDHs). PEs registered in the State of New Mexico (NM) are required to obtain at least four PDHs in ethics biennially [1].

As part of its contractual obligation to the US government, the Laboratory provides a mandatory one-hour general ethics course (on topics such as conflict of interest, raising and resolving ethical issues, and standards of conduct and business ethics) online to all workers annually. These courses enable PEs to claim two PDHs biennially, leaving a gap of two PDHs every two years.

LANL is located in rural Northern NM, approximately 100 miles from the nearest commercial airport. There is little in the way of vendor-provided engineering-specific ethics training available in the area, and budget constraints have made supporting travel for training difficult. Distance education courses, while widely available, do not address company-specific ethics policies or federal requirements impacting the practice of engineering at LANL.

Given the large target population (about 120) who need ethics PDHs on a biennial basis, it was decided that in-house delivery of engineering-ethics training that could be used to fulfill PDH requirements was the preferred solution. Because of the recurring nature of the requirement and the static nature of the information (i.e., the core principles of engineering ethics are relatively constant), it was determined that workers should be exposed to an initial, in-depth training followed by annually updated refresher training (which is defined as a "short-term course aimed at recall and reinforcement of previously acquired knowledge and skills" [2]).

Developed in accordance with the Systematic Approach to Training (SAT), the initial training covers the elements of the NM *Code of Professional Conduct – Engineering and Surveying* (NMAC); ethical obligations to the engineering profession and other professionals; and various federal legal requirements, most especially export control law, that have the potential to impact the practice of engineering at the Laboratory. It has been delivered both in classroom and online settings. Although the initial training does incorporate some case-based "test your knowledge" exercises, it is primarily a lecture- or presentation-based pedagogical model.

Research in the field of "andragogy" (the art of teaching adults [3]) dating back to the 1970's, however, suggests that lectures, and especially lectures in which the same information is repeated, may not be the ideal instructional model for adult learners. Therefore, as we designed the engineering ethics refresher training, we looked to other instructional designs.

2. RATIONALE FOR SELECTION OF THE CASE MODEL APPROACH FOR CONTINUING EDUCATION

Knowles' model, which has evolved over the years (his seminal work is now in its seventh edition [4]) contains six core andragogical principles that influence instructional design decisions for adult learners:

1) the adult learners' need to know not only the subject matter, but also to understand the underlying why, what, and how
2) the tendency toward movement from dependency upon an instructor to greater autonomy and self-directedness as learners age
3) the prior experience of the learner – particularly their mental models – and the need to use techniques that incorporate the adult learners' experience base as an integral part of instruction
4) the adult's orientation to learning being problem-centered and contextual
5) the dependency of adults' readiness to learn on the developmental phases associated with the various roles that they play/have played in their (professional) lives
6) the basis of adults' motivation to learn being in intrinsic value and personal payoff

Case studies are an ideal method for adult instruction. In the case method, knowledge is acquired while dealing with a real-life problem and not in isolation of its context [5], consistent with the fourth of Knowles' principles listed above. "Although the case method does not actually provide real experiences, it is personal as it puts the burden of thinking on the learners and arouses their interest by making them active participants [6]." This characteristic is responsive to both the adult learner's need for self-directedness and for engaging his/her own experience base. While development of skills generally requires an element of actually doing the skill-based activity, the case method provides the opportunity for skill development through presentation of different cases that exercise the same skill over a period of time [5]. Finally, because cases involve real people with real problems, they are also more likely to stimulate adult learners than are subject matter-based lectures or texts [7]. The narrative format encourages learner engagement [5].

3. DESIGN OF THE ENGINEERING ETHICS CASE STUDIES

Clark [6] describes case studies as being of one of two forms. The first type uses short and specific situations in which the problem is apparent. The learner is asked to demonstrate his/her problem solving ability by applying principles that have been taught previously. The second type focuses on appreciating different perspectives on a situation. This type provides complex information that requires deep analysis and focuses on problem identification as well as finding solutions. The Food and Agriculture Organization of the United Nations (FAO) (see reference [5]) refers to these two types as "caselets" and "comprehensive cases," respectively.

The case studies used in the LANL project are more like Clark's [6] first type, as they are typically short and require the learner to apply the knowledge gained from the initial training. Many of the case studies used in the refreshers do, however, share one characteristic of Clark's second type, namely, the need to consider multiple perspectives on the problem. As the case study unfolds, the learner may be asked to take the position of the involved worker, coworkers, consultants, or managers, exercising Knowles' [4] principle regarding an adult's readiness to learn being dependent upon the learner's phase of development in various roles.

Because the refreshers are delivered online, one of the integral aspects of the case method, discussion with a group of co-learners [5], is lost. Recognizing that much of the value of the discussion is in the feedback provided to the learners to positively reinforce learning [5], the courseware was designed with branching, which takes the learner to different paths through the material. Selection of the "best" response leads either additional segments of the case or to a new case. Selection of a non-optimal response leads to feedback as to why the response is not the best option and, in some cases, the opportunity to further explore the rationale underlying the "best" response by answering additional questions.

Learning Objectives

Cases were selected to reinforce the learning objectives developed for the initial training. These included:

- Enhancing learners' knowledge of ethical conduct expected of engineering professionals and of how to apply this knowledge in situations requiring ethical judgment; subject matter associated with this objective specifically addressed the NMAC *Code of Professional Conduct* as well as the codes of conduct of the major engineering professional societies

  The NMAC rules address five topics:
  1) Protection of public safety, health, welfare, and property

2) Specialization and the performance of services only in specific areas of competence
3) Issuing public statements
4) Professional relationships with an employer or client
5) Solicitation of professional employment

- Familiarizing learners with business ethics principles
- Familiarizing learners with ethical conduct regarding authorship and publication
- Enhancing learners' knowledge of how to use and protect information in an ethical manner
- Providing learners with information on where to go for additional resources on ethics and ethical conduct

Adapting the Cases

While we tried to stay true to the details of the cases, it was sometimes necessary to adapt them to bring out salient features of the NMAC, LANL policies, or relevant laws and regulations. Generally, we begin by presenting a high-level, factually accurate summary of the case, with the only adaptation being removal of the names of involved individuals and substitution of names like "Eddie Engineer" and "Mike Manager" to enable students to track the participants through successive presentation of the case without compromising individual privacy.

Additional details about the case are presented in subsequent "frames." The case is doled out in small increments, with questions probing various ethical principles embedded within each segment. Fictitious situations or characters may be introduced to allow the scenario to explore aspects that were not present in the real case. Introduction of fictitious characters also facilitates having the learner take the position of actors other than those who were directly involved.

4. METHOD

For the 2010 and 2011 courses, "success" was defined in terms of the participation rate among members of the target audience (throughput), percent of correct responses (as a surrogate for knowledge transfer/retention), and informal feedback from participants. Due to a limitation in LANL's course delivery mechanism, no formal feedback was obtained.

Throughput and correct response rate were also monitored for the 2012 refresher, and formal feedback was solicited as well. The survey used was a modified version of Thalheimer's [9] learner survey. Unlike many "smile sheets," which ask general questions about the learning experience, this survey format asks learners to respond to specific learning points covered in the learning intervention. We used the learning objectives for the refresher training (shown in Table 1) as the key learning points to survey

against. Note that these learning objectives are a roll-up of the objectives used in the initial training; for example, the initial objectives related to authorship and protection of information are included in the more general topic of business ethics.

Capturing data about the *value* of individual key concepts provides more meaningful information about changes that should be made in future learning interventions [9]. In addition to addressing general ratings, the evaluation form also asks two critical questions related to how likely the concepts learned will be utilized on the job and how likely the concepts will be shared with others. This provides information regarding whether the training is likely to have an impact where it was intended.

Modifications to Thalheimer's [9] basic structure included questions related to participant preferences regarding case-based learning as compared to other instructional methods along the andragogical factors suggested by Knowles [4] and questions related to the utilization and value of asynchronous discussion augmentation of the online cases. It was hoped that this would allow us to validate our conclusion that a case-based model is the most appropriate method for delivering the educational experience to our target population and to gauge the effectiveness of threaded dialogue in improving the richness of the learner's experience and the quality of the feedback provided.

5. RESULTS

Throughput and Correct Response Rate

Throughput and correct response rates for the 2010 and 2011 refreshers have been discussed extensively elsewhere [10] and are not repeated here. The participation rate for 2012 was similar to that of past years (~49%). The correct response rate was significantly lower (at 68.8%, F = 4.41, p = .019) for the 2012 refresher than was observed in 2010 or 2011. Two of the 2012 questions had particularly low percent correct rates – one with just 25% correct and the other with 33.9%. With these removed, the 2012 results are still lower than in previous years, but the result is not statistically significant (mean correct response rate = 75.9%, F = 1.82, p = 177).

Course Evaluation

Of the 59 trainees who took the 2012 refresher, 29 completed the survey, for a response rate of 49%. Table 1shows participants' ratings of the value of the specific information contained in the learning objectives. Data in each cell is the percentage of respondents providing the response. The most common response across all learning objectives was that the materials "provided a nice reminder."

When asked to rate the overall value of the learning experience, 72.4% rated it as valuable or very valuable; only 3.4% provided a rating of little or very little value.

TABLE 1. VALUE OF SPECIFIC INFORMATION

Learning Objective	Rating				
	Most people already know this	*I already use these concepts regularly*	*Provided a nice reminder*	*Deepened earlier understanding*	*Concepts were new to me*
Making better decisions when faced with ethics-related situations	3.6%	7.1%	57.1%	32.1%	0.0%
Being knowledgeable regarding the Rules of Professional Conduct that apply to Professional Engineers licensed in the State of New Mexico	3.4%	13.8%	41.4%	41.4%	0.0%
Knowing how to identify and resolve business situations requiring ethical judgment	3.4%	13.8%	48.3%	31.0%	3.4%
Knowing where to go to get help when I am unsure about my best course of action	0.0%	17.2%	41.4%	34.5%	6.9%

The questions aimed at assessing the likely impact of the training each provided a range of likelihoods from 0% to 100%, and advanced in increments of 10 units. For the question regarding using the information on the job, 65.4% said that they were at least 70% likely to use the information on the job; only 10.2% gave a likelihood rating of 30% or less. For the question about sharing the information with a coworker or friend, the results were less positive – 48.1% expressed a 70% or greater likelihood of sharing, while 13.8% said that they were 30% likely or less.

Value of the Discussion Board

Survey respondents were also asked about their participation in and valuation of the discussion board that was introduced in 2012. While only 20.7% reported visiting the discussion board, 100% of those who did visit rated the experience as being of average or greater value.

Instructional Delivery Using Case Methods

Respondents were asked to rate how well case methods support each of Knowles' [4] andragogical principles when compared to traditional lecture- or presentation-based instruction. Table 2 shows their responses to these questions; data in the cells is the percentage of responses. The methods were viewed as equally supporting most of Knowles' principles by a plurality, if not a majority, of respondents. There were two principles for which the case method was seen by the majority of respondents as providing better support: the tendency toward movement from dependency upon an instructor to greater autonomy and self-directedness and the orientation toward learning as being problem-centered and contextual. In no case was the presentation-based method of instruction viewed as best supporting the andragogical principles by a plurality of respondents.

Comments on the case method received in response to an open-ended question were consistently positive:

- "The case method puts a real world perspective on the lessons and, especially when consequences of failure to behave ethically are demonstrated, it makes the lesson have meaning."
- "For this subject matter, case studies seem to be more meaningful."
- "For web-based instruction, I prefer case studies."

TABLE 2. RATINGS OF INSTRUCTIONAL DELIVERY METHODS

Adult Learning Instructional Design Principle	Ratings		
	Better supported by the case method	*Supported equally well by the case method and by presentation-based instruction*	*Better supported by presentation-based instruction*
The need to know not only the subject matter, but also the why, what, and how underlying it	37.9%	44.8%	17.2%
The tendency toward movement from dependency upon an instructor to greater autonomy and self-directedness	58.6%	37.9%	3.4%
The need to incorporate the learner's experience base as an integral part of the instruction	37.9%	51.7%	10.3%
The orientation toward learning as being problem-centered and contextual	55.2%	37.9%	6.9%
The need to incorporate the various roles that the learners play/have played in their professional lives	41.4%	51.7%	6.9%
The basis of the learner's motivation being in the intrinsic value of the learning and personal pay-off	24.1%	69.0%	6.9%

The only negative comments we received had to do with learners being uncomfortable with the lack of a definitively right or wrong answer for many of the scenarios: "Ethics can be black/white, but sometimes it is gray (or striped or polka-dotted)... these gray areas are the hardest thing for engineers to come up with the 'right answer.'" These were consistent with comments that had been received informally in prior years.

6.   DISCUSSION

Correct Response Rate

The question with the 75% failure rate in the 2012 refresher was a multiple choice question where more than one answer was correct; most respondents (66.7%) selected the more obvious of the two correct responses and missed the more subtle correct response. This may reflect a problem with the question stem, which is worded as follows: "Which of the NMAC 16.39.8.9 Rules of Professional Conduct **best** apply to this situation?" [emphasis added]. Use of the word "best" may have implied that there was a single correct answer, even though the answers themselves clearly conveyed the option of multiple responses (e. g., option d. in the response set was "a and c"). This result may also be reflective of lack of attention on the part of respondents, since the answer that was most often selected was option a. – respondents may not have read through the entire response set before selecting their answer. The 2013 version of the refresher contains several questions with similarly worded stems where the correct answer is an option having multiple answers and where the distractor most likely to be incorrectly selected is located somewhere other than at option a. We will examine patterns of responses in the 2013 refresher responses to determine if similar trends exist.

The question that had the 66.1% failure rate portrayed a situation in which a vendor had invited an engineer at a manufacturing firm to play golf at an exclusive country club. The initial question about the case simply asked if it would be ethical for the engineer to accept the invitation. The case files of the National Society of Professional Engineers (NSPE) contain many cases of this type, and the

NSPE Board of Ethical Review has uniformly deemed the offering and acceptance of such social exchanges to be ethical as long as there is no *quid pro quo* and the nominal value of the gift is not excessive in the context of the industry's culture. Both conditions were considered to have been satisfied in the case as initially presented [11]. However, accepting such a gift would be a violation of LANL's Code of Conduct, which prohibits the acceptance of gifts valued at more than $20 per occurrence or $50 cumulatively throughout the year. Our speculation is that respondents answered this question from the LANL point of reference. This hypothesis is being tested in the 2013 refresher, which contains a question about acceptance of a gift that makes explicit the frame of reference as the NSPE Code, then asks what the respondent would do if confronted with the same situation as a LANL employee.

Course Evaluation

The result that the most common rating of the value of the specific information contained in the learning objectives was that it "provided a nice reminder" is not surprising given that the case studies were intended to refresh knowledge gained through prior training.   The results on the two questions related to the likely impact of the training were positive – trainees generally reported a high probability that they would use what they learned in their job and, to a lesser extent, that they would share what they had learned with their coworkers. The differences in the two sets of impact ratings were not statistically significant ($F = 4.47\ E\text{-}06$, $p = .998$).

Comments indicating discomfort with the "squishiness" of ethics cases validated our previous conjecture that trainees' discomfort with the lack of a definitively right or wrong answer for many of the scenarios was due, not to an inherent weakness in the case method (as had been suggested by the FAO [5]), but to the nature of the ethical dilemmas.

Value of the Discussion Board

The low self-reported rate of visits to the discussion board was disappointing. Even more disappointing was the fact that the only comments posted on the board were the ones "planted" by the instructor. Breadcrumbs confirmed that the site had had visitors, but those visitors "lurked."   Knowles' [4] principle, that the basis of the adult learner's motivation comes from the intrinsic value of the learning and personal pay-off, had led us to speculate that learners would be motivated to participate in the discussion board if they found that it had intrinsic value. Although all those who visited the discussion board said that it was of average or greater value, they did not participate. Perhaps this was due to a lack of perceived personal payoff – learners did not receive any additional credit for participating, so may have opted to do the minimum amount of work required to get the PDUs.   To test this hypothesis, we are in the process of developing an incentive of 1 additional PDU for substantive discussion board input for roll-out early in FY15.

Instructional Delivery Using Case Methods

Case methods were viewed as better meeting adult learners' needs (as compared to traditional presentation-based instruction) in two regards:

- Moving from dependency upon an instructor to greater autonomy and self-directedness, and
- Having the learning be problem-centered and contextual

It is somewhat surprising that the methods were viewed as equally supporting two-thirds of Knowles' principles by at least a plurality of respondents. We expected case studies to be better at meeting learners' needs regarding incorporating their experience base as an integral part of the instruction and incorporating the various roles that the learners play or have played in their professional lives. The latter is especially true

in view of the fact that the case studies were explicitly designed to have the trainees consider the cases from a variety of points of view – involved worker, coworker, manager, etc. – reflective of different roles they may have played.

7. CONCLUSIONS

The selection of the case model approach was driven not by the subject matter to be taught but rather by the target audience – the adult learner. Therefore, there is the potential to apply case model-based learning in other professional disciplines, such as health care, law, and public accountancy, in which there are continuing education requirements placed on adult learners. In addition to sharing the characteristics of adult learners, like our target population these audiences are all motivated to complete continuing education for personal payoff – maintenance of the license or certification needed for continued employment in their profession – which reflects Knowles' final andragogical principle [4].

Because the engineering ethics refresher courses are delivered online rather than in a classroom, one of the integral aspects of the case method, namely discussion with a group of co-learners [5], is lost. As a result, online presentation of cases lacks the richness and feedback associated with cases presented in the classroom.. It was hoped that an asynchronous discussion thread could be added to the training experience to provide some of that richness. While our attempts at augmenting the online training with a discussion forum failed, we still believe that this approach is promising and are in the process of examining inducements that would encourage trainee participation on the discussion boards.

8. RESOURCES

[1] American Society of Mechanical Engineers Ethics Center http://www.asme.org/NewsPublicPolicy/Ethics/Ethics_Center.cfm

[2] National Society of Professional Engineers Board of Ethical Review http://www.nspe.org/Ethics/EthicsResources/BER/index.html

[3] New Mexico State Board of Licensure for Professional Engineers and Professional Surveyors Disciplinary Actions http://www.sblpes.state.nm.us/discipline.html

[4] Online Ethics Center for Engineering and Research http://www.onlineethics.org/Resources/Cases.aspx

[5] The American Lawyer http://www.americanlawyer.com/current_issue.jsp

[6] The Center for the Study of Ethics in Society http://ethics.tamu.edu/pritchar/an-intro.htm

9. REFERENCES

[1] New Mexico Administrative Code (NMAC). Title 16 Occupational and Professional Licensing Chapter 39 Engineering and Surveying Practitioners.

[2] BusinessDictionary.com, "What is refresher training? Definition and meaning," 2010. [Online]. Available: http://businessdictionary.com/definition/refresher-training.html [Accessed 9 April 2012].

[3] M. S. Knowles, **The Adult Learner: A Neglected Species**. Houston, TX: Gulf Publishing Company, 1973.

[4] M. S. Knowles, E. F. Holton III, and R. A. Swanson, **The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development.** Boston, MA: Elsevier, 2011.

[5] Natural Resources Management and Environment Department, "Management of agricultural research: a training manual. Appendix 2: the case method," 1997. [Online.] Available: http//www.fao.org/docrep/W7500E/w7500e0b.htm [Accessed 9 April 2012].

[6] D. Clark, "Case method (case study)," **Knowledge Jump Production**, 23 September 2010. [Online.] Available: http//www.nwlink.com/~donclark/hrd/history/case.html

[7] Answers.com, "Case method: information from Answers.com," 2012. [Online.] Available: http://www.answers.com/topic/case-method

[8] J. B. Kardon, "Bridge collapse and the duty to warn," **Online Ethics Center for Engineering**, 24 November 2010. [Online.] Available: http://www.onlineethics.org/Resources/Cases/24355.aspx [Access 15 August 2011].

[9] W. Thalheimer, "New design for my smile sheet," **Will at Work Learning**, 9 July 2008. [Online.] Available: http://www.willatworklearning.com/2008/07/smile-sheets-th.html [Accessed 27 April 2009].

[10] H. A. Hahn, "Adapting the case model approach for delivery of engineering ethics professional development units," in **Proceedings of the 16th World Multi-Conference on Systemics, Cybernetics, and Informatics**, Orlando, FL, 2012.

[11] R. Pritchard, "Golfing," [Online]. Available: http://ethics.tamu.edu/Portals/3/1992Cases/ConflictsOfInterest/ethics.tamu.edu_pritchar_golfing.pdf [Accessed 16 August 2012].

Cyber-attacks and attack protection

Petr HRUZA
Department of Military Management and Tactics, University of Defence
Brno, 662 10, Czech Republic

and

Radovan SOUSEK
Jan Perner Transport Faculty, University of Pardubice
Pardubice, 53333, Czech Republic

and

Stanislav SZABO
Faculty of Transport science, Czech Technical University in Prague
Prague, 128 00, Czech Republic

ABSTRACT

Dependence on cyber space is not a frequent topic of discussion; the reason might be that people fear possible abuse of cyber space or they just do not understand the center of the problem. Cyber space is used more and more – which increases the dependence. The dependence can be abused in case of a terrorist or cyber-attack. Now, this is the space where future wars can take place. Together with the increase of using the IT we can mark an increased risk of misusing these technologies. Attacks aimed against the IT assets and networks are becoming a worldwide phenomenon and their impact result to extensive economic damage throughout both public as well we private sectors. What is the cyber-attack and what are anticipated targets of such attacks? How we can protect the cyber-space? Answers to all these questions are subject elaborated in our article.

Keywords: Cyber Attack, Cyber Security, Scenario and Target of Cyber Attack, Warfare, Cyberspace.

1. INTRODUCTION

Nowadays, more than ever before, we have to consider the slogan "information is worth of gold", no organization exist that does not pose some important information (no matter if commercial entity or state administration). Remarkable increase of utilization of both IT assets and software in current world leads, on one hand, to establishing the information community, acceleration of communication and extensive development of services. However, the dependency of the community on IT grows up because of this as well. With this growing dependency on IT a risk of misusing these technologies also increases, what can result to considerable damages.

A general trend in this area worldwide is to develop a quality protection of the IT systems against the attack that could result to their malfunctions. Attacks targeted against IT installations are a worldwide phenomenon and their impact result to extensive economic damages in both public and private sectors, in the national as well as global scale. In the case when the attack is aimed against the elements of critical infrastructure the security even the very existence of a particular country might be endangered.

Attacks against IT are gradually becoming more sophisticated and more complex. Assuring the cyber security in the individual countries is one of the key challenges of contemporary era. The infinity and ubiquity of cyber threats require the extensive international co-operation but also common effort when assuring the cyber security of the individual countries. The area of cyber security has been and will be one from decisive aspects of the security environment in advanced countries. Growing part of the economic activities is being shifted to the area of internet – the cyberspace. By establishing the social networks the most known cyberspace – the internet – becomes a substantial societal phenomenon, by what the community can be influenced significantly both positive as well as negative way.

Robert Mueller, the FBI Director (September 4, 2001 – September 4, 2013), stated at the conference on security in San Francisco in the spring 2012 the following: „*We are losing data, money, ideas and innovations. Together we have to find the way how to stop it… I am convinced that only two kinds of companies exist in this respect: those the hackers have already attacked and other that will be attacked. And these two groups are merging together very fast. I am worried that in a near future only such companies will exist whose systems have been attacked by hackers already and others whose systems will be attacked by hackers again.*"

When talking about cyber warfare we have to think hard and decide if we talk about a virtual battlefield within the Internet or about a real physical battlefield. Recently the cyber space has been militarized. Does that mean that developed countries are afraid of cyber warfare? Cyber warfare would happen without soldiers, weapons, tanks or rockets. Instead, attacks would be conducted by groups of computer warriors (or hackers). Is cyber warfare a real threat? Can cyber warfare paralyze the world or just a particular country? In these days nobody can imagine cyber warfare of a world scale. Probably it would be a world war of unforeseeable consequences. The risk of cyber warfare is

and will be one of the most important topics in the field of information security in the following years.

2. CYBER-ATTACKS

A cyber-attack takes place in cyberspace. Cyberspace can be understood as a metaphor describing virtual (non-physical) environment created by interconnecting computer systems in the network. In cyberspace there is the same interaction between subjects as there is in the real world – but without necessity of physical activity. Information is shared either in a real time or it can be delayed, people can buy things, share experience, search contents, conduct research, work, or play. [6]

A cyber-attack is an activity carried out by an attacker, aim of whom is to get, modify, or destroy data (information, influence in a negative way or take over the control of some elements of cyberspace infrastructure). Cyber-attacks happen more and more frequently and are better organized. Dealing with their consequences is more and more expensive. This damage can reach a level that can endanger prosperity, security and stability of a country or an organization (possibly an individual). The fundamentals of cyber system security are supervision of cyberspace system infrastructure elements, security incidents management, regular audits, cyberspace systems vulnerability assessment, or carrying out penetration tests. [7]

Cyber-attack affects key features of a technology. This can endanger lives of people, or seriously endanger a process of production, technology, or services. That is why a cyber-attack can be defined as an attack on the structure of information technologies. The aim of such attack is either damaging or gaining sensitive or strategically important information. It is very often connected with politically or militarily motivated attacks. [8]

In these days organized crime in cyberspace provides more profit but lowers the risk for a hacker to be traced and tried. Ministries and government agencies all over the world face hundreds or thousands of attacks every day. Their security is threatened by hackers sponsored by money coming from foreign countries and organizations. They also focus on international companies which very often are not, in case of a cyber-attack, willing to cooperate with the police or FBI. The reason is they are afraid of long and demanding investigation. [4]

An effort to limit the attacked subject's activities (making their web sites unavailable, so called "hacktivism") or political activism (changing information, disinformation, scaremongering) which is connected with, so called, "perception management" (influencing public opinion using manipulative information chosen for target subjects) can be found among the cyber-attacks targets. These targets are different from common targets of hackers, crackers, or electronic sprayers who usually attack web sites in order to demonstrate their abilities. In cyber warfare we talk about coordinated activities purposely targeting other countries or a group of people.

Probably the most famous cyber-attacks are the ones on SCADA systems using the Stuxnet worm. The most famous published attack delayed start-up of Iranian Bushehr nuclear plant. The attack was focused on uranium enrichment factory in Natanz where fuel for this power plant was being prepared. The

Stuxnet worm spread all over the world; apart from sixty thousand cases in Iran we know about a million cases in China. [5]

3. CYBER-ATTACK SCENARIO

If we start to talk about the cyber-attack targets, most people think that they are not concerned by it, rather, that it is a government administration issue, respectively, an issue for institutions involved in the national defence or national security. Such way of thinking, however, pose a serious security risk in the area of the cyber threats.

Cyber-attack, in general, has to be hidden, or kept under cover until the last moment at least. The real target must not become apparent until the sufficient conditions to attack it effectively are not created. Here we can obviously see a parallel to common military tactics, and a hacker, even he/she is to conduct the attack by other weapons, is the same fighter as the soldiers in the field in this sense. Fighting principles are then similar to those of conventional war. Therefore, for future conflicts we have to be prepared to situations like beyond the classical war conflict we will also have to face the attacks against critical communication and information elements of a given country. As an example we can name the attacks of the Anonymous hackers group that attacked hundreds of the Chinese official websites early in April 2012. They achieved to change contents of these sites. Hackers attacked them as a protest against the censorship of internet enforced by Beijing officials.

Within the cyber combat we will encounter attacks aimed to e-government, i.e. against applications linking the government ministries with the population of a given country. These attacks will be masked and will come via home PCs of common private users in a particular country attacked. It is expected that all the applications of ministries will be attacked and put inoperative; also penetration to the steering information systems of ministries will also occur. Other group of professional hackers is to conduct attacks using tactic of parallel attack aimed to all banks in the country, including the national bank. This will result to the total collapse of trading and blocking the nationwide banking systems; another stage is expected to encompass the continuing extensive cyber-attack against crucial ministries and medical care systems aimed to the effort to eliminate their information systems completely (police database systems, tactical and information systems, medical care, etc.). Partial elimination of the information role of the state cyberspace is also expected to occur as well as the energy sources shutdowns (e.g. by imitation of nuclear power station accident, remote takeover of administration, etc.). Disruption the integrated rescue system will result to general panic together with elimination of the general mass media.

Another task of cyber warriors is anticipated to be to inflict the last crushing attack to the national communication system (mobile phone operators, internet services providers, etc.) and against the crucial points of the existing networks. This will result to interruption of communication, security and support role of the national administration cyberspace. Possible variant of the cyber-attack will be just the simple disruption of the national monetary and financial system together with disinformation campaign.

4. CYBER WARFARE

When we start talking about cyber warfare targets, most people assume this has nothing to do with them. They believe this is their government's issue, or better said, an issue for an institution related to a country's defense or security. However, this way of thinking can be very risky under the terms of cyber threats.

Cyber warfare can be perceived as an effort of a country to penetrate into computers or computer network of another country in order to damage or violate its integrity. Cyber security glossary defines cyber warfare as "*Use of computers and the Internet to wage a war in cyberspace. System of extensive, often politically motivated, related and mutually provoked organized cyber-attack and counterattacks*". [2, p.58]

Cyber warfare can be perceived as a use of computers and information technologies to carry out acts of war at governments and large organizations levels. Cyber warfare starter can be an individual, organization, or a state institution. There are a lot of different kinds of cyber warfare – from specialized hacker attacks to generally aimed attacks aim of which is to eliminate a particular service or paralyze critical infrastructure of the attacked country. An attack that completely cuts off the connection to the Internet is the highest level of cyber warfare.

Today's cyber-attacks are primarily carried out in order to achieve information concerning diplomatic, economic, and military programs. Paralyzing critical infrastructure of a particular state can represent a secondary target. Naturally, a cyber-attack must be hidden all the time, or, at least, until the last moment. The real target must not be revealed until sufficient forefront allowing an efficient impact is created. Parallel with common war tactics can be seen here. Although hackers use different weapons, in this sense they are the same warriors as common soldiers in the field. The principles of a combat are similar to a classical conventional war. That is why in future conflicts we can expect that beside a standard war conflict there will also happen attacks on significant communication and information points of a given country. [8]

5. CYBERSPACE PROTECTION TASKS

To provide the cyber security and to assure the respective rights for information self-determination via access to working services of the information community, it is necessary to process the information on the occurrence of the cyber security issues from as big number of resources as possible. For the same reason it is necessary to coordinate the protective measures. Services of the information community are characterized by their networking capabilities and even a very small element of the network can influence remarkably its elements, no matter if being close or on the geographical distance.

According to Veselý [7] these are the following tasks:
- Coordination of countermeasures for IT security incidents in the critical infrastructure.
- Gathering the information on serious security incidents.
- Coordination of filling the security gaps in the critical computer systems.

- National critical infrastructure must not be internet-dependant fully, but has to be capable of operation in case when a cyber security crisis occurs.
- Have an emergency plan of actions to be taken in case of becoming a victim of a cyber-attack.
- Resolutions of information-security incidents in cooperation with owners and operators of impacted parts, telecommunication operators, internet services providers and government administration components.
- Generation and distribution of skills on selected areas of information and cyber security among public.
- Generation of capabilities to defend against cyber-attacks.
- Achieving international consensus on the standards of behaviour in the cyberspace.
- Reduction of vulnerability of government systems and critical infrastructure elements.
- Cooperation with foreign organizations and presentation of the Czech Republic in the field of information security at international level.
- Support of education of professionals in the area of the cyber security.
- Increasing the enforceability of law in the area of cyber security.
- Improving prevention and build-up of general awareness.
- Improving the awareness in private sector.
- Generate the administration professionals in the area of IT security systems.
- Understanding the tactics of attackers, their techniques and procedures enabling conversion the protective measures to suitable forms.
- Be ready to prevent the attack or to respond as fast as possible in case of being compromised.
- Prioritize the prevention, but maintain focus also to detection and suitable response to attack.
- Keep assured that the suppliers in the area of critical infrastructures are not compromised and pose sufficient backup measures in case if their systems are disrupted.

The US Congress House of Representatives approved an Act on cyber security in April 2012 that enables the government and private sectors to cooperate when resolving the protection against hackers. By this new Act the US security components are empowered to share the confidential information on internet threats to private companies that can then better protect their networks against the attacks of hackers. This cooperation should work vice versa as well, what causes worries of those who struggle for protection of personal data. They anticipate that the security components will be enabled to gather large volumes of sensitive date on non-government entities what is otherwise restricted by the US law.

6. POSSIBLE CYBER WARS OF THE FUTURE

Possible cyber wars of the future are the reason for all of us to feel discomposed. In spite of traditional warfare, which requires a huge amount of sources (weapons, personnel, equipment), cyber warfare just needs somebody with the right knowledge, computer technology; somebody who wants to cause confusion. The enemy can be anywhere; even outside the nation or an organization. A strong attack can be carried out just by a group of hackers using standard computers.

Another horrifying aspect of cyber warfare is that a cyber-attack can come as a part of a coordinated attack, or it can be just a

sick idea of a wicked hacker who just wants to have a good time. Regardless what the attacker's motive is, cyber-attacks can cause huge financial losses. Many countries are desperately unprepared to face these unexpected cyber-attacks.

First of all it is necessary to realize the aspects of cyber warfare and their comparison with traditional aspects (conventional warfare, diplomacy, etc.). This kind of warfare can be more acceptable for public than the use of conventional warfare means. Although it is possible to avoid losing lives and damaging property by the use of cyber warfare, yet there is still risk of indirect danger. On the border between state and private sectors, there is a critical infrastructure. Here we talk about distributive water conduit, electric energy, air traffic control and other systems that are crucial for all countries. It is just the attack on critical infrastructure that can lead to paralyzing the network used by health service, water management and other vitally important elements of the country. During a conflict the enemy can make use of our dependence on cyber space and get a strategic advantage. It is supposed that cyber warfare will precede conventional warfare. That is why it is necessary to define critical communication and information structures in order to set priorities for cyber defense of the country. [9]

7. CALCULATING THE PROPORTION OF FORCES AND MEANS IN CYBER WARFARE

Calculating the proportion of forces and means in cyber warfare Various tactical calculations can be a part of combat planning. On of such calculations is proportion of forces and means. Proportion of forces and means is processed as an auxiliary document especially in decisive phases of task (combat task) accomplishment. Proportion of forces and means can be processed in quantitative or qualitative ways. When talking about quantitative proportion of forces and means the processors quote number of deployed (anticipated) enemy and own forces and means divided into required categories. When talking about qualitative proportion of forces and means quality of deployed forces and means is taken into consideration. The quality is included into final proportion of forces and means in a given category via so called "combat potential". The fact is that considered enemy forces and means (as for especially quantity but also quality) are stochastically assessed. When expressed mathematically, we talk about a sum of qualitative and quantitative data concerning forces and means of both belligerents, which, when compared, can provide us with a concept of combat possibilities of both belligerents and of a possible combat result.

Looking for a mathematical formula that would be able to calculate proportion of forces and means in the field of cyber warfare has become a new trend in the last few months. How can proportion of forces and means of cyber armies involved in a cyber war conflict be calculated? Is it realistic to carry out this calculation at all? In fact, it is nearly impossible. In case of a cyber space war it is not quantity that counts but quality. A hundred trained soldiers from special cyber units can fail facing one gifted adversary. We can talk about numerical superiority; however, it does not have to be crucial in cyber spaces. Is a hundred or a thousand specially trained soldiers too many or not enough? [9]

I think that cyber military units can be compared using following three criteria:

- Ability to carry out an attack in cyber space (offensive potential).
- Ability to avert an attack (defensive potential).
- Dependence on cyber environment (dependence).

Ability to restore key systems, existence of spare or back-up systems, emergency plans, etc. can be considered to be other aspects to be used for more detailed comparison. In case of cyber warfare we rather talk about potential possibilities of proportion of cyber forces calculation because concrete information is not available. The reason is that any statement made by a future attacker can inform the adversary about his possible potential weaknesses and can provoke trial actions or can lead to losing public confidence. A future attacker issuing a declaration concerning his own network perfect security can provoke trial action carried out by both potential attackers and random nongovernmental people (e.g. hackers) aim of whom is to discover possible weaknesses. [9]

8. CONCLUSIONS

Cyber-attacks will soon become a bigger threat than terrorism. Both public and private sectors are trying hard to protect their critical information and communication infrastructures from one of the biggest threats of the 21st century – cyber-attacks. The main task of a national cyber defence is to protect national communication and information infrastructure and systems from cyber-attacks.

With the emergence of targeted attacks and advanced persistent threats, it is clear that a new approach to cyber security is required. Traditional techniques are simply no longer adequate to secure data against cyber-attacks. Advanced persistent threats and targeted attacks have proven their ability to penetrate standard security defenses and remain undetected for months while siphoning valuable data or carrying out destructive actions. It is obvious that even the most secure computer networks are susceptible to attack.

Cyber-criminals, as well as the advertising agencies are currently very capable to utilise the current trends and events. They know what is currently "in" and misuse this information to internet frauds. Very often they utilize the method of social engineering, a current phenomenon among the computer threats.

Finding a mathematical formula to calculate proportion of forces and means in the field of cyber warfare has become a new trend in all armies.

9. REFERENCES

[1] Richard A. CLARKE, Robert K. KNAKE. **Cyber War - The Next Threat to National Security and What to Do about It**. USA: Ecco Press, 2010, 290 s. ISBN 9780061962233

[2] JIRÁSEK, Petr, Luděk NOVÁK and Josef POŽÁR. **Cyber Security Glossary**. The second updated edition. Prague: The Police Academy of the Czech Republic in Prague, 2013, 200 s. ISBN 978-80-7251-397-0. [in Czech].

[3] **Halted '03 Iraq Plan Illustrates U.S. Fear of Cyberwar Risk**. The New York Times, USA, 2009. [online] [2014-01-28].

<http://www.nytimes.com/2009/08/02/us/politics/02cyber.html?ref=cyberwar&_r=0l>.

[4] **Jak by mohl vypadat kybernetický útok na ČR**. [online] [2014-01-30].
<http://tm.m.idnes.cz/blog/clanek.248339.idn>. [in Czech].

[5] **Kybernetické útoky**. [online] [2014-02-10].
<http://www.ceskatelevize.cz/porady/10121359557-port/739-kyberneticke-utoky/>.[in Czech].

[6] McQUADE III., Samuel. **Encyclopedia of Cybercrime**. Westport: Greenwood. 2008.

[7] VESELÝ, Josef. **Cyber attacks**. In Cyber attacks to information systems. Prague: The Police Academy of the Czech Republic in Prague, 2012. ISBN 978-80-7251-371-0. [in Czech].

[8] HRŮZA, Petr. **Cyber Security**. The first edition. Brno: University of Defence. 2012, 90 p. ISBN 978-8-7231-914-5. [in Czech].

[9] HRŮZA, Petr. **Cyber Security II**. The first edition. Brno: University of Defence. 2013, 100 p. ISBN 978-80-7231-931-2. [in Czech].

[10] NĚMEC, V., VITTEK, P., KRULA, L. **Security management tools and UML standard at international airport**; In: 9th international scientific conference - New trends in aviation development; Gerlachov, Slovak Republic, 2010, ISBN 978–80–553 0475–5.

[11] NĚMEC, V. **Examinations of aircraft staff in Czech Republic**, New development trends in aeronautics, Košice, Slovak Republic, ISBN 80-8073-520-4

[12] KRAUS, J. - NEMEC, V., FAJČIK, P.: **The Use of Wireless Sensor Network for Increasing Airport Safety**; In: MAD – magazin of Aviation Development, Volume: 1, Issue: 5, September 2013, s.: 16-19; ISSN 1805-7578

[12] FUCHS, P., SOUSEK R., et all.: Dopravní infrastruktura jako prvek kritické infrastruktury státu, Košice 2011, ISBN 978-80-8928256-2

A Browser-based Cybersecurity Software Web Training Environment

Chengcheng Li
School of Information Technology
University of Cincinnati
Cincinnati, Ohio, 45220, USA

ABSTRACT

Cybersecurity plays an increasingly important role in today's business. Providing secure services and defending critical digital assets have become assimilated into every core operation process that the success of the business depends on. The ubiquitous availability of cybersecurity and hacking software through the WWW is the major force behind the exploding increase of incidents of cybercrimes. One of the top objectives of college cybersecurity curriculums is to teach students to proficiently use the available software tools. Most of these tools are free, open-sourced, and command-lined based. Although without easy-to-use graphic interfaces, these software tools are able to accomplish almost any security tasks in all the security domains if the correct parameters and syntaxes are used. This paper introduces a browser based online training system for learning some of the most popular cybersecurity tools. These tools are carefully selected from a Linux security distribution and a Windows based system security suite. Students learn the correct syntax and the rich collection of parameters of each tool on a simple web interface. The generated commands are executed on two remote servers and the results are displayed in real-time on the client side. Various security scenarios are created on the server side by using a set of virtual machines utilized as the targeted victims. Thus, there are no security threat exposure and additional required computing recourses other than a web browser on the client side.

Keywords: Cybersecurity, eLearning, Online Education, Remote Laboratory, Web Application

1. INTRODUCTION

There is a severe shortage of cybersecurity experts defending enterprise networks and government digital assets [1][2]. A huge gap exists between the security job market needs and the college graduate skills [3]. With the booking job market in cybersecurity, the unemployment rate of professionals with practical work experiences is reaching the lowest point in history [4]. However, lack of hands-on skills and experiences of adopting emerging technologies such as cloud computing, mobile, cognitive and social networking are employers' major concern of hiring new graduates [5]. Traditional Computer Science programs alone do not prepare a security workforce with sufficient skill sets for today's wide range of security practices. Information Technology and security professional studies programs in 2-year and 4-year colleges complements with traditional CS and CE programs for preparing the graduates to be qualified security technologies, instead of scientists and engineers. The graduates from the IT programs usually have strong hands-on skills in system deployment and administration and specialties in secure practices in every core processes of business, highly demanded skills by the industry and government sectors. Industry certification training and intern and co-op programs further expose IT students to industrial grade hardware and software tools and allow them to obtain practical experience before graduation. A careful design of practical training activities associated with hands-on experiences on a set of security tools are critical for students to build up their portfolios and develop their specialties in cybersecurity. The hardware and software tools in a cybersecurity degree program should be broad but also need to be specialized to cover the entire security intelligence cycle from information collection of targeted system and remote networks, sophisticated hacking actives, to defending network and data assets based on corporate polities, standards and government compliances. Special skills in system and network forensics, vulnerability analysis, incident response, penetration testing, intrusion detection and prevention, and secure code auditing, all require comprehensive understanding and proficiency of

using the available software tools. For example, the new Certified Ethical Hacking (CEH) certificate training, a well-accepted mid-level security professional certification by cybersecurity industry, provides trainees with more than 1000 software tools that covers more than 20 domains and security practices. Most of these tools are free open-source software. These tools are usually built upon simple command-line interfaces and do not need installation so that the overall file size of them is small for good portability, easy replication, and fast transfer on the internet. It is because the security software are hacking tools in nature; and portability and functionality far overweigh usability for hacking tools [6]. The universal availability of these free hacking tools on the internet is one of the reasons that the number of incidences of hacking and cybercrimes has been increasing in an alarming rate [7].

A main task required for cybersecurity specialists is to be proficient in choosing and using the security tools. The IT graduates are not required to program and create new tools as CS or CE graduates. But they need to set up security investigation and defense systems with appropriate tools for the right tasks. It is very important for security practitioners to test and choose a large selection of tools, and then collects the preferred ones to create their own tool kit for the tasks in their job descriptions. Selecting the right tools based on their costs and functionalities is an important training objective in the security curriculums.

Based on the security professional training needs, the government, academia, and the industry are creating training material and virtual online environments. The US Department of Homeland Security has more than 800 hours of training material available online through its National Initiative for Cybersecurity Careers and Studies program [8]. The NSF supported Advanced Technological Education training centers also created a great amount of cybersecurity materials [9]. Most of these materials require either installation of a new software-based environment to simulate and demonstrate security concepts, or require hosting virtual machines in isolated sandboxed environments, such as using Virtualbox, VMware, and Netlab systems [10]. These system setups hinder general public adoption and make remote access and e-learning inconveniently.

Teaching security software tools may not need sophisticated system setup and additional software other than what are already available from the vast online resources and open source software community. This paper introduces a novel approach to test security tools on remote servers on web browsers. This browser based approach eliminates the requirement of additional software and hardware installation.

The biggest obstacle of learning many of the security tools is understanding the syntax and using the parameters on the not-so-user-friendly command-line interfaces. For example, there are more than 20 parameters of a simple "ping" command. Memorizing and mastering the combination of these parameters can be a frustrating experiences. However, various parameters lead to rich functionalities of the security tools, making them accomplish many different tasks. This paper provides a learning environment for user to test different parameters of selected tools by following examples and tutorials completely through a web browser.

2. TRAINING ENVIRONMENT DESIGN

The goal of the web training environment is to allow students to execute the security tools remotely via a web browser, so no additional software is needed on the client side. The tools are installed and executed on the server, therefore, no security concerns and computation resources are needed on the client side. The outputs of the executed programs on the server are displayed in real time on the browser window for students to view, download, and analyze. Based on various security tasks, different servers and mock targeting systems are set up on the server side.

To display server-side output in real-time, a CherryPy web service infrastructure and an Apache-PHP web service were initially selected for fast development and deployment for this project. This study found that both CherryPy and PHP infrastructures produced desired results. Apache-PHP solution is discussed and demonstrated in this paper because it is the most popular web service solution on the internet.

To conduct system analysis, security practices, and forensics, two web servers are set up on two popular OS's, a Window 7 and a Backtrack 5 Linux VM.

Many of the security tools are OS independent. But for system analysis, Microsoft does have their specialized tools. For example, *sysinternals.com* listed a variety of Microsoft tools that can only be run on a Windows-based system. For computer forensics, there are also different procedures, utilities, and file systems on Windows and Linux. Backtrack 5 is a specialized Linux distribution for penetration testing, computer forensics, intrusion detection, and a large number of other cybersecurity tasks. Many security tools are already installed, configured, and categorized on Backtrack 5. Kali Linux is a new Linux distribution for security and considered improved version of backtrack 5. We can use Kali or any other Linux distribution as well since the installation of individual security tools is a relatively straightforward process. To conduct network security and forensics analysis, four target network VMs are set up for the servers to collect information, generate attacks, and remotely monitoring. These four VMs, a Windows XP, a Windows 7, a Windows 8, and a Ubuntu 12.4 server, are general purpose client systems and servers providing database, web, file, email, and other services as illustrated in Diagram 1. The Linux and Windows web servers have two interfaces that connect to the Internet and the VLAN with the target VMs.
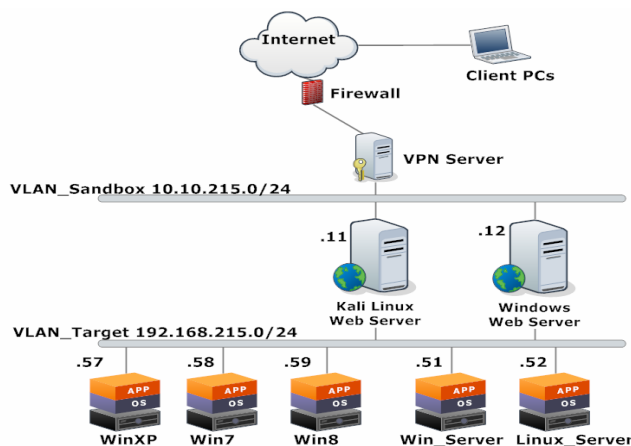
Diagram 1. Remote system design

This web training environment can serve general cybersecurity education purposes in a number of IT courses. The web interface should be intuitive and easy for students to learn. It should also be able to handle flexible user input so that the students can try a collection of security tools. It provides a tutorial system through which the students receives help and

recommendations on choosing the tools and parameters, and learn a complete security investigation process from information gathering, hacking and defense, and data traffic and log file analysis. To serve all these objectives, the web interface should include instructional examples and clear explanations of the tools and the associated parameters. The web interface should also be flexible for students to try different parameters and generate real-time views of execution results so that students can acquire proficiency in using these tools and practical hands-on skills in an efficient manner. Diagram 2 depicts the web interface designed for the training system. Users can select the embedded security tools by clicking the options on the main dropdown list to populate the command-parameter textbox on the top left. Explanations and examples of using the tools and correct parameters are given at the left pane of the interface. Users have the options to type the target machines or network or add other necessary parameters to complete the command in the second target textbox. The purpose of this design is to guide students to generate the correct commands and parameters for specific security tasks. The Execute button launches the generated command on the remote server. The executed command is then displayed on the top; and the results of the commands are displayed in real-time on the right pane. Some tasks can be done quickly, such as displaying some system information. Some tasks can last over a period of time. For example, packet collection for traffic analysis and network monitoring can produce a big amount of data and last from minutes to days. Students can stop the on-going task anytime and harvest the outputs from copying the result from the output pane or download the log or other text files generated by the commands and conduct off-line analysis.
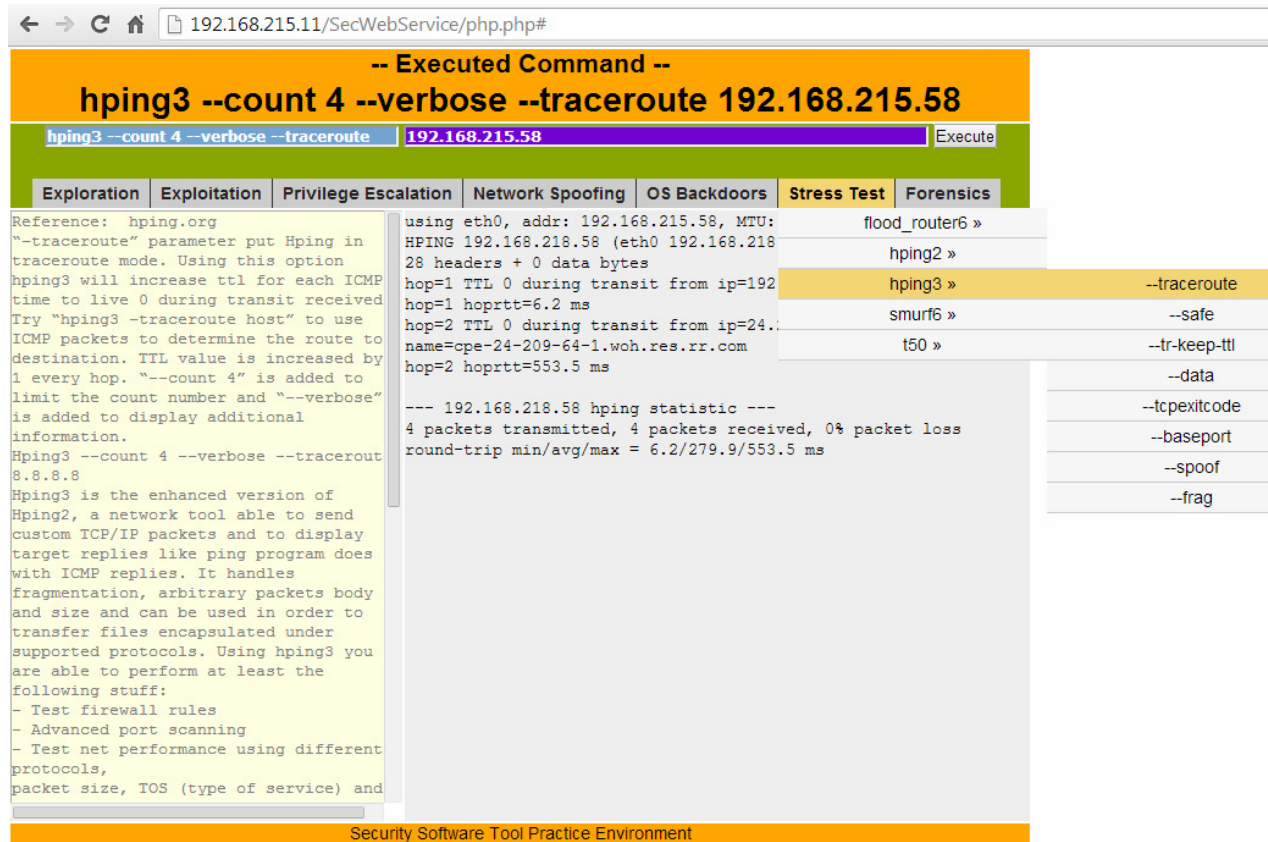
Diagram 2. Web Interface of the Online Training Environment

Exploration	Exploration	Privilege Escalation	Network Spoofing	OS Backdoors	Stress Test
Ping	Fasttrack	Acccheck	DNSchef	DBD	Flood_router6
Traceroute	SQLmap	Hexorbase	Fack_router6	HOtpatch	Hping2
Whois	Oscanner	Ncrack	Fiked	SBD	Hping3
TCPdump	Websploit	SQLdict	Intercepter_ng	Trixd00r	Smurf6
NMAP	Hondyd	Dsniff	Yersinia	Msfpayload	T50

Table 1. Secure tools selected from Backtrack 5

3. TOOL SECLETCTION

The security tools selected for this web project are all command-line based and one-time-execution software applications. They are carefully selected to represent many security domains from information gathering to forensics as shown in Diagram 1. The tools selected to be embedded on the Linux web server follows the categories defined by Backtrack 5. These tools are shown in Table 1.

There are specialized tools for Windows system analysis. For example, the tools added to the information collection category includes PStools, a system diagnosis tool set for Windows.

There are eight command-line based tools within PStools. There are also many other tools that can be potentially added to either the Windows and the Linux web training interfaces.

4. STUDENT INVOLVMENT

Student involvement is critical for selecting the tools and creating the tutorial contents. Students were given a project, worth of 20 grade points in an undergraduate security course, to select 3 tools and provide examples and recommendations of using the correct parameters for corresponding security tasks. They are also responsible of designing the security scenarios and system setups on the server

178

and target VMs; so the desired outcomes can be displayed when the commands are corrected generated and executed by the system. By doing this project, most students became proficient in using at least three security tools and started a good practice of building up their own security toolkit for their future security professional career. Students in this project generated tutorial materials in required text format and write a project report on the design of the security scenarios used to demonstrate the tools. The instructor in this project created the interfaces and implemented the security scenarios designed and reported by the students.

5. CONCLUSIONS AND FUTURE WORK

This paper introduced a browser based online training system for learning available software tools in a variety of security domains. Currently, the web servers are only open to the students who register in the selected security courses, with required VPN access to the university's sandbox system. Public access to the web services can be made possible in the future after careful redesign and test so that this education system will not be maliciously exploited by hackers or criminals for hacking and attacking targets on the internet. This training environment has the potential to be an effective teaching tool for K-12 and public security education once it's made public after penetration tests are performed and firewall policies are implemented.

6. REFERENCES

[1] "Cisco 2014 Annual Security Report", **Cisco Systems, Inc**., Jan., 2014

[2] Amber Corrin, "Is There a Cybersecurity Workforce Crisis?", **Workforce Management, The Business of Federal Technology**, http://fcw.com/articles/2013/10/15 , Oct., 2013

[3] John Connolly, "IBM Helps Bridge the Cyber Security Skills Gap by Partnering with More Than 200 Global Universities", **IMB Press Room**, http://www-03.ibm.com/press/ , Nov., 2013

[4] Robert Lemos, "Security Job Market 'Rocking,' But Pressures   Rise", http://www.darkreading.com/management/security-job-market-rocking-but-pressur/240152570 , **Dark Reading – A TechWeb Publication**, Manhasset, NY, April, 2013

[5] Amiet Goldman, "Bridging the Cyber Security Skills Gap, cyber-security, Training, Security", **IBM Training**, http://www.-304.ibm.com/connections , Nov. 2013

[6] Pham, D.V.; Syed, A.; Mohammad, A.; Halgamuge, M.N., "Threat analysis of portable hack tools from USB storage devices and protection solutions," **2010 International Conference on Information and Emerging Technologies (ICIET)**, pp.1-5, June, 2010

[7] Lesk, M., "Cybersecurity and Economics", **IEEE Security & Privacy**, vol.9, no.6, pp.76-79, Dec., 2011

[8] Paulsen, C.; McDuffie, E.; Newhouse, W.; Toth, P., "NICE: Creating a Cybersecurity Workforce and Aware Public", **IEEE Security & Privacy**, vol.10, no.3, pp.76-79, June 2012

[9] Perez, L. et al., "Information assurance education in two- and four-year institutions", **Proceedings of the 16th annual conference reports on Innovation and technology in computer science education**, pp. 39-53, June, 2011

[10] Gao Xiaopeng; Wang Sumei; Chen Xianqin, "VNSS: A network security sandbox for virtual computing environment," **2010 IEEE Youth Conference on Information Computing and Telecommunications (YC-ICT)**, pp.395-398, Nov. 2010

A Rational Choice Theory Perspective of Deploying Honeypots to contain the Insider Threat

Keshnee PADAYACHEE
School of Computing, University of South Africa
Pretoria, Gauteng 0003, South Africa

ABSTRACT

The aim of this paper is to provide an understanding of the insider threat using rational choice theory. Specifically, the decision-making process of a maleficent insider in the presence of a luring honeypot is reviewed. Understanding these decisions may assist in designing technological solutions that are deployed at opportune decision points to contain the insider threat. To test the premise, an insider threat mitigation strategy derived from the rational choice modeling perspective is presented.

Keywords: Rational Choice Theory, Honeypots, Deterrent Control, Insider Threat and Compliance.

1. INTRODUCTION

According to the CyberSecurity Watch Survey [1], 33% of respondents believe that the insider threat is becoming more costly as insiders are becoming increasingly more sophisticated. Examples of attacks include unauthorized extraction, duplication or exfiltration of data, tampering with data, deletion of critical assets, etc. [2]. The term 'insider threat' refers to any individual who works in an organization and uses the authority granted to them for illegitimate gain (see Schultz [3] for a detailed discussion of the term). Honeypots may be a way of containing the insider threat [4] that is, limiting the damage caused by an insider by deflecting the malicious insider to a honeypot. However, it is problematic to use a honeypot to prosecute an individual and this type of lawsuit could damage a company's reputation. Hence it is more practical to contain the insider threat and guide the insider to compliance. In this paper, rational choice theory (RCT) modeling is employed to devise a strategy to lure and contain the insider threat using honeypots. This type of analysis assists in determining at which decision points to activate interventions to redirect the insider threat to compliance.

A honeypot is an information system resource whose value lies in the unauthorized use of that resource [5]. Using honeypots should confirm an insider's actions and potentially analyze their motivations and their resources [4]. Recently there has been a trend towards recommending honeypots or decoys to counter the insider threat (see [6] and [7]). However, in general there are several ethical and legal considerations that need to be taken into account. Hence in this article honeypots are reviewed within the broader context of insider threat mitigation. In a sense a honeypot is a 'pre-e-crime' detection strategy where the insider is given the opportunity to commit a crime controlled by insider threat mitigation. In this pre-e-crime scenario, RCT is used to determine intervention points for insider threat mitigation thus guiding the insider to compliance.

Compliance with information security is an adaptive response, while non-compliance is a maladaptive response [8]. 'Deterrent controls are intended to discourage individuals from intentionally violating information security policies or procedures' [9] and discourage a maladaptive response, whereas positive-reinforcements encourage an adaptive response. Compliant information security behavior refers to the set of core information security activities that have to be adhered to by end-users to maintain information security as defined by information security policies [10]. Pahnila et al. [11] considered both positive-reinforcement and negative-reinforcement in terms of compliance. They considered sanctions, 'threat appraisal', 'coping appraisal' and normative beliefs as negative-reinforcements while 'information quality', 'facilitating conditions', rewards and habits as positive-reinforcements [11]. Some of these factors are relevant to this paper, however aspects such as coping appraisal and threat appraisal are not. These are based on the assumption that an insider will comply because they are concerned that their actions will jeopardize the organization's information security [11]. However, this is not a concern for the insider threat whose sole intention is maleficence. While sanctions are derived from General Deterrence Theory (GDT), normative beliefs are based on the perception of peer behavior [11]. This paper will only consider negative-reinforcement from the GDT perspective [12]. D'Arcy and Hovav [13] considered the countermeasures that deter internal systems' misuse, namely awareness of security policies, monitoring, preventative software and training. GDT has its roots in RCT. GDT is moderated by certainty of detection, severity of punishment and the swiftness of detection [14]. GDT focuses more on the cost of the crime, while RCT considers both the cost and the benefits of the crime [15].

RCT is relatively basic; it assumes that a criminal makes decisions based on a cost-benefit analysis [16]. RCT involves increasing the perceived effort of crime, increasing the perceived risks, reducing the anticipated rewards and removing the excuse for crime [17]. Recently there has been a trend towards using RCT to explain the insider threat. For instance, Li et al. [18] used RCT to examine an employee's intention to comply. In their study, Bulgurcu et al. [19] also applied RCT to understand insiders. The authors considered the benefits of non-compliance. According to Willison [20], RCT is highly appropriate to understanding insider threat. He argues that these types of theories complement existing security strategies and help to identify offender behavior and their associated criminal choices. These insights are useful in developing safeguards.

An effort was made in this study to employ RCT to develop a strategy to contain the insider threat. The rest of the paper is structured as follows: Section 2 contains a discussion of the concept of a honeypot as a lure to contain the insider threat. In

Section 3, RCT is examined with regard to its suitability in understanding the insider threat. Rational choice modeling is used in Section 4 to model the decision-making process of an insider in the presence of a luring honeypot and the paper concludes with Section 5 and possible future research opportunities.

2. DEPLOYING HONEYPOTS AS A LURE TO CONTAIN THE INSIDER THREAT

According to Spitzner [5], honeypots are more than just a computer or a physical resource. A honeypot may be anything from a Windows program to an entire network of computers. According to Spitzner [4], there are several advantages to using honeypots for the insider threat:
1) Small data sets: Only collect data when an attacker is interacting with them; hence easier to analyze and manage.
2) Reduced false positives: They are superior to other forms of detection strategies because an interaction with a honeypot is guaranteed to be illicit.
3) Catching false negatives: Other techniques may fail if there is a new type of attack, unlike a honeypot, as any interaction with one implies an anomaly.
4) Encryption: It does not matter if an attack is encrypted, the honeypot will capture the activity.
5) Highly flexible: Honeypots are adaptable from a simple honeytoken which may be embedded in a database to an entire network of computers.
6) Minimal resources: Honeypots require minimal resources.

According to Bowen et al. [21], the following are properties that honeypots (decoys) must have in order to bait an insider:
1) Believable: Decoys must appear to be authentic.
2) Enticing: They must be attractive to the insider.
3) Conspicuous: They must be visible to the insider.
4) Variability: Decoys should not be easily identifiable.
5) Non-interference: A decoy should not hinder, obstruct or impede normal operation.

Gupta et al. [22] identify the characteristics of luring in terms of a context honeypot. A context honeypot is used to identify a probable privacy violator. However, certain principles hold in general:
1) A suspected user can be lured only once by the same technique.
2) False negatives must be close to zero – for example, a criminal who is not lured escapes monitoring.
3) False positives should be kept low – for example, an innocent user is lured and is kept under observation for no reason.
4) Lure data will remain static until confirmation/elimination of suspicion over a suspected user – the data disclosed to the suspect must remain static so as not to raise the suspicions of the insider.
5) The suspected user must be oblivious of the luring.
6) Lure data will be different for each suspected user.

It is important for a luring honeypot to maintain the characteristics identified by Gupta et al. [22] and Bowen et al. [21]. Bowen et al. [21]) also specify that a decoy should be detectable (i.e. decoys must generate an alert). However, a honeypot will not be effective if the insider threat decides not to select a honeypot if they recognize that the luring honeypot is in fact a trap. Hence interactions with a honeypot should be detectable to the system administrator but not to the insider threat.

According to Spitzner [4], honeypots have several disadvantages. There is a risk that an attacker may use a honeypot to harm other systems. Honeypots are only of value when an attacker interacts with them and only capture actions related to this activity. Additionally, there are legal and ethical challenges. Spitzner [23] provides a comprehensive review of legal challenges associated with honeypots: entrapment, privacy and liability. Entrapment may be used as a defense to escape prosecution. Honeypots may violate an individual's right to privacy. There is also the issue of liability if an organization's honeypot is used to cause others harm [23]. The ethical question centers on whether it is ethical 'to pose a computer system as something it is not' [24]. Hence, the current research also considers integrating controls such as deterrent controls and positive-reinforcement with honeypots to guide the insider to compliance. In the next section, the concept of a luring honeypot is considered from an RCT perspective.

3. THE RATIONAL CHOICE PERSPECTIVE

The basis of RCT is that criminals formulate an assessment of the costs and benefits prior to and whilst committing an offence. An individual does not only consider the formal sanctions such as imprisonment but also the informal sanctions such as embarrassment [15]. Furthermore, the individual considers both the tangible benefits of a crime such as financial gain as well as the intangible benefits [15].

RCT is classified under the basic category of opportunity theories of crime as described by Felson and Clarke [17]. This consideration of RCT as an opportunity theory provides a rationale for examining luring honeypots under the same reflection. The premise of a honeypot is that it is a device deployed to provide an individual with malicious intent with an opportunity to commit a crime. According to Felson and Clarke [17], 'opportunity plays a role in causing in all crime'.

In order to consider the insider threat as a crime of opportunity, it is practical to reflect on some of the principles of opportunity theory as discussed by Felson and Clarke [17]. There are also some properties that may attract, detract or generate more crime [17]. In this context honeypots are attractors, whereas deterrents are detractors. For the insider threat, committing a major offence can result in other minor offences being committed and vice versa [17]. Sometimes a minor offence may provide a 'camouflage' for major offences. Honeypots can therefore provide a gateway to discovering major offences and deterrent controls may deter a minor offence thereby removing 'pre-criminal conditions' to a much more serious offence. The properties of value, inertia, visibility, accessibility (VIVA), based on the routine activities theory, provide a point for evaluating objects that are suitable targets [25]. Hence in order for a honeypot to be an effective lure, aspects such as high value, low inertia, high visibility and easy accessibility have to be considered as well. In a physical crime scenario, low inertia refers to an object that is transportable [17]. Perhaps, in the sense of cyberspace, low inertia could refer to data that can be easily disseminated on a network. A crime prevention method may provide the added benefit of diffusion [17]. The benefits of diffusion may occur when an insider assumes that the deterrent controls in one system may be applicable to other systems too.

It is crucial to recognize that a deterrent may result in maleficence being displaced to another part of the system.

In this study RCT was employed to understand the decisions of an insider in the presence of a luring honeypot. This involved considering controls that may inhibit the insider from committing maleficence. These controls influence the insider to perceive that the effort or risk of committing such an offence is more than they have estimated. According to Felson and Clarke [17], criminals rarely have a complete understanding of the costs involved. Controls that increase the awareness of information security policy may assist in 'removing the excuses' (i.e. rationalizations) of an insider. Willison [26] postulates that 'if offenders can be prevented from rationalizing and excusing their criminal actions in specific settings, they will be open to feelings of guilt and shame', thus preventing further crime.

RCT is not beyond reproach. Rational choice theorists often defend the theory by stating that criminals are not 'purely rational' as the notion of rationality is 'bounded' by factors such as time, ability and values [15]. Critics of RCT assert that rarely do criminals formulate a complete assessment of the costs and benefits prior to committing a crime [15]. It is clear that RCT is not the best tool to predict criminal behavior, rather it is a tool to predict the possible decision points of a criminal act. At a pragmatic level, one of the motivations for opting for RCT for this study was to leverage the decision models for RCT provided by Clarke and Cornish [27]. It is evident that the RCT perspective may be used as a 'heuristic device or a conceptual tool' rather than a theory [28]. These decision models expose the rational choices that a criminal makes at every stage of a crime. In the next section, these models are reformulated to consider the insider context.

The current research 'deconstructs' the methodology of using honeypots as lures to contain an insider and considers how an insider would view such a decoy and how they would react from a rational theory perspective within a holistic mitigation strategy.

4. RATIONAL CHOICE MODELING

Clarke and Cornish [27] assert that the models for rational choice do not have to be complete but merely 'good enough' to provide empirical directions for query or to derive policy. The models are derived from criminology and were developed for considering burglaries. In this paper these models are reapplied from a fresh perspective considering the insider threat. The processes of crime involvement are initial involvement, event, continuance and desistance. The models are schematic representations of the key decision points in criminal behavior. The decision models are not flowcharts. It is not necessary to model all four processes. The continuing involvement model was not considered because it focuses on the success of the crime.

Initial Involvement
The first process is the initial involvement model (see Figure 1). Here the insider represents a unification of previous learning and background factors (apathy, low self-control, opportunistic behavior, etc. [29]). The dimensions of the previous learning factor are derived from Bowen et al. [6]. They recommend that honeypots or decoys be designed levels of sophistication that

characterize the insider threat, ranging from insiders with a low level of sophistication to the highly privileged. An insider at a low level of sophistication relies on what can be discerned from a cursory scan, while a highly privileged insider threat will know that there are decoys in the system and will attempt to disable or avoid them. The previous learning also includes past deviant behavior [29] which is a predictor for future deviant behavior. The generalized needs are derived from the motivations of malicious insiders and range from espionage, sabotage, terrorism, embezzlement, extortion, bribery, corruption, and ignorance to apathy [30].
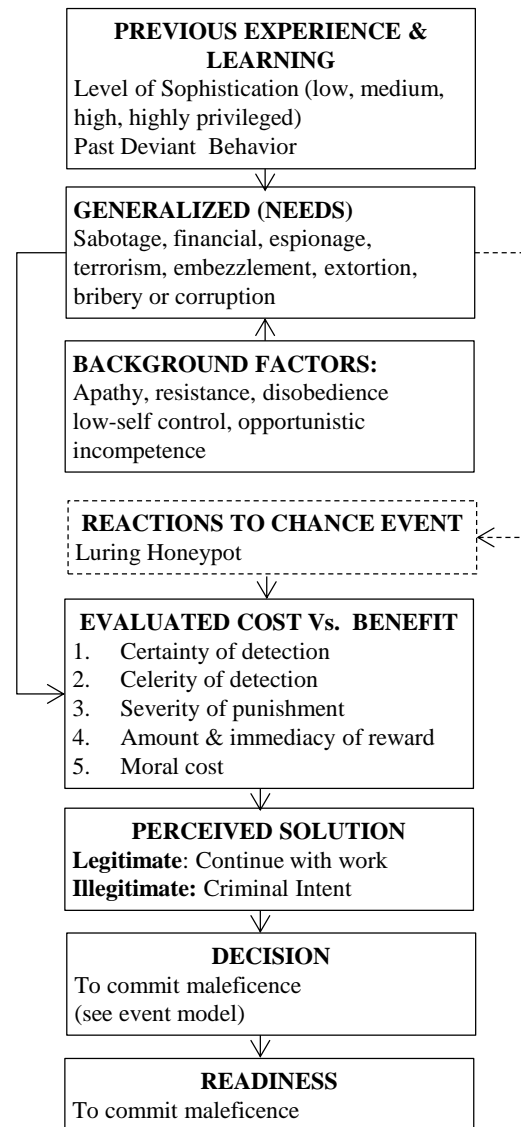
Figure 1: An initial involvement model precipitated by a luring honeypot

The insider may have malicious intent and decides whether they are ready to commit maleficence. However, the insider may be presented with an opportunity by a luring honeypot to commit an offence. At this juncture the insider will evaluate the 'cost versus benefit', and determine their readiness to commit the offence. 'Readiness' implies that the insider is lying in wait for the perfect opportunity to commit an offence.

There are two important decision points in this model: firstly, the readiness to commit maleficence and, secondly, the decision to commit maleficence, which may be precipitated by the luring honeypot. The decision to commit maleficence may be made simultaneously with the chance event or the chance event may precipitate the decision to commit maleficence.

The significance of the initial model is that it may be used for input into the properties that need to be taken into account to develop a truly effective honeypot lure. These input properties may be used to define the visibility, inertia, value and accessibility of the lures. Additionally the frequency of the honeypot lures (i.e. how often an insider should be targeted by lures) is another input property that needs to be considered. These five properties will need to be adapted to various degrees depending on the previous experience and learning; generalized needs and background factors. For instance, a highly sophisticated insider may be lured by a honeypot that appears to be a technical challenge. Hence, lowering the accessibility of the lure may provide this technical challenge. It may also be pragmatic to model lures based on the insider's motivation. For example, an insider that is motivated by financial gain will be lured by a honeypot that appears to be of high value. An insider's background may influence how they react to a lure. For example, insiders who are opportunistic or have low-self-control may need to be targeted more frequently by honeypot lures as they are more susceptible to crime. These factors may be determined by profiling.

The Event Model
The event model shows the sequence of decision-making nodes involved when an insider decides to accept the luring honeypot (see Figure 2). The alternative route, where a user decides to commit the offence without being lured, is not considered as it eliminates the honeypot in the process. In the event model, the user chooses the honeypot lure, trusting it to be real data, because it is believable, enticing, etc.

The significance of this model lies in the fact that it demonstrates that the opportunity for intervention is at the point where the insider has decided to accept the lure. At this juncture deterrent controls may be deployed to increase the perceived effort or risks of committing the crime. The lure is attacked if it is believed to be of high value, has high visibility, low inertia (e.g. easily forwarded in an e-mail) and is easily accessible. While the event model intends to attract the insider; the aim of applying deterrent controls is to subtly 'train' the insider to be compliant and furthermore the deterrent controls should target the 'supposed offence' being committed.

The aspects that deter non-compliance include sanctions, monitoring and policies, and technological controls such as usage control deterrents [29]. Usage control deterrents include conditions and obligations on the usage of data [31] and intend to deter rather than deny access. Note, the original concept of usage control deterrents did not consider how these conditions or obligations may be used from an RCT perspective to increase the costs associated with maleficence. Usage control deterrents may be used to increase the perceived risk and effort associated with the offence and may diminish the excuses for committing maleficence. They also increase awareness about the legitimate uses of organizational data and increase the effort required to access the data by stipulating conditions and obligations on the use of the data. These conditions and obligations should be aligned with information security policies.
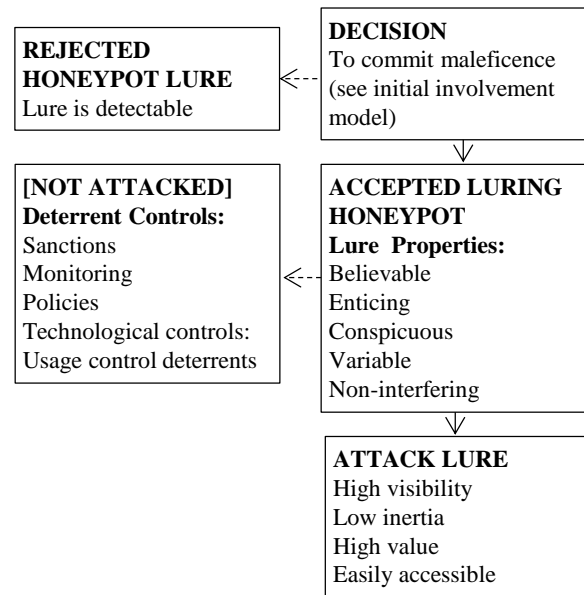
Figure 2: An event model integrating luring honeypots and deterrent controls

Desistance Model
This model (see Figure 3) is based on the aversive experiences of an insider during the course of offending (i.e. the deterrents) and variations in positive-reinforcement. Positive-reinforcements may include improved 'information quality' (i.e. improving the perceived usefulness of information); 'facilitating conditions' (i.e. providing a supportive work environment); rewards and mandating good security habits [11]; and organizational commitment. These positive-reinforcements may encourage the insider in the best case scenario to comply or in the worst case scenario to consider other crimes (displacement).

The rejected alternatives show areas of discontentment. For example, an insider may be disgruntled due to working long hours. Wortley [32] describes several types of situations that result in maladaptive behavior. The conditions that correlate with the virtual world include frustrations caused by failures of equipment and services, and invasion of privacy. Organizations need to be aware of these types of conditions as by reducing areas of discontentment, insiders will be less likely to engage in maleficence.

In this derivation, it is envisaged, that containment events via honeypots and the deterrent events will contain and deter a maladaptive response. On the other hand, positive reinforcement events will increase the likelihood of an adaptive response. Desistance implies either an end to all maleficence activity or displacement (i.e. another type of crime). It is assumed that once an insider accepts one honeypot lure, they will be presented with more honeypot lures for the purposes of reconnaissance.
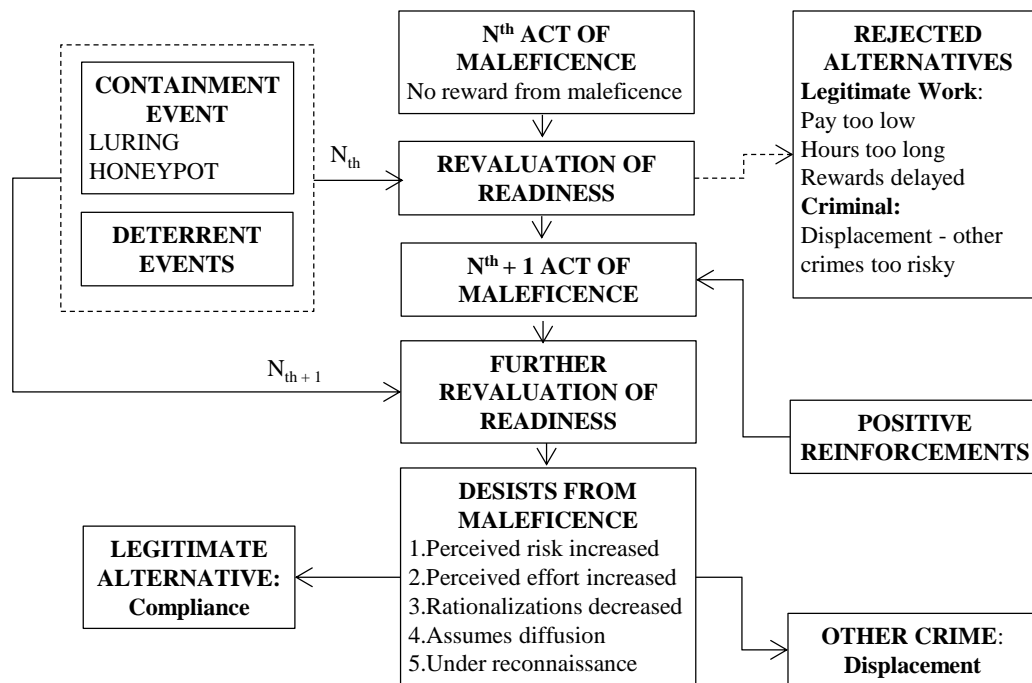
Figure 3: Desistance model integrating honeypots into a holistic mitigation strategy

In this context it is anticipated that deterrent controls and positive-reinforcements may result in malicious insiders desisting from further offences. Desistance involves the insider considering the following: the costs of the crime (i.e. perceived risks and perceived effort); the justifications for the crime (i.e. rationalizations); diffusion and currently being under reconnaissance.

The significance of this model lies in the fact that it shows junctions where organizations can intervene if the insider ignores the deterrence strategy and continues with attacking the honeypots. This implies that the insider requires further interventions which may require the use of positive-reinforcements to encourage compliance.

The other important juncture considering the rejected alternatives that is those negative aspects in the work environment that are precipitating the insider to continue with attacking honeypots despite the interventions. Organizations also need to consider the possibility that interventions have caused the insider to displace the crime. Hence there is a need to have contingency plans for this outcome.

Implications for Practice
The models are not intended to be all-encompassing. The models derived here may be used as a proactive mitigation strategy which seeks to change behaviors and motivations. This technique of viewing polarizing concepts of luring, deterring and positive-reinforcement may offer a strategy for organizations deploying honeypots without the legal pitfalls.

According to Clarke and Cornish [27] this type of modeling helps focus on a specific crime and helps develop a policy as it breaks the crime process down into a smaller components

which allows for policies to target these specific components. This type of analysis helps the development of measures to increase the effort of offending and decrease the associated rewards and defenses.

5. CONCLUSIONS

In this study rational choice modeling was used to understand how to effectively lure and contain the insider threat from further maleficence. It is based on a containment strategy where the insider is deflected to honeypots instead of attacking real data while deterrent controls and positive-reinforcements attempt to negotiate them into compliance. It is based on an approach where insiders are lured successively and re-conceptualizing their notions about the costs of committing the associated offence by discouraging maladaptive responses and encouraging an adaptive response. Future research will involve implementing and evaluating the strategy that was derived. The rational choice models derived offer a broader perspective in terms of considering the psychological and organizational components, and where interventions may be deployed.

6. REFERENCES

[1] CSO MAGAZINE, USSS, CERT, and Deloitte. "2011 Cybersecurity Watch Survey: Organizations need more skilled cyber professionals to stay secure", www.cert.org/archive/pdf/CyberSecuritySurvey2011.pdf. (accessed 12 March 2013).
[2] M. B. Salem, S. Hershkop, and S. J. Stolfo, "A survey of insider attack detection research: Beyond the Hacker", **Insider Attack and Cyber Security**, Advances in

Information Security S. J. Stolfo, S. M. Bellovin, A. D. Keromytis, S. Hershkop, W. S. Smith and S. Sinclair, eds., pp. 69-90, New York: Springer US, 2008.

[3] E. E. Schultz, "A framework for understanding and predicting insider attacks", **Computers & Security**, Vol. 21, No. 6, 2002, pp. 526-531.

[4] L. Spitzner, "Honeypots: Catching the insider threat" in 19th Annual Computer Security Applications Conference (ACSAC 2003), Las Vegas, USA, 2003, pp. 170-179.

[5] L. Spitzner, "Honeytokens: The other honeypot", http://www.securityfocus.com/infocus/1713 (accessed 12 March 2012).

[6] B. M. Bowen, M. B. Salem, S. Hershkop, A. D. Keromytis, and S. J. Stolfo, "Designing host and network sensors to mitigate the insider threat", **IEEE Security & Privacy**, Vol. 7, No. 6, 2009, pp. 22-29.

[7] J. White, and B. Panda, "Implementing PII honeytokens to mitigate against the threat of malicious insiders" in IEEE International Conference on Intelligence and Security Informatics (ISI'09), 2009, p. 233.

[8] A. Vance, M. Siponen, and S. Pahnila, "How Personality and Habit affect Protection Motivation", Workshop on Information Security and Privacy (WISP 2009), Phoenix, Arizona, 2009, pp. 14-21.

[9] H. F. Tipton, "Types of information security controls", **Handbook of Information Security Management**, M. Krause and H. F. Tipton, eds., pp. 1357-1366, Boca Raton: CRC press, 2007.

[10] M. Chan, I. Woon, and A. Kankanhalli, "Perceptions of information security in the workplace: Linking information security climate to compliant behavior", **Journal of Information Privacy and Security**, Vol. 1, No. 3, 2006, pp. 18-41.

[11] S. Pahnila, M. Siponen, and A. Mahmood, "Employee's Behavior towards IS Security Policy Compliance" in Proceedings of the 40th Hawaii International Conference on System Sciences Hawaii, 2007, p. 1561.

[12] G. E. Higgins, A. L. Wilson, and B. D. Fell, "An Application of Deterrence Theory to Software Piracy", **Journal of Criminal Justice and Popular Culture**, Vol. 12, No. 3, 2005, pp. 166-184.

[13] D. D' Arcy, Hovav, A., "Does one size Fit All? Examining the differential effects of IS Security Countermeasures", **Journal of Business Ethics**, Vol. 89, 2009, pp. 57-71.

[14] D. D'Arcy, and T. Herath, "A review and analysis of deterrence theory in the IS security literature: making sense of disparate findings", **European Journal of Information Systems**, Vol. 20, 2011, pp. 643-658.

[15] G. F. Vito, J. R. Maahs, and R. M. Holmes, **Criminology: Theory, Research, and Policy**, 2nd edition, Boston: Jones & Bartlett Learning, 2006.

[16] D. B. Cornish, and R. V. Clarke, "Understanding crime displacement: An application of rational choice theory", **Criminology**, Vol. 25, No. 4, 1987, pp. 933-947.

[17] M. Felson, and R. V. Clarke, Opportunity makes the thief: Practical theory for crime prevention, Police Research Series Paper 98, Policing and Reducing Crime Unit, Research, Development and Statistics Directorate, London, 1998.

[18] H. Li, J. Zhang, and R. Sarathy, "Understanding compliance with internet use policy from the perspective of rational choice theory", **Decision Support Systems**, Vol. 48, No. 4, 2010, pp. 635-645.

[19] B. Bulgurcu, H. Cavusoglu, and I. Benbasat, "Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness", **MIS Quarterly**, Vol. 34, No. 3, 2010, pp. 523-548.

[20] R. Willison, "Understanding the perpetration of employee computer crime in the organisational context", **Information and Organization**, Vol. 16, No. 4, 2006, pp. 304-324.

[21] B. M. Bowen, M. B. Salem, S. Hershkop, A. D. Keromytis, and S. J. Stolfo, "Baiting inside attackers using decoy documents" in Security and Privacy in Communication Networks: 5th International ICST Conference (SecureComm 2009), Athens, Greece, 2009, pp. 51-70.

[22] S. K. Gupta, R. G. Damor, A. Gupta, and V. Goyal, "Luring: A framework to induce a suspected user into a context honeypot" in IEEE Second International Workshop on Digital Forensics and Incident Analysis (WDFIA 2007), 2007, pp. 55-64.

[23] L. Spitzner, "Honeypots: Are They Illegal?", http://www.symantec.com/connect/articles/honeypots-are-they-illegal (accessed 12 March 2013).

[24] M. E. Kabay. "Honeypots, Part 4: Liability and ethics of Honeypots",http://www.networkworld.com/newsletters/2003/0519sec2.html (accessed 13 March 2013).

[25] L. E. Cohen, and M. Felson, "Social change and crime rate trends: A routine activity approach", **American Sociological Review**, Vol. 44, No. 4, 1979, pp. 588-608.

[26] R. Willison, and J. Backhouse, "Opportunities for computer crime: Considering system risk from a criminological perspective", **European Journal of Information Systems**, Vol. 15, No. 4, 2006, pp. 403-414.

[27] R. V. Clarke, and D. B. Cornish, "Modeling offenders' decisions: A framework for research and policy", **Crime and Justice**, Vol. 6, 1985, pp. 145-185.

[28] D. B. Cornish, and R. V. Clarke, "The rational choice perspective", **Environmental Criminology and Crime Analysis**, R. Wortley and L. Mazerolle, eds., pp. 21-47, Uffculme, UK: Willan Publishing, 2008.

[29] K. Padayachee, "Taxonomy of compliant security behavior", **Computers and Security**, Vol. 31, No. 5, 2012, pp. 673-680.

[30] K. Nance, and R. Marty, "Identifying and visualizing the malicious insider threat using bipartite graphs," in 44th Hawaii International Conference on System Sciences (HICSS), Koloa, Kauai, Hawaii, USA, 2011, pp. 1-9.

[31] K. Padayachee, "An aspect-oriented approach towards enhancing optimistic access control with usage control ", Computer Science University of Pretoria (Doctoral-Thesis) Pretoria, 2009.

[32] R. Wortley, "A classification of techniques for controlling Situational Precipitators of Crime", **Security Journal**, Vol. 14, 2001, pp. 63-82.

AUTOMATIC CLASSIFICATION OF TOBACCO LEAVES USING IMAGE PROCESSING TECHNIQUES AND NEURAL NETWORKS

Edison Pereira Dachi[1], Alexandre Gaspary Haupt[1] and Andre F. S. Guedes[1]

[1]Image Processing Laboratory, SENAI College of Technology - 8450, Assis Brasil Avenue - Zip Code 91140000, Porto Alegre, Rio Grande do Sul, Brazil.

ABSTRACT

Agriculture smoke in Brazil and in the world is an activity that has bustling economy, generating direct and indirect employment. This activity begins at planting and the selection goes to the leaves. Following the steps of planting, harvesting and drying, the leaves go through a selection step, where they are classified. The dried leaves differ in color and size, presenting characteristics that impact product quality and consequently the price of sheets. Despite the importance of classification, this process is still done manually and therefore is subject to many errors in addition to impact on workers' health due to the visual effort, often in a location with inadequate lighting. This paper proposes the automatic classification of tobacco leaves from parameters such as coloring and size (length and width) of the leaves. These characteristics are related to the degree of maturity of the sheet and its position at the foot of smoke. The manual classification is accomplished through a visual process often by people without experience. The classification, even when performed by experienced classifiers, does not guarantee the homogeneity of the results due to fatigue after working hours without staining pattern, adequate lighting without control and without calibration. This results in the need for a new classification when the leaves arrive in the tobacco industry. This results in prices that often do not correspond to that expected by the farmer. This paper proposes an experimental method for automatic classification of tobacco leaves using image processing techniques to extract information from the leaves. The OTSU method was used to automatically determine the threshold level of gray to separate the object from the background image. The classification, from the extracted image parameters, was performed using a backpropagation neural network layers 2 and knowledge of an expert to train the network by Levenberg-Maquardt method. 60 samples of tobacco leaves belonging to three different classes for training, to testing and validation of the proposed method were used. 20 samples with known classes were used to train the network, other 20 samples were used to test the results and another 20 samples to validate the results. An index of success of 87% in the classification of leaves was obtained. It is believed that the results can be improved using other color representation space such as HSV and lighting control in the image acquisition environment.

Keywords: classification, image processing, neural networks, tobacco leaves.

INTRODUCTION

Tobacco agriculture in Brazil and other countries of the world like China and the United States is divided into activities ranging from planting until the arrival of the leaves in the tobacco industry. The process begins with the planting of tobacco seedlings in appropriate beds. After a period of growth the plants are transported to the crop in situ. When they reached the final stage of growth your plants reach maturity and are ready for harvest. After harvest the tobacco leaves are placed on greenhouse to the stage known as "toasting". At this stage the leaves are dried and reach its final state. Leaves "toast" are not all alike and have different characteristics that impact the quality of the final product: the cigarette. So the process of classification of tobacco leaves is necessary. This classification is based on leaf color and size (length and width). These characteristics depend on the degree of ripeness of leaf and region where you collected was the leaf [1][2].

The classification is performed manually by experienced and non experienced people, adults and children. Thus classification errors are inherent to this type of procedure, even when performed by experienced workers due to lack of a standard and by repetitive work, often performed in environments with poor lighting. Apart from hazardous working conditions and fatigue of those performing this task 10 h per day on average[3][4].

It is known that, due to these errors, a new classification is performed in the tobacco industry. This results in prices that often do not correspond to that expected by the farmer because the classification process is not standardized.

Thus, we identify the problem of the traditional method for classification of tobacco leaves: the lack of homogeneity in the results with many misclassifications.

This paper proposes an automatic classification of tobacco leaves, using techniques of image processing for feature extraction from leaves and neural networks trained with knowledge of patterns used by experts.

EXPERIMENTAL DETAILS

The objective is to develop "scripts" using the MatLab program for image processing and classification using a neural network backpropagation from images of tobacco leaves with samples obtained after the roasting process. In this section we present all implementations used on the job. All scripts in this work were developed in MatLab version 6.1 environment.

Figure 1 illustrates the steps of the classification of tobacco leaves used in this work process.
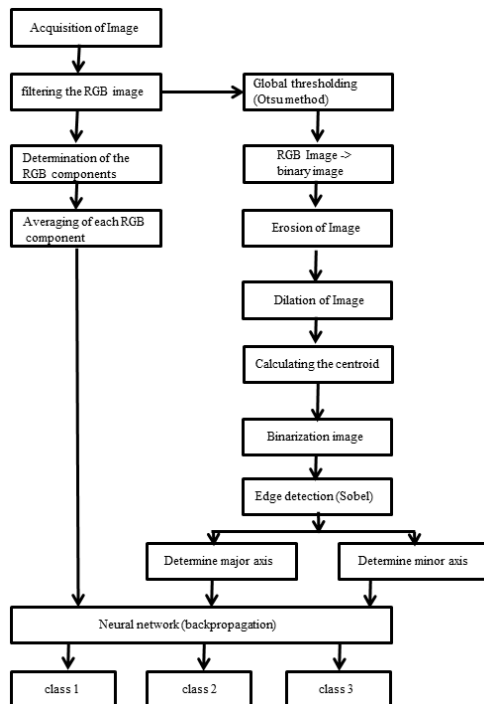
Figure 1. Flowchart of methodology for classification of tobacco leaves.

For image acquisition we used a digital camera mirage, Steel model and 3.0-megapixel. In the best color reproduction and "daylight" fluorescent lamps were used. Figure 2 illustrates an image of roasted tobacco leaf.

Figure 2. Original image of tobacco leaf obtained after the roasting process.

A feature used by the experts to classify tobacco leaves is color. To train the neural network this feature is also used, but as colors are not constant, is required decomposition into its Red, Green and Blue components, to determine the average for each of the colors on the leaves. Figure 3 illustrates the breakdown of the leaves in its R, G and B components [5].
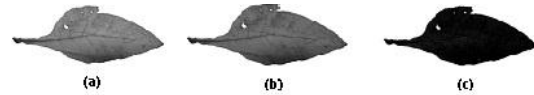
Figure 3. Decomposing the original image in R, G and B components: (a) Red Layer (b) Green Layer (c) Blue Layer.

To eliminate possible noise acquisition and standardize the image, we applied a Gaussian filter with $\sigma = 0.5$ to each of the three R, G and B components leaving these more homogeneous for next step that consists in calculating the average values of each color. To calculate the average of each color only the pixels belonging to the region of interest, we used a marker. The idea is to separate the background pixels of the region of the interest image of the leaves. Thus they used white pixels as pixels "background" and pixels belonging to the region of each R, G and B. The procedure is scanning the image. If the pixel is white, dismisses it. If it belongs to the region of interest, enter the calculation of the average. The average of each of the three colors is determined by equation 1 [6][7][8].

$$\mu = \frac{\sum_{i=1}^{n} Vpxp}{Npxp}$$

Eq. 1

where:
μ = value of mean color;
V_{pxp} = pixel value of the region of interest;
N_{pxp} = number of pixels in the region of interest.

To determine the centroid of leaf and length and width leaf, separated the leaf background.

The separation was performed by image binarization, where the threshold is determined by the Otsu Method.

The binary image reduces the image in two different regions, making the process more reliable extraction of features, simplifying the calculation of the centroid and the size of the leaves, as well as reducing the processing time.

Otsu's method for determining the binarizing threshold (thresholding) involves iterating through all possible threshold values and calculates a propagation measure for the levels of pixels on either side of the threshold. The goal is to find the threshold value at which the sum of the scattering of foreground and background is at its minimum. Figure 4 shows the binary image after applying the binarization with threshold Otsu [8][9].

Figure 4. Binary image of the leaf after application of thresholding with the threshold value determined by the Otsu Method.

The number of pixels of the leaf area was determined from the binary image. For this the image is scanned to count the number of black pixels. Obtaining the number of pixels of the leaves, here called N_{PXP}, is important for determining the average color of each layer and the RGB leaves (C_X, C_Y) center coordinates. Equations 2 and 3 show, respectively, the coordinates of the center of the leaves were obtained [10].

$$Cx = \frac{\sum_{1}^{N} x}{Npxp}$$

Eq. 2

$$Cy = \frac{\sum_{1}^{M} y}{Npxp}$$

Eq. 3

where:
C_X - abscissa of the centroid;
C_Y - ordinate of the centroid;
N_{PXP} - number of pixels of the leaves;
x - position of abscissa of the pixels that belong the object;
y - position of pixels ordered that the object belongs.

Failures such as holes in the leaves caused by insects, sunburn, etc. can influence the determination of width (W) and length (H) of the leaves. Aiming to fill "gaps" in the image that may mask the result, the solution adopted was closing consisting of erosion followed by dilation operations on binary image operations. Some structural elements have been tested and the best results were obtained with the structural element in line, 15 width, and angle of 90°, completely filling the "holes" in the image. The Matlab functions were used [4]:

se = strel('line',15,90);
bw2 = imerode(bw,se);
bw3 = imdilate(bw2,se);

After the closing operation was necessary binarization the image again. The next step was to obtain the contour of the binary image necessary for determining the length and width of the leaves.

bw = im2bw (bw3);
hy = fspecial('sobel'); %Filtro Sobel
hx = hy'; % Atribuio a matriz hx, a matriz transposta de hy
Iy = imfilter(double(bw), hy, 'replicate');
Ix = imfilter(double(bw), hx, 'replicate');
gradmag = sqrt(Ix.^2 + Iy.^2);

The contour was obtained by the Sobel method is used to determine the major axis, calculating the largest distance between pixels belonging to contour. In this case we chose an exhaustive search, finding the pixels and calculating the greatest distance.

To fix the position of the leaves and make its most aligned with the abscissa axis, the angle between these axes was determined and rotating-image in the value of this angle.

To determine the minor axis calculated the greater distance between the pixels in the direction perpendicular to the line formed by the major axis passing through the center coordinates of the leaves.

Obtained the values means RGB colors, major and minor axes of tobacco leaves, it was used as input data to the neural network to classify tobacco leaves.
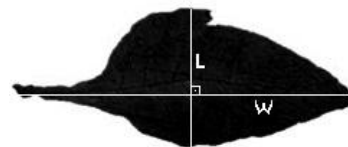
Figure 5. Determination of the minor axis (L) leaf from the line through the centroid oriented 90° with respect to the major axis (W).

The class of a leaves of tobacco may be associated with the position of the leaf tobacco plant. The color and size of leaves are also

indicative of its class. Figure 6 illustrates the classes used in this study set from characteristics evaluated by experts as follows: The B class arising leaves located in the upper part of the plant is darker green color. Its dimensions are smaller and its more elongated leaves. The C class consists of leaves located in the middle part of the tobacco plant, with more yellow and larger than the leaves of the B class staining lengths, with a little more rounded. The X class consists of leaves at the bottom of the tobacco plant, having a lighter color than the leaves of C class. Its dimensions are more rounded and reduced in relation to C class [11].
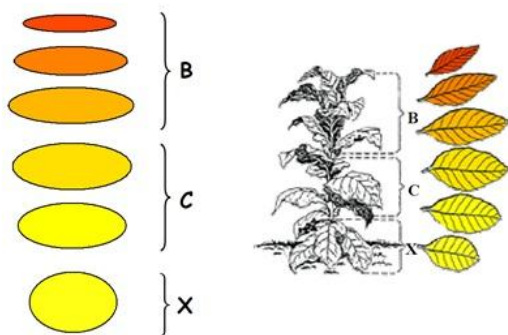
Figure 6. Classes in accordance with the position of the leaves, its dimensions and coloring.

Initially, to test the methods and backpropagation neural networks, artificial leaves was used. Table 1 shows the characteristics of color and size of 9 samples created for training artificial neural network. The characteristics were obtained from information from an expert.

Table 1 Samples used for training artificial leaves.

Sample	W (px)	L (px)	Color			Class specialist
			R	G	B	
1	170	51	255	72	1	B
2	140	36	255	129	0	B
3	150	41	255	183	0	B
4	205	140	255	224	0	C
5	187	98	255	255	0	C
6	192	123	225	223	0	C
7	130	80	255	250	13	X
8	140	91	255	253	24	X
9	152	105	255	255	21	X

To classify the tobacco leaves was used a backpropagation neural network with 4 input parameters, 10 hidden layer and 3 outputs, each representing one of three classes: B, C and X. Table 2 shows the result obtained with artificial images.

Table 2 Samples used for testing artificial leaves.

Sample	W (px)	L (px)	Color			Class Rede Neural
			R	G	B	
1	160	41	255	72	1	**B**
2	152	33	255	129	0	**B**
3	201	115	255	224	0	**C**
4	209	119	255	255	0	**C**
5	145	90	255	250	13	**X**
6	135	102	255	253	24	**X**

To verify the results was used 40 samples of tobacco leaves in their different classes. Leaves were classified using expert knowledge. Table 2 shows the characteristics of color and size classification of each sample.

Table 3 Samples used for training the neural network.

Sample	W (px)	L (px)	AVERAGE COLOR			CLASS
			R	G	B	
25	357	154	106	78	46	**B**
16	372	166	112	80	34	**B**
19	394	140	120	73	36	**B**
20	383	165	118	82	41	**B**
17	441	273	115	130	30	**C**
26	428	253	106	143	46	**C**
18	397	240	126	110	55	**C**
14	402	262	119	122	51	**C**
22	331	140	118	139	80	**X**
27	346	153	110	145	77	**X**
35	324	141	115	135	82	**X**
34	310	137	117	149	73	**X**

DISCUSSION

The processing of the test file of artificial images showed good results for classification, through the trained network, as shown below in Table 4.

Table 4 Samples of artificial leaves used for the test.

Sample	L	W	Color			Result			D
			R	G	B	RN			
						B	C	X	
1	160	41	255	72	1	**1,0**	-0,9	-1,0	B
2	152	33	255	129	0	**1,0**	-1,0	-1,0	B
3	201	115	255	224	0	-1,0	**1,0**	-1,0	C
4	209	119	255	255	0	-1,0	**1,0**	-1,0	C
5	145	90	255	250	13	-1,0	-1,0	**1,0**	X
6	135	102	255	253	24	-1,0	-1,0	**1,0**	X

The Figure 7 shows the learning curve obtained at the end of the training 12 samples, each containing four prototype class.
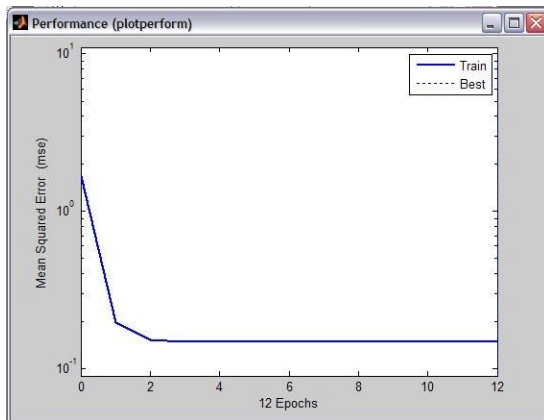
Figure 7. Learning curve for the RN - backpropagation

Thus, as shown in Table 1, using artificial images were obtained 100% accuracy in classifying the six images used for testing. This ensures the functioning of the implemented algorithm under controlled situations, ie, with no problems as noise, shadows, out of focus, reflexes, etc.

As shown in Table 5, with using real images were obtained 87% correct classification of the 16 pictures used in the test file.

Table 5 Results obtained in the classification of real samples of the test file.

Sample	W	L	Average Color			Result RN	Result Desered
			R	G	B		
01	357	154	106	78	46	B	B
02	372	166	112	80	34	C	B
03	357	154	106	78	46	B	B
04	394	140	120	73	36	B	B
05	383	165	118	82	41	X	B
06	441	273	115	130	30	C	C
07	428	253	106	143	46	C	C
08	397	240	126	110	55	B	C
09	402	262	119	122	51	C	C
10	428	253	106	143	46	C	C
11	331	140	118	139	80	X	X
12	346	153	110	145	77	X	X
13	324	141	115	135	82	X	X
14	310	137	117	149	73	B	X
15	310	137	117	149	73	B	X
16	310	137	117	149	73	X	X

The topology of the neural network in training and performance are presented in Figure 8. In Figure 9 we present the learning curve in the real image training.
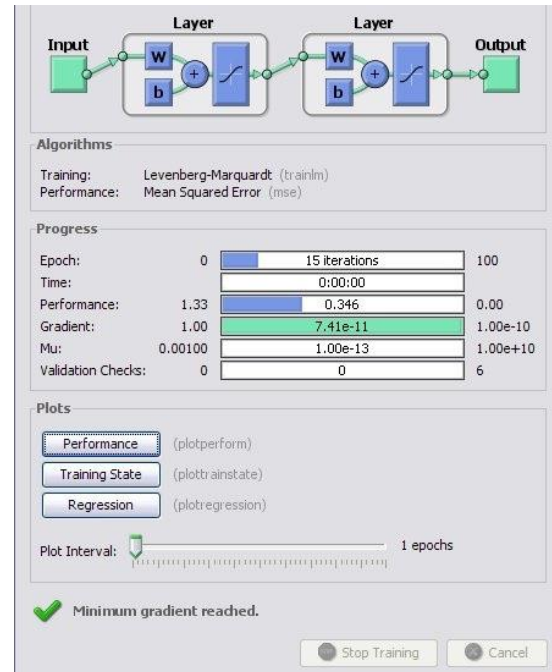
Figure 8. Topology of RN and Performance Training.
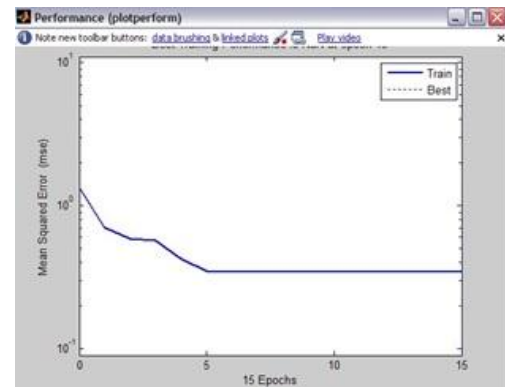
Figure 9. Learning curve in the real image training.

The confusion matrix, Table 6, shows the successes and mistakes made by the assessed hypothesis. In this case, it is concluded that the precision is 87%.

Table 6 Confusion Matrix.

	B	C	X
B 5	5	0	0
C 5	1	4	0
X 6	1	0	5

CONCLUSIONS

As demonstrated using artificial images obtained 100% accuracy in classification of 6 images used on the test file, containing the 3 classes. This ensures the functioning of the implemented under controlled situations, ie, with no problems as noise, shadows, out of focus, reflexes, etc. algorithm. In real samples was obtained 87% of correct classification. Thus it is considered that the objectives have been achieved and that the results are fully satisfactory.

ACKNOWLEDGMENTS

REFERENCES

[1] G.S. Amaral, *Sistema de Posicionamento Dinâmico para um Pequeno Veículo Flutuante*, x, 60 pp , 29,7 cm (UERJ, Engenharia Eletrônica, 2008). Projeto de Graduação – Universidade do Estado do Rio de Janeiro.

[2] A. R. Weeks. *Fundamentals of Eletronic Image Processing*, (1996). IEEE Press Artit Yawootti and Pakorn Kaewtrakulpong. A Machine Vision System for Thai Flue-Cured Tobacco Classification. Dissertação de Mestrado. Department of Control System and Instrumentation Engineering, 2006.

[3] R.C. Gonzalez; R.E. Woods. *Digital Image Processing*, 3rd edition. Pearson, 2010.

[4] Mathworks Inc. *Matlab users Manual*. Prentice Hall, 2010.

[5] The Lab. Book Pages. Otsu Thresholding. Site da internet. Disponível em:<http://www.labbookpages.co.uk/software/imgProc/otsuThreshold.html>.Acesso em: 15 nov. 2013.

[6] F. Zhang; X. Zhang. *Classification and Quality Evaluation of Tobacco Leaves Based on Image Processing and Fuzzy Comprehensive Evaluation.* Disponível em: http://www.mdpi.com/journal/sensors.

[7] D.A. Forsyth; J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.

[8] Mubarak Shah. *Fundamentals of Computer Vision.* Disponível em: <http://www.cs.ucf.edu/courses/cap6411/book.pdf>. Acesso em: Acesso em: 17 nov. 2013.

[9] R. Szeliski. *Computer Vision: Algorithms and Applications.* Springer, 2001.

[10] R. O. Duda; P.E. Hart, and D. G. Stork. *Pattern Classification*, Wiley Interscience, 2001.

[11] S. S. Haykin. *Redes Neurais* - 2ed. Bookman, 2001.

A Feature-Level Fusion for Multimodal Biometrics

Ren-He Jeng and Wen-Shiung Chen
VIP-CCLab., Dept. of Electrical Engineering
National Chi Nan University, Taiwan
E-mail: wschen@ncnu.edu.tw

and

Lili Hsieh
Dept. of Information Management
Hsiuping University of Science and Technology, Taiwan

ABSTRACT

Due to inherent properties in each biometric, none of any single biometric can guarantee 100% authentication accuracy by only itself. Information fusion is the key step in multimodal biometric recognition systems. Previous researches have confirmed that fusion at low-level considered to be effective, but difficult. In this paper, a feature-level fusion mechanism based on shuffle coding rule is proposed and three types of fusion are addressed. Some competitive experimental results show that the proposed feature fusion mechanism is very advantageous.

Keywords: Biometrics, Multimodal, Information Fusion, and Feature-Level Fusion.

1. INTRODUCTION

Previous works [1]-[6] organized multimodal biometrics into two major approaches, pre-classification and post-classification. Pre-classification refers to combining information on senor- or feature- level. The post-classification combines classification information, such as matching sources and decision.

The previous researches confirmed that fusing information at early stages will be much effective than at later stages. Since the source data and feature data contain rich information, fusing information at pre-classification is more useful than at post-classification [6]. Despite its improving accuracy, multimodal biometric has several drawbacks. First, since recognition is a serial steps, it may appear error propagation problems. Second, combining non-homogenous features is difficult and even impossible, because the relationship of each feature spaces may not be known. Third, concatenating two feature vectors result in a feature vector with huge dimensionality leading to the 'curse of dimensionality' problem. Therefore, few researchers have studied integration at the feature level. Besides, it is too rigid to combine information at post-classification. We will consider both the implementation complexity and information quantity. Therefore, we focus mainly on feature-level fusion.

A good feature subset would contain most information and least noise under a feature selection criterion. However, it cannot guarantee to obtain the optimal feature subset. In feature selection methods, the filter-based method is universal and a fast way in removing most of non-critical noises. It could work as a pre-selector to evaluate feature subsets in advance for reducing scale of searching. Whereas, the wrapper-based method may be viewed as a part of learning algorithm, which evaluates feature subsets based on the clustering error rate. The wrapper-based method may achieve a higher accuracy and find a smaller feature subset, nevertheless more computation time must be paid in evaluating feature subsets. Feature selection methods are always based on a rigorous mathematical derivation and are usually applied to high dimensional complex problems. In many cases, the real underlying components can be more naturally imagined as sparse vectors.Feature combination method is a simple and intuitive method in combining features, however not universal. Since the feature-level fusion could be promising approaches to improve recognition performance by combining various biometric traits, in this paper a new information fusion mechanism based on a coding scheme with shuffle operation at feature level, called Shuffle Coding-based Feature-Level Fusion (SC-FLF), is introduced to combine the features from several different modalities. The proposed SC-FLF approach is actually viewed as an extension to feature combination, in which each combination of features can be represented by codes. Hence, the feature codes can be concatenated in series easily. Specifically, one of the advantages of the SC-FLF mechanism is its universality.

2. THE PROPOSED FEATURE-LEVEL FUSION

2.1 OVERVIEW OF THE METHOD

In our proposed approach, feature extractors extract feature vectors by using different algorithms and quantizing scales. Since feature spaces are uncorrelated, it is difficult or even impossible to combine them. We suppose that even two distinct groups of features are not correlation, the distance of a pair of features are correlation. To describe the distances, the quantization scales should be properly considered. Therefore, our approach normalizes features in distinct independent spaces by using different normalization parameters. Since feature spaces are independent, it can reserve the relativity between original data in new feature spaces.

In this section, a fusion mechanism, called shuffle coding-based feature-level fusion (SC-FLF), is to combine two (or more) feature vectors to form a new feature via "shuffle" operation. The SC-FLF mechanism consists of the following processes: (i) feature extraction, (ii) dimension adjustment, (iii) Feature re-mapping, and (iii) Shuffle coding fusion, as shown in Figure 1.
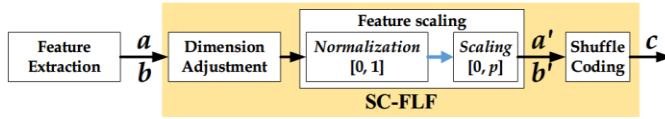
Figure 1. The framework of SC-FLF mechanism.

Let $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ be two feature vectors, obtained by any different feature extraction algorithms. We select the interpolation rule as the candidate to adjust feature vectors. The available fused data a and b are adjusted into the length of $d = \max(m,n)$. Then, the feature scaling is a method used to normalize and up-remapping the vectors as $a' \in \mathbb{R}^d$ and $b' \in \mathbb{R}^d$. in the range $[0, p]$. Finally, a shuffle coding rule generates a new binary codes $c_i \in \{0,1\}$ length p as $c \in \mathbb{R}^{p \times d}$.

2.2 Feature Re-mapping

The system first extracts the feature vector from each trait no matter how the modalities are same or different. It is actually difficult, or even impossible, to fuse them since the feature vectors extracted by different methods always are in different feature spaces. The feature re-mapping process that re-maps the feature vectors onto the same feature space by using the re-scaling operation is obligatory to avoid a disregard for the feature vector with a smaller range of value. It can become the basis for combining the features.

This implies that a proper normalization rule should be considered. The proposed approach normalizes the feature vectors in distinct independent spaces by different normalization parameters. For example, the dimensions of the feature vectors a and b are, respectively, m and n. Since the normalization parameters are different for each vector, there are $m+n$ available quantization parameters in total. After up-scaling the normalized vectors, the independence between two raw data vectors may be well reserved in the new feature space. Consequently, after re-mapping a and b with the distinct normalization parameters independently, the new feature space well preserves their own distributions.

Consider the feature vector $f \in \mathbb{R}^n$ containing n feature elements x_i. The feature element f_i, for $i = 1, 2, \ldots, n$, is re-scaled into a decimal real value of f'_i, within the range of $[0, p]$. In this paper, three re-scaling rules, such as max-min, fuzzy, and max-mean, are suggested and described as follows:

Max-min: Max-min re-scaling is a process by transforming the data measured in its engineering units into decimal values between 0.0 and 1.0, and then performing the up-scaling with the parameter of p. This re-scaling rule ensures that the extreme values are constrained within a fixed range. The re-scaled values are computed by

$$f'_i = \left\lfloor \left(\frac{f_i - f_{min}}{f_{max} - f_{min}} \right) \times p \right\rfloor,$$

where $f_{max} = \max\{f_i\}$, and $f_{min} = \min\{f_i\}$ for all i are, respectively, the maximum and the minimum of the elements of the feature vector f.

Fuzzy: Fuzzy re-scaling is defined as follows:

$$f'_i = \left\lfloor \left[\frac{1}{2} + \frac{1}{2} \sin \left(\frac{\pi}{f_{max} - f_{min}} \left[f_i - \frac{1}{2} \left(f_{max} - f_{min} \right) \right] \right) \right] \times p \right\rfloor.$$

Max-mean: The value f'_i represents the distance between the raw data and the population mean. The f'_i is close to zero when the raw data is below the mean. The max-mean re-scaling is given by

$$f'_i = \left\lfloor \left(\frac{f_i - \mu}{M} + \frac{1}{2} \right) \times p \right\rfloor,$$

where the mean value $\mu = 1/n \sum_{i=1}^{n} f_i$ and $M = \max\left(f_i - \mu \right)$.

2.3 Fusion Strategy: Shuffle Coding

After adjusting the feature vector dimensions and quantizing the feature vectors, a fusion algorithm based on shuffle coding scheme is designed to produce the composite feature code C_s from the trait data a' and b' (actually, a and b). The length of the binary code (i.e., shuffle code) c_i is p for each fusion feature. If $a' \leqq b'$, this algorithm returns a binary code $c_i \in \mathbb{R}^p$, in which

$$c_{ij} = \begin{cases} 1 & a'_i \leq j \leq b'_i \\ 0 & others \end{cases}, \forall i = 1, 2, \ldots, d, j = 1, 2, \ldots, p.$$

The code shall be set to "1" from bit- a'_i to bit- b'_i, and the other bits are set to "0." That is, the code bits from bit-$(a'_i$-1) to bit-$(b'_i$-1) are set by "1," and the other bits are set by "0." Conversely, if $a' > b'$, it returns a binary code c_i by

$$c_{ij} = \begin{cases} 1 & \left(j < a'_i + 1 \right) \vee \left(j > b'_i - 1 \right) \\ 0 & others \end{cases}, \forall i = 1, 2, \ldots, d, j = 1, 2, \ldots, p.$$

Our approach produces the code that is set to "0" from bit-b'_i to bit-a'_i, and the other bits are set to "1." In this formula, the MSB (most significant byte) always represents a value of 1. The MSB in an p-bit binary number represents a value of 2^{p-1} decimal. In summary, our approach can be written as:

$$c_i = \begin{cases} [2^{b'_i} - 2^{(a'_i-1)}]_2 & a'_i \leq b'_i \\ [(2^p - 1) - (2^{a'_i} - 2^{(b'_i-1)})]_2 & a'_i > b'_i \end{cases},$$

where $c_i = (c_{i1}, c_{i2}, \ldots, c_{ip}) \ \forall c_{ij} = \{0,1\}$, and the operation $[]_2$ stands for binary conversion. Finally, according to the quantized values of all input feature vectors, the feature code vector is obtained by

$$C_s = \left[c_1, \cdots, c_i, \cdots, c_d \right].$$

Figure 2 shows some S-FLF codes for different classes, in which there are two samples for each class.

Table 1 The accuracy of S-FLF Type 1.

		H-F	H-U	H-L	H-R	F-H	F-U	F-L	F-R	U-H	U-F	L-H	L-F	R-H	R-L
local-global	n1	5.6	2.2	1.7	3.0	6.2	6.7	5.4	6.1	3.1	6.0	4.3	8.7	5.4	9.3
	n2	6.2	2.7	2.3	3.7	8.3	6.7	6.6	6.9	4.2	7.2	5.6	10.7	5.9	10.1
	n3	4.5	2.7	3.0	3.0	4.0	4.1	5.2	4.5	3.2	6.2	3.0	6.3	5.2	6.4
global-local	n1	6.1	3.8	4.4	5.6	5.2	5.3	7.9	8.4	2.3	5.9	2.0	5.2	2.6	6.1
	n2	7.3	4.8	5.4	6.3	6.2	6.7	11.3	11.0	2.6	5.4	2.0	5.6	3.1	6.2
	n3	4.4	2.6	3.0	5.1	4.3	5.0	6.4	6.9	1.9	4.4	2.6	4.1	2.7	4.8

Table 2 The accuracy of S-FLF Type 2.

		H	F	U	L	R
local-global	n1	1.8	12.4	4.2	4.0	6.0
	n2	1.9	13.2	5.2	5.1	6.7
	n3	2.9	9.0	4.2	4.1	4.6
global-local	n1	2.2	10.6	5.2	6.2	6.3
	n2	2.6	12.4	5.4	5.9	6.8
	n3	2.4	8.7	3.9	3.5	5.5

Table 3 The accuracy of S-FLF Type 3.

		H.F	H.U	H.L	H.R	F.U	F.L	F.R	H.F.U	H.F.L	H.F.R
local-global	n1	1.67	0.07	0.39	1.27	1.74	1.96	2.71	0.32	0.35	0.90
	n2	1.39	0.34	0.41	1.63	2.05	2.07	2.90	0.52	0.31	0.70
	n3	0.78	0.06	0.36	1.02	1.34	1.34	2.02	0.01	0.28	0.36
global-local	n1	1.76	0.30	0.35	0.78	1.82	1.25	2.16	0.33	0.26	0.46
	n2	1.37	0.68	0.68	1.47	2.73	2.18	2.33	0.33	0.32	0.72
	n3	1.10	0.07	0.28	1.23	1.28	1.04	2.38	0.01	0.02	0.37

(a)

(b)
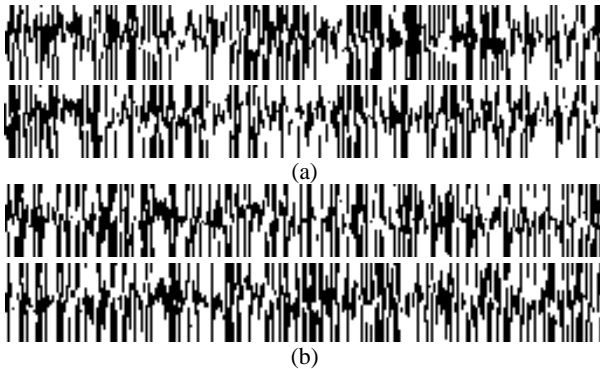
Figure 2. Some examples of shuffle coding operation.

3. EXPERIMENTAL RESULTS

In this paper, we combined three different biometrics, such as face [7], iris [8], and palmprint [9]. The performance was evaluated by equal error rate (EER), which is the value where the false accept rate (FAR) and false reject rate (FRR) are equal.

According to a variety of combinations of feature vectors for generating S-FLF codes, three different types of S-FLF fusion scheme are addressed, as shown in Figure 3. (i) Type 1 - Uni-Biometric Mode: Single Trait - Double Features: Two distinct features *a* and *b* extracted from the "one" (same) trait with different feature extractors are fused, as shown in Figure 3(a). (ii) Type 2 - Bi-Biometric Mode: Double Traits - Double Features: Two distinct features *a* and *b* extracted from ``two'' (different) biometric traits with same or different feature extractors are fused, as shown Figure 3(b). (iii) Type 3 - Tri-

Biometric Mode: Multiple Traits - Multiple Features: Type 3, as shown in Figure 3(c), is actually extended from Type 2. Two distinct features *a* and *b* are extracted from "one" (same) trait with different feature extractors. The SC-FLF scheme fuses these features for each trait into a binary code independently using Type 2. Thus, these SC-FLF codes would not interfere with each other and can be concatenated easily. At the end a concatenation operation is performed to fuse all of the SC-FLF Type 2 codes into a single code.

(a) Type-1

(b) Type-2

(c) Type-3
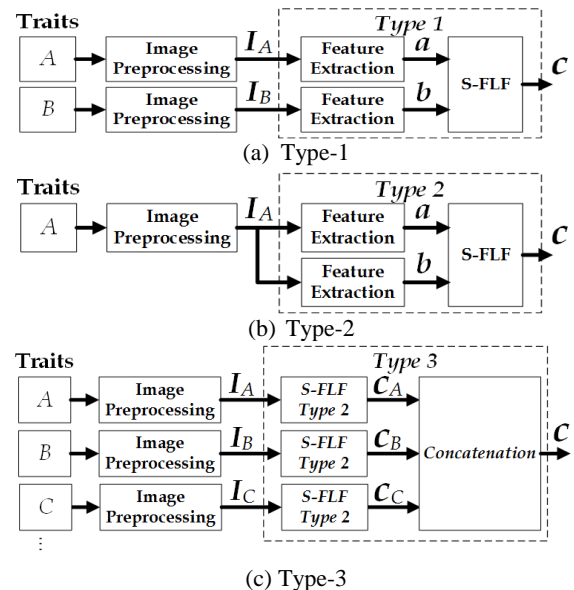
Figure 3. The frameworks of three types of S-FLF scheme.

As far as we know, a global feature provides the statistical information of an entire image while a local feature may provide some structural information. In this work, some local and global features will be fused by using the proposed SC-FLF scheme in the experimental multimodal biometric recognition system. Any suitable feature extraction methods for extracting the global and local features may be used. In the experiments, we adopt two simple feature expressions. The local feature extractor computes the mean values of each 8×8 patch on an enhanced image as the local features. The global features used here may be extracted from gray-level image and Local Binary Pattern (LBP) image, respectively.

In Table 1, the left symbol and right symbol in each of biometric trait combination respectively are referred to as feature vectors a and b to be fused by $S(a, b)$. The algorithms such as "max-min," "fuzzy," and "max-mean" are denoted by "n1," "n2" and "n3," respectively. All experiments are tested on the scales 8×8 of traits on three modalities are tested in the experiments.

In Figure 4, the difference values of quantized vectors *max-min* and *max-mean* are almost normally distributed. The quantized vectors (*fuzzy*) concentrate on the right-hand side (i.e., positive values). According Table 1 and Table 2, the experimental results and the distributions in Figure 3 shows that the feature a must be smaller than the feature b for a better recognition accuracy.
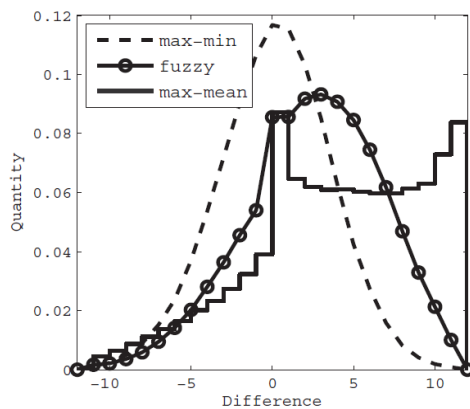
Figure 4. The distribution of the difference in features.

The SC-FLF Type 2 extracts two independent features separately using two distinct feature extractors from a single trait. Table 2 shows that, similarly, Type 2 also achieves better performances in the combinations of ``L-G'' and ``G-L'' for any of modalities. Our approach is workable and provides a great potential for fusing multiple features information of single modality at the feature-level.

The SC-FLF Type 3 is actually a parallel type of multiple Type 2, which may work on multiple traits from multi-modalities. The SC-FLF Type 3 concatenates multiple binary codes in which each of them is generated by Type 2 fusing the features a and b extracted from a single modality. Our approach uses the most suitable parameter p.

4. CONCLUSION

This work proposed an information fusion scheme at feature level for multimodal biometrics. Unlike previous works, our approach describes the relationship between two feature spaces. It does not need too much resource and is simple. The experimental results reveal that the proposed tactics can improve the performance further.

REFERENCES

[1] A. Ross, K. Nandakumar, and A. K. Jain, **Handbook of Multibiometrics**, 1st ed. Springer, 2006.

[2] B. Bhanu and V. Govindaraju, **Multibiometrics for Human Identification**, 1st ed. Cambridge University Press, 2011.

[3] R. S. Blum and Z. Liu, **Multi-sensor Image Fusion and Its Applications**, CRC Press, 2006.

[4] A. Ross1, A. Jain and J. Z. Qian, "**Information fusion in biometrics**," International Conference on Audio- and Video-Based Person Authentication, vol. 1, 2000, pp. 354-359.

[5] C. Sanderson and K. K. Paliwal, "**Information fusion and person verification using speech and face information**," Research Paper IDIAP-RR 02-33, 2002.

[6] A. K. Jain, K. Nandakumar, and A. Ross, "**Score normalization in multimodal biometric systems**," Pattern Recognition, vo. 38, no. 12, pp. 2270-2285, 2005.

[7] FEI Face Database. http://fei.edu.br/~cet/facedatabase.html.

[8] H. Proença and L. A. Alexandre, "**UBIris: A noisy iris image database**," International Conference on Image Analysis and Processing, pp. 970-977, 2005. http://iris.di.ubi.pt/

[9] W. S. Chen, Y. S. Chiang, C. M. Hsieh, S. W. Shih and Y. H. Chiu, "**Biometric Verification with Hybrid Features of Palm-print and Hand Geometry based on Wavelet Trans-form and Statistical Moments**," IPPR Conference on Computer Vision, Graphics and Image Processing, 2005.

Super-resolution in Infrared Imaging Systems Affected by Nonuniformity

Guillermo MACHUCA[a,*]**, Pablo MEZA**[a]**, and Sergio TORRES**[b]

[a]Departamento de Ingeniería Eléctrica
Universidad de La Frontera
Temuco, Chile
[b]Departamento de Ingeniería Eléctrica and Center for Optics and Photonics
Universidad de Concepción
Concepción, Chile
*g.machuca01@ufromail.cl

ABSTRACT

In this paper an analysis for super-resolution (SR) techniques using imagery affected by nonuniformity (NU) noise is presented. In general, the SR techniques are based on an acquisition mathematical model that does not consider NU noise. The applicability for standard SR techniques is illustrated by applying the SR algorithms over images affected by simulated NU noise. Further, result on real infrared (IR) imagery corrupted by NU is included. To this end, a laboratory prototype, based on a CEDIP Jade-UC IR camera, has been implemented to acquire IR data in the range of 8-12 [μm]. A mobile platform has also been constructed to simulate the scene's shifts required by the SR techniques.

Keywords: Superresolution, Infrared Focal Arrays, Nonuniformity, Infrared Imaging Systems, Fixed-Pattern Noise.

1. INTRODUCTION

Practical IR imaging systems are limited in terms of their spatial resolution, which is directly related with the size of their Focal-Plane Array (FPA) [1]. This is due to the costs and difficulties involved in the design and implementation of larger IR detector arrays [2]. In this regard, using SR image processing techniques increases the spatial resolution of an IR imaging system, without modifying the hardware of the camera [3]. In this work a sequence of low resolution (LR) IR images from the same scene, with small shifts in the spatial domain, are used to generate a high resolution (HR) image [4]–[6]. The SR process can be separated in two stages: i) the registration and ii) the reconstruction. In the registration process, the parameters describing the translation and rotation between the LR images are estimated; this is accomplished by considering LR images as basis [6]–[8]. In the reconstruction process, by using the parameters obtained from (i), the HR image is estimated using different techniques for reconstruction [5], [9], [10]. Further, the traditional model used for representing the image acquisition is:

$$Y_k = D_k \cdot B_k \cdot M_k \cdot X + n_k, \qquad (1)$$

where $k = 1, \ldots, N$ is the image index from a set with cardinality N, and Y_k and X are the LR and HR image, respectively. M_k is the matrix that captures the shifts between

images, B_k is the defocusing matrix, D_k represent the decimation matrix and n_k is the additive random noise [4], [5]. This model is widely used for representing the digitalization of a scene in the visible (VIS) range.

For uncalibrated IR imaging systems, the acquired imagery is also corrupted by another type of noise [2], [11]. The noise, better known as NU noise, is due to the nonuniform response produced by each detector when stimulated by the same amount of irradiance [12], [13]. Also, if the parameters defining the NU noise are constant during the operating time of the IR camera, the noise can be considered fixed in time as a Fixed-Pattern Noise (FPN) [1], [14], [15]. The model described in Eq. (1) does not consider the presence of NU noise, which affects the performance of the SR algorithms during registration and reconstruction of the HR image.

When the NU compensation is overlooked, both image registration [6], [7], [16], [17] and reconstruction [5], [9], generates a HR image with the presence of a pronounced NU noise [3], [8], [15]. To the best of our knowledge, a few methods have been proposed to compensate for NU noise during the SR process. Armstrong *et al.* shows that Non-Uniformity Correction (NUC) improves the accuracy of registration, and consequently, enhances the quality of HR reconstruction [14]. Hardie and Droege proposed a maximuma posteriori (MAP) algorithm for estimating a noise-free HR image from a set of LR images corrupted by NU noise [15].

In this article, an evaluation of traditional SR techniques using simulated and real IR images corrupted by NU noise is performed. In this regard, two image quality metrics were included: the Root Mean-Squared Error (RMSE) and the roughness metric.

The main goal is to analyze the capability of SR techniques when using IR imagery corrupted by NU.

The rest of this article is structured as follows. In Section 2, we briefly review the observation model and the algorithms used for registration and reconstruction. Next, in Section 3, the experimental results are presented and the accuracy of the estimated HR IR images, when affected by NU noise, is highlighted. Finally, in Section 4 we present the conclusions of our results, and lastly, we indicate prospective paths for future work.

2. OBSERVATION MODEL

The acquisition process correct representation is illustrated in Figure 1, which relates a set of LR images with a corresponding desired HR image.
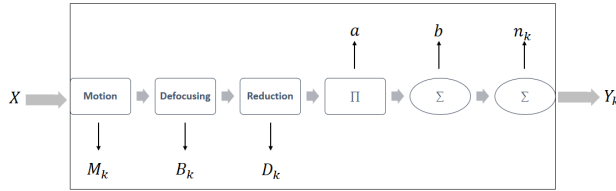
Fig. 1: Block diagram of an IR imaging system. The HR scene is acquired by the IR camera, modifying the signal during each stage of the transduction process.

From Fig. 1, were X is the HR input captured by the sensor. The scene translation and rotation are represented by the matrix M_k. The two dimensional signal is affected by blurring mainly due to the camera lens and atmospheric conditions, which is modeled by the matrix B_k. The transduction process is carried out by the FPA, which converts the incoming radiation into an electrical current. However, due to the small size of the array, only a portion of the scene is captured. Hence, the matrix D_k represents the decimation process carried out by the FPA. The NU noise is mathematically represented by a multiplicative and additive noise term, a and b, respectively [13], [15]. It should be remarked that both parameters are independent from the image index. Finally, the electrical current is amplified and converted into digital count by the readout electronic. In this stage, the random noise modeled by n_k appears. Therefore, the following pixel-based model is extracted:

$$Y_k = a \cdot D_k \cdot B_k \cdot M_k \cdot X + b + n_k. \quad (2)$$

Under the conditions presented in the Eq.(2), the performance of the registration and reconstruction algorithms will be analyzed.

2-A. Registration Methods

The registration algorithms studied in this work are the: i) Keren, ii) Lucchesse, iii) Marcel, and iv) Vanderwall algorithms. Keren *et al.* [16], [18] propose an algorithm defined in the spatial domain based on approximations with the Taylor series. The shifts in space are represented by a first order function that includes the spatial displacement and rotation on the image. Lucchesse and Cortelazzo [7] propose an algorithm defined in the frequency domain, computing the rotation angle between LR images exploiting the properties of the Fourier transform. The rotation is calculated by comparing the Fourier spectrum of the LR images with a reference image spectrum. Consequently, the translation parameters are estimated by computing the ratio between phases. Marcel *et al.* [17] used a polar representation of the magnitudes in the frequency domain to estimate the rotation angle through phase correlation. Likewise, they performed a counter-rotating Fourier

transforms and again applied a phase correlation technique to retrieve the translation vector. In Vanderwall *et al.*. al [6], the authors propose an algorithm in the frequency domain using the Fourier transform of the LR images and estimate the rotation angle in polar coordinates. Also, the computation of the image displacement is obtained from the phase difference.

2-B. Reconstruction Methods

The algorithms used for reconstructing the HR image are: i) nonuniform interpolation and ii) robust super-resolution. The nonuniform interpolation [9] seeks to reconstruct a HR image by mapping LR images, based on the parameters given in the registration stage, creating a new grid of points. Therefore, it performs a weighting of points, assembling the new HR image. This type of algorithm is highly sensitive to noise and blur. The robust algorithm [4], [5] follows an inverse imaging problem approach, using the model presented in Eq. 1. The estimation of the HR image is based on an iterative least-mean-square algorithm. These two types of reconstruction algorithms were chosen because they differ in the HR reconstruction process.

2-C. Image Quality Metrics

The metric for evaluating NU on SR algorithms were selected. The first metric is the RMSE, and it is used to measure the distortion of an image with respect to a reference image. It is worth mentioning that the images are in the same dynamic range. The mathematical representation of the metric is:

$$RMSE(X, \widehat{X}) = \left[\frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \left(X(m,n) - \widehat{X}(m,n) \right)^2 \right]^{\frac{1}{2}},$$
$$(3)$$

where $X(m,n)$ is the reference image and $\widehat{X}(m,n)$ is the estimated HR image. Also, $M \times N$ correspond to the image dimension. If the value of the RMSE is close to zero, the metric indicates that the estimated image is similar to the reference.

The second metric is the roughness (ρ), which measures the content of an image detail in the spatial domain. Unlike the RMSE, the ρ metric does not need a reference image. The mathematical formulation of the metric is presented as follows:

$$\rho(\widehat{X}) = \frac{\|h \otimes \widehat{X}\| + \|h^T \otimes \widehat{X}\|}{\|\widehat{X}\|}, \quad (4)$$

where h is the difference filter $[1 \ -1]$, the symbol $\otimes$ is the convolution and $\| * \|$ is L1 norm. A ρ value close to zero indicates that the image is relatively smooth.

3. EXPERIMENTAL RESULTS

In this section, we present the result for simulated NU noise in order to test the SR techniques under different known levels of noise power. From a noise-free HR image, a set of LR images with small shifts in the spatial domain is generated. According to the model presented in Eq. 2, the same NU noise has to be added on each LR image. Furthermore, by using an experimental setup based on a long-wave IR camera, a set of

real IR images affected by NU is generated. The IR data set is corrected by using a black body calibrator device to compare the SR techniques under ideal conditions.

3-A. Experimental Analysis Using Simulated NU Noise

The performance of the registration and reconstruction algorithms mentioned in Sections 2-A and 2-B, respectively, is tested over two data sets of simulated LR images corrupted by NU noise. The NU noise is generated by assuming a Gaussian distribution with two different variance values, which are defined according to the image's dynamic range (8 bits). Tables I and II list the performance results of HR images when affected by NU noise with a variance of 5% and 10%.

| | | Noise variance σ^2 | | | |
| | | 5% | | 10% | |
Registration	Reconstruction	ρ	RMSE	ρ	RMSE
Keren	Interpolation	0.3072	5.2384	0.3719	6.9089
	RobustSR	0.4223	6.0507	0.5524	7.5877
Lucchese	Interpolation	0.2505	7.0597	0.2710	8.3665
	RobustSR	0.3936	7.3939	0.5139	8.7609
Marcel	Interpolation	0.2196	7.0601	0.2778	8.6342
	RobustSR	0.3878	7.8620	0.5485	8.8782
Vanderwall	Interpolation	0.2587	6.9175	0.3657	6.9232
	RobustSR	0.4046	7.6836	0.5577	8.0010

Table I: Data set 1: SR results for simulated images affected by NU noise with a variance of 5% and 10%.

| | | Noise variance: σ^2 | | | |
| | | 5% | | 10% | |
Registration	Reconstruction	ρ	RMSE	ρ	RMSE
Keren	Interpolation	0.1196	3.9689	0.2003	6.2136
	RobustSR	0.2088	4.9445	0.3588	7.4383
Lucchese	Interpolation	0.1106	4.2014	0.1863	6.3415
	RobustSR	0.1965	5.2178	0.3370	7.4423
Marcel	Interpolation	0.0876	5.3115	0.1367	7.3028
	RobustSR	0.1788	5.5202	0.3030	7.4983
Vanderwall	Interpolation	0.1120	4.3253	0.2258	5.6063
	RobustSR	0.1867	4.8963	0.3677	7.0985

Table II: Data set 2: SR results for simulated images affected by NU noise with a variance of 5% and 10%.

Observing the results of Table I, with a NU noise variance of 5%, the lowest ρ value is obtained for the Marcel registration algorithm using interpolation reconstruction (0.2196). Further, the HR image generated with the Marcel algorithm plus the robust SR reconstruction show the highest results in this category with a values of 0.3878. In particular, the RMSE value closer to zero was 5.2384 for the Keren algorithm reconstructed by interpolation. For the images with a NU noise variance of 10%, the best ρ value was 0.2710 for the Lucchese algorithm using the interpolation reconstruction. The same case, but using the robust reconstruction achieved a ρ value of 0.5139. The best RMSE result was 6.9089 obtained by the algorithm proposed by Keren using the interpolation reconstruction. The registration performed by Keren and reconstruction using robust SR obtained a value of 7.5877.

In Table (II), for the images corrupted by NU noise with a variance of 5%, the best estimated HR image scored a ρ of 0.0876, which was obtained by the Marcel algorithm using the interpolation reconstruction. In terms of RMSE,

the algorithm proposed by Keren exhibits a value of 3.9689 using the interpolation. On the other hand, the algorithm by Vanderwall *et al.* scored a value of 4.8963 using the robust SR. For the case of NU noise with a variance of 10%, the Marcel algorithm showed a ρ value of 0.1367 using interpolation. The best RMSE values, 5.6063 and 7.0985, were obtained by the Vanderwall algorithm using interpolation and robust SR reconstruction, respectively.

A naked eye evaluation between the estimated HR images is depicted in Figs. 2 and 3. Both show that the presence of NU noise in the LR images persists on the estimated HR images. A closer examination of Figs. 2(c) and 2(d) shows small differences. In the letters section of the images, NU noise presents fewer artifacts in Fig. 2(d). This is due to the pre-filtering process performed by the Vanderwall algorithm. Figures 2(e) and 2(f) are the images reconstructed by robust SR. These estimated HR images show a greater level of detail, but with a stronger level of NU noise than the HR images produce with interpolation. The main difference between these two reconstruction processes is that the interpolation softens the estimated HR images during the weighing of the points.

For the Fig. 3, all the estimated HR images exhibit a higher image degradation than Fig. 2. This is due to the low variability in scene information. Comparing the images in Fig. 3(c) and 3(d), the Vanderwall algorithm produces an estimated HR image worse than that produced by the Keren algorithm. Moreover, the images in Figs. 3(e) and 3(f) show an even higher degradation, which indicates that the low variability in the scene information affects the robust SR estimation.

3-B. Experimental Analysis Using Real NU Noise

To measure the performance of the SR algorithms over real IR imagery, a laboratory set up has been implemented. It is composed of a CEDIP Jade-UC IR camera with a spatial resolution of 240×320 pixels, with a spectral response between 8-12 [μ m],and 14 bits digital output. In addition, detector material is an uncooled a-Si microbolometer. The reference IR source used for calibration is a Mikron black body radiator, which operates between 0-150°C and with a thermal resolution of 0.1^{o}C. Also, the scene shifts required for the SR estimation is performed by mounting the IR camera over two motorized linear stages.

| | Compensated Noise | |
Registration	Reconstruction	ρ
Keren	Interpolation	0.1121
	RobustSR	0.1480
Lucchese	Interpolation	0.3149
	RobustSR	0.2059
Marcel	Interpolation	0.0805
	RobustSR	0.1484
Vanderwall	Interpolation	0.1146
	RobustSR	0.1718

Table III: IR Data Set 1: SR results for real IR images compensated for NU noise using black body calibrators.

From Table III, the best ρ values were obtained for the Marcel algorithm using interpolation reconstruction (0.0805)
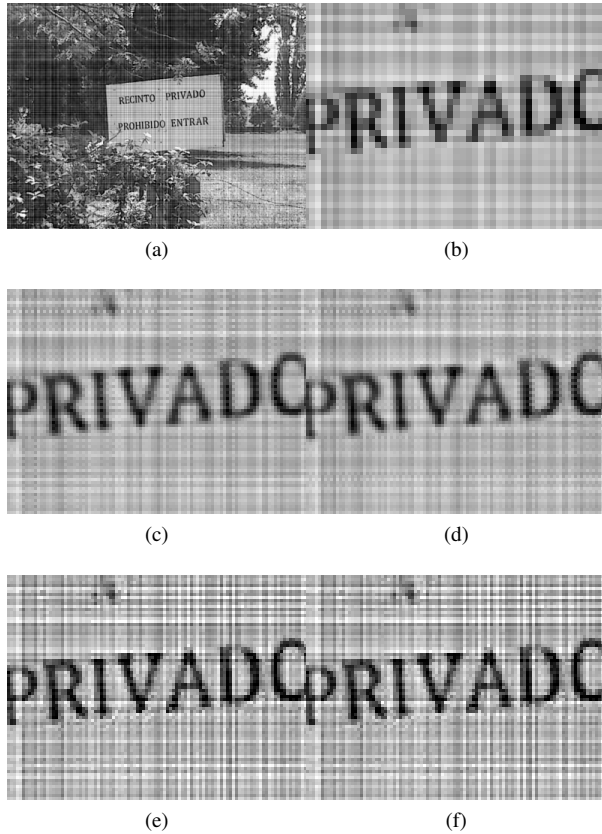
(a)   (b)

(c)   (d)

(e)   (f)

Fig. 2: Data set 1: SR results for simulated images affected by NU noise with a variance of 10%. (a) LR image, (b) HR image obtained with nearest neighbor algorithm, (c) HR image obtained with Keren algorithm and interpolation reconstruction, (d) HR image obtained with Vanderwall algorithm and interpolation reconstruction, (e) HR image obtained with Keren algorithm and robust SR reconstruction, and (f) HR image obtained with Vanderwall algorithm and robust SR reconstruction.

(a)   (b)

(c)   (d)

(e)   (f)

Fig. 3: Data set 2: SR results for simulated images affected by NU noise with a variance of 10%. (a) LR image, (b) HR image obtained with nearest neighbor algorithm, (c) HR image obtained with Keren algorithm and interpolation reconstruction, (d) HR image obtained with Vanderwall algorithm and interpolation reconstruction, (e) HR image obtained with Keren algorithm and robust SR reconstruction, and (f) HR image obtained with Vanderwall algorithm and robust SR reconstruction.

and for the Keren algorithm using the robust SR (0.1480). Also, in Table IV the best ρ values were obtained by the Keren algorithm using a interpolation reconstruction (0.1707) and by the Vanderwall algorithm using the robust SR (0.3018).

In Fig. 4(a) the original noise-free LR image is presented. Figure 4(b) corresponds to the HR image using the nearest neighbor technique. The images 4(c) and 4(d) are the estimated HR images using the Keren and Vanderwall algorithms for registration , respectively, and the interpolation technique for reconstruction. Figure 4(c) exhibits more details than Fig. 4(d), which is smoothed by the pre-filtering process performed by the Vanderwall algorithm. The images 4(e) and 4(f) are the estimated HR images using the Keren and Marcel algorithms for registration , respectively, and the robust SR technique for reconstruction. Both images appears to be similar in terms of spatial details. By comparing the images 4(c) and 4(e), it seems that the interpolation reconstruction achieves a
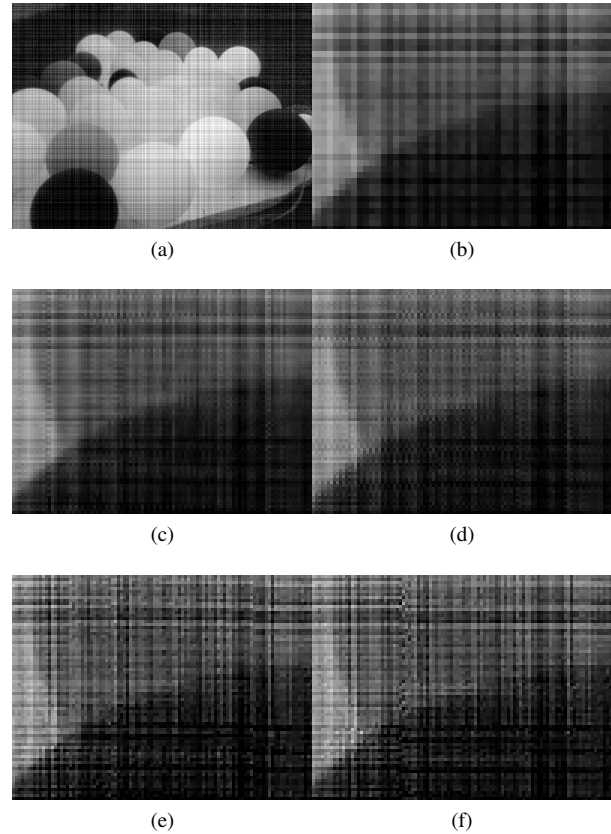
better quality HR image than the robust SR. In particular, the interpolation produces an apple with clearer details. Finally, all the estimated HR images shown an increase in detail compared to Fig. 4(b).

The IR data set affected by NU noise is presented in Fig. 5. Figures 5(a) and 5(b) correspond to the original corrupted LR image and to the HR image produced by the nearest neighbor technique, respectively. The images 5(c) and 5(d) are the estimated HR image using the Keren and Vanderwall algorithm for registration , respectively, and the interpolation technique for reconstruction. The images 5(e) and 5(f) are the estimated HR image using the Vanderwall and Marcel algorithm for registration, respectively, and the robust SR technique for reconstruction.

All the SR techniques performing poorly due to the presence of NU noise. However, the interpolation produces smoother images than the robust SR due to the low-pass filtering effect

	Actual Noise	
Registration	Reconstruction	ρ
Keren	Interpolation	0.1707
	RobustSR	0.3491
Lucchese	Interpolation	0.4399
	RobustSR	0.3488
Marcel	Interpolation	0.1786
	RobustSR	0.3153
Vanderwall	Interpolation	0.1793
	RobustSR	0.3018

Table IV: IR Data Set 2: SR results for real IR images corrupted by NU noise.

induced when weighting pixels. It should be noted that, each reconstruction technique produce an estimated image with a different noise structure than the original NU pattern.

4. CONCLUSIONS

In this article, an evaluation was conducted for different techniques of SR on captured images with different noise intensity. The first test was performed with simulated data, adding noise with a vairance of 5% and 10%. As for the second test, IR images were used with real NU noise. This was done using a laboratory prototype based on a CEDIP Jade-UC IR camera, which had been implemented to acquire IR data in the range of 8-12 [μm].

For the case of simulated data with a variance of 5% and 10%, the best results were obtained by the Vanderwall algorithm for the first simulation and the Keren algorithm for the second simulation, both reconstructed by interpolation. Moreover, for the case of the IR data, the Keren algorithm using interpolation has exhibited the best results for the first test. For the second test, all the SR techniques performed poorly due to the presence of NU noise. However, the Marcel algorithm using robust SR obtained very good results for the two tests test using real IR imagery.

Looking at the values obtained by the metrics, it can be seen that increasing noise intensity affects performance of the SR techniques.

As a future work, the spatial variability induced by small shifts in the scene required by the SR techniques could be used to compensate for the NU noise, since NUC algorithms have the same requirement.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Mark D. Nelson, Jerris F. Johnson, and Terrence S. Lomheim, "General noise processes in hybrid infrared focal plane arrays," *Optical Engineering*, vol. 30, no. 11, pp. 1682–1700, 1991.

[2] Dean A. Scribner, M.R. Kruer, and Joseph M. Killiany, "Infrared focal plane array technology," *Proceedings of the IEEE*, vol. 79, no. 1, pp. 66–85, Jan 1991.
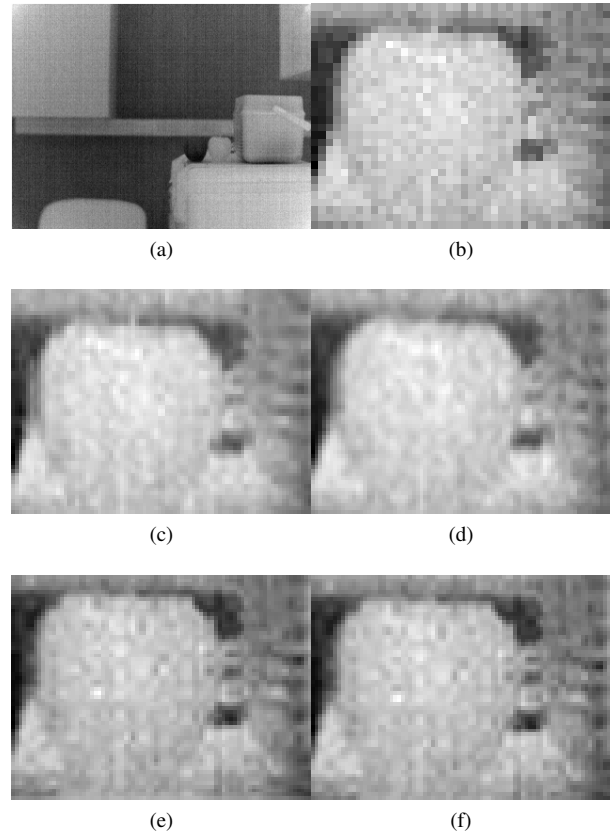
Fig. 4: IR Data Set 1: SR results for real IR images compensated for NU noise using black body calibrators. (a) LR image, (b) HR image obtained with nearest neighbor algorithm, (c) HR image obtained with Keren algorithm and interpolation reconstruction, (d) HR image obtained with Vanderwall algorithm and interpolation reconstruction, (e) HR image obtained with Keren algorithm and robust SR reconstruction, and (f) HR image obtained with Marcel algorithm and robust SR reconstruction.

[3] Russell C. Hardie, Kenneth J. Barnard, John G. Bognar, Edward A. Watson, and Ernest E. Armstrong, "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Optical Engineering*, vol. 37, no. 1, pp. 247–260, 1998.

[4] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *Image Processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1327–1344, 2004.

[5] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. I–645–I–650 vol.1.

[6] Patrick Vandewalle, Sabine Susstrunk, and Martin Vetterli, "A frequency domain approach to registration

(a)   (b)

(c)   (d)

(e)   (f)
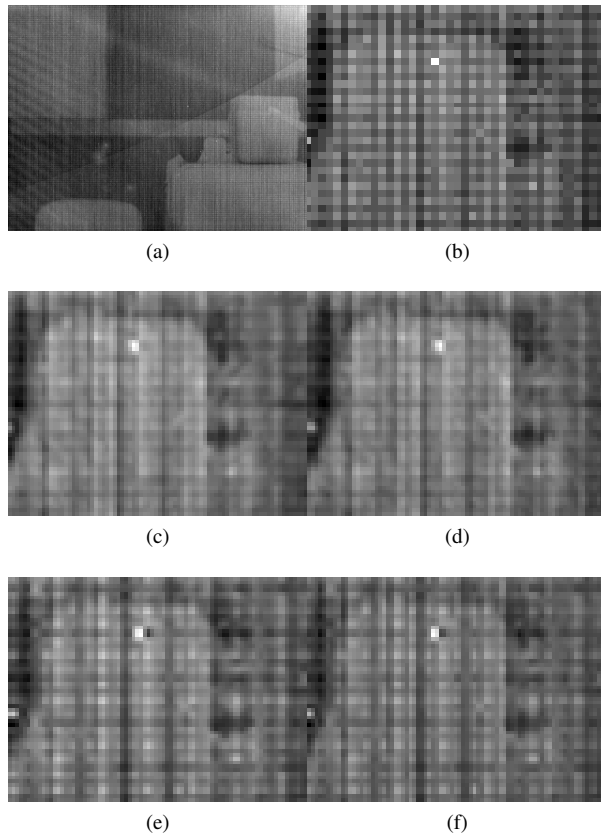
Fig. 5: IR Data Set 2: SR results for real IR images corrupted by NU noise. (a) LR image, (b) HR image obtained with nearest neighbor algorithm, (c) HR image obtained with Keren algorithm and interpolation reconstruction, (d) HR image obtained with the Vanderwall algorithm and interpolation reconstruction, (e) HR image obtained with the Vanderwall algorithm and robust SR reconstruction, and (f) HR image obtained with the Marcel algorithm and robust SR reconstruction.

of aliased images with application to super-resolution," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, pp. 071459, 2006.

[7] Luca Lucchese and G.M. Cortelazzo, "A noise-robust frequency domain technique for estimating planar roto-translations," *Signal Processing, IEEE Transactions on*, vol. 48, no. 6, pp. 1769–1786, 2000.

[8] Russell C. Hardie and Kenneth J. Barnard, "Fast super-resolution using an adaptive wiener filter with robustness to local motion," *Opt. Express*, vol. 20, no. 19, pp. 21053–21073, Sep 2012.

[9] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-resolution image reconstruction: a technical overview," *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 21–36, 2003.

[10] Nirmal K. Bose, Michael K. Ng, and Andy C. Yau, "A fast algorithm for image super-resolution from blurred observations," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 226–226, Jan. 2006.

[11] Chao Zuo, Qian Chen, Guohua Gu, and Xiubao Sui, "Scene-based nonuniformity correction algorithm based on interframe registration," *J. Opt. Soc. Am. A*, vol. 28, no. 6, pp. 1164–1176, Jun 2011.

[12] A. F. Milton, F. R. Barone, and M. R. Kruer, "Influence of nonuniformity on infrared focal plane array performance," *Optical Engineering*, vol. 24, no. 5, pp. 245855–245855–, 1985.

[13] M. Shehadeh and O. Kuybeda, "Robust nonuniformity correction in infrared images based on global motion," in *Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of*, 2008, pp. 275–279.

[14] Ernest E. Armstrong, Majeed M. Hayat, Russell C. Hardie, Sergio N. Torres, and Brian J. Yasuda, "Nonuniformity correction for improved registration and high-resolution image reconstruction in ir imagery," 1999, vol. 3808, pp. 150–161.

[15] Russell C Hardie and Douglas R Droege, "A map estimator for simultaneous superresolution and detector nonunifomity correction," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 089354, 2007.

[16] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on*, 1988, pp. 742–746.

[17] B Marcel, M Briot, and R Murrieta, "Calcul de translation et rotation par la transformation de fourier," *Traitement du Signal*, vol. 14, no. 2, pp. 135–149, 1997.

[18] Chong Fan, Jianya Gong, Jiangjun Zhu, and Lihua Zhang, "An improvement approach based on keren sub-pixel registration method," in *Signal Processing, 2006 8th International Conference on*, 2006, vol. 2, pp. –.

Denoising of SPECT-Image Sinogram-Data Before Reconstruction

Sicong YU, Hamed HAMID MUHAMMAD

School of Technology and Health STH, Royal Institute of Technology KTH, Stockholm, Sweden

ABSTRACT

Nuclear medicine images have low signal-to-noise ratio (SNR) due to several physical limitations which degrade the image quality considerably. In this study, the Gaussian filter and the patch confidence Gaussian filter (PCG) were used to improve the image quality for Single Photon Emission Computed Tomography (SPECT). The new approach applies these filtering methods on the acquired 2D-projections before reconstructing the image. The new approach was evaluated on a SPECT dataset and the performance was compared with several conventional methods presented in the literature.

Keywords: Low SNR, SPECT, Sinogram, 2D Projections, PCKNN, Patch Confidence Gaussian Filter PCG.

1. INTRODUCTION

Nowadays, a lot of filters are used to reduce the noise in reconstructed SPECT-images in order to improve the image SNR to fulfill the demands of clinical diagnosis and body functional research [1, 2, 3]. On the other hand, SPECT-image noise is initially generated during the acquisition process of gamma photons and added to the acquired projections which form the sinograms of the slices that can be reconstructed. During the scanning process, each gamma-photon detector measures the intensity of discrete incident photons over a certain time interval. The photons arrive on the sensor in a random and independent way and generate photon noise which can be described by the Poisson distribution (also called Poisson noise) corresponding to the statistically expected incident photon counts [5].

In conventional SPECT imaging systems, the images are reconstructed from the acquired sinogram data, then these images are denoised with various kinds of filters (i.e. post-reconstruction denoising is performed) in order to reduce Poisson noise and to enhance the image quality. Therefore, reconstructing the noisy sinogram data will spread out the distortion effect of the noise over the whole reconstructed image. Furthermore, since none of the existing reconstruction methods is a simple aggregation of the detected signals, the contamination of the reconstructed image with noise will occur and and result in an unpredictable noise type which in its turn depends on the chosen reconstruction algorithms.

Therefore, in this work, a novel denoising approach is proposed to apply the denoising filters directly on the 2D projections to reduce (and ideally eliminate) the Poisson noise before reconstructing the image (i.e. pre-reconstruction denoising is performed here). The Gaussian filter and the patch confidence Gaussian filter (PCG), which is a special case and simplified variant of the patch confidence K-nearest neighbor filter (PCKNN) [8], are utilized for this purpose. This approach is tested and evaluated on a SPECT dataset (2D gamma images or projections of a patient) and the resulting reconstructed images have been compared with the corresponding results of a number of existing conventional image-denoising methods.

2. METHOD

In the human body, the same type of tissue shows specific physical, chemical and physiological characteristics, such as absorption and accumulation of isotope-marked molecules and the attenuation coefficient which affects the penetration of gamma photons through this type of tissue. These factors generate the main information that can be captured by

the SPECT imaging technique. Therefore, regarding the denoising aspect for this kind of images, the statistical properties of surrounding voxels in SPECT images are expected to be similar. This makes it feasible to consider denoising approaches for SPECT images where a certain neighborhood around the currently filtered voxel is processed to contribute to the resulting voxel value. Therefore, in this work, the Gaussian [6] and the PCG filters are utilized for denoising purposes as follows in this section.

Gaussian Filter

When using the Gaussian filter [7], the whole image area will be modified by convolution with a Gaussian function given as follows:

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

where x is the distance from the origin (i.e. the central element) of the filter kernel in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

Patch Confidence Gaussian Filter

The new filter proposed and coined as the PCG filter in this work is (as discussed previously in this section) a special case of the patch confidence K-nearest neighbor filter (PCKNN) [8, 9]. The PCG filtering approach can be seen as a combination of Principal Component Analysis (PCA) and Gaussian filtering.
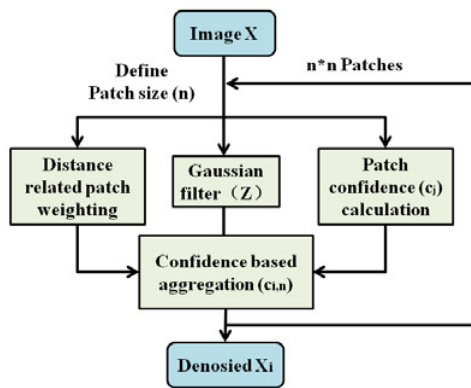
Figure 1. PC-Gaussian algorithm

The basic framework of the PCG filter is shown in Figure (1) and the filtering algorithm can be described according to the following steps:

1. Determine the patch size (n*n) depending on the original image size in addition to the size of the details of interest.

2. Patch Denoising: for each pixel Z inside the patch, apply Gaussian Filtering to the selected patch area (Z*):

3. $z^* = \frac{1}{\sum_{z_j \in KNN(z)} \omega_j} \sum_{z_j \in KNN(z)} \omega_j z_j$

    (2) Patch Denoising Confidence: compute a similarity measure showing the influence of the patch on the original pixel X as follows:

    $$c_j = \frac{(\sum_{z_j \in Gaussian(z)} \omega_j)^2}{\sum_{z_j \in Gaussian(z)} \omega_j^2} \qquad (3)$$

4. After calculating all denoised patches and the confidence factors of the patches, the value of the currently filtered voxel can be set as:

    $$x_i = \frac{\sum_1^{n^2} c_i x_{i,n}^*}{\sum_1^{n^2} c_{i,n}} \qquad (4)$$

    where $c_{i,n}$ is the confidence of patch n among the n$\times$n patches which contain pixel X_i.

3. EXPERIMENT AND RESULTS

A set of SPECT images was used to validate the proposed methods. The dataset was provided from Karolinska Institute Hospital Solna, Stockholm, Sweden. It consists of 32 2D-projections with 128-slices in each projection. The median filter [10], Gaussian filter and the proposed PCG filter are implemented and used in our proposed approach, where these denoising filters are applied to the SPECT projections as a preprocessing step before the image reconstruction step. The Maximum Likelihood Expectation Maximization (MLEM) image reconstruction method [11, 12] is used with 20 iterations to produce all SPECT images to be able to compare the performance of all filtering approaches. In addition, a reference image (denoted as undenoised) for each slice (there are totally 128 slices) is generated and reconstructed using the MELM reconstruction

algorithm only without additional denoising. Furthermore, these reference images are denoised using the median filter for comparison with the results of the proposed approach in this work.

Some slices are chosen for visual comparison of the reconstruction results.

Figures (2), (3) and (4) show comparisons of the reconstructed images for slices 48, 49 and 50, respectively. In these three figures, the reconstruction results are presented as follows: (a) is the undenoised image; (b) is the post-reconstruction denoising result using an additional median filter; (c), (d) and (e) are the pre-reconstruction denoising images obtained when using an additional median filter, Gaussian filter and PCG filter, respectively.

In these figures, two ischemia spots on the left ventricle (i.e. small regions of the heart muscle with reduced blood supply) can be easily recognized when using the two pre-reconstruction denoising approaches with the PCG filter and the median filter.

Visual inspection and comparison of the resulting images presented in these three figures shows that the other approaches are not as accurate as these two.

Furthermore, a quantitative evaluation and comparison of these four image denoising and reconstruction approaches is performed by employing the signal-to-noise ratio (SNR) of slice 50, as shown in Table (1).
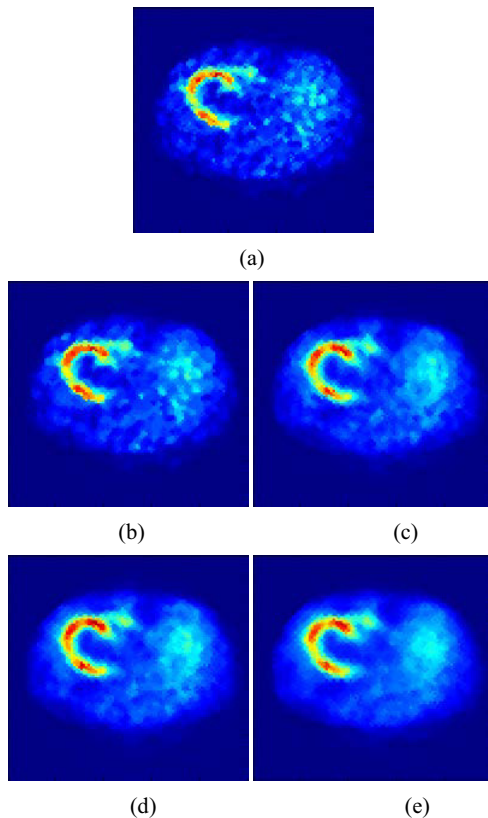
Figure 2. Reconstructed SPECT images of slice 48
(a) undenoised image; (b) post-reconstruction denoising using median filter; (c) pre-reconstruction denoising using median filter; (d) Pre-reconstruction denoising using Gaussian filter; (e) Pre-reconstruction denoisng using PCG filter.
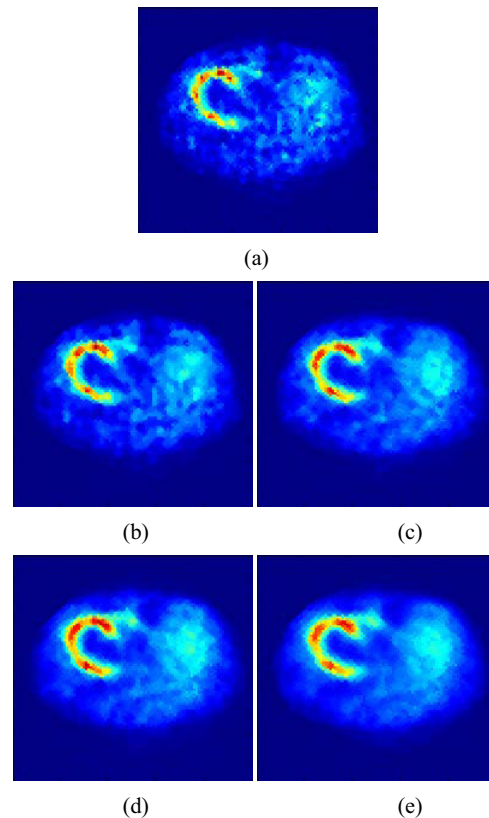
Figure 3. Reconstructed SPECT images of slice 49
(a) undenoised image; (b) post-reconstruction denoising using median filter; (c) pre-reconstruction denoising using median filter; (d) Pre-reconstruction denoising using Gaussian filter; (e) Pre-reconstruction denoisng using PCG filter.

Table (1) shows that among all four approaches, the pre-denoising reconstruction using the PCG filter can produce an image for slice 50 with the best SNR value (of 31.4 dB). Combining qualitative and quantitative performance evaluation indicates that the proposed approach using pre-denoising with a PCG filter can give the best reconstruction results.

(a)

(b)   (c)

(d)   (e)
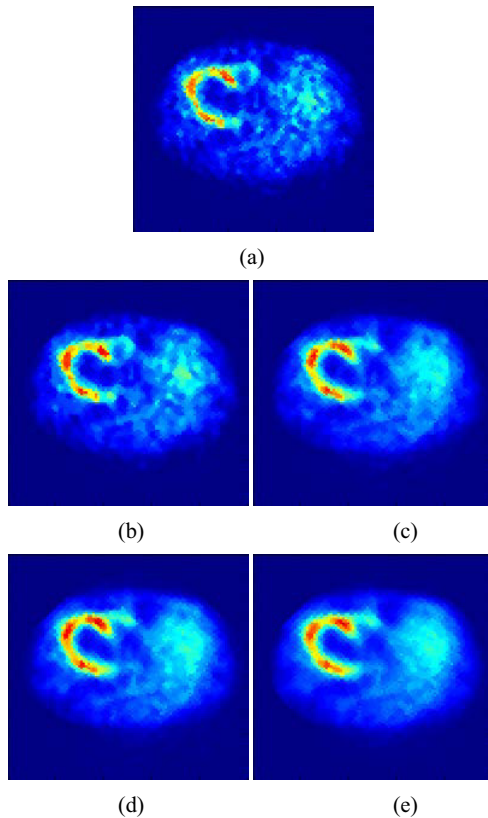
Figure 4. Reconstructed SPECT images of slice 50 (a) undenoised image; (b) post-reconstruction denoising using median filter; (c) pre-reconstruction denoising using median filter; (d) Pre-reconstruction denoising using Gaussian filter; (e) Pre-reconstruction denoisng using PCG filter.

Table 1. SNR of slice 50 using all four approaches.

MLEM reconstruction method		SNR (dB)
Undenoised		17.8
Pre-denoising method	Median filter	25.4
	Gaussian filter	30.8
	PCG filter	**31.4**
Post-denoising method	Median filter	21.4

The SNR is computed (in dB) and defined as:

$$SNR = 20 \log_{10} \frac{M_A}{S_A} \qquad (5)$$

where M_A is the mean intensity value within the object of interest region and S_A is the standard deviation of the intensity values of the same region.

4.   DISCUSSION

In a SPECT imaging system, the noise is generated and added to the projections while acquiring them. Thus it sounds logical to try to purify these projections from noise before using them to reconstruct the corresponding SPECT images. In addition to that, the statistical properties of a group of voxels within a small neighborhood in a SPECT 2D-projection are expected to be of similar nature. Therefore, it also sounds logical to expect that it is efficient to use a denoising approach which considers and processes a certain neighborhood of voxels. Therefore, the PCG filter is proposed and utilized to denoise the 2D-projections taking into account the influence of a neighborhood on the filtering process.

The final images that are reconstructed using this approach can give solid evidence that vital diagnostic information is enhanced and preserved.

5.   CONCLUSIONS

Enhancing the quality SPECT images is a challenging task because of the poor resolution and low SNR of these images. The aim of image denoising is to enhance and sharpen the edges and to smooth the inner part of each segment that is supposed to be uniform. The filters are always producing results that are related to the average of the information in a certain way [13]. Therefore, after denoising, the resulting image will be smoothed including the edges that will be faded away. In SPECT imaging used for cardiology and oncology, disease-related spots in SPECT images that have abnormal intensities

(brighter or darker than healthy and normally functioning tissues) represent the vital diagnostic information that medical doctors wish to detect and recognize in these images. This goal could be achieved in this work by implementing and applying the PCG filter to the 2D projections (i.e. sinogram data) before performing the reconstruction task. The results show that it is possible to efficiently reduce the noise in the SPECT image while preserving and enhancing the useful diagnostic information.

Future work will focus on optimizing and testing the proposed algorithm on more SPECT data as well as Positron Emission Tomography (PET) data since both types of data suffer from the same type of limitations and can be enhanced by the same type of algorithms.

6.   ACKNOWLEDGMENT

The authors would like to thank Prof. Bjorn-Erik Erlandsson (KTH, Stockholm, Sweden) for his support.

7.   REFERENCE

1.   M. J. Tapiovaara and R.F. Wagner. "**SNR and noise measurements for medical imaging: I. A practical approach based on statistical decision theory**." *Physics in medicine and biology*, vol. 38, no.1, pp. 71, 1993.

2.   M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, M. S. Verani, "**Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart**", *Circulation. 2002*; vol. 105, pp. 539-542.

3.   Blankenberg, G. Francis, "**Monitoring of Treatment-Induced Apoptosis in Oncology with PET and SPECT**", *Current Pharmaceutical Design,* vol. 14, no. 28, October 2008, pp. 2974-2982(9).

4.   A. Rahmin, H. Zaidi, "**PET versus SPECT: strengths, limitations and challenges**", *Nuclear Medicine Communications 2008,* vol. 29, pp.193-207.

5.   G. Healey, R. Kondepudy, "**Radiometric CCD Camera Calibration and Noise Estimation**", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 16, no. 3, pp. 267-276, 1994.

6.   R. Chakre, P. Thapa, R. Siddharth, U. Baruah , "**An Experimental Study on Image De-Noising Filters**", *International Journal of Emerging Technology and Advanced Engineering,* vol. 3, no. 2, pp. 549-553, 2013.

7.   I. T. Young, L. J. Van Vliet, "**Recursive implementation of the Gaussian filter**", *Signal processing*, vol. 44, no. 2, pp. 139-151, 1995.

8.   C. V. Angelin, E. Debreuve, Mi. Barlaud, Fellow IEEE, "**Patch confidence k-nearest neighbors denoising**", *IEEE ICIP, Hong Kong: Hong Kong(2010)*.

9.   A. Buades, B. Coll, J. Morel, "**A non-local algorithm for image denoising**", *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE*.

10.   R. Chakre, P. Thapa, R. Siddharth, U. Baruah , "**An Experimental Study on Image De-Noising Filters**", *International Journal of Emerging Technology and Advanced Engineering,* vol. 3, no. 2, pp. 549-553, 2013.

11.   L. A. Shepp, Y. Vardi, "**Maximum likelihood reconstruction for emission tomography**", *IEEE Transactions on Medical Imaging,* vol. MI-1, no.2, October 1982.

12.   S. Vandenberghe, Y. D'Asseler, R. Van de Walle, T. Kauppinen, M. Koole, L. Bouwens, K. Van Laere, I. Lemahieu, R.A. Dierckx, "**Iterative reconstruction algorithms in nuclear medicine**", *Computerized Medical Imaging and Graphics* 25 (2001), pp. 105-111.

13.   C. V. Angelino, E. Debreuve, M. Barlaud, "**Image restoration using a Gaussian-variant of the mean-shift**", 978-1-4244-1764-3/08*, 2008 IEEE*.

Speaker Independent Emotion Recognition from Speech using Combination of Different Classification Models

Manav Bhaykar, K. Sreenivasa Rao, Manjunath K E

School of Information Technology

Indian Institute of Technology Kharagpur, India - 721302

manavbhaykar2008@gmail.com, ksrao@iitkgp.ac.in, ke.manjunath@gmail.com

Abstract—In this study, we have analysed speaker independent emotion recognition from speech using combination of different classification models. Classification models considered in this study are Auto-Associative Neural Networks (AANN) and combinations of Gaussian Mixture Model-Hidden Markov Model (GMM-HMM), AANN-GMM and HMM-AANN. IITKGP-SESC and IITKGP-SEHSC emotional speech corpora are used for carrying out these studies. The emotions considered in this study are anger, disgust, fear, happy, neutral, sarcastic and surprise. Mel Frequency Cepstral Coefficients (MFCCs) are used as features for building Emotion Recognition models. Improved speaker independent emotion recognition performance is observed using combination of different classification models.

*I*ndex Terms— Emotion recognition, GMM-HMM, AANN-GMM and HMM-AANN, IITKGP-SESC, IITKGP-SEHSC

I. INTRODUCTION

Human beings express their feelings through emotions. Presence of emotions makes speech more natural. Speech signal contain information like intended message, speaker identity and emotional state of speaker. An important issue in speech emotion recognition is to determine a set of important emotions to be classified by an automatic emotion recognizer. Many researchers agree with the palette theory [1]. It is very difficult to classify all types of emotions present in speech. However, in this work, we will consider seven emotions namely anger, disgust, fear, happy, neutral, sarcastic and surprise.

In literature, several works have been contributed for emotion recognition. Excitation source features [2] [3] and vocal tract system features [4] [3] [5] are mainly used for emotion recognition. Several classification models are explored for developing emotion recognition system such as Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Auto-Associative Neural Networks (AANN) [6] [7] and SVM.

In previous work [8], we have explored GMM and HMM for emotion recognition. Form the results, it is observed that speaker dependent emotion recognition perform better than that of speaker independent emotion recognition. Hence in this paper we focused on improvement of speaker independent emotion recognition by exploring AANN model and combining the results of GMM, HMM and AANN models for developing emotion recognition systems. Indian Institute of Technology Kharagpur Simulated Emotional Speech Corpus (IITKGP-SESC) [9] and Indian Institute of Technology Kharagpur Simulated Hindi Emotional Speech Corpus (IITKGP-SEHSC) [10] speech databases are used for performing emotion recognition study. The effectiveness of Mel Frequency Cepstral Coefficients (MFCCs) have been explored for speaker independent emotion recognition.

Rest of the paper is organized as follows. In Section II classification models used in this work are discussed. Section III describes the architecture of emotion recognition system. Section IV provides the details of emotional speech corpus used for building emotion recognition system. Section V describes detailed procedure for combining the recognition scores of two models. In Section VI experimental results are discussed. The summary and conclusion are presented in Section VII.

II. CLASSIFICATION MODELS

In this section, we discuss about the details of HMM, GMM and AANN models.

A. Hidden Markov Model

The Hidden Markov Model (HMM) is a doubly stochastic process, which consists of a first-order Markov chain whose states are hidden from the observer. Each state is associated with a random process, which generates the observation sequence. Hidden states of the model capture the temporal structure of the data. There are five components present in the HMM namely number of hidden states in HMM, number of observation symbols per state, state transition probability distribution, observation symbol probability distribution in each state and initial state probability distribution [11], [12].

In this paper, HTK (HMM toolkit) is used to implement the HMM [13]. HTK is primarily designed for building HMM-based speech processing tools, in particular recognizers. Thus, much of the infrastructure support in HTK is dedicated to this task. HTK tools are designed to run with a traditional command-line style interface. In each state of HMM, observation symbol probability distribution are modeled using Gaussian components. Emotion recognition performance critically depends on number of states and numbers of Gaussian components per state. Therefore, we have explored different variations of HMM models such as 4 state and 5 state HMM with varying number of Gaussian components per state.

B. Gaussian Mixture Model

The Gaussian Mixture Model is one of the statistically mature method for unsupervised clustering [14]. The complete Gaussian Mixture Model is parametrized by the mean vector, diagonal of covariance matrix and the mixture weight of each component. These parameters are collectively represented by the following notation .

$$\lambda = \{w_i, \vec{\mu_i}, \Sigma_i\} \ , \qquad i = 1, 2, ..., M \qquad (1)$$

where, M indicates number of Gaussian components, w_i is the weight, $\vec{\mu_i}$ is the mean vector and Σ_i is the covariance matrix of ith component. The mixture weights satisfy the constraint that, $\Sigma_{i=1}^{M} w_i = 1$. These parameters are initialized using $k-$means clustering on the training set of feature vectors, with $k = M$.

A well known Expectation Maximization (EM) algorithm is used for finding the maximum likelihood estimates of the parameters. EM is an iterative method that alternates between performing an expectation (E) step, which computes an expectation of the log likelihood with respect to the current estimate of distribution and a maximization (M) step which computes the parameters that maximize the expected likelihood found in the E step. These parameters are then used for E step of next iteration. Once the parameters are re-estimated by EM algorithm, the training phase is completed. Now, given a test vector $\vec{x}$ of dimension D, the score is generated by the equation (2).

$$p(\vec{x}|\lambda) = \sum_{i=1}^{M} w_i b_i(\vec{x}) \qquad (2)$$

where, $b_i(\vec{x})$ is the component density of ith component, given by the equation (3)

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \ exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu_i})' \Sigma_i^{-1} (\vec{x} - \vec{\mu_i})\right\} \quad (3)$$

The GMM can have several different forms depending on the choice of covariance matrices. The model can have one covariance matrix per Gaussian component or one covariance matrix for all Gaussian components in a cry model or a single covariance matrix shared by all cry models. The covariance matrix can be full or diagonal. Here, diagonal matrices are primarily used. The choice of model configuration (number of components, full or diagonal covariance matrices and mixture weights) is often determined by the amount of data available for estimating the GMM parameters and how the GMM is used in a particular application.

Model parameters w_i, μ_i and Σ_i are estimated using Maximum likelihood (ML) procedure. Model parameter's likelihood is been maximized through iterative expectation maximization (EM) algorithm. The performance of EM algorithm depends on the initial model estimation s. To achieve good initial estimate of model, vector quantification technique is used. In every iteration the posterior probabilities for the ith mixture is computed using [14] :

$$Pr(i|x_t) = \frac{w_i P_i(x_t)}{\sum\limits_{j=1}^{M} w_j P_j(x_t)} \qquad (4)$$

The model parameters are updated according to the following expressions [14] :

New mixture *weight* of $i^t h$ Gaussian component is

$$\overline{w_i} = \frac{\sum\limits_{i=1}^{T} Pr(i|x_t)}{T} \qquad (5)$$

New *mean* vector of $i^t h$ Gaussian component is

$$\overline{\mu_i} = \frac{\sum\limits_{i=1}^{T} Pr(i|x_t)x_t}{\sum\limits_{i=1}^{T} Pr(i|x_t)} \qquad (6)$$

New *covariance matrix* of $i^t h$ Gaussian component is

$$\overline{\sigma_i^2} = \frac{\sum\limits_{i=1}^{T} Pr(i|x_t)|x_t - \overline{\mu_i}|^2}{\sum\limits_{i=1}^{T} Pr(i|x_t)} \qquad (7)$$

While estimating model parameters, we had used diagonal covariance matrices instead of full covariance matrix.

C. Auto-associative neural network

AANN models are basically feed-forward neural network (FFNN) models, which try to map an input vector onto itself, and hence the name auto-association or identity mapping [15], [16]. It consists of an input layer, an output layer and one or more hidden layers. The number of units in the input and output layers is equal to the dimension of the input feature vectors. The number of nodes in one of the hidden layers is less than the number of units in either the input or output layer. This hidden layer is also known as dimension compression layer. The activation function of the units in the input and output layers is linear, whereas in case of hidden layers it is either linear or nonlinear. The performance of AANN models can be interpreted in different ways, depending on the application and the input data. If the data is a set of feature vectors in the feature space, then the performance of AANN models can be interpreted as linear or nonlinear principal component analysis (PCA) or capturing the distribution of input data [17], [18], [19].

Determining the network structure is an optimization problem. At present, there are no formal methods for determining the optimal structure of a neural networks. The key factors that influence the neural network structure are learning ability of a network and capacity to generalize the acquired knowledge. From the available literature, it is observed that, 5 layer symmetric auto-associative neural networks, with three hidden layers have been used for different speech tasks. The first and the third hidden layers have more number of nodes than the input or output layer. Middle layer (also known as dimension compression layer) contains less number of units [20], [21]. In this type of network, generally the first and third hidden layers are expected to capture the local information among the feature vectors and the middle hidden layer is meant for

capturing global information. Most of the existing studies [22], [20], [21], [23], [24], [25] have used the 5 layer AANNs with the structure N_1L-N_2N-N_3N-N_2N-N_1L, for their optimal performance. The general structure of the 5 layer AANN is shown in the Fig. 1. Here N_1, N_2 and N_3 indicate the number of
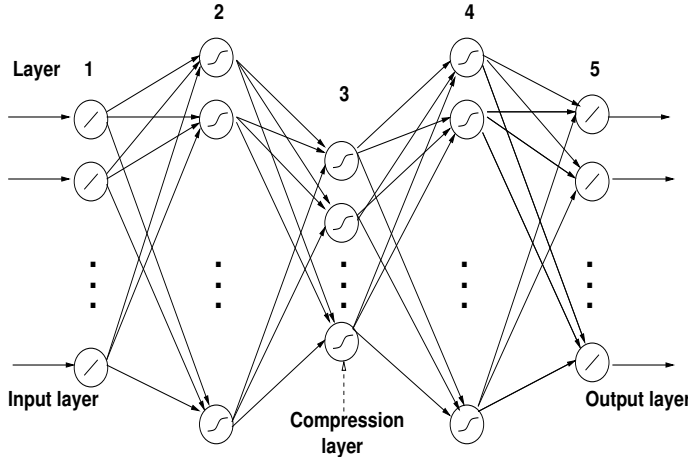
Fig. 1. General structure of 5 layer AANN.

units in the first, second and third layers respectively, of the symmetric 5-layer AANN. Usually N_2 and N_3 are derived experimentally, for achieving the best performance in the given task. From the existing studies, it is observed that N_2 is in the range of 1.3 to 2 times of N_1 and N_3 is in the range of 0.2 to 0.6 times of N_1. For designing the structure of the network, we have used the gross results of the existing studies and experimented with few structures for finalizing the optimal structure. The final network structure of 19L-38N-8N-38N-19L is chosen.

III. ARCHITECTURE OF EMOTION RECOGNITION SYSTEM

System architecture used for emotion recognition is shown in Fig. 2. The architecture has two phases namely training and testing phase. In training phase, feature extraction is the first step. MFCCs are extracted from speech utterances of known emotions. Emotion recognition models (GMMs or HMMs or AANNs) are trained using the MFCC feature vectors. In testing phase, feature vectors corresponding to the test utterances are given input to all trained models to find out the emotion present in that utterances. The outputs of all emotion models are given to decision system to decide the emotion present in the test utterance. The decision system hypothesizes the emotion of the test utterance based on the maximum among the outputs of emotion models.

IV. EMOTIONAL SPEECH CORPORA

For characterizing the emotions, either for synthesis or for recognition, a suitable emotional speech database is a necessary prerequisite. The design and collection of emotional speech corpora mainly depends on the research goals. For example a single speaker emotional speech corpus would be enough for the purpose of emotional speech synthesis, whereas recognizing emotions needs a database with multiple speakers and various styles of expressing the emotions. In this work, two emotional speech databases, IITKGP-SESC and
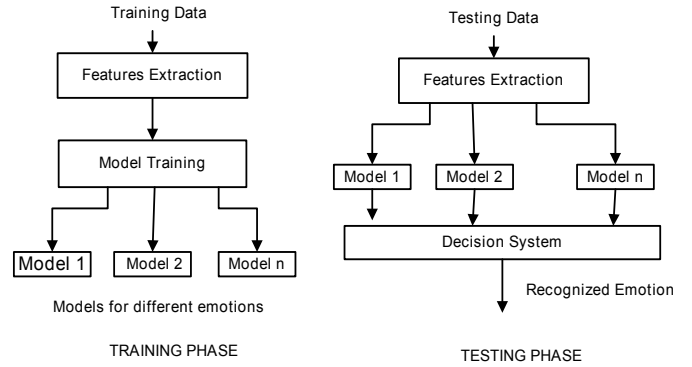
Fig. 2. Architecture of emotion recognition system.

IITKGP-SEHSC are used for emotion recognition. IITKGP-SESC is a Telugu speech database recorded using 10 (5 male and 5 female) professional artists from All India Radio (AIR) Vijayawada, India [9]. The eight emotions considered for recording this database are anger, disgust, fear, happy, neutral, sadness, sarcastic and surprise. Fifteen emotionally neutral, Telugu sentences are chosen as text prompts for the database. Each of the artists has to speak 15 sentences in 8 given emotions in one session. The number of sessions recorded for preparing the database is 10. The total number of utterances in the database is 12000 ($15 text prompts \times 8 emotions \times 10 speakers \times 10 sessions$). Each emotion has 1500 utterances. The total duration of the database is around 7 hours. The sessions were recorded on alternate days to capture the variability of the human speech production system. Recording was done in such a way that each artist had to speak all sentences at a stretch in a particular emotion. This provides coherence among sentences for a specific emotion category. IITKGP-SEHSC is a Hindi speech database recorded using 10 (5 males and 5 females) professional artists from All India Radio (AIR) Varanasi, India [26]. Both Hindi and Telugu databases are recorded in similar manner in all aspects.

V. PROCEDURE FOR COMBINING THE SCORES OF TWO MODELS

The following sequence of steps are used for combining the scores of two models.

A. Posterior probability values from both model:

When a speech utterance is given to a classification model, it returns seven posterior probability (likelihood) values (one from each emotion model) related to similarity with each emotion model.

B. Normalize the posterior probability values:

As we collected the posterior probability values from two different models, each model have different range of values and we have to combine those values, therefore we need to normalize the probability values given by different models in the same range of 0 to 1. For normalizing the value, first we have to find the minimum (low) and maximum ($high$) value

out of 7 values obtained from each model. i^{th} probability value ($val(i)$) can be normalized ($nor\_val(i)$) as follow

$$nor\_val(i) = \frac{val(i) - low}{high - low} \qquad (8)$$

C. Assign weights for each model:

After we calculate the normalized probability values then next task is to assign the weights to each model. Weight should be assigned in such a way that sum of weights of both model is equal to 1, For example $wt1$, $wt2$ are the weights assigned for *model_1* and *model_2* respectively, and condition is

$$wt1 + wt2 = 1$$

D. Calculate the combined results:

The combined probability values of a particular emotion are calculated using normalized values and assigned weights of each model. This is done as per the equation shown below.

$$Comb\_val(i) = wt1 \times md1\_val(i) + wt2 \times md2\_val(i) \quad (9)$$

where ($Comb\_val(i)$) is combined value of ith emotion, $md1\_val$ and $md2\_val$ are normalized values of model 1 and model 2 respectively and $wt1$ and $wt2$ are weights of model 1 and model 2 respectively.

Using above procedure, we have combined the results of two classification models (*model_1* and *model_2*). After combining the scores for each of the model, emotion model having highest value is hypothesized as recognized emotion.

VI. RESULTS AND OBSERVATIONS

In this paper, we focused on speaker independent emotion recognition. IITKGP-SESC and IITKGP-SEHSC emotional speech corpora are used to carry out emotion recognition performance. In each case, each of the speakers's data was used for testing and remaining 9 speaker's data were used for training the emotion models. Both emotional speech corpus contain 10 speaker's data. Therefore, for each emotional speech corpus 10 cases were considered for evaluating emotion recognition performance with respect to each speaker. In speaker independent emotion recognition, training and testing data were 90% and 10% respectively. The test data consists of 150 speech utterances per emotion.

A. Emotion Recognition performance using AANN

In our previous work [8], we explored GMM and HMM models for speaker independent emotion recognition. In this paper we explored Auto-Associative Neural Networks (AANN). System architecture discussed in [8], is used to develop emotion recognition system using AANN. An individual AANN model is created for each emotion. We used five layered AANN structure of networks, 19L-38N-8N-38N-19L where L- Linear units and N- Non-linear units. Speaker independent emotion recognition performance is tested on both Hindi and Telugu emotional speech databases. 19 dimensional MFCC features extracted from speech are used for characterizing emotions. Table I shows the emotion recognition performance using AANN on speaker 3 of Hindi language. After

completion of all the cases it is observed that AANN gives poor recognition performance. The values in Table I indicates the number of utterances classified against each emotion. The following observations can be made based on results given in Table I. The diagonal entries corresponds to number of utterances correctly classified emotions, whereas other entries denotes the number of misclassified utterances. Each column indicate indicate the number of utterances classified into the corresponding emotion. It can be observed that most of the test utterances are classified into 3 emotions: anger, disgust and neutral. Also it can be noted that the recognition performance of sarcastic and fear are very poor.

TABLE I.   EMOTION RECOGNITION PERFORMANCE USING AANN ON SPEAKER 3 OF HINDI EMOTIONAL SPEECH CORPUS
(Ang=Anger   Dis=Disgust   Hap=Happy   Neu=Neutral   Sar=Sarcastic   Sur=Surprise   AVP=Average Recognition Performance)

	Ang	Dis	Fear	Hap	Neu	Sar	Sur	AVP
Ang	50	9	2	7	76	0	6	
Dis	21	82	0	3	38	0	6	
Fear	36	21	8	1	41	0	43	
Hap	8	13	0	45	75	0	8	32.79%
Neu	20	0	2	4	114	0	10	
Sar	18	47	0	12	66	2	5	
Sur	34	31	1	10	31	0	43	

Average speaker independent recognition performance using AANN on Telugu emotional speech database is 22.40% and for Hindi emotional speech database is 31.36%. Even though average recognition performance is low using AANN but after combining the results of AANN with results of GMM and HMM, we observed improved emotion recognition performance.

B. Emotion Recognition performance using combination of GMM and HMM models

According to the obtained results of speaker independent emotion recognition using GMM and HMM in [8], it is observed that GMM perform better in case of happy and surprise emotions and HMM performs better in case of disgust, fear and neutral emotions. By combining the results of GMM and HMM, we might get the better recognition performance in case of all emotions.

Using the combination of models technique discussed in section V, emotion recognition performance is tested for each speaker of Telugu emotional speech corpus. After completing all 10 tests (one for each speaker), we changed weights of model in such way that 9 cases were generated. Table II shows the average emotion recognition performance using GMM-HMM combination model on Telugu emotional speech database.

From the results, it is observed that combined model of GMM-HMM improved average recognition performance, but in most of the cases combined model completely ignored the anger and surprise emotions. Average recognition performance of anger and surprise emotions is below 15%. For disgust, fear and neutral emotions individual emotion recognition performance is more than 60%.

We temporarily excluded the anger and surprise emotions, and recognition performance checked using remaining

TABLE II.   AVERAGE EMOTION RECOGNITION PERFORMANCE (IN %) USING GMM-HMM COMBINED MODEL

Case	Weight factor for GMM model	Weight factor for HMM model	Average Recognition Performance(%)		
			Telugu DB all Emotions	Telugu DB (excluding Anger & Surprise)	Hindi DB all Emotions
1	0.9	0.1	15.86	18.46	46.34
2	0.8	0.2	17.48	20.66	48.34
3	0.7	0.3	22.29	28.14	50.6
4	0.6	0.4	26.09	34.51	51.65
5	0.5	0.5	28.41	38.14	51.70
6	0.4	0.6	31.13	41.69	51.80
7	0.3	0.7	36.1	49.72	51.34
8	0.2	0.8	35.62	48.16	50.73
9	0.1	0.9	31.08	45.58	50.52

TABLE III.   AVERAGE EMOTION RECOGNITION PERFORMANCE USING COMBINED MODEL OF AANN-GMM ON HINDI EMOTIONAL SPEECH CORPUS

Case	Weight factor for GMM model	Weight factor for AANN model	Average Recognition Performance(%)
1	0.9	0.1	43.80
2	0.8	0.2	44.52
3	0.7	0.3	43.74
4	0.6	0.4	43.32
5	0.5	0.5	41.17
6	0.4	0.6	39.02
7	0.3	0.7	36.89
8	0.2	0.8	33.95
9	0.1	0.9	32.16

5 emotions. Much better results are observed. Table II show the emotion recognition performance excluding anger and surprise emotions. The highest average emotion recognition performance is obtained, when a weight of 0.3 for GMM and a weight of 0.7 for HMM is used. We have also used the combined model technique on Hindi emotional speech corpus. The last column of Table II shows the Average emotion recognition performance using combined model on Hindi emotional speech corpus.

C. Emotion Recognition performance using Combination of GMM and AANN model

We also explored combined model of GMM and AANN for emotion recognition. Combined model technique discussed in section V is used to combine results of GMM with AANN. Nine different types of combinations were explored by changing the weight of AANN from 0.1 to 0.9.

Speaker independent emotion recognition performance is tested using combined model of GMM-AANN for both languages using all 9 types of weights. After performing all the combinations, improved recognition performance is observed especially in case of Hindi language. Results are more scattered than Telugu language. In Telugu language recognition performance is improved. Most of the results are biased towards happy and surprise emotions because recognition results of GMM are more biased toward anger and surprise emotions and results of AANN models biased toward happy and surprise emotions. After using combined model of AANN-GMM on Telugu language overall recognition for some speakers are increased but for some speakers it is decreased. But in case of Hindi language for all speakers improved recognition rate is observed. Table III shows the average emotion recognition performance using combined model of AANN-GMM on Hindi language. Each entry in table is average of 10 results (10 cases of speaker independent recognition).

D. Emotion Recognition performance using Combination of AANN and HMM model

Combined model of AANN and HMM is also explored for emotion recognition. Results of AANN and HMM is combined in similar way as results of GMM-HMM and GMM-AANN

TABLE IV.   AVERAGE EMOTION RECOGNITION PERFORMANCE USING COMBINED MODEL OF AANN-HMM ON TELUGU EMOTIONAL SPEECH CORPUS

Case	Weight factor for AANN model	Weight factor for HMM model	Average Recognition Performance(%)
1	0.9	0.1	21.70
2	0.8	0.2	21.78
3	0.7	0.3	24.47
4	0.6	0.4	26.18
5	0.5	0.5	31.34
6	0.4	0.6	32.96
7	0.3	0.7	36.46
8	0.2	0.8	39.16
9	0.1	0.9	37.04

are combined. Speaker independent emotion recognition performance is tested using combined AANN-HMM model. Nine different types of combinations were explored by changing the weight of HMM from 0.1 to 0.9.

After performing all the results using combined model of AANN-HMM, it is observed that when we increase the weight of HMM model more than 0.7 then recognition results are biased toward disgust, fear and neutral emotions in case of Telugu language. Results are biased towards these 3 emotions because as we seen in [8], recognition results using HMM are more biased towards disgust, fear and neutral emotions. When a weight of 0.5 is used for both models (AANN and HMM) results are observed to be uniformly distributed for all emotions. In case of Hindi emotional speech corpus, using combined model of AANN-HMM slight improvement of performance is observed. Table IV shows the average emotion recognition results on telugu emotion speech corpus using combined model of AANN-HMM. Each entry in table is the average of 10 cases. Even though highest average recognition performance is 39.16% using a weight 0.2 for AANN and a weight of 0.8 for HMM, but in this case results are biased towards disgust, fear and neutral emotions. Using a weight of 0.4 for AANN and a weight of 0.6 for HMM results are uniformly distributed for all emotions.

VII. Summary and conclusion

In this paper, speaker independent emotion recognition performance is analysed using AANN and combination of GMM, HMM and AANN models. Emotional speech databases of two languages were used for conducting experiments. Average emotion recognition performance is improved using combination of models. AANN gives poor emotion recognition performance. When the results of AANN are combined with GMM and HMM average emotion recognition performance is observed to be improved.

References

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias and J. G. Taylor, "Emotion recognition in human computer interaction," *IEEE signal processing magazine*, vol. 18, pp. 32–80, Jan 2001.

[2] K. Sreenivasa Rao and Shashidhar G Koolagudi, "Characterization and recognition of emotions from speech using excitation source information," *International Journal of Speech Technology (Springer)*, vol. 16, pp. 181–201, 2012.

[3] Shashidhar G Koolagudi and K. Sreenivasa Rao, "Emotion recognition from speech using source, system and prosodic features," *International Journal of Speech Technology (Springer)*, 2012.

[4] Shashidhar G. Koolagudi and K. Sreenivasa Rao, "Emotion recognition from speech using sub-syllabic and pitch synchronous spectral features," *Speech Communication, Elsevier.*, vol. 15, pp. 495–511, 2012.

[5] Shashidhar G. Koolagudi and K.Sreenivasa Rao, "Exploring speech features for classifying emotions along valence dimension," *in IEEE International Advanced Computing Conference, Thapar University, Punjab,India, IEEE Xplore*, pp. 1292–1296, Mar 2009.

[6] K. Sreenivasa Rao, V.K.Saroj, Sudhamay Maity and Shashidhar G. Koolagudi, "Recognition of emotions from video using neural network models," *Expert Systems with Applications, Elsevier*, vol. 38, no. 10, pp. 13 181–13 185, 2011.

[7] Shashidhar G. Koolagudi and K. Sreenivasa Rao, "Neural network models for capturing prosodic knowledge for emotion recognition," *in 12th International Conference on Cognitive and Neural Systems, Boston, MA, USA*, pp. 14–17, may 2008.

[8] Manav Bhaykar, Jainath Yadav and K. Sreenivasa Rao, "Speaker dependent, speaker independent and cross language emotion recognition from speech using gmm and hmm," *Natinal Comference on Communication, IIT Delhi*, Feb 2013.

[9] Shashidhar G. Koolagudi, Sudhamay Maity, Anil Kumar Vuppala, Saswat Chakrabarti and K. Sreenivasa Rao, "IITKGP-SESC: Speech database for emotion analysis," *Springer-Verlag Berlin Heidelberg*, vol. 40, pp. 485–492, 2009.

[10] Shashidhar G. Koolagudi, R. Reddy, Jainath Yadav and K. Sreenivasa Rao, "IITKGP-SEHSC : Hindi Speech Corpus for Emotion Analysis," in *International Conference on Devices and Communications (ICDeCom)*, Feb. 2011, pp. 1 –5.

[11] L. R. Rabiner, "A tutorial on hidden markov models and selected application in speech recognition," *IEEE Proceedings*, vol. 77, pp. 257–286, Feb 1989.

[12] Y. Zhao and B. H. Juang, "Stranded gaussian mixture hidden markov models for robust speech recognition," *in proceedings of the ICASSP,Kyoto, Japan*, 2012.

[13] Steve Young, et. al., Cambridge University Engineering Department, *The HTK Book (for HTK Version 3.4)*, 2009. [Online]. Available: http://htk.eng.cam.ac.uk

[14] D. A. Reynolds, "Speaker Identification and Verification Using Gaussian Mixture Speaker Models," *Speech Communication*, vol. 17, pp. 91–108, 1995.

[15] B. Yegnanarayana, *Artificial Neural Networks*. New Delhi, India: Prentice-Hall, 1999.

[16] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New Delhi, India: Pearson Education Aisa, Inc., 1999.

[17] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*. Newyork: John Wiley and Sons, 1996.

[18] M. S. Ikbal, H. Misra, and B. Yegnanarayana, "Analysis of autoassociative mapping neural networks," USA, 1999, pp. 854–858.

[19] S. P. Kishore and B. Yegnanarayana, "Online text-independent speaker verification system using autoassociative neural network models," Washington, USA., Aug. 2001, pp. 1548–1553 (V2).

[20] S. Desai, A. W. Black, B. Yegnanarayana, and K. Prahallad, "Spectral mapping using artificial neural networks for voice conversion," vol. 18, no. 5, pp. 954–964, Apr. 2010.

[21] K. S. Rao and B. Yegnanarayana, "Intonation modeling for indian languages," vol. 23, no. 2, pp. 240–256, Apr. 2009.

[22] C. K. Mohan and B. Yegnanarayana, "Classification of sport videos using edge-based features and autoassociative neural network models," *Signal, Image and Video Processing*, vol. 4, no. 1, pp. 61–73, Nov. 2008.

[23] L. Mary and B. Yegnanarayana, "Autoassociative neural network models for language identification," in *International Conference on Intelligent Sensing and Information Processing*. IEEE, Aug. 24 2004, pp. 317–320.

[24] K. S. Rao, "Application of prosody models for developing speech systems in indian languages," vol. 14, pp. 19–33, 2011.

[25] K. Sreenivasa Rao, "Role of neural network models for developing speech systems," *SADHANA (Springer)*, vol. 36, pp. 783–836, Oct. 2011.

[26] Shashidhar G. Koolagudi and Ramu Reddy and Jainath Yadav and K. Sreenivasa Rao, "IITKGP-SEHSC : Hindi speech corpus for emotion analysis," in *International Conference on Devices and Communication*, 2011.

Brain Dynamics under Mobile Phone Radiation – A Wavelet Power Approach

C. K. Smitha ,
Department of Electronics & Instrumentation Engg.
College of Engineering, Vadakara,
Kerala, India- 673105, *smi_c_k@yahoo.com,*

and

N. K. Narayanan
Department of Information Technology,
Kannur University, Kerala, India- 670567,
nknarayanan@gmail.com

ABSTRACT

Frequent exposure of human body to electromagnetic fields is a growing concern in present lifestyle. It raises questions about the effect of electromagnetic field on brain using mobile phone. The paper analyze the effects of mobile phone radiation on brain using EEG. Wavelet Transform is an efficient tool for analyzing non linear signals like EEG. Power of the signal provides important information about the state of the signal, which leads to the state of brain at that condition. We used statistical analysis and ANN classifier to analyze the feature parameter, the normalised power of all bands of EEG. The obtained result shows that brain signals are undergone some changes while using mobile phone. This demonstrates transformation in the activities of brain due to radiation.

Keywords: EEG, Mobile phone radiation, Wavelet Power, Student's t-test, ANN classifier

1. INTRODUCTION

In recent years usage of mobile phone has increased drastically**.** There is an increased concern about adverse effects of mobile phone radiation on the nervous system also. A large number of investigations [1-5] were conducted to study the effects of mobile phone radiation. In almost all the studies conducted using EEG, the signal is approximated to a linear signal. Linear methods work properly only for stationary signals, but assumptions of stationarity required are ignored while using these linear algorithms.

Electromagnetic radiation can be classified into ionizing radiation and non-ionizing radiation, based on whether it is capable of ionizing atoms and breaking chemical bonds. Mobile phones generate a modulated radio frequency electromagnetic field (RF-EMF), which is a form of non-ionizing radiation. Mobile phone radiation is unable to cause ionizations in atoms or molecules. However, it is unknown whether mobile phone radiation could affect cellular and physiological activities by other mechanisms. A strong electric field, depending on its frequency, might warm up tissues or disturb the neuronal functions. Thermal effects are based on energy absorption from the field to the tissue, which causes the oscillation of molecules. The radio waves emitted by a GSM handset have a peak power of 2 watts, and CDMA handset has below 1 watt.

The rate at which radiation is absorbed by the human body is measured by the Specific Absorption Rate (SAR). The maximum power output from a mobile phone is regulated by the mobile phone standard and by the regulatory agencies in each country. The SAR limit permitted over a volume of 1 gram of tissue is fixed as 1.6 W/kg, in India.

James C. Lin reported that pulse-modulated microwaves from cellular phones may promote sleep and modify human brain activity [6]. Aruna *et al* [7] in a study using EEG analysis, concluded that GSM mobile phone has larger effect on brain compared to CDMA phones.. In another research by H.D Costa *et al* [8] concluded that full power mode exposure may influence human brain activity than standby mode. J L Bardasano [9] and colleagues concluded a study by stating that use of a protective device can reduce the effect of mobile phone radiation. A study on Influence of a 900MHz signal with gender on EEG by Eleni Nanou [10] and colleagues reported difference in spectral power due to gender. In another study by Hie Hinrikus *et al* [11], reported the increase in beta rhythm due to radiation. During the past few years, there has been a considerable interest in the study of fractal dimension of EEG

In the previous work by the authors, signal complexity is measured using Higuichi's method by considering EEG signals from all electrodes as a whole [12]. And in [13], three methods were used to find FD. The change in FD is analyzed using statistical methods in paper [14]. In the paper [15], hundred samples were selected from each data set, FD is calculated using Higuichi's, Katz and k-NN method and analyzed using an F - test. In [16], the relative Wavelet energy of all bands was used as feature parameter for the analysis using ANOVA test and t-test. The result shows that there are some changes in the FD while using a mobile phone.

In this paper method of data acquisition, preprocessing, methods of normalized Power using wavelet method are provided in section 2. In section 3, different results obtained and interpretation of the result is included. The conclusion and scope of further work are discussed in section 4

2. MATERIALS AND METHODS

EEG waveforms are generally classified according to their frequency, amplitude, and shape, as well as the sites on the scalp at which they are recorded. The most familiar classification uses

EEG waveform frequency (e.g.,Gamma , Beta > 13 Hz , Alpha 8-13 Hz, , Theta 3.5-7.5 Hz & Delta 3 Hz or less). Information about waveform frequency and shape is combined with the age of the patient, state of alertness or sleep, and location on the scalp to determine significance.

Data Acquisition

Thirty five healthy individuals 18 male and 17 female of different age groups having mean age and standard deviation of 39.8±11.8 were participated in this study. EEGs were recorded from EEG Lab under the Neurology Department of Malabar Institute of Medical Sciences Hospital, Calicut, India is using Galileo N.T machine. EEG of the volunteers is recorded by keeping mobile phones at two different positions of head for 5 minutes each, 1) near to ear/auricle (Au) and 2) at Cz position. This procedure is repeated using two different mobile phones with different SAR values. SAR of the phone 1 is 1.3W/Kg and for phone 2 is 0.987 W/Kg.

In EEG recording, electrodes and their proper function are crucial for acquiring high quality data. The conventional electrode setting named 10–20 system with commonly used scalp electrodes of Ag–AgCl disks, less than 3 mm in diameter, is used

Preprocessing

Unwanted signals or artifacts are removed by visual inspection and by filtering. A notch filter is used to remove 50 Hz line frequency due to a fault in grounding or imperfect balancing. Wavelet algorithm using threshold filtering is used to de-noise the signal. Fig -1 shows the plot of the EEG signal of a subject before and after wavelet thresholding
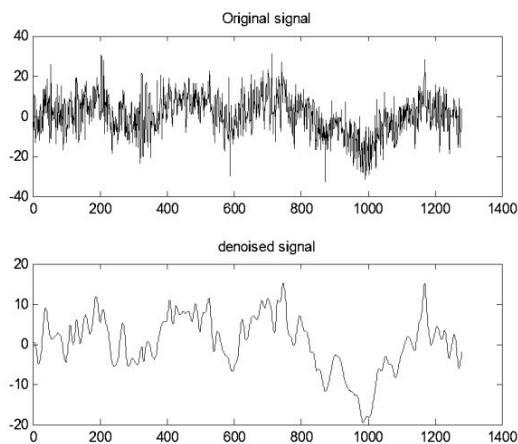
Fig -1 Plot of the EEG signal of a subject before and after wavelet thresholding

Feature Extraction

Wavelet Transform : Wavelet transform is a transform or remap of a signal that provides more information than the original [17-19]. The dyadic sampled wavelets are related to multi resolution analysis and the transform can be efficiently computed using filter banks.

The EEG signal is decomposed into progressively finer details by means of multi-resolution analysis using complementary low (H0) and high (H1) pass filter. After a first level decomposition, two sequences representing the high and low resolution components of the signal are obtained. The low-resolution components are further decomposed into low and high

resolution components after a second level decomposition and so on. The multi-resolution analysis, using five levels of decomposition, yields six separate EEG sub-bands.

After fifth level of decomposition, the components retained are cA5 (0-4 Hz), cD5 (4-8 Hz), cD4 (8-16 Hz), cD3 (16-32 Hz), cD2 (32-64Hz) and cD1 (64-128 Hz). The coefficient, cD4 is further decomposed to get alpha (8-12Hz) and beta-1(12-16Hz) ; cD2 is decomposed to get particular band of gamma (32-48Hz). These are equivalent to physiological classified bands such as δ, θ, α, low β , high β & γ. Table -1 shows the frequency bands corresponding to different decomposition levels.

TABLE- 1
FREQUENCY BANDS CORRESPONDING TO DIFFERENT
DECOMPOSITION LEVELS

Decomposed Signals	Frequency bands	Decomposition level
D1	64-128 (Noise)	1
D2	32-64Hz (Gama)	2
D3	16-32Hz (Beta2)	3
D4	8-16Hz (Alpha+low Beta)	4
D5	4-8 Hz (Theta)	5
A5	0-4Hz (Delta)	5

In the present study, Daubechi's wavelet is selected as mother wavelet, due to the analogy with EEG waveform. Wavelet coefficients are calculated using DB4 in 5 levels, Power is calculated for all bands. Figure 2(a) and (b) shows the detailed and approximate wavelet coefficients of an EEG signal at rest and EEG on using mobile phones.
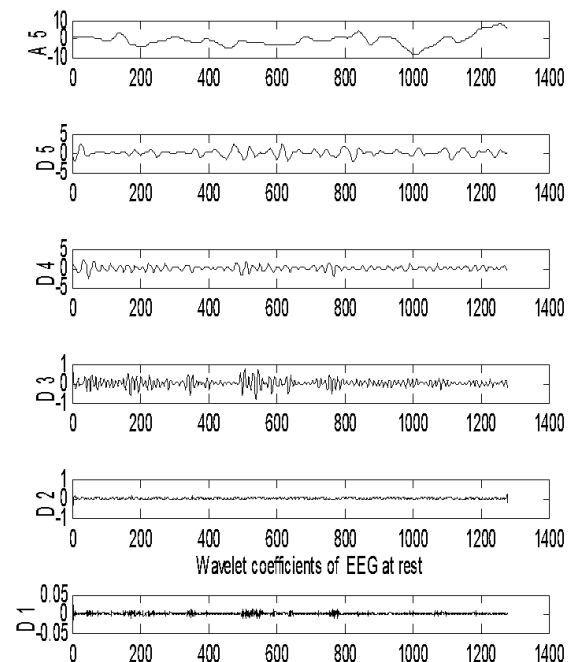
Fig 2(a) Plot of Detail coefficients of EEG signal of a subject at rest
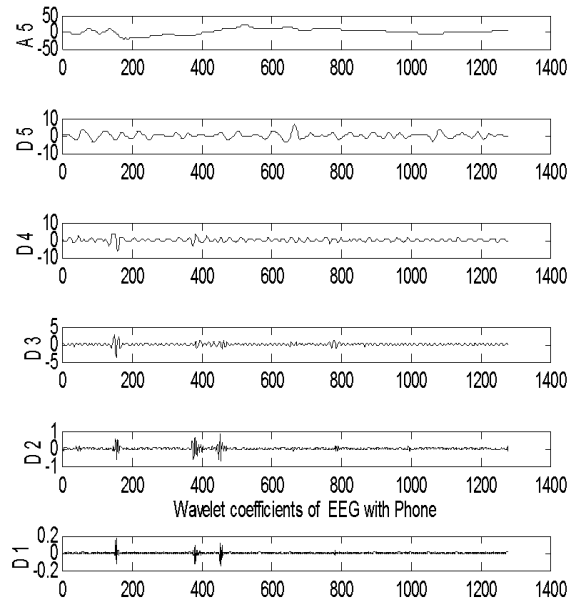
Fig 2(b) Plot of Detail coefficients of EEG signal of a subject with phone positioned near to the ear.

Wavelet coefficients of all detail sub bands from *n* levels are used in the formulation of wavelet Power. It is assumed that each band will demonstrate a distinctive range of Power values. The Power reflects the strength of the signal. Power is

$$P = \frac{1}{2N+1}\sum_{-N}^{+N}|x_j|^2 \quad j = 1,2 \dots N$$

Where x_j Is the discrete value of EEG signals. The normalised value of power is used as feature parameter. The normalized value of each band is calculated as the ratio of Power of individual band to the total power of all bands.

Statistical Method : Student's t- test [21, 22] is used to analyze the data sets of feature values of normalized power of different bands of EEG. In general, if the same measurement method was used, two samples from a population have same statistical parameters like mean, variance etc. Hypothesis testing is used for depicting inferences about a population, based on statistical evidence. In Student's t-test, t_{stat}, t_{crit}, and P-value is calculated. The Null hypothesis is rejected if t_{stat},> t_{crit}, and P-value < α (confidence level).

Artificial neural network classifier: An artificial neural network, is a mathematical model inspired by biological neural networks [23]. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases a neural network is an adaptive system changing its structure during a learning phase. Neural networks are used for modeling complex relationships between inputs and outputs or to find patterns in data. The learning ability of ANN allows acting as the human brain. If a specific task is given to solve, and a class of functions, learning means using a set of observations to find which solves

the task in some optimal sense. ANNs are widely used in the biomedical field for modeling, data analysis and diagnostic classification. The most frequently used training algorithm in classification problems is the back propagation (BP) algorithm. ANN classifier is used to verify the difference in feature parameter while using a mobile phone at two conditions

3. RESULT AND DISCUSSION

Analysis using normalised power

EEG data were preprocessed to reduce line frequency artifacts, de-noised and re-referenced. Data from electrode-2 is used for analysis. Average values of normalized power of Wavelet coefficients of all bands of EEG for each subject under all conditions were tabulated. Table -2 shows average values of normalized Power for both conditions separately. Fig -3(a) –(e) shows the graph of mean values of feature parameters

TABLE 2
MEAN OF NORMALISED POWER VALUES

Para meters	Rest	Ph-1 at Au	Ph-2 at Au	Ph-1 at Cz	Ph-2 at Cz
Delta	0.989	0.986	0.984	0.982	0.983
Theta	0.008	0.01	0.011	0.011	0.012
Alpha	0.003	0.004	0.004	0.005	0.005
Beta	0.00024	0.00032	0.00038	0.00228	0.00054
Gama	0.00003	0.00005	0.00005	0.00055	0.00006
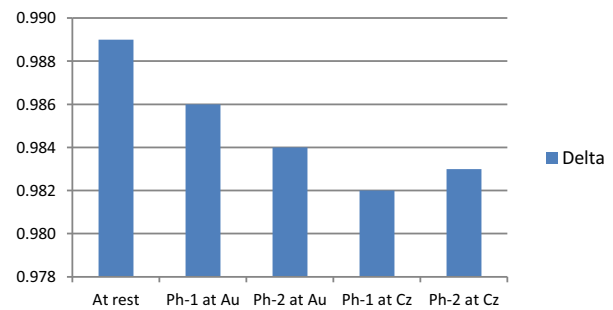
Fig 3-(a) Bar chart of mean value of normalized Power of Delta band
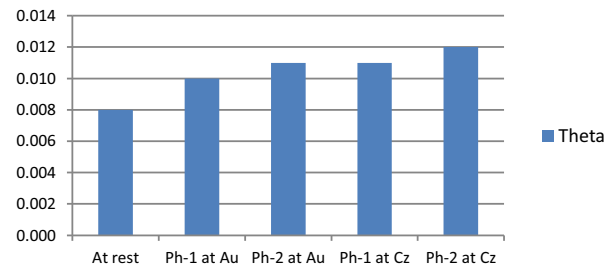
Fig 3-(b) Bar chart of the mean value of normalized Power at Theta band

215

Mean Power of Alpha Band
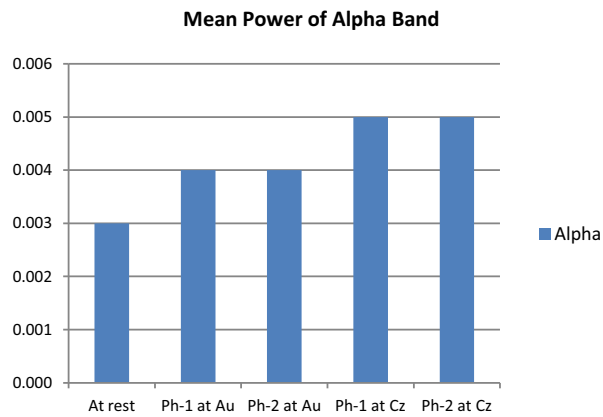
Fig 3-(c) Bar chart of the mean value of normalized Power at Alpha band

Mean Power of Beta Band
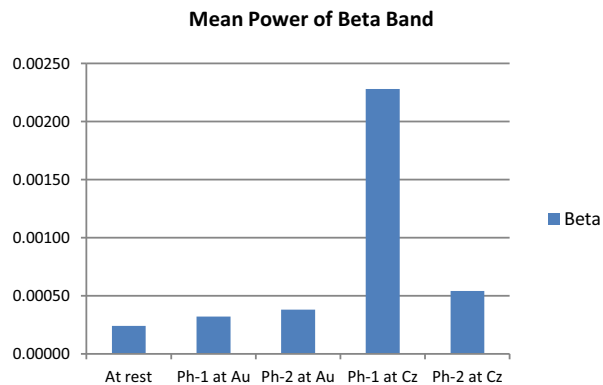
Fig 3-(d) Bar chart of the mean value of normalized Power of Beta band

Mean Power of Gamma Band
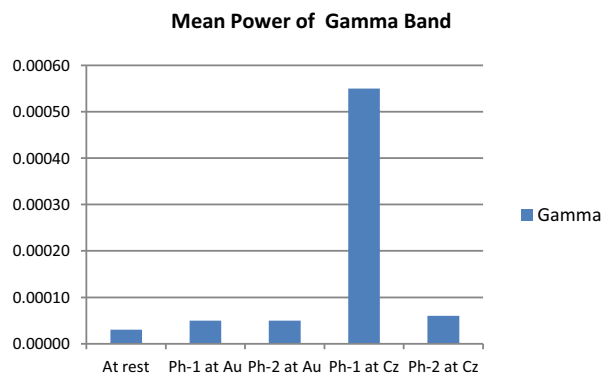
Fig 3-(e) Bar chart of the mean value of normalized Power at Gamma band

STATISTICAL ANALYSIS OF SAMPLES : Student's paired t-*test* is used to compare the above mean data obtained for each condition. The values of t_{stat}, P_{value} and t_{crit} are tabulated in table- 3.

TABLE 3
RESULT OF STUDENTS PAIRED T-TEST

Cond ition	Signifi cant values	Delta	Theta	Alpha	Beta	Gamm a
Ph-1 (Au)	t_{stat}	2.29	2.461	1.637	1.857	0.6
	P value	0.028	0.019	0.111	0.07	0.553
Ph-2 (Au)	t_{stat}	3.157	2.889	3.468	4.208	0.609
	P value	0.003	0.0067	0.0014	0.0002	0.5468
Ph-1 (CZ)	t_{stat}	2.287	2.822	3.186	1.144	1.509
	P value	0.029	0.008	0.003	0.261	0.1405
Ph-2 (CZ)	t_{stat}	3.66	3.37	3.835	3.402	1.456
	P value	0.0008	0.0018	0.0005	0.001	0.154

. Here the two samples for comparison are selected in such a way that the first sample is set of data from rest position and second as a set of data with the phone.Critical value of t (t_{crit}) is 2.03. Here from Table-3, the first value of column 3, 2.29> 2.03, and P-value, 0.028 < α so the Null hypothesis is rejected. It shows 100% rejection of the hypothesis in both conditions for delta and theta bands, shows values of data sets under test are different. In alpha band, hypothesis is accepted while keeping phone-1 in Auricle position, values of power are same as that at rest. Similarly for beta band, the hypothesis is accepted for both the cases of phone -1. For gamma band, the hypothesis is accepted for all the cases. This depicts the effect of mobile phone radiation except for gamma band and use of phone-1(GSM) in Beta band. The results are shown in fig 4(a) –4(e)

Analysis of Delta Band Power using t- test
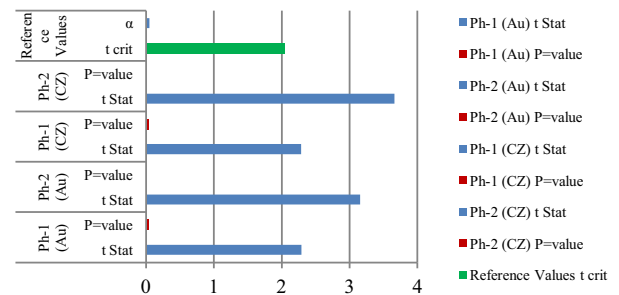
Fig 4-(a) Result of t-test using Power of Delta band as feature parameter

Analysis of Theta Band Power using t- test
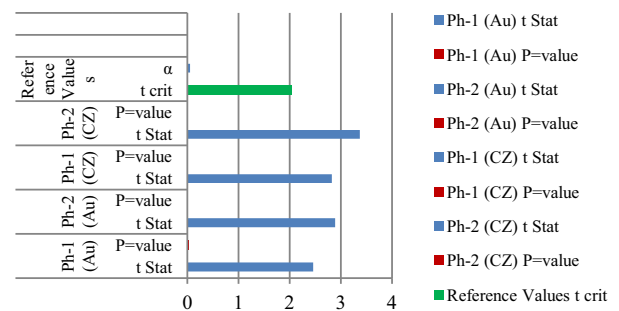
Fig 4-(b) Result of t-test using Power of Theta band as feature parameter

Analysis of Alpha Band Power using t- test
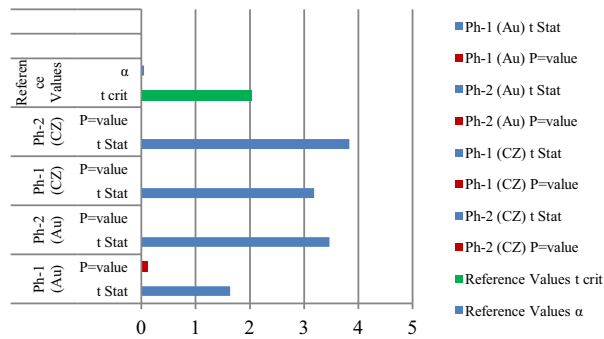
Fig 4-(c) Result of t-test using the power of the Alpha band as feature parameter

Analysis of Beta Band Power using t- test
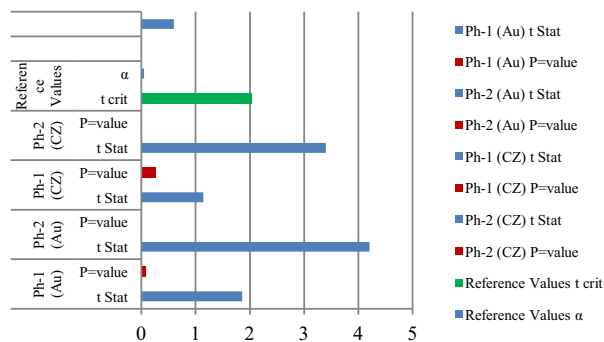
Fig 4-(d) Result of t-test using the power of the Beta band as feature parameter

Analysis of Gamma Band Power using t- test
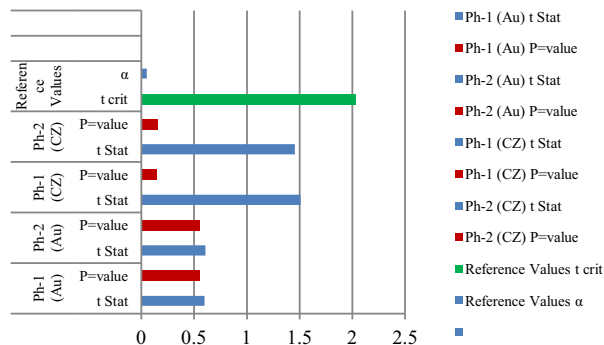
Fig 4-(e) Result of t-test using the Power of Gamma band as feature parameter

Alpha waves have been thought to indicate both a relaxed awareness without any attention or concentration. Here all the data are taken in a relaxed situation with eyes closed, So the change is more important.

Analysis using ANN

The method of classification with ANN is used to check whether the EEG data with phone and without a phone is classified. The impact of radiation is sufficient to change the feature parameters, the ANN classifier clsssifies the data into two. Here the data sets are divided into two, with phone and without a phone. The feature parameters are given to an ANN back propagation classifier for classification.

A five layer feed-forward ANN classifier with 2000 iterations were used for classification. The architecture of the feed-forward ANN used in this work is: Five neurons for input layer, 10 neurons in hidden layer and two neurons in the output layer. A set of data set of 37 subjects in each condition was used for training the neural network. The data set with phone at different positions are grouped together and compared with data set for the signals at rest. The percentage of correctly identified patterns is 89% , which is done using Nuclass 7.

It can be interpreted as the data for different conditions are atypical, since the classifier can classify the dataset with 89% accuracy. So it is proved that effect of radiation is evident while using a mobile phone. Accuracy is the ratio of the number of correct decisions to the total number of cases.

4. CONCLUSION

In this study, the data set includes data for 6 conditions, namely rest, with phone-1 and phone 2 at Cz position and at auricle position. Normalized power of all bands is calculated, tabulated and analyzed using student's paired t-test. The result shows that there are some changes in the normalized power while using mobile phone especially delta theta and alpha ranges. The result obtained using t- test shows, that mean are different, it depicts that the data sets are different, not from a population. This shows that the samples behave as they are coming from different population. All the data of same subject are recorded in succession under same conditions, so it is evident that there are some changes in brain dynamics while using mobile phones.

The data sets are grouped as with phone and without phone and feature parameters are inputted to ANN classifier. ANN classifier classified the data set into two with 89% accuracy, this also points to the change in EEG while using mobile phone.

The effect of radiation may vary; due to gender difference, age difference, mode of usage of phone has to be further investigated. The relationship with age, mode of use cannot be determined since the data set is small.

REFERENCES

[1]. Hardell L, Nasman A, Pahison A et al , " Use of cellular telephones and the risk for brain tumours: a case-control study. 1999,Int J Oncol 15:113–116

[2]. Hardell L, Hallquist A, Mild H et al , " Cellular and cordless telephones and the risk for brain tumors, 2002, Fur J Cancer Prev 11:377–386

[3].   Min Kyung Chu , Hoon Geun Song, Chulho Kim and Byung Chul Lee, " Clinical features of headache associated with mobile phone use: a cross-sectional study in university students", BMC Neurology 2011

[4].   Ashok Agarwal, Fnu Deepinder, Rakesh K.Sharma, Geetha Ranga, and Jianbo Li, "Effect of cell phone usage on semen analysis in men attending infertility clinic: an observational study", 2008 American Society for Reproductive Medicine, Published by Elsevier Inc. 124-128

[5].   IARC/A/WHO   (2011)   Classifies   radiofrequency electromagnetic fields as possibly carcinogenic to humans. Press release.

[6].   James C Lin, "Cellular Telephone Radiation and EEG of Human Brain',IEEE Antennas and Propagation Magazine. Vol 45, No -5 October 2003.

[7].   ArunaTyagi1,Manoj Duhan and Dinesh Bhatia,"Effect of Mobile phone radiation on brain activity GSM Vs CDMA" IJSTM Vol. 2, Issue 2, April 2011

[8].   H. D'Costa, G. Trueman, L. Tang, U. Abdel-rahman, W. Abdel-rahman, K. Ong and I. Cosic" Human brain wave activity during exposure to radiofrequency field emissions from mobile phones", Australasian Physical & Engineering Sciences in Medicine Volume 26 Number 4, 2003

[9].   J. L. Bardasano, J Alvarez- Ude, I Guiterrez , R. Goyal, "New Devices Against Non- thermal Effects from Mobile Telephones", 2005 Springer Science, The Environmentalist, 25, 257-263, 2005.

[10].   Eleni Nanou, Vassilis Tsiafakis, E. Kapareliotis, "Influence of the Interaction of a 900 MHz Signal with Gender On EEG Energy: Experimental Study on the Influence of 900 MHz Radiation on EEG" 2005 Springer Science, The Environmentalist, 25, 173–179, 2005.

[11].   Hie Hinkirikus, Maie Bachmann, Ruth Tomson and Jaanus Lass ,"Non-Thermal Effect of Microwave Radiation on Human   Brain"   ,2005   Springer   Science,   The Environmentalist,   25, 187–   194, 2005

[12].   C. K. Smitha and N. K. Narayanan" Effect of mobile phone radiation on brain using EEG analysis by Higuichi's fractal dimension method ", Proc. SPIE 8760, International Conference on Communication and Electronics System Design, 87601C (January 28, 2013); doi 10.1117/12.2012177

[13].   C. K. Smitha & N.K.Narayanan, "Study of Brain Dynamics under Mobile Phone Radiation Using Various Fractal Dimension   Methods",   Proce.CSIPR2013,International Conference on Signal processing, image processing and

pattern   recognition   (February   7   2013).doi 10.1109/ICSIPR.2013.6497942

[14].   C. K. Smitha & N.K.Narayanan,   " Effect of Mobile Phone Radiation on EEG Using Various Fractal Dimension Methods" in International Journal of Advancements in Research & Technology , Volume 2, Issue5, May-2013 IJOART (ISSN 2278-776

[15].   C. K. Smitha & N.K.Narayanan "Effect of Mobile Phone Radiation on EEG Third International Conference on Advances in Computing and Communications (ACC-2013) (29-31 August 2013) DOI 10.1109/ICACC.2013.28

[16].   C. K. Smitha & N.K.Narayanan Accepted paper on " Effect of Mobile Phone Radiation on Brain using Wavelet Energy The   15th   International   Conference   on   Biomedical Engineering (ICBME-2013 to be held in 4-7 December 2013, Singapore

[17].   Omerhodzic, S. Avdakovic, A. Nuhanovic, K. Dizdarevic "Energy Distribution of EEG Signals: EEG Signal Wavelet-Neural Network Classifier" International Journal of Biological and Life Sciences 6:4 2010

[18].   Tai Nguyen-Ky, Student Member, IEEE, Peng (Paul) Wen, Yan Li, Member, IEEE, and Robert Gray "Measuring and Reflecting Depth of Anesthesia Using Wavelet and Power Spectral Density", IEEE Transactions on Information Technology in Biomedicine, Vol. 15, NO. 4, July 2011

[19].   Abdullah Al Jumah "Denoising of an Image Using Discrete Stationary Wavelet Transform and Various Thresholding Techniques" Journal of Signal and Information Processing, 2013, 4, 33-41 doi:10.4236/jsip.2013.41004 Published Online February 2013 (http://www.scirp.org/ journal/jsip)

[20].   R. Harikumar, C. Ganesh Babu,T. Vijayakumar, " Performance Analysis of Elman Neural Networks as Post Classifiers for Wavelet Transforms Based Feature Extraction using Hard and Soft Thresholding Methods in the Classification of Epilepsy Risk Levels from EEG Signals"

[21]   Irwin Miller and John E Freund, "Probability and Statistics for Engineers" 4th Edition, Prentice Hall, 1985

[22]   A Stuart and M G Kendall, "Statistical Inference and Statistical RelationshipThe advanced theory of Statistics vol-2', Hafers Press ,1986

[23]   Robert J Schalkoff, "Pattern Recognition: Statistical Structural and Neural Approaches ",John Wiley &Sons Inc 1992.

Database or Ontology Driven Applications?

Ladislav BURITA, Vojtech ONDRYHAL
Department of Communication and Information Systems, University of Defense
Brno, Czech Republic

ABSTRACT

The article presents two approaches to the creation of applications aimed at working with information: an approach driven by a database system or the one driven by ontology. The first approach is introduced by the application for processing the information about university science and research (IS-SID), and the second approach is represented by the application supporting cooperation of military universities (MilUNI). The article describes the software used for the creation of both applications and presents the applications through their structure and functions. The approaches are compared, which results in stipulating their potential benefits and problems of the specific applications development.

Keywords: database, ontology, information processing, Postgres, IS-SID, MilUNI, ATOM

1. INTRODUCTION

Over the almost twenty years our research team has analyzed, designed, created and implemented a number of systems the primary aim if which was to process data and provide information to users. The development environments of these systems, their data models and principles of operation have been changing, and thus it is appropriate to summarize and assess them to identify the main trends in their development.
In our experience two important trends in data processing and information systems have taken shape: 1) Systems with a database basis, and 2) Ontology driven systems. Both approaches mostly pertain to the development of web applications.

The article aims to describe the database and ontology driven approaches and to give examples of their applications. It describes their infrastructure and tries to compare them. In the database oriented direction of the systems development it is an application for data processing and compilation of information on university science and research - Information System for Scientific Information Division (IS-SID). In the ontology driven direction of the systems development it is an application for military universities cooperation.

2. RELATIONAL DATABASE MANAGEMENT SYSTEM

Relational database management system (RDBMS) has been the platform for data storage and retrieval for more than 40 years. The platform is highly successful; many information systems have been created with the RDBMS data layer (Figure 1).
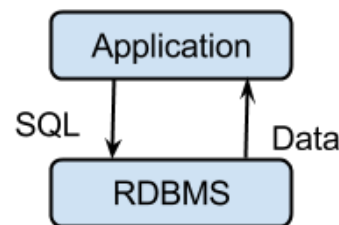
Figure 1 Applications with the database layer

IS-SID background database system is Postgres, but any RDMBS with SQL language support can replace it. Early applications with database background were developed using structured methodologies. In such methodologies the data are the core of the application; usually the simple create, read, update and delete (CRUD) operations are available. In today's applications, the focus is more on the user interaction and data structures are usually created based on interactive and iterative methodologies with UML support.

The selected RDBMS PostgreSQL is one of the leaders in non-commercial DBMS, it is comparable with most commercial systems. Its SQL implementation strongly conforms to the ANSI-SQL:2008 standard. It has full support for subqueries, read-committed and serializable transaction isolation levels. And while PostgreSQL has a fully relational system catalogue which itself supports multiple schemas per database, its catalogue is also accessible through the Information Schema as defined in the SQL standard.

The PostgreSQL boasts sophisticated features such as Multi-Version Concurrency Control (MVCC), point in time recovery, tablespaces, asynchronous replication, nested transactions (savepoints), online/hot backups, a sophisticated query planner/optimizer, and write ahead logging for fault tolerance. It supports international character sets, multibyte character encodings, Unicode, and it is locale-aware for sorting, case-sensitivity, and formatting. It is highly scalable both in the sheer quantity of data it can manage and in the number of concurrent users it can accommodate. There are active PostgreSQL systems in production environments that manage in excess of 4 terabytes of data [4].

3. INFORMATION SYSTEM FOR SCIENTIFIC INFORMATION DIVISION

The goal of the project is to support day-to-day activities associated with research and development at university. The following processes are implemented in the system:
- The process of building the register of the scientific results by university members.

- The register is used for the export into national register of scientific results.
- Management provides news and links to the relevant resources.
- Projects management aspects like funding and budget preparation is supported.
- The support for material, books and business trips using requests based system.
- The main processes are supported by many simpler and technology processes and files like the user management or logging system; see an example in Figure 2.
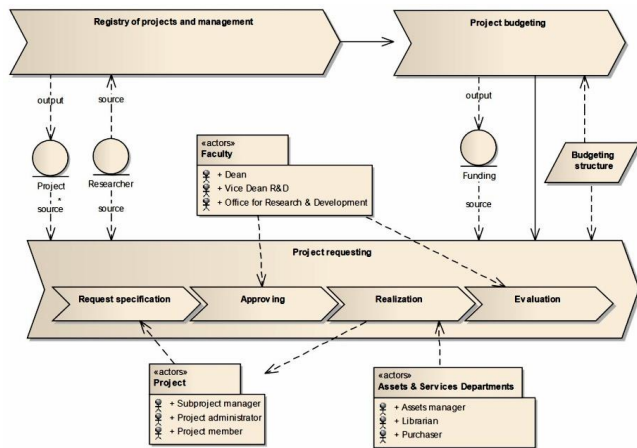
Figure 2 The requesting process

IS-SID is a web based application. The database layer serves only as data storage. The application logic is moved to the web application server. User experience is enhanced using Rich Internet Application techniques (interactive forms, dialogs and screens).

In Figure 3, a list of packages and application classes on the server is provided. The system components follow the MVC pattern (Model-View-Controller). The models (the right part of the figure) correspond with data tables on the server. The model classes provide an object oriented approach to data tables in the database. A detailed view of the user interface is presented in Figure 4.

4. SOFTWARE ATOM FOR THE DEVELOPMENT OF ONTOLOGY DRIVEN SYSTEMS

For the creation of ontology driven applications we use ATOM (Aion Topic Maps engine) software (SW) [1]. One of the goals and benefits of the ATOM is to support the implementation of projects of knowledge systems, especially effective development of powerful web applications. This has necessitated some extensions or specification of the Topic Maps (TM) standard, such as changes in the processing of occurrences of classes, work with associations and development of the user interface.
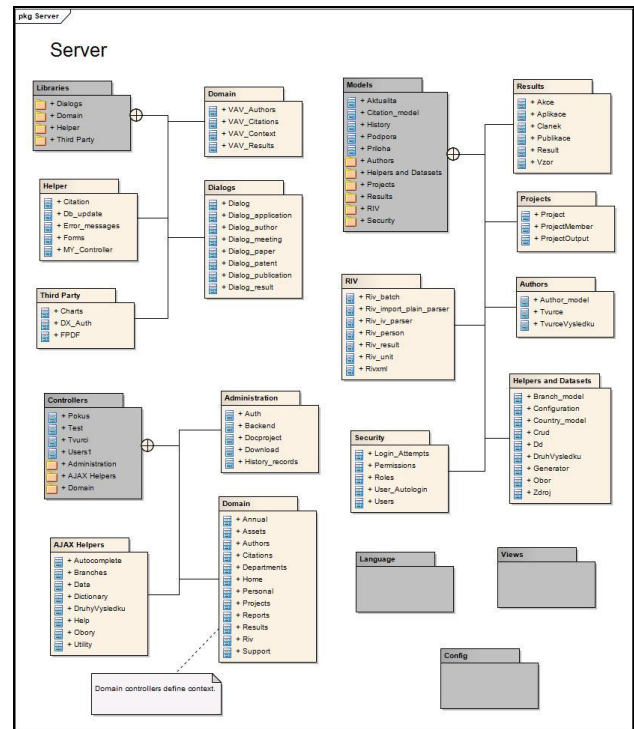
Figure 3 IS-SID Packages diagram

Figure 1 Rich user interface example

ISO Standard 13 250: Topic Maps

The TM model consists of three basic elements: topic, association between topics, and occurrences of the topics [2]. The TM is standardized in ISO/IEC 13250:2003. The topic contains a denominated subject of interest. Each topic represents just one subject. It is a place in TM where all known information on the given subject is available by means of relations and occurrences. The subject is a part of the real world, which is described in TM. Each subject is represented by

one topic and it can be anything: a person, thing, entity, process, etc. Associations represent relationships between topics; they are bidirectional, and express relationship between subjects. Occurrences are formed by information relevant to a given topic; they can refer to information or they might just contain it.

Implementation of TM in ATOM SW

A simple set of basic features for TM Internal Occurrences has been renamed to a Variant type property. The following data types have been further complemented:

- Code and Ident – for identification of the entity.
- Group Tree – a simple built-in taxonomy.
- Selection – forms a one-level code list, applied cardinality 1: N and M: N.
- Text – text content in XHTML; ATOM provides the built-in text editor.
- Picture, File – internal storage of images and files.
-
- In the definition of associations the following changes were finalized:
- Order (Sort) – each association can provide a structure of embedded occurrences.
- Hierarchy – a special type of association for Parent-Child relations.

Characteristics of ATOM SW

ATOM is a SW for sharing data with co-workers, customers, or friends via web browsers; it is a non-programming web database SW that does not require special knowledge and can be used for intranets with more powerful features than typical solution like SharePoint or box.com; it can be used as a construction kit for building web applications with powerful information retrieval, and for various encyclopedias, dictionaries, knowledge bases in applications where wiki approaches are not sufficient enough.

The complete ATOM solution includes three layers (environments):

1. ATOM Studio: Ontology Designer, user administration, batch data in/output module.
2. Data Editor: Includes data into database via ontology.
3. User Portal: Approach to knowledge system.

At first, create the ontology of your problem domain in the Ontology Designer. You can design the ontology by drawing, which is similar to drawing on a flip board, or writing it through filling in forms and using pre-prepared templates and adjusting them to your needs; see Figure 5.
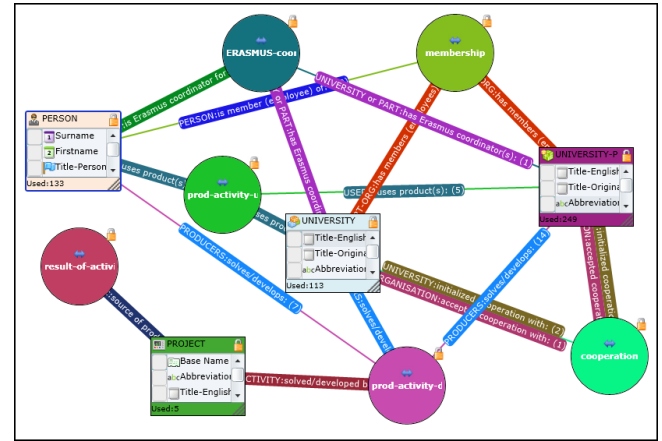
Figure 5 ATOM Studio - ontology design
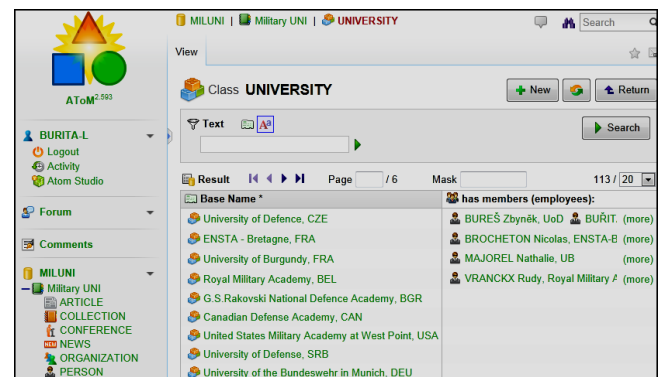
Figure 6 ATOM Data Editor – including and editing data

Immediately, you can enter data through the forms which are generated on-fly from the ontology in the Data Editor, see Figure 6.

And after that you are able to work with data, perform information retrieval or even visualize them and complete access by user portal.

5. THE MILUNI - ONTOLOGY DRIVEN APPLICATION

The content of an application is defined in the ontology. The ontology definition is an important procedure that follows an analysis of the domain of interest. The most important part of the ontology consists of classes that formulate the term structure of the domain. The MilUNI ontology contains the following set of classes:

- ARTICLE (DOCUMENT hierarchy that's divided according to the structure and type)
- COLLECTION (collection of articles)
- PERSON (member of a university, author of a paper …)
- ORGANIZATION (other organizations outside the university)
- UNIVERSITY (structure of hierarchy)
- UNIVERSITY- PART (faculties and departments)
- CONFERENCE (specialized/scientific conference)
- STUDY PROGRAMME (educational, research, development)
- PRODUCT (result of an activity)

- PROJECT (single activity)
- VENUE (place, e.g. conference centre)
- NEWS (current articles for the portal)
- GEO TREE (group tree with the continent, country and city via CIA World Factbook [3])

Content of the application

The MilUni is prepared in a complex shape; see the user interface in the editing mode in Figure 7. The main feature is a user friendly access to the information about the structure of the system, its main educational areas, the program of the faculties' education, list of departments, research and conference activities, etc. MilUNI includes data from public sources on about 100 universities and their 300 organizational parts (faculties, departments). They are situated in 40 countries and 130 cities.
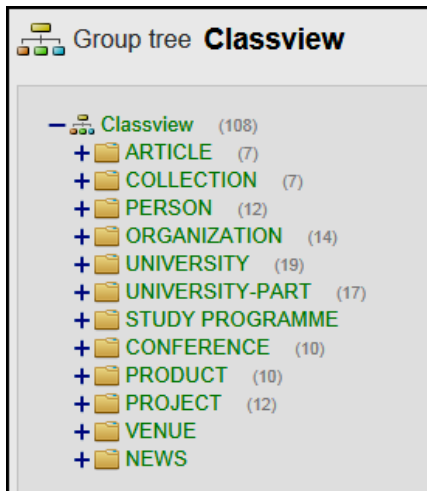
Figure 7 User interface - class view
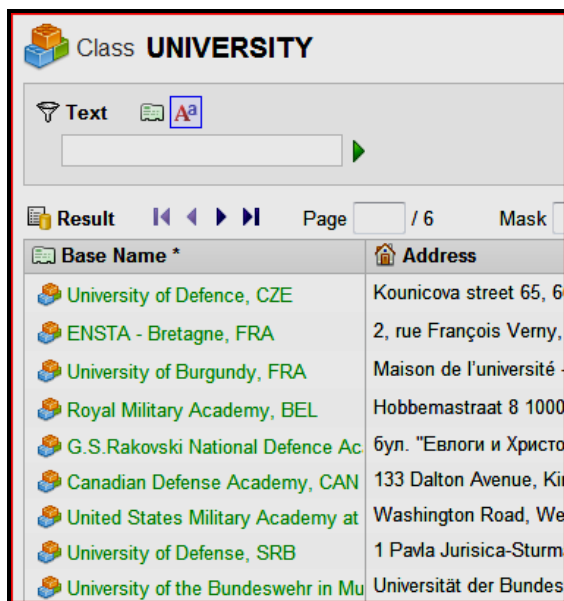
Figure 8 Set of universities
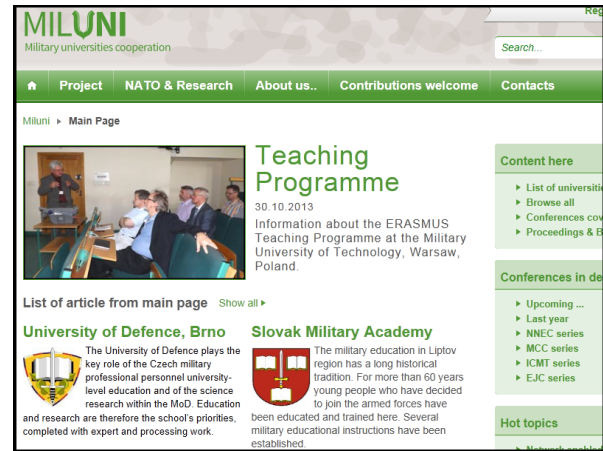
Figure 9 User Portal – main page

The most common way to obtain the required information is browsing the knowledge base of a selected class, such as UNIVERSITY; see an example in Figure 8. The filter can be used, let us assume, for the ENSTA university. Required data about the university can be obtained, then its field of study, the list of its academic staff and students, and perhaps even their publications at conferences.

The application has, according to the rights set for users, two forms. As a portal for a common user to get the information needed and as an ATOM SW interface for a qualified user for entering and editing data, or as a system administrator for editing ontology and managing users.

User Portal

The User Portal (UP) covers data, information, and knowledge in the system to shield users from the details of implementation. The UP is prepared as a typical portal template that is designed for any similar type of applications, see Figure 9, 10 and 11.

Figure 10 User Portal – search results

United States Naval Academy

United States Naval Academy is the undergraduate college of naval service. Students attend the Academy for four years, graduating with Bachelor of Science degrees and commissions as ensigns in the Navy or second lieutenants in the Marine Corps.

Founded in 1845 by Secretary of the Navy, George Bancroft, the Academy started as the Naval School on 10 acres of old Fort Severn in Annapolis. Since then, the history of the Naval Academy has reflected the history of the United States. As our country has changed culturally and technologically, so has the Naval Academy. In only a few decades, the Navy has moved from a fleet of sail and steam-powered ships to a high-tech fleet with nuclear-powered submarines and surface ships and supersonic aircraft. The Academy has changed, too, giving midshipmen the up-to-date academic and professional training they need to be effective naval officers in their assignments after graduation.

Every day, as the undergraduate college of the naval service, the United States Naval Academy strives to accomplish its mission to develop midshipmen "morally, mentally, and physically."

Moral Education

Moral development is a fundamental element of all aspects of the Naval Academy experience. As future officers in the Navy or Marine Corps, midshipmen will soon be responsible for the priceless lives of young Sailors and Marines and multi-million dollar equipment. From Plebe Summer through graduation, the Naval Academy's Officer Development program is a four-year integrated continuum that focuses on the attributes of integrity, honor, and mutual respect. This program helps midshipmen develop a clearer sense of their own moral beliefs and the ability to articulate them. Honor is emphasized through the Honor Concept of the Brigade of Midshipmen-a system originally formulated in 1951 which states, "Midshipmen are persons of integrity: they stand for that which is right." These

Figure 11 User Portal – page with detail information

6. COMPARISON OF BOTH APPROACHES

The comparison of these two approaches is executed in the creation of the application data structure, in the development environment used and in the form of the application user interface. The evaluation is not based on the comparison of both approaches (in terms of which one is better or worse), but as a commentary on each of the options so that the readers could compare the differences themselves.

Creation of the application data structure

The data structure in the database is flat, only simple associations between tables, 1:1 and 1:N, are supported. The database schema is a product of standard methodologies, e.g., structured (Entity Relationship Diagram, Relational Data, Model) or iterative (Use Case Realization). The custom application then defines more complex associations between data elements.

In the ontology driven application the data structure is stored in the ontology. It is defined by classes with characteristics (class attributes) and associations between classes. However, the data structure also consists of other elements, such as group- trees (predefined classifiers or taxonomies); e.g., geographical layout (continent – country – city).

The creation of ontology is a complex task and its successful accomplishment lies in the quality analysis and in good knowledge of the subject area (domain of interest). Some partial design or verification procedures for creating ontologies can be partially automated. Ontologies of functional applications are under constant development in accordance with the changes in the subject area as well as in the application objectives.

The implementation process of ontology in the ATOM Studio is quite an easy and clearly stated task if the previous analytical phase has been successfully managed. The ATOM SW huge advantage lies in the wide variety of data types for class characteristics and the possibility of virtually any organization of associations. Another significant benefit is the ability to

make changes in the ontology, which is immediately reflected in the data. The flexibility of ontology is an important factor in achieving dynamic applications.

Application development environment

The application development is independent on the database management system. The interface between application and data storage is strictly defined; it is based on SQL Language. Developer skills are required to build even a simple application. But any web technology can be used to develop an application. In our case the combination of the following technologies is used: JavaScript/jQuery UI, PHP/CodeIgniter.

The development environment of the ontologically driven application based on SW ATOM is a stable, but closed modular system in which the user is not focused on the development of the application itself, but only on its content and structure. The instances of classes connected by relationships and appropriately supplemented by group- trees, hierarchies, and other elements, form the knowledge base of the applications.

The content of the user portal is built in the ontology in such a way that the different parts of the portal are filled with the content: the main page, the page of the resulting search list in the knowledge base and the page of the detailed view; all according to the user requirements.

User interface of the application

The developer and user roles in the database approach are strictly separated. The developers write codes in special tools (code editors). The resulting application supports the user needs, but the users do not have any access to the application code or data structures outside the developer application.

The user interfaces of an ontology driven applications are different for different user roles in the system. The knowledge engineer works in the ATOM Studio (see Figure 4), manages the ontology and structure of the portal. The creator of the knowledge base (data editor) uses the environment ATOM Data Editor (see Figure 5) for inserting the class instances and integrating them into relationships. The user of the application accesses the knowledge base when searching for relevant information through the portal (see Figure 8) that is automatically generated from the ontology.

7. CONCLUSIONS

Before summarizing the outcomes gained from the comparison of the two approaches to creating applications, let us present Table 1 showing the initial facts in both approaches. It is apparent that the database access is a long established and successfully implemented, while the ontological approach is relatively new, still seeking its massive practical implementation.

Table 1 Database and ontology approaches

Criterion	Database	Ontology
Maturity	Mature. Many applications have been created based on RDBMS.	New

Community	Large communities for most RDMBSs	Small communities
Standardization	SQL Language standardization, Custom extensions for most of RDBMS	Standardization in progress. Separated standards for each technology (Topic map, OWL)

The criteria subjected to comparison show the differences in the definition of the data structure. While the database schema allows only a table structure with simple relationships, the ontology defined structure allows variety of classes and relations of any cardinality.

The development environment of both applications is quite different. While the database approach is based on the creation of a custom application, the ontology application SW is closed. The user interface of the database access is determined by the developed application. The ontology driven application can change the interface by the fulfillment of the ontology.

In conclusion, it seems that we are trying to compare incomparable approaches. Although they are entirely different, their resulting applications can fulfill both comparable and different objectives.

8. REFERENCES

[1] AION CS company and software Aion Topic Maps Engine (AToM), AION CS, Zlín, Czech Republic 2013, URL:<http://www.aion.cz/>.

[2] PEPPER Steve. The TAO of Topic Maps, 2012, URL: <http://www.ontopia.net/ topicmaps/ materials/tao.html>.

[3] The World Factbook, Central Intelligence Agency, USA 2013, URL: <https://www.cia.gov/library/publications/the-world-factbook>.

[4] PostgreSQL About. Available online. URL: <http://www.postgresql.org/about/>.

224

A Matrix-Based Prediction Method for Spatial Trajectories

Wen-Chen HU
Department of Computer Science, University of North Dakota
Grand Forks, ND 58202-9015, USA

Naima KAABOUCH
Department of Electrical Engineering, University of North Dakota
Grand Forks, ND 58202-7165, USA

and

Hung-Jen YANG
Department of Industrial Technology Education, National Kaohsiung Normal University
Kaohsiung City 80201, Taiwan

ABSTRACT

This paper proposes a kind of location-based research, spatial trajectory prediction, which is to predict the track of a subject's future movements. The proposed method works as follows. The mobile user sends his/her current trajectory along with several dummy trajectories to the server by using a 3D trajectory matrix, which encodes a set of trajectories. The server restores the trajectories from the 3D matrix and matches the restored trajectories to the saved trajectories. The predicted trajectory is found as the trunk of the tree, which is built by superimposing the matching results. The server then sends the predicted trajectories back to the user, who will apply the target predicted trajectory to a real-world problem such as the example mentioned in this paragraph. Preliminary experimental results show the proposed method successfully predicts spatial trajectories based on current and previous trajectories. User privacy is also rigorously protected by using a simple method of dummy trajectories.

Keywords: Location-Based Services, Spatial Trajectories, Spatial Trajectory Prediction, Matrices and Privacy.

1. INTRODUCTION

This research proposes a novel spatial trajectory prediction based on the previous trajectories because human travel patterns normally have an inertia feature, e.g., people are attracted by interesting locations or landmarks such as parks and malls. The proposed method is simple and effective by using an innovative matrix representation. Trajectory processing then becomes matrix processing, which is well documented and includes plenty of tools and software. The trunk of a tree (a superimposed matrix) is then treated as a predicted trajectory. At the same time, user privacy is rigorously preserved by using a simple method of dummy (false) trajectories. Spatial trajectory prediction can provide useful information to mobile users. For example, the proposed trajectory prediction, based on the previous users' travel routes, is able to show nearby or distant gas stations or interesting places such as shopping malls and landmarks.

The proposed research predicts the trajectories based on the current and previous trajectories. Its accuracy is high because of the consistence of user traveling patterns. In addition, a trajectory is a list of locations. The proposed research is made easy by using a matrix representation, which facilitates the trajectory storage, indexing, transmission, and matching. Trajectory processing then becomes matrix processing, which is simpler and faster than list-of-location processing. Additionally, instead of sending trajectory matrices between the clients and server, this research uses a 3D trajectory matrix, which stores a set of encoded trajectories, for transmission, so processing time is shortened. User privacy is usually a big headache for location-based services. It becomes simple and effective by using our method. This research sends the user's current trajectory along with several dummy trajectories to the server, so the server is not able to tell the correct user trajectory among the trajectories it receives. The server then tries to extend the trajectories based on the current and previous trajectories. The predicted trajectories, including the target predicted trajectory, are sent back to the user for the intended application. Preliminary experiment results show the proposed methods are effective and user-privacy is rigorously preserved.

The rest of this paper is organized as follows. Section 2 gives the background information of this research including two themes: (i) spatial trajectory prediction, and (ii) privacy protection of spatial trajectories. Section 3 introduces the proposed system. The data structure, trajectory matrices, used in this project is explained in Section 4. Section 5 gives algorithms of the proposed methods. The last section summarizes this research.

2. BACKGROUND AND LITERATURE REVIEW

A location-based service is a service based on the geographical position of a mobile handheld device [1]. Two of the LBS examples are finding a nearby ethnic restaurants and locating a nearby store with the best price of a product. Popular LBS include mapping and navigation, search and information, social networking, entertainment, and tracking [2]. A nice introduction of LBS technologies and standards is given in the articles [3][4]. Two themes related to this research are given as follows:

1) *Spatial trajectory prediction*: Trajectory prediction has been applied to a variety of areas such as location-based services, traffic control, and networks. There are roughly two approaches to trajectory prediction: probability-based approach and learning-based approach. This subsection

discusses major methods of each approach. Some related trajectory research, but not trajectory prediction, can be found from the articles [5][6].

2) *Privacy preservation of spatial trajectories*: Various methods of user privacy preservation of spatial trajectory computation have been proposed [7], but most of them do not meet our need, consistent user identities. There are two methods, spatial cloaking and dummy locations/trajectories, has the feature of consistent user identities and are introduced in this subsection. Other related privacy preservation methods can be found in the article [8].

3. STRUCTURE OF THE PROPOSED SYSTEM

One of the location-based services, spatial trajectory prediction, receives a great attention these days because of its usefulness and popularity. For example, spatial trajectory prediction can be used to send location-based advertisements/announcements or find better network routes. However, it also suffers the privacy problem of most location-based services. This research proposes a privacy-preserving spatial trajectory prediction, so users can use this function without concerns. This section explains the proposed system by giving (i) the structure of the proposed system, (ii) the trajectory matrices for a trajectory representation, and (iii) the method of privacy-preserving trajectory prediction.

Figure 1. The system structure of the proposed system.

This research predicts spatial trajectories based on current and previous trajectories. A 2D trajectory matrix is used to represent a trajectory, a sequence of locations. A mobile user sends his/her current trajectory along with several dummy trajectories to the LBS server by using a 3D trajectory matrix, which is built by superimposing 2D trajectory matrices. Trajectories are then predicted by the server, and are sent back to the user for an LBS application. Figure 1 shows the system structure of the proposed system consisting of two parts: server-side and client-side subsystems. It also includes two phases: training and testing, where the training phase is used to fill up the LBS Database with trajectory data and the testing phase is used by the application, predicting trajectories. The server performs most tasks of this method including collecting and saving sets of trajectories and the later predicting trajectories, which will be detailed in the next section. It includes the following components, of which the first three components are used during the training phase:

- *Trajectory data collection*: The first step is to collect trajectory data. This step can be done before the app is put to use or it continues collecting and updating the trajectories while the app is used.
- *Trajectory data preparation*: Not all collected trajectories are usable for this research and the trajectories may need to be processed before being put to use [9]. In addition, the GPS data is usually not stable or consistent. This step may include noise removal, deletion of rarely-used trajectories, or trajectory completion.
- *Trajectory pattern analysis and discovery*: Before the trajectories are entered to the database, the system may try to find useful trajectory patterns such as the most popular trajectories or likely destinations of the trajectories.
- *LBS Database:* The trajectories, lists of locations, are then saved in the LBS Database including two tables: `Trajectories` and `Locations`, whose schema are given in Figure 2.

Figure 2. The schema of the two tables `Trajectories` and `Locations` in the LBS Database.

- *LBS Manager:* It receives trajectories from users via 3D trajectory matrices, which will be discussed in the next section. The LBS Manager sends the result trajectories back to the users after prediction discussed in the Section 5 Spatial Trajectory Prediction.

On the other hand, the client-side subsystem is made simple on purpose. It includes two major tasks: (i) sending and receiving trajectories to and from the server, and (ii) applying the predicted trajectory to a location-based services. By using this approach, user privacy is preserved at the same time because the target trajectories are sent along with several dummy trajectories, so it is not possible to associate the users to specific trajectories.

4. TRAJECTORY MATRICES

Spatial trajectories, lists of locations, are the major data to be processed in this research. Trajectory/list management is normally not easy and complicated. Some of trajectory indexing and retrieval methods can be found from the article [10]. This research adapts the data structure, matrices, for the trajectory representation. The representation, called the trajectory matrices, facilitates the trajectory storage, indexing, transmission, and processing. This section will introduce the trajectory matrices.

Map Tile System

A map is usually rendered from a number of map tiles for convenience and speed. A map tile system includes many tiles of different zoom levels. By using map tiles, a map can be panned and zoomed easily and quickly. For panning, some of the map tiles instead of the whole map are replaced. For zooming, different zoom-level tiles can be retrieved more efficiently because lengthy tile searches can be avoided if map tiles are used. A map is usually rendered from a number of map tiles for convenience and speed. A map tile system includes many tiles of different zoom levels. By using map tiles, a map can be panned and zoomed easily and quickly. For panning, some of the map tiles instead of the whole map are replaced. For zooming, different zoom-level tiles can be retrieved more efficiently because lengthy tile searches can be avoided if map tiles are used. Several methods are used to find and display the correct maps. For example, Figure 3 shows a map using pixel coordinates [11]. The pixel at the upper-left corner of the map has pixel coordinates (0,0) and the pixel at the lower-right corner of the map has pixel coordinates (2047,2047). One of the advantages of pixel coordinates is that the specific locations can be pinpointed quickly. At the same time, using an image to represent a world map of a specific zoom level is not realistic because the image size would be huge. Therefore, a map usually consists of a number of map tiles. For example, Figure 3 shows a world map consisting of 64 map tiles from $(0,0)_t$ to $(7,7)_t$ and each tile contains $\left(\frac{2048}{8}\right) \times \left(\frac{2048}{8}\right) = 256 \times 256$ pixels. Given a pair of pixel $(x,y)_p$ coordinates, the corresponding tile coordinates $(a,b)_t$ containing that pixel is given by Equation 1:

$$a = \left\lfloor \frac{x}{256} \right\rfloor \quad and \quad b = \left\lfloor \frac{y}{256} \right\rfloor \tag{1}$$

On the other hand, most locations use latitude and longitude. Examples of how to convert latitude and longitude to tile numbers can be found from the CloudMade web site [12].
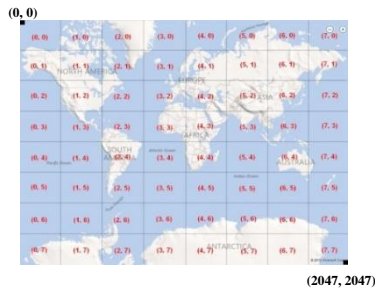
Figure 3. A world map consists of 64 map tiles.

Traditional map panning and zooming is complicated and slow because they need an algorithm to find the correct map locations. Map panning and zooming are made easy by map tiles. Figure 4 shows an example of map tiles of three zoom levels. For panning and zooming, the following methods are used:

- *Zooming*: Zooming in the tile x will result in four tiles $x0$, $x1$, $x2$, and $x3$ and zooming out will do the reverse. For example, zooming in the tile 32 in Figure 4 will bring in the tiles 320, 321, 322, and 323. On the other hand, zooming out the tiles 320, 321, 322, and 323 will result in the tile 32.
- *Panning*: There is no need to replace the whole maps; instead, only certain tiles are replaced. For example, panning the map including the tiles 320, 321, 322, 323 right may drop the tiles 320 and 322 and bring the tiles 330 and 332 in.
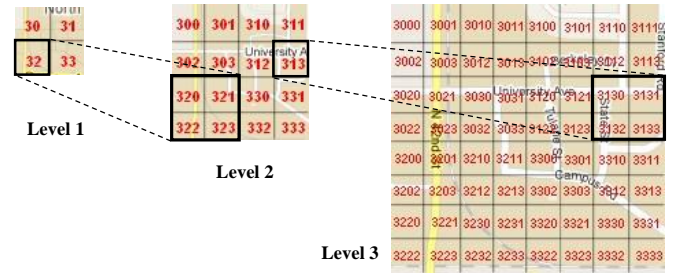
Figure 4. Map tiles of three zoom levels.

Trajectory Matrices

Today's map services could offer great details of the Earth. For example, the Google has a tile of 256×256 pixels. For the highest zoom level 24, each pixel size at equator is 9.3mm, so the tile covers an area of $256 \times 9.33mm \times 256 \times 9.33mm = 2.39 \times 2.39m^2$, which is detailed enough for our purpose.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} \tag{2}$$

A spatial trajectory is a list of locations and lists are normally not easy to manage. This research uses a matrix of map tiles to represent a trajectory. Matrices are a well-known mathematical subject and plenty of algorithms and methods of matrix processing are available. The notation of an $m \times n$ matrix A is given as above. The $(i,j)^{th}$ entry of the matrix A is written as $a_{i,j}$. Figure 5 shows a typical spatial trajectory and its corresponding map tiles and 10×10 trajectory matrix.

(a)

(b)

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(c)
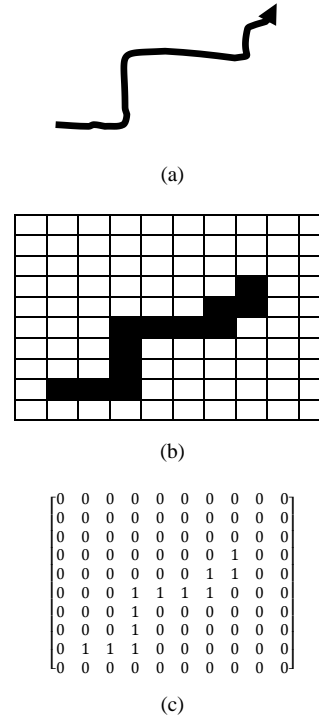
Figure 5. An example of trajectory matrix representation: (a) a spatial trajectory, (b) the corresponding map tiles, and (c) the corresponding 10×10 trajectory matrix.

If the trajectory passes through a map tile, the tile is colored black as shown in Figure 5.b. Figure 5.c is a sample trajectory matrix, whose entry values are either 1 (on) if the corresponding map tile is black or 0 (off) if the corresponding tile is blank. However, the values do not have to be 1 or 0. They could be any values/weights. Equation 3 shows a generic matrix notation for the trajectory p:

$$T_p = \begin{bmatrix} t_{1,1} & t_{1,2} & \dots & t_{1,n} \\ t_{2,1} & t_{2,2} & \dots & t_{2,n} \\ \dots & \dots & \dots & \dots \\ t_{m,1} & t_{m,2} & \dots & t_{m,n} \end{bmatrix}, \qquad where\ t_{i,j} = w_{i,j,p} \qquad (3)$$

where $w_{i,j,p}$ is the weight of the trajectory p on the map tile coordinated at $[f_x(i,l), f_y(j,l)]$. The functions $f_x(i,l)$ and $f_y(i,l)$ gives the x and y coordinates of the map tile at the entry $t_{i,j}$, respectively. The argument l is the physical location of the entry $t_{1,1}$ and is used to find the corresponding map tile at $[f_x(1,l), f_y(1,l)]$. For example, the map tile coordinates are $[f_x(1,l)+i-1, f_y(1,l)+j-1]$ for the entry $t_{i,j}$. Most trajectories can be represented by sparse matrices and many matrix optimization methods and representations can be found in the literature [13].
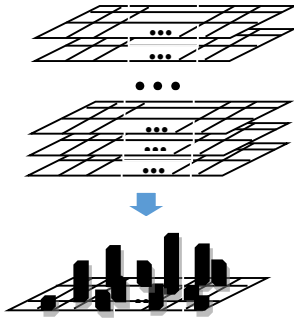
Figure 6. An example of a 3D trajectory matrix including a set of spatial trajectories and the corresponding 3D trajectory matrix.

3D Trajectory Matrices
A set of spatial trajectories are collected and saved in the LBS Database. A trajectory is a list of locations and managing lists is always a difficult task. This research requires transmissions of trajectories between the users and server. A 3D matrix representation is therefore proposed to facilitate the spatial trajectory storage, transmission, and processing. Figure 6 shows an example of 3D trajectory matrix storing a set of trajectories. A generic 3D trajectory matrix notation is given as follows:

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,n} \\ d_{2,1} & d_{2,2} & \dots & d_{2,n} \\ \dots & \dots & \dots & \dots \\ d_{m,1} & d_{m,2} & \dots & d_{m,n} \end{bmatrix},$$

$$\qquad (4)$$

$$where\ \ d_{i,j} = [d_{i,j,1}, d_{i,j,2}, \dots, d_{i,j,h}]\ \ and\ \ d_{i,j,p} = w_{i,j,p}$$

where each entry $d_{i,j}$ is an array of weights of the trajectories, h is the number of trajectories, and $w_{i,j,p}$ is the weight of the trajectory p on the map tile coordinated at $[f_x(i,l), f_y(j,l)]$. Again, each entry $d_{i,j}$ is likely a sparse one-dimensional matrix. Other data structures, such as linked lists, can be used to store the weights and thus save the space. Each 3D matrix encodes many

trajectories and several 3D matrices are sent, so the server is not able to associate trajectories to the user and privacy is preserved.

Trajectory Matching
Trajectory matching is difficult because it usually requires location-by-location matching. The trajectory matching is made easy by using our trajectory matrices. Various definitions of similarity are defined. This research uses the following formula to measure the similarity between two matrices. For example, $R = C\Delta S$ is the result $m \times n$ matrix R of matching the current $m \times n$ trajectory matrix C to the stored $m \times n$ trajectory matrix S:

$$r_{i,j} = c_{i,j}\Delta s_{i,j}$$

$$= \begin{cases} \dfrac{|C \cap S|}{|C|}, & if\ (c_{i,j} = 0\ and\ s_{i,j} \neq 0)\ or\ (i = i_e\ and\ j = j_e) \\ 0, & otherwise \end{cases} \qquad (5)$$

where $i \leq m$, $j \leq n$, $|C|$ is the number of entries of the matrix C being on, $|C \cap S|$ is the number of both corresponding entries of C and S being on, and (i_e, j_e) is the entry for the map tile of the current location (or the final location of the current trajectory). Figure 7 gives an example of the trajectory matching, where $|C| = 9$ and $|C \cap S| = 6$.

5. SPATIAL TRAJECTORY PREDICTION

This research proposes a method of privacy-preserving spatial trajectory prediction based on current and previous trajectories because a generalization of the past user travel trajectories usually indicates a trajectory trend. A discussion of the proposed algorithm is given in this section.

> *Privacy-Preserving Spatial Trajectory Prediction*
> 1. *(Client) Send a 3D trajectory matrix encoding the current trajectory along with several dummy trajectories to the LBS server.*
> 2. *(Server) Restore trajectories from the 3D trajectory matrices.*
> 3. *(Server) For each of the restored trajectories, perform the Steps 4-6.*
> 4. *(Server) Match the restored trajectory to the stored trajectories to generate a weighted matrix.*
> 5. *(Server) Superimpose the matching result matrices.*
> 6. *(Server) Find the trunk of the tree, the predicted trajectory, in the superimposed matrix.*
> 7. *(Server) Send a 3D trajectory matrix encoding the predicted trajectories to the client.*
> 8. *(Client) Restore the trajectories from the 3D matrix.*
> 9. *(Client) Start using the predicted trajectory.*

Figure 7. An algorithm of the proposed spatial trajectory prediction.

Proposed Method
An algorithm of the proposed method is given in Figure 7. The LBS Service records the current trajectory and sends it along with several dummy trajectories to the server by using a 3D trajectory matrix. The server then initiates the following steps: (i) restoring the trajectories from the 3D matrix, (ii) matching the restored trajectory to the saved trajectories, (iii) superimposing the result matrices after the matching, (iv) the superimposed matrix containing a weighted tree and the trunk of the tree being treated as the predicted trajectory, and (v) the predicted trajectories sent back to the user for a location-based application. User privacy

may be violated at two places in this method: (i) user sending his/her trajectory to the server and (ii) the server sending the predicted trajectory to the user. User privacy is preserved by using this simple approach because the server is not able to tell the user trajectory from the dummy trajectories submitted together. On the other hand, the target predicted trajectory is retrieved from several predicted trajectories on the client side, so it is not possible to associate trajectories to the user other than the user himself/herself.

Algorithm of the Spatial Trajectory Prediction
The proposed method builds a tree (rooted at the end location of the current trajectory) from the previous trajectories. It then takes the trunk of the tree as the predicted trajectory. A path is a sequence of vertices connected by edges and a simple path is a path without repeated vertices. For the tree $T = (V, E)$, a trunk is a simple path $P = (V', E')$ minimizing $\delta(P) + w(P)$ where

$$\delta(P) = \sum_{v \in V} d(v, P) = \sum_{v \in V} \min_{u \in V'} d(u, v), \tag{6}$$

$$w(P) = \sum_{e \in E'} w(e), \tag{7}$$

V is a set of vertices, E is a set of edges, $V' \subseteq V$, $E' \subseteq E$, $w(e)$ is the weight of the edge e, and $d(u, v)$ is the distance between the nodes u and v [14][15]. Since the matrix does not include edges, this research uses $d(e) = w_v$, the weight of the end node or tile v. A predicted trajectory is usually a path among a set of paths that users like to take or is the most popular one and the trunk is the core of a tree and has a minimal weight of total edges. It is the main reason that the trunk of the tree (superimposed trajectories) is treated as the predicted trajectory. Experiments or proofs may be needed to prove that this approach correctly predicts the trajectory we want.

6. CONCLUSIONS

Nowadays there are more smartphones and tablet computers than PCs and servers. The omnipresence of smartphones and tablet computers makes location-based services like Foursquare extremely popular. This paper proposes a kind of location-based research, spatial trajectory prediction. For example, assume you have several friends who are supplying their mobile locations to you constantly. Now, you want to reach any one of them, but the problem is their locations are dynamic instead of static. Using a method of spatial trajectory prediction, we may predict our friends' forthcoming locations and find the ones who are close to us in the next moment. This research proposes a simple but effective method for spatial trajectory prediction. Two of the many contributions made by this project are

- Trajectories, lists of locations, are normally difficult to manage. Trajectories in this research are represented by an innovative matrix representation, the trajectory matrices, which facilitate trajectory storage, indexing, transmission, and processing.
- The proposed method makes user privacy preservation simple and robust by sending the target trajectory with several dummy (false) trajectories. Therefore, the server is not able to associate the trajectories with the users.

In addition, human travel behavior is useful and has been applied in many applications such as city and street design and planning. This research studies various issues related to human travel behavior; e.g., what are the frequent routes from location A to location B and which part of city has the highest probability of travel anomalies?

7. REFERENCES

[1] K. Kolodziej and J. Hjelm. *Local Positioning Systems: LBS Applications and Services*, CRC Taylor & Francis, 2006.

[2] S. J. Vaughan-Nichols. Will mobile computing's future be location, location, location? *IEEE Computer*, 4(2), 14-17, February 6, 2009.

[3] S. Wang, J. Min, and B. K. Yi. Location based services for mobiles: technologies and standards. *Proceedings of the IEEE International Conference on Communication (ICC)*, Beijing, China, May 19-23, 2008.

[4] S. Steiniger, M. Neun, and A. Edwardes. *Foundations of Location-Based Services*. 2006. [Online]. Available: http://www.geo.unizh.ch/publications/cartouche/lbs_lecturenotes_steinigeretal2006.pdf

[5] X. Cao, G. Cong, and C. S. Jensen. Mining significant semantic locations from GPS data. *Proceedings of 36th International Conference on Very Large Data Bases*, pages 1009-1020, Singapore, September 13-17, 2010.

[6] M. Duckham, S. Winter, and M. Robinson. Including landmarks in routing instructions. *Journal of Location Based Services*, 4(1), 28-52, March 2010.

[7] C.-Y. Chow and M. F. Mokbel. Privacy of spatial trajectories. In Y. Zheng and X. Zhou (Eds.), Computing with Spatial Trajectories (pp. 109-141). New York: Springer, 2011.

[8] C. A. Ardagna, G. Livraga, and P. Samarati. Protecting privacy of user information in continuous location-based services. *Proceedings of the 15th IEEE International Conference on Computational Science and Engineering (CSE 2012)*, Paphos, Cyprus, December 5-7, 2012.

[9] W.-C. Lee and J. Krumm. Trajectory preprocessing. In Y. Zheng and X. Zhou (eds.), *Computing with Spatial Trajectories* (pp. 3-32). New York: Springer, 2011.

[10] K. Deng, K., Xie, K, Zheng, and X. Zhou. Trajectory indexing and retrieval. In Y. Zheng and X. Zhou (eds.), *Computing with Spatial Trajectories* (pp. 35-59). New York: Springer, 2011.

[11] Microsoft. *Bing Maps Tile System*. 2014. [Online]. Available: http://msdn.microsoft.com/en-us/library/bb259689.aspx

[12] CloudMade. *The CloudMade Developer Zone*. 2014. [Online]. Available: http://developers.cloudmade.com/

[13] S. Pissanetzky. *Sparse Matrix Technology*. London: Academic Press, 1984.

[14] Y. Li, S. Peng, and W. Chu. Optimal Algorithms for Finding a Trunk on a Tree Network and its Applications. *The Computer Journal*, 52(4), 268-275, 2009.

[15] S. Peng and W.-t Lo. Efficient algorithms for finding a core of a tree with a specified length. *Journal of Algorithms*, 20(3), 445-458, 1996.

How does ubiquitous game-based system influence students' learning motivation

Tsung-Yu Liu

**Department of Multimedia and Game Science, Lunghwa University of Science and Technology,
Taoyuan County 33306, Taiwan, R.O.C.**

ABSTRACT

This study investigated how ubiquitous game-based system influence learning motivation for junior high school students. The English course was conducted on a campus by using a ubiquitous learning system called the Handheld English Language Learning Organization (HELLO). Two group students participated in the learning activities prescribed in a curriculum by separately using ubiquitous game-based learning and non-gaming learning. During the experiment, a survey, and interviews were conducted for the students. The evaluation results of the learning motivation demonstrated that incorporating ubiquitous games into the English learning process could achieve a better learning motivation than using non-gaming method.

Keywords: Augmented reality, Collaborative learning, game-based learning, Task-based learning, Ubiquitous learning system.

1. INTRODUCTION

English is the most popular language in the world and has become the most important foreign language in many non-English-speaking countries. Developing effective and efficient approaches to increasing practical opportunities in actual contexts, and thus, improving students' English learning motivation has become an extremely important research topic. Conventional listening and speaking learning approaches, in particular, have several drawbacks. The first of which is that students lack sufficient opportunity to practice conversation with their English teachers, classmates, and native English speakers. Secondly, schools lack appropriate English learning tools (including software and hardware) for coaching individual listening and speaking. Students have to rely on books and audio CDs as their major learning materials, which leads to deficiencies in spoken English.

Advances in ICT technology have created a new learning model called ubiquitous learning (u-learning). U-learning has afforded a new way to infuse learning into daily life. This research strives to devise low-cost methods of designing an effective English learning system and adopt appropriate learning strategies in designing an effective game-based curriculum. This study attempts to address the major research question: How ubiquitous games influence English learning motivation?

2. METHODOLOGY

Tests were used to evaluate the significant difference in the learning motivation for different learning intervention. Qualitative data were collected from interviews and were used to understand the students' opinions and explain the experimental results.

Curriculum design

The motivation theories have been applied to the curriculum design. McMahon argued that students' active motivation will push them to strive for better performance, achievement, and ability [1]. This study applied Keller's attention, relevance, confidence, and satisfaction (ARCS) motivation model, a model useful for the creation of various student-centric instructional tasks [2]. By using an effective learning system and creating interesting learning activities, we provided students with opportunities for self-learning and collaborative learning. As the course content is designed to be closely related to life experiences, students can really perceive the importance of learning. Besides, we also expect to offer students with opportunities to accomplish challenging tasks, build their confidence, and gain a sense of satisfaction from task accomplishment. For the above objectives, we developed game-based learning activities that make the course more attractive, increase students' life experiences and confidence, and provide students with challenging tasks so that they can experience a sense of satisfaction from overcoming them.

In the Game Generation, computer games comply with the children's contemporary needs, habits and interests [3]. Educational games encompass educational objectives and subject matter that have the potential to enable learning more learner-centered, easier, enjoyable, interesting, efficiency and effective [4], [5], [6]. Yu, Chang, Liu, and Chan reported that the use of a game for high school English learning could increase students' satisfaction of the learning experience [7]. Game-based learning (GBL) is designed to combine learning and game playing, so it will improve the ability of the player to retain education subjects and apply them to the real world. Therefore, this study adopted the GBL learning tactic in its curriculum design in order to effectively engage learning interest and increase learning motivation.

In ubiquitous learning (u-learning), ubiquitous computing occurs all around the learner, whether or not they are aware of it. Liu, Tan, and Chu [8] argued that the nine characteristics of u-learning as being permanency, accessibility, immediacy, interactivity, situation, calmness, adaptability, seamlessness, and immersion. Moreover, ubiquitous games (UGs) are developed using ubiquitous technology and game science; in the real environment, players can use devices or equipment at any time and location to play interactive games involving a portion of physical objects and a portion of virtual ones, which allow them to feel personally and physically involved in the games. Employing UGs in the course could enable more interesting, motivated, and effective learning as well as increase immersive and collaborative learning experience. Therefore, this study made use of ubiquitous games in designing the curriculum.

In addition, task-based language learning (TBLL) focuses on asking students to complete meaningful tasks using the target language. The characteristics of TBLL are interaction, student-centered focus, meaningful materials, fluency language production, learning in the real world, and clear learning goals [9]. Willis pointed out that, in TBLL, students can learn by doing [9]. Kiernan and Aizawa further argued that second language acquisition is best promoted through task-based learning [10]. On the other hand, collaborative learning can improve the cognitive activity of students [11] and increase learning motivation and satisfaction [12]. Omaggio suggested that effective language teaching should provide more practice opportunities in real situations and should guide students to complete a task collaboratively [13]. Therefore, this study also used collaborative TBLL learning tactics in developing learning activities to afford opportunities for competition, enable successful learning, and eventually provide satisfaction to the students.

The proposed game-based English course included eight topics related to the library, health clinic, auditorium, computer classroom, laboratory, store, classroom, and playground zone. An eight-week experiment was conducted during class time, and a 45 minute course was conducted each week. The learning goals of this curriculum were as follows: to enhance English learning, to increase English learning interest and motivation through the designed learning games and to enable students to learn in a real environment. The students in the experimental group used gaming learning approach (using HELLO), while the students in the control group used a non-gaming learning approach (using printed materials and CD players). The two groups used the same course content although the interfaces they used during the classes were different. The curriculum was named 'My Campus', and was designed in five activities [14], [15].

In preparation activity (Week 1), the teachers conducted a pre-test to both groups. In 'Campus Environment' activity (Week 2–3), the control group used printed materials and audio CDs to learn during their free time. The experimental group employed the HELLO to play a ubiquitous game in which they used smart phones to practice listening and speaking. In 'Campus Life' activity (Week 4–5), the control group used printed materials with a zone-related map and audio CDs to learn during class time. The experimental group employed the HELLO to perform a treasure hunt game during class time. In 'Campus Story' activity (Week 6–7), the control group employed a voice recorder to collaboratively perform a story in the classroom. The experimental group employed the HELLO to collaboratively perform a story in an actual context. In evaluation activity (Week 8), the teachers conducted a survey to both groups.

System architecture
In order to enable learning at any time and location and effectively increase practicing opportunities and improve English learning, this study adopted the u-learning system called HELLO[14], [15]. Fig. 1 illustrates the architecture of HELLO [14], [15]. HELLO possesses the following functionalities: Students' learning portfolios can be uploaded into the Evaluation Database (EDB) of the HELLO server, making them available for teachers to review. Teachers input materials and assessments into the Content Database (CDB) through the Content Agent, Assessment Agent, and Push Agent. Each student has a mobile device with which he or she can communicate with the HELLO server. From these mobile devices, students can access materials via a wireless local area network (WLAN). Students utilize a u-Browser tool to download papers, news, learning games, English comics, English songs, listening materials and conversational materials from the HELLO server. They then use the u-Browser tool to play, listen to and watch learning materials. Students can use the u-Test tool to take tests and evaluate their learning progress immediately. Students utilize the u-Speaker to talk to a virtual learning tutor (VLT) which is in the form of an animated-speaking agent which appears on the mobile device. Each student takes a PDA phone near a zone which is attached to a 2-D bar code. The u-QRcode tool on the PDA phone uses the phone camera to photograph the barcode and interpret the image as data. This data is used to access learning material from the server relevant to the location, and display it on the PDA phone.

Procedures
During the preparation phase, explained the experimental purpose, goals, outlines, and evaluation methods to the two groups. The teachers demonstrated how to use the HELLO functions, and introduced the learning activities to the experimental students before they began.

During the 'Campus Environment' activity, the experimental group used the HELLO to execute self-learning game. Each student in the experimental group had a PDA phone installed, with u-Tools for English learning. The u-Tools included several tools which can be used to access self-study English songs, listening materials, and conversational materials from the HELLO server via the WLAN. When the students launched the game, a campus map appeared on the screen of each PDA phone. This map contained numerous zones, each of which was clearly marked. Fig. 2 depicts the 'Campus Environment' computer game [14], [15]. Students moved the character into the learning zone, and the u-Browser then opened zone-related materials. For instance, when a student selected the zone, 'Library', a library appeared on the PDA phone. The student could then choose the movies in order to practice an English conversation or watch an English movie clip. The key aspect of these options is that they enable students to learn without the constraints of time and place, and without having to visit a real library. In contrast, the students in the control group learned zone-related audio conversations by using CD/MP3 players and printed materials during their free time.

During the 'Campus Life' activity, the students in the experimental group used the HELLO to conduct a treasure hunt game, which was designed based upon the ubiquitous game-based learning strategy. Students in the experimental group were asked to practice listening and speaking related to the learning zones. Each student used a PDA phone installed with u-Tools, and followed a guide map on the screen to play the learning game. In order to approach the learning zones, each student followed the guide map on his or her PDA phone, which was equipped with a video camera and hooked up to the WLAN, in order to complete the learning process. For instance, when approaching the real 'Library' zone, a student could use his or her PDA phone to take a picture of the 2-D bar code beside the library, and then decrypt the 2-D bar code. The detected identification of the barcode was then sent to the HELLO server, which located the student and returned situation-related conversational material to the student's PDA phone. The VLT was then superimposed with the zone video on the PDA screen.

The student could then practice a library-related conversation with the VLT, just as he or she would talk with an actual person in a simulated way. The VLT played the role of speaker 'Tutor' and the student played the role of speaker 'Learner'. The VLT spoke the first sentence, and then the student spoke the next sentence following the prompt of conversation sentences in sequence. The conversation between the VLT and the student can be stored into a PDA phone by means of an embedded software recorder, and can then be uploaded onto a server for instructors to grade. The students were thus able to access context-aware content related to locations, enabling context-aware learning. Upon completing a conversation with the VLT in a particular zone, the student was given a hint relating to the next zone. Then he or she proceeded to the next zone, and continued until all of the assigned zones had been visited. Meanwhile, the students in the control group continued to use CD/MP3 players and printed materials to learn conversations in the classroom.

During the 'Campus Story' activity, the designed collaborative learning activity was a story relay race, which was designed based upon a collaborative TBLL strategy. In the beginning, the students could listen to several sample stories, after which they were asked to edit a story collaboratively. Students in the experimental group used the HELLO to play the story relay race game (named 'Campus Story' ubiquitous game). Each team in the experimental group had to select five zones on the map, and then each member had to visit one zone and create a piece of a story about each zone. Each member orally recorded the piece of the story on the PDA phone. Upon successfully completing a piece of a story in a given zone, each member handed his or her baton (PDA phone) to the next member, who listened to the previous story piece and walked to the next zone, continuing in this manner until all of the team members had passed through their five selected zones. In contrast, the students in the control group completed stories by using digital voice recorders in the classroom.

During the evaluation phase, a survey was conducted with the experimental group students upon completion of the course. A survey containing 16 questions divided into five groups was administered, and in order to understand the students' learning motivation, in-depth interviews were conducted upon completion of the survey.

Data collection
The survey was administered to students for examining learning motivation and learner satisfaction. Based on the Keller's ARCS modal, we developed English learning motivation scale,. This survey comprised four sub-scales: attention (Group A), relevance (Group R), confidence (Group C) and satisfaction (Group S) with a total of 23 questions which are listed in Table 5. Responses to all questions were on a five-point Likert-scale, from 5 for "strongly agree" to 1 for "strongly disagree." The Cronbach's α for each sub-scale was 0.86, 0.87, 0.86, and 0.90, respectively, and for the survey was 0.91. These alpha coefficients exceeded 0.85, which confirmed the internal consistency of the survey.
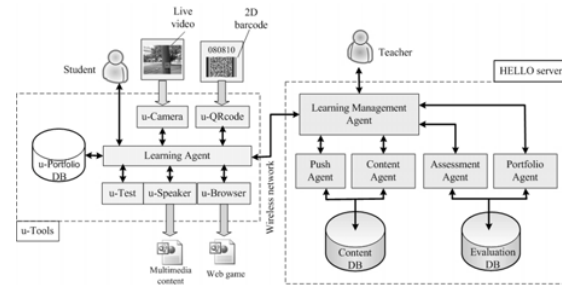
Fig. 1 System architecture of HELLO.
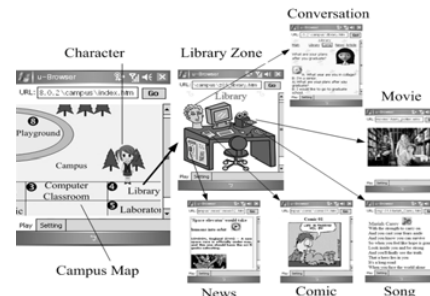
Fig. 2 Self-learning game entitled 'Campus Environment'

Table 5. The survey questions

Item	Question
A1	The themes of the learning materials draw my attention.
A2	The manner in which the learning materials are presented helps me focus my attention.
A3	I can concentrate on the learning activities.
A4	The learning activities can arouse my curiosity.
A5	The 'Campus Environment' activity in this course is interesting to me.
A6	I find the 'Campus Life' activity in this course interesting.
A7	I find the 'Campus Story' activity in this course interesting.
R1	I can link the content of this course to the knowledge that I am already familiar with.
R2	The content of this course is linked to my daily experiences.
R3	The content of this course is worth learning.
R4	The 'Campus Environment' activity in this course has been very helpful to me.
R5	The 'Campus Life' activity in this course has been very helpful to me.
R6	The 'Campus Story' activity in this course has been very helpful to me.
C1	The progressing method of learning activities meets my expectations.
C2	I can control my progress in the learning activity.
C3	I am confident that I can accomplish all the activities.
C4	I am confident that I can apply what I learn from this course to my daily life.
S1	I enjoy the 'Campus Environment' activity.
S2	I enjoy the 'Campus Life' activity.
S3	I enjoy the 'Campus Story' activity.
S4	I am satisfied with my learning achievement in the 'Campus Environment' activity.
S5	I am satisfied with my learning achievement in the 'Campus Life' activity.
S6	I am satisfied with my learning achievement in the 'Campus Story' activity.

Data analysis

The SPSS (Statistical Package for the Social Science, version 15) was used to score the data and answer the research questions. The significance level was set at p<0.05 for all statistical analysis, as is standard practice. This study adopted Cronbach's α coefficient in order to evaluate the internal consistency reliability of questionaire. Cronbach's α coefficient ranges between 0 and 1, and Nunnaly (1978) suggests that 0.7 is an acceptable minimum reliability coefficient. The learning motivation and learner satisfaction of the statistical results of the questionnaire were obtained using a one-way ANOVA.

3. RESULTS AND DISCUSSIONS

A total of 64 valid questionnaires were submitted, with a response rate of 100%. The results (p. > 0.05) of Levene's test for equality of variances indicate that the assumption of the homogeneity of variances in each item is satisfactory. Accordingly, an ANOVA was then performed.

Responses to item A1 indicated that there wasn't a significant difference in the attractiveness of contents of learning materials: F=0.15, p=0.699>0.05. Responses to item A2 indicated that there was a significant difference in the attractiveness of the presentation of learning materials: F=11.675, p=0.001<0.05. Responses to item A3 indicated that there was a significant difference in the attractiveness of the active nature of the learning activities: F=17.75, p=0.000<0.05. Responses to item A4 indicated that there was a significant difference in students' curiosity toward the active nature of the learning activities: F=28.459, p=0.000<0.05. Responses to item A5 indicated that there was a significant difference in students' interest in the 'Campus Environment' activity: F=24.591, p=0.000<0.05. Responses to item A6 indicated that there was a significant difference in students' interest in the 'Campus Life' activity: F=21.502, p=0.000<0.05. Responses to item A7 indicated that there was a significant difference in students' interest in the 'Campus Story' activity: F=7.757, p=0.007<0.05. Students in the experimental group pointed out that games could enhance their concentration and encourage their learning through play, and therefore they thought the learning activities were interesting.

Responses to item R1 indicated that there wasn't a significant difference in the link between the curriculum and students' knowledge: F=0.109, p=0.743>0.05. Responses to item R2 indicated that there wasn't a significant difference in the link between the curriculum and students' daily experiences: F=0.038, p=0.847>0.05. Responses to item R3 indicated that there wasn't a significant difference in thinking that the course is worthy of learn: F=1.066, p=0.306>0.05. Responses to item R4 indicated that there was a significant difference in thinking that the 'Campus Environment' activity is helpful to students: F=27.941, p=0.000<0.05. Many students in the experimental group thought that the 'Campus Environment' ubiquitous game could help their listening skill. Responses to item R5 indicated that there was a significant difference in thinking that the 'Campus Life' activity is helpful to students: F=21.462, p=0.000<0.05. Responses to item R6 indicated that there was a significant difference in thinking that the 'Campus Story' activity is helpful to students: F=5.327, p=0.024<0.05. Students in the experimental group stated that completing a task

collaboratively in a real context encouraged them to accrue more creations than they did in the classroom.

Responses to item C1 indicated that there was a significant difference in believing that the progressing method of learning activities meets students' expectations: F=16.634, p=0.000<0.05. Students in the experimental group indicated that these learning activities were a real learning experience which they could not possibly gain from textbooks and audio CDs. Responses to item C2 indicated that there wasn't a significant difference in controlling the progress of learning activity: F=0.615, p=0.436>0.05. Responses to item C3 indicated that there wasn't a significant difference in having the confidence to accomplish all activities: F=0.958, p=0.332>0.05. Responses to item C4 indicated that there was a significant difference in having the confidence to apply what they learn from this course to students' everyday life: F=9.945, p=0.002<0.05. In the interviews, students in the experimental group stated that they rarely practiced speaking with English teachers, and therefore, they lacked confidence when talking with teachers. The HELLO enabled them to talk with a VLT which could encourage them to gain confidence to speak back.

Responses to item S1 indicated that there was a significant difference in students' enjoyment of the 'Campus Environment' activity: F=12.003, p=0.001<0.05. Students in the experimental group stated that the game was interesting compared with a textbook and they could practice listening through rich multi-media English learning resources, such as conversations, songs, and movies. Responses to item S2 indicated that there was a significant difference in students' enjoyment of the 'Campus Life' activity: F=23.250, p=0.000<0.05. Most of the students in experimental group felt that the 'Campus Life' ubiquitous game was interesting, and they seemed to be immersed in the learning situation during the activity. Responses to item S3 indicated that there was a significant difference in students' enjoyment of the 'Campus Story' activity: F=10.886, p=0.002<0.05. Most of the students in experimental group enjoyed the TBLL game because they could complete a common task collaboratively in the story relay game in the real context, which was an interesting experience. Responses to item S4 indicated that there was a significant difference in being satisfied by students' achievement in the 'Campus Environment' activity: F=16.634, p=0.000<0.05. Responses to item S5 indicated that there was a significant difference in being satisfied with students' achievement in the 'Campus Life' learning activity: F=24.885, p=0.000<0.05. Many students in the experimental group stated that they could practice the same conversation repeatedly until they were familiar with the learning content. Responses to item S6 indicated that there was a significant difference in being satisfied with students' learning achievement in the 'Campus Story' activity: F=6.043, p=0.017<0.05. Numerous students in the experimental group stated that practicing speaking in the real context could encourage them to speak out in public. Therefore, the statistical results of the survey also demonstrate that for the experimental group, using context-aware ubiquitous games in the learning process produced better learning outcomes and learning motivation than using the other two learning games. During the interview, students pointed out that ubiquitous game-based learning is motivating, but the non-gaming learning style is still very important. They thought that ubiquitous learning games could not be used alone without traditional teaching, and believed that ubiquitous game-based learning would be widely adopted in the future.

4. CONCLUSIONS AND FUTURE WORK

This study investigated how ubiquitous game-based course affects the learning motivation of English listening and speaking. A case study was performed with the participation of three high school teachers and 64 high school juniors. A survey and interviews were administered to the students following the experiment. The survey results indicate that the experimental group students gained better learning motivation for attention, relevance, confidence and satisfaction. According to the interviews, students in the experimental group thought the ubiquitous games interesting and useful for assisting them to learn, so they felt satisfied with the ULGs. Moreover, students pointed out that using the HELLO to conduct game-based learning can not only provide opportunities to practice, but also to engage in enjoyable experiences for assisting listening and speaking. Therefore, the students had a positive learning motivation toward using HELLO to aid language learning and were also satisfied with its effectiveness.

In future research, we will continue to work with high school English teachers in order to conduct further studies. We will set up multiple interactive touch screens at learning zones in campuses; thus, students can communicate with the virtual characters on the touch screens in English and further increase their opportunities to learn English.

5. ACKNOWLEDGEMENT

REFERENCES

[1] T. McMahon. Teaching for more affective learning: Seven maxims for practice. Radiography. 2006, 12(1): 34–44.

[2] J. M. Keller. Development and use of the ARCS model of motivational design," Journal of Instructional Development. 1987,10(3): 2-10.

[3] L. Henderson. Video games: A significant cognitive artifact of contemporary youth culture. [Online]. Available: http://www.diagra.org/dl/db/06276.11341.pdf. Last accessed December 19, 2008.

[4] M. Prensky. Digital Game-based Learning, New York: McGraw-Hill, 2001.

[5] M. Virvou, G. Katsionis, and K. Manos, "Combining software games with education: Evaluation of its educational effectiveness," *Educational Technology & Society. 2005*,8(2): 54–65.

[6] M. Papastergiou. Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation," *Computers & Education.* 2008, 52(1): 1–12.

[7] F. Y. Yu, L. J. Chang, Y. H. Liu, and T. W. Chan. Learning preferences towards computerized competitive modes," *Journal of Computer Assisted Learning.* 2002,18(3):341–350.

[8] T. Y. Liu, T. H. Tan, and Y. L Chu. Outdoor natural science learning with an RFID-supported immersive ubiquitous learning environment. *Educational Technology & Society.* 2009, 12(4): 161–175.

[9] J. Willis. *A framework for task-based learning,* London: Longman, 1996.

[10] P. J. Kiernan, and K. Aizawa. Cell phones in task based learning-are cell phones useful language learning tools?" *ReCALL. 2004,*16(1): 71–84.

[11] W.W. Hartup. *Having Friends, Making Friends, and Keeping Friends: Relationships as Educational Contexts,"* Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education, 1992. ED 345 854.

[12] E. Ushioda. *Learner Autonomy 5: The role of motivation,* Dublin, Ireland: Authentik, 1996.

[13] A. C. Omaggio, *Teaching Language in Context: Proficiency-oriented instruction*, Heinli, Boston. (1986).

[14] T. Y. Liu, "A context-aware ubiquitous learning environment for language listening and speaking," *Journal of Computer Assisted Learning.* 2009, 25(6): 515–527.

[15] T. Y. Liu, and Y. L. Chu. Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education.* 2010, 55(1): 630–643.

Unstructured Data Analysis for Marketing Decisions in Agri-food Sector

Roberto PAIANO, Adriana CAIONE, Anna Lisa GUIDO, Andrea PANDURINO

Salento University, Engineering Innovation Department via Monteroni Lecce, 73100, Italy

roberto.paiano@unisalento.it, adriana.caione@unisalento.it, annalisa.guido@unisalento.it, andrea.pandurino@unisalento.it

and

Monica FAIT, Paola SCORRANO

Salento University, Economic Sciences Department via Monteroni Lecce, 73100, Italy

monica.fait@unisalento.it, paola.scorrano@unisalento.it

ABSTRACT

The increase in business, the intense competition between companies, make the marketing an operation and a management mechanism essential for enterprises.

On the other hand, social networks are expanding rapidly and can help companies in collecting customers' feedback, establishing a brand presence, observing the way their brands are discussed and perceived.

This paper focuses on the advantages of analysing data extracted from unstructured sources for marketing purpose and shows the results obtained in a case study in the agri-food sector.

The innovative architecture proposed is realized through the integration of some existing java based tools and of one or many ontologies, in order to retrieve more frequent concepts from unstructured sources, suggest links of articles and images, detect the language used in the sources, suggest other concepts related to the research and filter the results obtained from the elaboration of the unstructured sources.

Keywords: Knowledge extraction, ontology, unstructured sources, marketing intelligence, agri-food sector.

1. INTRODUCTION

In the actual economic environment, characterized by sudden changes, globalization and intensification of competition in any sector, knowledge of cognitive and behavioural elements, that characterize the purchasing process [1] [2], is a fundamental aspect for companies in order to gain a competitive advantage.

To allow the market entry and get good and consistent profits, it is essential to anticipate and beat the competition. To do so, companies need to adopt innovative and automated techniques to study customers' needs, tastes and interests and to contribute additionally to the results obtained using the consolidated marketing tools.

The existing management systems allow an optimal management of structured information (data readily identifiable, organized into structures and interpretable by software tools); the same can not be said for unstructured information (data without an identifiable structure). Some examples of unstructured data are images, video, documents, web pages and text. Data in web pages, represented in a mark-up language such as HTML, are considered to be unstructured [3]. Documents are readable by humans, but difficult to be processed by traditional software tools. The difficulty depends on the intrinsic ambiguity of natural language. In enterprise and institution environment, it has become primary the use of techniques that can extract information of interest from lots of documents that must be managed daily.

Particular relevance assumes the ability to understand, highlight and extract most relevant unstructured content (Sentiment Analysis) to transfer value to business processes. The exploitation of unstructured information allows the monitoring of the target market (market sensing) and the acquisition of data useful for designing marketing activities (market insight) [4]. A study on the expansion of social network in marketing is done in [5] where are examined advantages and risks of the social network marketing, its potential success for business and the possibility to promote products and services via the social network platform.

In recent years, techniques for Information Extraction (IE) have been increasingly developed. Many are the models of text mining proposed [6] to acquire relevant concepts from unstructured sources [7]. For example in [8] many tools are described and compared. The analysis and tests show the importance that may result from the integration of measurement and functionalities of different tasks in the perspective of providing useful analytic data out of text.

These tools use the Natural Language Processing (NLP), a type of semantic search that, trying to get closer to the mechanism of human learning, returns results containing concepts semantically related to each other. Often these tools and techniques are supported by ontology that is a formal, explicit specification of a shared conceptualization [9]. The introduction of ontologies comes as part of the semantic web, from the need to have a language for domain representation that allows expressing the meaning of the documents in the network. The peculiarity of the ontology is the use of common words and concepts to describe and represent the domain of interest; this makes it understandable and usable by people, applications, databases, etc. in order to share a common knowledge concerning any domain.

Semantics and ontologies maximize the value of the process of information and document management, automate and simplify the processes of analysis and classification of information and documents according to taxonomic rules constantly updated.

The paper [10] shows a framework that is useful for finding potential relationship between consumers and mobile OS market activities using ontology. It helps providers offer the right product to the right consumer, by retrieving important information.

Besides tools/algorithms integration, another aspect of great importance is the multi-language support. It is argued in [11] in order to make the approach of knowledge extraction independent of the language in which the text is written.

This paper makes a combined use of information extraction tools from unstructured sources, mostly from blogs, forums and social networks, and of ontologies that describe the specific domain, in order to extract useful information for companies.

The paper is structured as follows. Section 2 gives a brief description of the existing tools of knowledge extraction used within the proposed architecture that is explained in Section 3. We detail each module that composes the architecture and the integration of tools and ontologies. Section 4 presents, whit a use case in the agri-food section, how the software works. Section 5 exposes the results obtained from the test and the conclusions are given in Section 5.

2. TOOLS OF KNOWLEDGE EXTRACTION

In this section we analyse the existing tools of knowledge extraction used in the architecture.

AlchemyAPI

AlchemyAPI (*http://www.alchemyapi.com*) technology uses natural language processing and machine learning algorithms to extract semantic metadata from a text.

The API endpoints are oriented to perform content analysis of web pages accessible from the Internet, html pages or textual content.

Among the features available there is the possibility of extraction of entities, concepts, text categorization, extraction of relations, language detection, extraction of words, sentiment analysis, text extraction, etc., We analyse those which it is made use within the semantic marketing intelligence software.

Language Detection

AlchemyAPI provides functionality for the recognition of the language of a text, HTML page or web-based content. It identifies more languages than other services of text analysis, with extremely high accuracy rates.

Keyword Extraction

AlchemyAPI is able to extract keywords from a textual, HTML or web-based content. Statistical algorithms and natural language processing technologies are used to analyse data, to extract keywords that can be used to index contents, to generate tag clouds, etc.

This processing is supported in different languages, and also enables the foreign language content to be classified and labelled.

Author Extraction

AlchemyAPI is able to extract publisher information from web pages. If an article of news or of blog specifies an author, it attempts to extract it automatically.

Zemanta API

Zemanta (*http://developer.zemanta.com*) is a content suggestion engine for bloggers and other creators of information.

It analyses the user-generated content (for instance a blog post), using the natural language processing and semantic search technology to suggest images, tags, and links to related articles.

It suggests content from Wikipedia, YouTube, IMDB, Amazon.com, CrunchBase, Flickr, ITIS, Musicbrainz, MyBlogLog, Myspace, NCBI, Rotten Tomatoes, Twitter, and Snooth Wikinvest, as well as Blog of other users Zemanta.

Zemanta is a service that connects well-known databases in a single-point solution to detect other content.

The software uses the extraction feature of articles and images.

Jsoup API

It is a Java library (*http://jsoup.org*) for HTML content processing . Provides an API for extracting and manipulating data. Using DOM, CSS, jQuery-like methods, parses the HTML from a URL, file, or a string, finds and extracts data, uses DOM and CSS selectors, manipulates HTML elements, attributes, and text.

3. KNOWLEDGE EXTRACTION ARCHITECTURE

In this section we describe the architecture that we propose. It can be distributed into three main layers as shown in Fig. 1.

Fig. 1 Architecture layers

Presentation Layer manages the interaction with the user. It also shows the results obtained by the core elaboration modules of the marketing intelligence software, such as concepts represented in tag cloud or tables with links to articles and images related to the research.

Integration and Elaboration Layer integrates the existing java based tools described in previous section in order to extract the most important and frequent concepts in the unstructured sources selected and the occurrence of the keywords of interest. It also suggests other articles and images useful for the research.

All information about the researches done is stored in a database through the interaction with the underlying layer.

Database and Ontology Layer stores data of the users, unstructured sources and keywords and of the researches done by the user and the results obtained.

Exploding the three layers we show the modules that compose the architecture (Fig. 2).
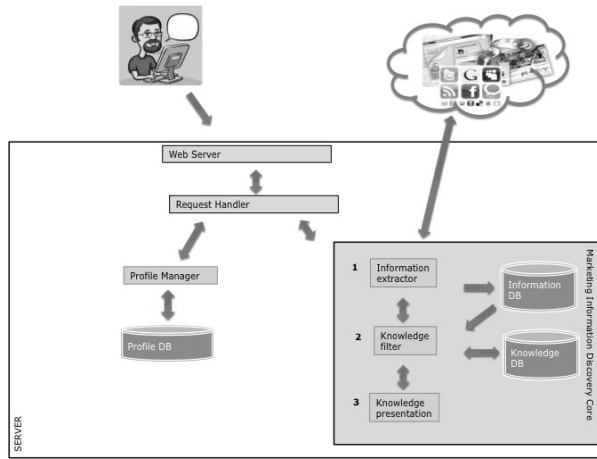
Fig. 2 Reference architecture

The core elements of the marketing intelligence system are:

1. Information Extractor: searches keywords within reference ontology in order to identify other related words. Searches the above words in the sources and calculates the occurrences. Uses the sources and keywords to search for other sources linked to the first;

2. Knowledge Filter: filters the extracted information based on the ontology;

3. Knowledge Presentation: displays this information to the user.

We now describe which modules compose the layer described before and we explain the role of each module that constitutes the architecture introduced.

Presentation Layer

The layer is a bridge between the user and the system. It is composed by:

- *Request Handler:* it receives the requests of the user. It interacts with the profile database in order to retrieve information of the user, or to store data related to him. On the other side, it forwards the requests to the module of marketing information discovery. Both this module of information discovery and the profile manager reply to the Request Handler with information of the authenticated user and of the outputs of the elaboration. The module, then, displays this information in a readable form for the user (for instance a tag cloud of relevant concepts or a table containing articles and images related to the research).

- *Profile Manager:* it is an intermediate level between Request Handler and Profile DB. It makes queries to the database to retrieve user's information and to store other ones.

Integration and Elaboration Layer

The table below summarizes the functionalities of each tool used and integrated in this layer.

API	Use
Alchemy	Extraction of key concepts from a textual content; Identification of the language used in the source; Extraction of the authors of the post
Zemanta	Tip of images and articles related to research

Jsoup	Extraction of pure content from web pages, providing as input the url of the source to be analysed; Extraction of links within the pages, useful for navigation.
WordReference	Translation of keywords in the language of the source

These functionalities are integrated in the modules:

- *Information Extractor:* it uses and integrates the existing java based tools, Jsoup, AlchemyAPI and WordReference, with the aim of analysing sources and keywords selected/entered by the user in the system. The inputs can be provided in different languages, in order to carry out researches related to countries and languages different from their own and get useful results. It is possible through the use of AlchemyAPI and WordReference. The module, indeed, detects the language of the sources and keywords, and if there are differences, it translates the latters in the language of the first ones.

- *Knowledge Filter:* it interacts with the Knowledge Database in order to extract other words or concepts related to the research that will be processed by the Information Extractor.

- *Knowledge Presentation:* it prepares the output to be presented to the user in the form of tag cloud or table of articles and images (extracted by Zemanta API), research results, etc.

Given as input one or more unstructured sources, one or more keywords, and the degree of the navigation for each source, the software can generate different outputs: a tag cloud, images and articles related to search.

Below (Fig. 3) we outline the logic and the integration of tools in order to generate a tag cloud containing the input keywords and the obtained concepts related to the domain, with the highest number of occurrences.
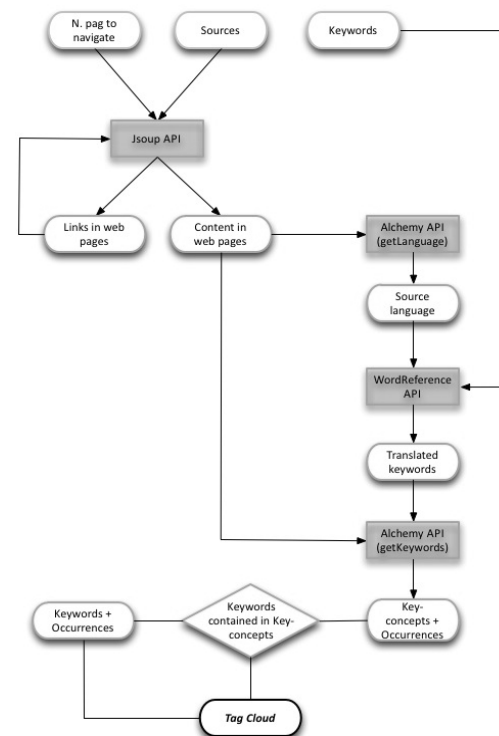
Fig. 3 Logic and integration of tools to generate a tag cloud

Through Jsoup, it is possible to extract pure content from a web page, given the address of the same. Besides extracting the content of the input sources, are extracted links (of number equal to that indicated by the user) contained in the page. Just as the first ones, these sources are processed through Jsoup to extract only the content.

The system checks whether it is possible to combine words entered with other semantically related by asking sector-specific ontology to extract other concepts connected to the first ones.

The content previously extracted from the source is passed to AlchemyAPI that detects the language used. It also identifies the language of the keywords. This information is used with WordReference API to translate keywords in the language of each source.

For each source the system (through AlchemyAPI) proceeds with the extraction of most recurrent concepts. The number of occurrences in the text of each concept is calculated.

In addition the system checks the presence or absence of the keywords selected/entered by the user. If the API has already extracted them, the value of its occurrences is increased; otherwise they are added to the list of words extracted. This list is used to create a tag cloud visible to front end.

The research results can be filtered on the basis of the ontology. Given the keywords, the system extracts the concepts related to them and the words that are not in this set are excluded from the result.

Zemanta API suggests contents such as images and articles (Fig. 4). This suggestion is done on the web in real time.
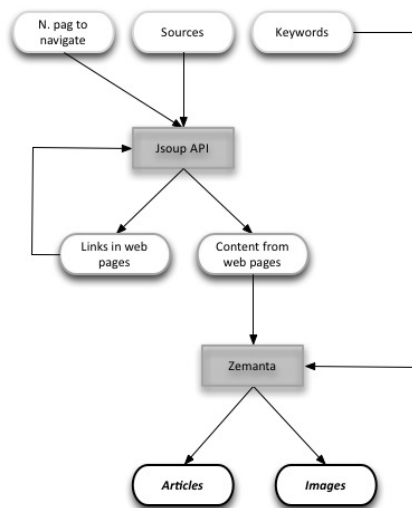
Fig. 4 Logic and integration of tools to obtain articles and images

Database and Ontology Layer

It is the layer of information storage. It includes two type of database:

- *Profile Database:* it stores users' information. Not only personal information, also it keeps track of the relations to transactions carried and to the users' unstructured sources.

- *Information Database:* it stores unstructured sources from which to extract the data and the keywords to search. These sources can be stored in shared manner or linked to the profile

of a user. The database also contains an historical of the researches performed by the users and the obtained results.

The layer integrates also one or more ontologies related to the sector of interest. An ontology, for example relative to the wine sector, was found on the web (*www.w3.org/TR/owl-guide/wine.rdf*), and reused. It can be integrated with other information over.

The ontology contains descriptions of hierarchies and categories of foods and wines (Fig. 5), along with restrictions concerning the association of particular instances.
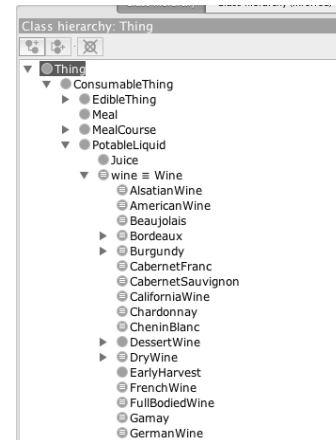
Fig. 5 Wine Ontology

In the proposed architecture, the respective module is:

- *Knowledge Database:* is the ontology of the specific agri-food field. The ontology is used in two direction:

- Before processing sources and keywords, to suggest other words related to the research;

- After processing sources and keywords, to filter the output deleting the words not related to the domain.

4. USE CASE IN AGRI-FOOD SECTOR

In this section we are interested in explaining with an example how the software works.

The software testing is carried out in collaboration with the Economic Sciences Department of the University of Salento. The focus fell on the olive sector.

A phase of identification of the input variables, such as sources and keywords, preceded the software test phase. We proceeded by identifying a foreign market more interesting, such as Australia, one of the new areas of consumption, and then we continued with the search of unstructured sources. The latter were classified by type (blogs, forums, producer, etc.) to identify individuals (bloggers, users, distributors, etc.) to extract information. The sources were provided by the Economic Sciences Department in response to a search on google.au preferring blogs, forums, reviews, magazines, web sites of producers of olive oil in Australia and the websites of the major exporters of oil in Australia.

As regards the choice of the keywords, we took those representing the three main categories of oil: *olive oil*, *virgin olive oil*, *extra virgin olive oil*.

The database has been populated with the detected sources and the keywords.

The experiment involved the sources of three major competitors in the Australian market (respectively Australian, Spanish and Italian competitors).

Primarily the sources and keywords relating to oil producers in the Australian market (Fig. 6) have been selected and processed.

Fig. 6 Sources and keywords related to oil producers in Australia

The output obtained is shown in the image below (Fig. 7). It contains also the keywords selected by the user and the words related to the first ones, extracted by the ontology.

Fig. 7 Tag cloud oil producers in Australia

Similarly, we selected sources and keywords for Italian and Spanish exporters of oil in Australia (Fig. 8 and Fig. 9 respectively).

Fig. 8 Sources and keywords Italian exporters of oil in Australia

Fig. 9 Sources and keywords Spanish exporters of oil in Australia

The portions of the tag clouds are showed in the images in Fig. 10 for Italian exporters of oil and in Fig. 11 for Spanish exporters of oil.

Fig. 10 Tag cloud Italian exporters of oil in Australia

Fig. 11 Tag cloud Spanish exporters of oil in Australia

5. USE CASE RESULTS

In order to interpret and evaluate the results obtained through the use of the software, we have followed an approach for quantitative analysis software in order to identify words with greater occurrence.

Words can be classified into three main categories:

- Cognitive: words within the sphere of knowledge, perception, sensory attributes [12];

- Context: words that refer to extrinsic features (price, method of production, etc.) [13];

- Experiential: words that indicate the consumption occasions, the expected benefits [14].

Starting from the analysis of the tag clouds, shown in the previous section, it is possible to observe that both producers and exporters of olive oil in Australia, uses the three keywords (olive oil, virgin olive oil, extra virgin olive oil). *Extra virgin* concept is more recurrent in the web sites of Australian producers. Indeed, frequent words are *Extra virgin olive oil, Australian Extra virgin, Label extra virgin.* Italian and Spanish exporters use often the generic concept of *olive oil*.

Focusing on the three main categories, we indicate the words with greater occurrence in web sites of producers and exporters of olive oil in Australia.

Category	Australian Exporters	Spanish Producers	Italian Producers
Cognitive	natural olive oil, olive oil premium, olive oil refined, organic, pure olive oil	acid, fluidity, fragrance, fruity, low acidity, aromatic, pure olive oil, light olive oil	acid, fluidity, fragrance, fruity, low acidity, aromatic, pure olive oil, light olive oil
	acid, fresh oil, extra light, light		delicious,

239

	olive oil, flavoured extra virgin		amazing, interesting, different
Context	environmentally friendly principles, ecologically sustainable production, good business practice, consistent quality yields, modern groves	olea europea quality, quality control pet bottles	
Experiential	fine Italian food, Italian food Australia, olive oil recipes, cooking effect, food safety, good diet plan		

We can observe that Australian exporters emphasize on aspects related to naturalness, purity and organoleptic qualities of the olive oil. On the other side, both Italian and Spanish producers accentuate the sensory attributes, in particular the flavour by Spanish producers.

Concerning the category of context, both the producers refer to standard element like quality and type of packaging. Australian exporters talk about modern cultivation and production of a technologically advanced.

At last there is a common thread in relation to the words of experiential, that is food and health benefits.

The analysis of the results leads to the affirmation that in addition to oil-exporting countries such as Italy and Spain, although local producers to satisfy the demand of consumption of the product. Therefore, the current operators and those who want to access this market must focus on enhancement and promotion of the production and consider the advent of new producers in Australia, the strategies initiated by non-traditional producers of olive oil and experiential and context elements on which implement marketing policies.

6. CONCLUSION

This paper proposes a software architecture of marketing intelligence that can extract, analyse and systematize unstructured information (current technological frontier) drawing from the web, in order to be able to support enterprises in the agro-food sector in identifying new markets and consumers' needs and trends. It goes beyond what is proposed in [10], integrating, in addition to ontologies specific for a sector, also existing tools of knowledge extraction and the management of the presentation of front end information to the user. The integration is useful because each tool has some strengths and weakness as described in [8] or has different characteristics.

It has provided useful results relative to the sector of Australian olive oil as the reference market, information on the product and on the actual operators in the sector, quantitative and qualitative data to help design marketing actions etc. This confirms the benefits for the business theorized within the paper [5].

Interesting is also the social aspect, a result of the sharing of information sources (at the discretion of the users) and the multi-language support that allows the use of unstructured sources and keywords in different languages.

The architecture can be used in any other scenario, not only in marketing, fundamental are the sources and keywords that are used as input for the software which will extract the information to be presented to the user.

7. REFERENCES

[1] J.P. Peter, J.C. Olson, **Consumer behavior and marketing strategy**, 4° ed., Irwin, New York, 1996

[2] M.R. Solomon, **Consumer behavior, buying, having, and being**, 6° ed., Pearson, Prentice Hall, New Jersey, 2004

[3] Pearson Education India, **Introduction to Database Systems**, cap. 17, 2008

[4] P. Kotler, K. L. Keller, F. Ancarani M. Costabile, **Marketing Management**, Pearson Education, 2012

[5] V. Bolotaeva and T. Cata, "Marketing Opportunities with Social Networks", **Journal of Internet Social Networking and Virtual Communities** [online], 2010

[6] M. Rajman and R. Besançon, "Text Mining - Knowledge extraction from unstructured textual data", **6th Conference of International Federation of Classification Societies**, 1998

[7] T. Wu, "Theory and Applications in Information Extraction from unstructured text", **Theses and Dissertations**, Paper 741, 2002

[8] A. Gangemi, "A Comparison of Knowledge Extraction Tools for the Semantic Web", **Proceedings of European Semantic Web Symposium (ESWC)**, Lecture Notes in Computer Science (LNCS), Springer, 26-30 May 2013, Montpellier, France, 2013

[9] R. Studer, R. Benjamins and D. Fensel, "Knowledge engineering: Principles and methods", **Data & Knowledge Engineering**, 1998

[10] Y. Nergis and A. Gülfem, "An Ontology-Based Data Mining Approach for Strategic Marketing Decisions", **The International Symposium on Management, Engineering and Informatics,** 2013

[11] D. Gerber and A.-C. N. Ngomo, "Extracting Multilingual Natural-Language Patterns for RDF Predicates", **Knowledge Engineering and Knowledge Management (EKAW)**, Lecture Notes in Computer Science, Springer, 8-12 October 2012, Galway City, Ireland, 2012

[12] J.-B. E. M. Steenkamp, "Conceptual Model of the quality Perception Process", **Journal of Business Research**, vol. 21, 1990

[13] G. Schamel, K. Anderson K., "Wine quality and varietal, regional and winery reputations: Hedonic prices for Australia and New Zealand", **The Economic Record**, vol. 79, n. 246, 2003

[14] A. Tragear, S. Kuznesof, A. Moxey, "Policy Initiatives for Regional Foods: Some Insights from Consumer Research", **Food Policy**, vol. 23, n. 5, 1998

Smart Technologies in Selected e-Europe Digital Agendas: Current   Status, Problems and Future Needs

Dusan Soltes

Faculty of Management, Comenius University,

Bratislava,820 05, Slovakia

ABSTRACT

Since its inception by the Lisbon strategy 2000 on e-Europe strategy and then its continuation through i2010 strategy up to now with the Horizon 2020 in support of the Innovative Europe 2020, the applications of smart ICT in the EU has passed a rather complex and sometimes also a quite controversial development .
In general it has been hampered by several key factors like e.g. the original Lisbon strategy adopted in year 2000 has been adopted by then only 15 EU member states while very soon the EU has been enlarged by 15 (2004) and then by another 2 (2007) mostly less developed CEECs and recently last   year by another one new member up to current EU 28.
The other and a very negative factor has been a lethal and still ongoing Euro crisis that since its start in year 2008 has substantially and quite negatively   influenced priorities and also financial means for the Union's strategies related to the future EU as the most advanced and innovative Union being based on the knowledge based economy and information society.
In the following parts of this paper we are dealing with the current status as achieved in the development of the future EU as the e-Europe as well as we are trying to find the ways and means how to accelerate the entire development in this respect within the selected ten sectors of the future e-Europe..

Keywords: e-Europe, e-Government, G2G, G2C, G2B, e-Health, e-ID, IST – Information society technology.

1.INTRODUCTION

In spite of the above existing problems and weaknesses in the development of the e-Europe there has been gradually going an increase in funding for the R&D activities including research for the ICT and/or IST and related smart technologies programs from around 35 bil. Euro for 6FP (2002-6) to 52 bil Euro for 7FP (2007-13). And now under the HORIZON 2020 it goes up to 70 bil. Euro for years 2014-20. What by itself is a steadily growth in financial support to the EU R&D but it remains still only fraction of the funding going to the most controversial CAP – Common Agricultural Policy that gets about the same amounts or in other words almost the half of the EU annual budget but … not for 5-7 years as it is in case of the support to R&D but for annual subsidies to the EU farmers.
However, in spite of these objective but also not so objective problems there has been achieved some evident progress regarding applications of smart ICT technologies for the needs of the future   electronic

Europe although there are still some gaps between the achievements of the "old" EU15 and the "new" EU12 and now also EU13. According to some key sectors the main results but also some problems as so far achieved and/or identified in the e-Europe development are as they are presented in the following parts of this paper. :

2. SELECTED   TEN KEY APPLICATION AREAS OF THE FUTURE e-EUROPE AGENDAS AND/OR SUBSYSTEMS

In this part of the paper   we are going to deal at least very briefly with some key application areas of the future e-Europe. Into   these   areas we have selected the following ten key application areas as follows:

- e-Government

- e-Signature

- e-Invoicing and e-Procurement

- e-Health

- e-Surveillance

- e-Inclusion

- e-Education

- e-Content and e-Libraries

- e-Knowledge and e-skills

- e-Infrastructure

2.1 e-Government

As far as the development of the future e-Government as one of the key application areas of the future e-Europe is concerned, the main results and challenges in this problem area have been within the following main three government services and that being G2C, G2B, G2G. The G2C stands for the governmental services of the future e-Government to citizens. The G2B means the services of the future e-Government to and/or for businesses. and the G2G represents the governmental agendas and interactions   between   and   among   the   various governmental agencies on the central as well as regional levels. In general we could state in this connection that even after the more than 14 years since the inception of the Lisbon strategy,   the results are still mixed ones exactly as it has been stated yet in 2004 in the well known W. Kok's mid-term evaluation report that later on has led to replacement of the Lisbon strategy by its less ambition version of i2010 and nowadays we have been looking forward in its current version till year 2020 under the HORIZON 2020 strategy adopted for years 2014-2020. As the main problems have been basically still the same ones i.e. that some fundamental technical preconditions have not yet been created like e.g. the e-Signature as the main access tool for becoming an authorized user for various e-Government applications. That we will deal in more details in the separate part of

this paper. But even more negative aspect of the entire concept of e-Government has been still the fact that there were not existing sufficient numbers of e-Government applications and services not only for citizens but as well as for businesses and also for interactions between various governmental agencies, ministries, etc. It would be really desired to have more relevant applications after the more than decade long "implementations" of various strategies related to e-Europe. As far as the Government services to citizens are concerned we could state that only now some main preconditions are created e.g. in many especially new EU member states including our country of Slovakia. In this connection it is worth to mention that e.g. only now has been launched the system of e-ID as a replacement of the still utilized system of classical plastic ID cards. Only now with the new by a chip equipped e-ID it will be possible for citizens to identify themselves in various ways and means of e-communications with various governmental agencies. But again it will be needed to implement as soon as possible the governmental applications that will enable citizens to arrange their needs from the government in the modern e-communications regarding e.g. issuing new passports, driving licencies, e-health or cadaster and other governmental services including e.g. also e-voting as a part of the e-Democracy, etc. The same also for various other e- applications regarding various permissions, authorizations, approvals, agendas that are integral part of the daily life of citizens who should have now an opportunity to arrange all those often bureaucratic demands without necessity to be running from an office to other offices, etc. For example only for the arrangement of a permission to build a new house it is estimated that it requires up to about 50 "signatures" of various governmental agencies what of course is also a source for a possible corruption, clientelism, favoritism, etc. All these and other negatives could be almost removed from the daily life if the particular contacts G2C would be carried out in the e-way of communications also for all other subjects not only citizens. At the moment most of data in this respect are collected mostly by individual governmental agencies rather than to be shared by various governmental agencies within the G2G from some common data storage facilities, etc.\

The same we could say also regarding the G2B as to open e.g. a new business even on the level of SME nowadays requires again various applications, permissions and authorizations from various governmental agencies. In the new modern e-G2B all these bureaucratic obstacles could be removed and the creation as well as operation of the modern e-businesses could be arranged through several steps within e-communication with the one-stop e-business service centers.

As for the G2G communications, the main task remains still the same as we have mentioned it above i.e. to remove lack of e-interactions between and among various

governmental agencies. Especially it is needed to remove the kind of autonomy in collection, storage and utilization of various data being collected by the individual governmental agencies from citizens, businesses, etc.

2.2. e-Signature

This very important e-tool for carrying out any and/or all fundamental e-activities has not been still generally available across the e-EU in the form that would meet the general requirements of the EU common market i.e. that it will be easily and equally available and functional across the entire single market of the Union. In different countries there was applied a different approach, so in principle there exist 28 different versions of e-signature. Some of them are offering it for free, in some other countries like e.g. in Slovakia it has been available only for a rather high fee of around 80 Euro. But there are still not yet so many applications as we have mentioned it above where to use it. Hence, especially in the case of SME, citizens, etc. it is rather too expensive if there are not available so many e-applications where e.g. SME could use it. The same situation is regarding G2C where is still relatively little e-agendas where the citizens could use their e-signatures, etc. Now there has been going e.g. in Slovakia a gradual implementation of the new e-ID with chips. It could be expected that the entire procedure will be then more simple and it can serve not only as an e-signature tool. It is also much cheaper than a "classical" e-Signature as it costs only 4.50 Euro so it is possible to expect that it will be more widely used than the existing e-signature. All that makes this problem area more closer to the practical needs of people as well as businesses especially those belonging to SME. All such applications like arranging e-ID, passports, driving and other documents, etc. are promised to be accessible electronically through this new e-ID alias e-signature. We have to only hope that it will be as being promised.

2.3 e-Invoicing and e-Procurement

E-Invoicing and e-Procurement are other main and very important preconditions for developing modern **e-**Business within the future EU digital internal e-market. There again has been achieved some progress on the national level of individual EU member states but just a very little regarding the unified "EU common e-digital market" of the EU28. The situation is similar like in the case of e-signature i.e. there are more or less working national systems but not the one for the needs of the EU future common digital e-market. One of the main problems is not only the technological one but also the language one as most countries are publishing their e-tenders only in the national languages and thus cutting off potential suppliers from other EU member states. Although it is clear that if it is published in one of 24 official but also national languages of the EU and not also

in one of the basic three working languages of the EU i.e. especially in English but also French and German then it is really difficult to consider such tenders as really ones that could be also EU acceptable. More strict legislation regarding also the language aspects is more than needed also in this problem area. As in many other similar problem areas just to have an EU directive and not more stronger EU regulation is most probably not the solution for achieving really and truly EU-wide solutions suitable for the future e-Europe.

2.4 e-Health

Without any doubts the e-health is one of the key sectors of the future e-Europe if we take into account the demographic development in the EU and especially very fast growing the share of the aging population on its overall population. Again as in other areas also in this one, some progress has been achieved on the national level especially in some most developed EU member states and especially in its Nordic group of states but there is again existing a big problem regarding the e-Health for the entire e-Europe. The main problems again are not in the technological aspects of its implementation but in the legislative and organizational ones. There has been existing already for years a kind of the "common" EU Health Insurance Card but in its classical plastic form only. It means that if the patient needs some medical care outside of its national territory the main problem is that foreign doctor has no information about the particular patient as the above plastic card contains no e-medical records, diagnosis, medications, etc. Hence for the foreign doctor it is sometimes too risky to offer any kind of medical services without this key medical information. Of course in case of life threatening cases some first emergency is normally provided but anything else without the proper medical e-documentation is very problematic and mostly rejected. It will be definitely needed to force the EU member states and their medical authorities to speed up their effort in creating a kind of unified and/or standardized EU e-medical records in the form that it will be easily acceptable within the entire EU! Then also all other related agendas like e-prescriptions, e-consulting and advisory medical services will be fully available to all EU citizens irrespectively where they need any kind of medical services or help.

2.5 e-Surveillance

Again it is a certain paradox of the entire e-Europe implementation strategy that although this specific sector of the future e-Europe originally has not been a part of it at all, in practice it is one of the most developed and according to many accounts one of the sectors being truly and fully developed across the entire EU. As in various other similar problem areas also in case of e-surveillance its enormous development and almost perfection has been achieved as a kind of secondary result of the ongoing

technological development in the modern smart ICT and their applications as it is in case of e.g. mobile phones, tablets, social networks, navigation systems, etc. Mostly it is so thanks to their enormous popularity among the people in general and the EU citizens as well. Moreover if we take into account that it is a citizenship with one of the highest standard of living in the world and thus having enough finances also for following and utilize the results of the latest technological development in this area. Hence even without any special interest and intentions of the operators and through them all interested parties either from the governmental as well as private sector they all have at their finger- tips enormous amounts of and very often even most sensitive personal, business, state/administrative information. As our ongoing research under the EU funded projects SMARTY, CONSENT, RESPECT have documented then it is only a question of legislative, ethical and administrative respects to what extent and if at all that very rich source of data would be used properly exclusively only for the purposes that data was officially recorded for, or it would be misused also for some other often discriminatory, non-ethical or even criminal purposes. In this connection again more stronger and unquestionable EU legislation on the protection of personal data, human dignity, privacy, confidentiality in communications, etc. would be needed to be enacted as soon as possible as otherwise there is a real thread that the generally adopted fundamental human rights and protection of personal data, etc. will become only a document that nobody will be respecting.

2.6 e-Inclusion

Thank to above technological progress being achieved especially in popularity of mobile phones but also tablets, smart phones, etc. as we have been dealing with them in the previous part of this paper also one of the critical and most important parts of all strategies on the e-Europe has become a very practical and relatively easy to be implemented. Although again as in various other similar cases the strategies since the first one i.e. regarding the EU/Lisbon strategy have not been quite clear how to achieve the full e-Inclusion i.e. that every citizen of the future e-Europe will become integral part of this modern information society where everybody will become e-included one. Thanks again especially to popularity of mobile phones and/or in particular of their smart phone versions it has become a common reality. From smallest kids up to the oldest senior citizens all of them are nowadays users of this mobile latest smart ICT and thus also an integral part of their e-inclusion into the contemporary modern information society of the EU. The only problem is that this natural and easy going process of "informatization" of society is not more supported by those who are for this e-Inclusion directly responsible i.e. the EU as well as national authorities, There is still not existing enough relevant applications and programs that

would be really fully and truly utilized their extremely big potential of these mobile as well as smart phones for all various needs of people in the case of e-Health, e-Education, e-Culture, e-Democracy, etc. There is not a short list of various applications in this respect but they are mostly results of business needs of producers and operators and in many cases more for their profit needs like e.g. various games, entertainment, etc. than for above practical needs of ordinary citizens in other they could to the full benefit from being an integral part of the e-Europe and all its potential e-benefits and e-services, etc.

2.7 e-Education

As mentioned in the previous part of this paper, the potential of the development in e-education especially regarding the so-called long-life education especially for adults as well as senior citizens has acquired through above mobile ICT quite new potential horizons. So far this potential has not yet been fully utilized especially in case of elderly and senior citizens for whom the e-education and/or better m-education is the most convenient ways together with the TV how to keep them up-to-date regarding the abruptly changing world. Hence in this respect there are quite big reserves and also potential to use the latest ICT for long life and/or various other forms of informal education. Otherwise, e-learning facilities for young people studying in various educational institutions are basically fully provided as the necessary technological basis has been widely available especially thanks to relatively cheap laptops and tablets and of course also smart phones. However also in this area of e-education there has been existing a serious problem regarding the availability of suitable teachers and educators. As in the most countries especially in the new EU member states the salaries of teachers are quite low ones it is no attraction especially for young people to teach at schools. After acquiring some practical skills and practice as the rule they quit their school jobs and go to work for private sector that has been paying much better salaries than it is in the school system.

2.8 e-Content and e-Libraries

Very closely related issue to the above problems regarding the e-education has been the problem of e-content and e-libraries. As we have mentioned also above there is still existing to some extent the lack of programs that would be supporting e-learning, life long and/or other forms of informal education i.e. educational forms especially intended for people who are already not a part of the regular formal educational systems at schools, etc. In this specific respect the role of the rich e-content and e-libraries are representing a very important and needed part of the e-Europe strategy. It is clear that some progress has already been achieved also in this respect but the more consistent progress has been to some extent negatively effected by the lack of funding for the

necessary staffing as needed for this kind of work. In most cases the practice in this problem area has been based on utilization of the work of some volunteers who in cooperation with librarians have been providing scanning of documents, materials, etc. of the future e-content and e-libraries. It is evident that such an important and to some extent also specific work especially in case of work with some historical or archive documents and/or other objects of interest like cultural items (paintings, sculptures, museum artefacts, etc.) require a bit more professional than just a volunteers capacities of students, etc.. Then again the initiative is in the hands of various private providers of various searching engines that for their commercial reasons are placing on Internet an e-content that in many cases is in direct contradiction to any elementary requirements for good habits, morale, etc. Again as in case of some other problem areas as we have mentioned that above it would require to get more support from the national administrations and governments not to underestimate this one of the key areas of the future information society. They have to create all necessary also financial and professional preconditions in order it would be secured for the benefits of the future information society that without the e-content and e-libraries cannot be existing. Otherwise they would be dominated by the e-content not for benefits of people but for various private providers and their only mostly profit oriented interests including such criminal e-content like regarding pornography, drugs, violence, terrorism, etc. of which the Internet is unfortunately too full regarding its e-content.

2.9 e-Knowledge and e-Skills

It is absolutely necessary that the tasks of e-Europe strategy in the above areas of e-Content and e-Libraries have to support first of all, all various forms of e-content that represents the best what has been achieved in all various forms and categories of the social, economic, cultural, social, etc. areas of the human and societal activities. After all it was one of the key strategic objectives of the original Lisbon strategy that has as one of its main objectives defined the future EU as the most advanced knowledge based economy and information society in the world, Although this objective then later has been to some extent subdued as unrealistic and has been replaced by some less ambitious strategic objectives, it is clear that the process of supporting the spreading and dissemination of the best knowledge and skills from all various sectors of the socio-economic life has to be still in the forefront of the EU as well as national authorities regarding their objectives regarding the future information society. In this respect the first and most important source of such best knowledge and skill we see in the projects and programs and their results as achieved by the research and development projects being funded by the EU under its Frame work programs up the latest one 7FP that was running as the main vehicle of the

EU funded research in years 2007-13. For the future it has been reorganized as the HORIZON 2020 R&D strategy for years 2014-20. Under these previous 1-7FPs there were completed really many so to say thousands of very successful projects with in many cases most relevant results for potentially very important improvements in various aspects of the socio-economic reality in the future EU. The main problem was that until now there have not been found a system that would guaranty that these remarkable R&D results would be fully utilized and implemented also in the practice on the EU as well as national level of individual member states. From our own long year experience from working permanently on various EU funded projects since adoption of the Lisbon strategy i.e. since year 2000 we have learnt that most or at least many member states rather apply for the EU funding for their own national programs and projects than would apply for the less funding that would be needed for using the results of the successful EU funded projects that are available for an immediate utilization and implementation in the national conditions from the EU Repository of all successfully completed EU funded projects. Again a more strict regulation would be needed that would force the member states to use these results rather than to claim EU funding for their various national projects or initiatives. That is in our opinion the best way how to achieve that we would not have in the EU so many various national versions of e-Signature, e-Procurements, e-Invoicing, etc. as we have characterized these problems in the previous parts of this our paper.

2.10 e-Infrastructure.

It is absolutely clear that all the above mentioned problem areas of the future e-Europe and of course also many other of them not to mentioned here in order to be developed properly are fully depended on the existence and some availability of suitable modern latest ICT technological basis. Or as it has been characterized in all EU development strategies it has to be based on the modern technological "backbone" based on the cheap and widely available Internet and the same also regarding the mobile phone networks with again cheap roaming across the entire EU, etc. In this respect all EU strategies on the development of the future modern e-Europe have stressed first of all also the necessity to build a EU wide high speed broadband Internet that would cover all EU member states in full of its availability by the target year 2020 under the HORIZON 2020.strategy. Thus the entire e-Europe strategy has to have at its disposal also its inevitable e-Infrastructure. It is clear it would be much better if such an e-infrastructure has been already available in full right now, but as the popular saying goes it is better if it is late than never.

3. In conclusion

In conclusion we would like just to state that as we have presented it in the previous parts of this paper it is clear that since the inception of the original Lisbon strategy on e-Europe some evident progress has been achieved in implementation of all various areas of this strategy. In many cases as we have tried also to point it out it is not needed much more effort to achieve the objectives in full. But in any case it would require a better coordinated approach especially from the EU member states regarding e.g. their willingness to use the results of the EU funded R&D projects rather than to emphasize their own "ambitions" to build the future e-Europe more through various national initiatives than to accept and fully utilize results of the EU funded projects. After all, they are financed from the EU budget that is primarily based on the contributions from all EU member states! On the other hand in many cases it would help to the EU e-Europe strategy if also the EU itself would be supporting its own strategies also by more strict legislation that would force the member states to be more interested in achieving the Union objectives than only in their own national "subsystems" of the future e-Europe as the whole!

4. REFERENCES

.

[1] J. Cannataci et al: EU/FP7-SSH-2009-A CONSENT – Collaborative project (small or medium scale focused project on Consumer sentiments regarding privacy on user generated kontent (UGC) services in digital economy, Annex I – Description of Work, University of Central Lanceashire, Preston, UK September 2009

[2]Cannataci, J. et al: EU/FP7-SSH-2009-A CONSENT Project, WP3 – Mapping Privacy Settings, D 3.1. Survey Report, University of Malta, La Valetta, October 2010

[3]//www.smart.law.muni.cz

[4] D. Soltes at al: EU/FP7/SMART Project, Work Packages WP2-8 with thé results as achieved in thé Slovak Republic gradually in years 2011-13, FM UK Bratislava, March 2013

[5] D. Soltes at al: EU/FP7-SMART Project, Work Package 10 – The Group Discussions for Slovakia, Bratislava, April 2013

[6] Fundamental Human Rights of the Citizen of the EU, European Commission Press, Brussels 2010

Awareness and Impact of Vietnamese Security Concerns in Using Online Social Networks

Narumon SRIRATANAVIRIYAKUL, Anna Lyza FELIPE, Anna SHILLABEER,
Ondris PUI, Mathews NKHOMA, Quyen Ha TRAN, Thanh Kim CAO
RMIT University
Ho Chi Minh City, Vietnam

ABSTRACT

Vietnam is ranked as the 15th in the world based on the number of Internet users with up to 39.7 million users, which accounts for 43% of Vietnamese population according to Internet Live Stats 2014. More than 50% of Vietnamese Internet users are participating in Online Social Network (OSN). This research will investigate the awareness and impact of Vietnamese security concerns in using OSNs by testing Vietnamese users' intention to use OSN sites and its relationship to users' concerns on privacy, trust and security. Using Structural Equation Modeling (SEM), the findings revealed that 'privacy' correlates with 'security' but these two variables do not have a significant impact on users' trust. Moreover, only 'trust' and 'security' affect users' intention to use OSN.

Keywords: Online social network, privacy concerns, security, trust, intention.

1. INTRODUCTION

Over the past 40 years social interaction enabled through a network has been standardized and opened for all through a worldwide network of interconnected computers: the Internet. This concept has had a great impact on the way we live, think, behave, and has revolutionized the way we communicate with each other. The Internet has been an essential medium for communication, sharing information, and above all a platform for social networking among individuals despite of their geographic location. Online Social Network (OSN) sites, such as Facebook, MySpace, Twitter, or LinkedIn, have rapidly gained popularity with millions of users all over the world, many of whom, particularly the younger generation, have grown up fully wired to the Internet, and consider OSNs more than just a means of keeping in touch but an integral part of life.

According to Boyd & Ellison [1], social networking sites enable users to create profiles, share content and contribute in social networking activities. Social networking sites offer a widespread selection of purposes and hundreds of them share the same basic feature such as a personal profile, which represents the user's identity [2]. When creating a profile users have to answers a series of questions related to their personal information such as age, gender, birthday, relationship status, occupation, interest, or contact details. Exposing of such personal information either can be by "system default" to the general public or it can vary depending on the user's discretion. The effectiveness of linking to people and allowing an individual to keep in touch and maintain a real relationship to people makes the OSN popular. OSNs make it easier for individuals to create new social connections based on the information they provide on their profiles. Making people

connected and revealing information also makes OSN an effective recruitment tool for many purposes.

One of the features of OSN providers is to allow users to have flexibility to choose which information is publicly disclosed or create a restricted or limited data set for audiences to view. Customized private setting limits grouped viewers to connect to the users account. However, there are still many users who are either negligent or unknowledgeable of what information is exposed on their profiles or how to use this feature. Exposing personal information to strangers or just an acquaintance often poses privacy or security threats to users who can be stalked, or have their identity or credit card or other sensitive information stolen.

The researchers developed an interest of users' online security and privacy concerns since there are still few studies regarding these issues. The number of Internet users in Vietnam in particular, has increased more than 15 times since 2000. About 32 million Vietnamese are connected to the Internet and 53% of Internet users have a chance of being attacked online with or without knowing the threat [3, 4]. This investigation will lead on creating an awareness and impact of Vietnamese security and privacy concerns using OSNs.

2. PRIVACY, TRUST, AND INTENTION IN ONLINE SOCIAL NETWORKS

OSN providers keep track of all user interactions and information and store the data in a server for analysis to improve their networking platforms. Several researchers have investigated the threats OSN users might face if such personal information being stored is leaked. Krasnova et al. [5] pointed out several OSN security issues including digital dossier collection by unauthorized parties, cyber stalking and cyber bullying. Gao et al. [6] had also surveyed all possible OSN security attacks, dividing them into four main categories: privacy breach; viral marketing, such as spams and phishing; network structural attacks such as Sybil attacks (one user can create and control several accounts and identities and link with each other in order to promote credibility and reputation enough to reach their attack goals); and malware attacks. Such possible security threats and attacks that users might face on the Internet often determine the users' perception of trustworthiness of any Internet website or service. These threats can be classified as being either concerned with privacy or trust and the users tolerance of these two constructs can affect their intention to use networking sites.

Privacy
Privacy and security are two concepts that often go together and are often used interchangeably. However, privacy and security should be treated as two separate variables. In the OSN context,

privacy concerns refer to the users' awareness of an OSN's handling of their personal information, such as the kind of data that is being collected, where the information will be stored, or how it will be used. This is usually disclosed when a user creates a site profile. The OSN's responsibilities in regards to maintaining privacy are also informed in the policy and made available to users. In contrast, security, involves the technical practices and mechanisms that OSN providers employ in order to ensure that their users' personal data is being well handled and their privacy is being protected so that the users are protected from danger [7].

Privacy also involves an individual's process of regulating the extent to which their personal information is disclosed to others and their control over how such information is disseminated [8, 9]. Privacy is a broad and multidimensional concept, which should be evaluated and examined within a particular context and any applicable rules or policies [10]. The primary concern of privacy resides in an individual's right to control information about oneself [11]. In the context of OSN, privacy refers to a user's control over identity anonymity (the ability to stay anonymous online), a user's personal space privacy (the visibility of the user's online self-representation), and a user's communication privacy (data regarding the user's network connection, such as IP address, length of connection or user's other messages [12]).

Privacy concerns are related to an individual's awareness of the site or service provider's practices and handling of their personal data and the risk tolerance of sharing information about themselves [11, 13]. Several recent studies provide insights into privacy concerns towards the use of OSN's [14-17], but the findings are inconsistent because privacy is usually examined as a multidimensional variable and does not address the unique characteristics of OSN usage [5]. In the Krasnova group's 2009 study, two focus groups were conducted to explore Internet users' privacy concerns in attempt to provide more validated 'privacy concern' measurements in the OSN context [5]. The content analysis of the study showed that there are four main privacy concerns about OSNs: general accessibility (fear of personal information being viewed by unwanted parties, such as parents, bosses, or unauthorized parties like stalkers); social threats (other users' actions they cannot control, such as being tagged in a photo or posting humiliating content on the user's profile); organizational threats (the misuse of personal information by OSN providers and third parties such as online marketing agencies); and identity theft.

In the Krasnova group study, the focus groups consisted of German students whose culture is very different from Vietnam's. Furthermore, the focus group participants were mainly users of Facebook and StudiVZ (a popular social network platform among European students which has strong similarities to Facebook). Facebook's privacy control has been reinforced over the last few years and the privacy settings button now is made more visible to users so that they are aware of the possibility of controlling their privacy.

Based on the study of Sharbaug and Le Trang [18] in understanding of Vietnamese' online personal privacy revealed that the Vietnamese perceived privacy as a means of keeping their own personal information from individuals who might use such information for malicious purposes. The Vietnamese are more concerned about privacy threats from individuals (e.g., friends, colleagues, hackers, thieves) than from entities (e.g.,

government, corporations, and marketers). With regards to their security of their personal information, the Vietnamese believe that it is the individual's, and not the service provider's responsibility to preserve the information. This understanding led the researchers to examine the connections between Vietnamese users' privacy and security awareness with regard to OSNs definition of their responsibilities.

The technologies and systems employed by OSN providers to maintain users' digital safety is an important factor determining users' trust [19]. Jøsang et al. [20] also stated that when security mechanisms provided by OSNs protect users from malicious parties, the websites are deemed more reliable and trustworthy by users. In fact, several studies have examined the link between privacy concerns, perceived security, and trust in Internet shopping, and the findings show that consumers have low level of trust towards an e-commerce website if they fear their transactions are insecure or that those websites might share their personal data with third parties, or that their credit card information might be hacked and stolen [21]. Nevertheless, little is known about the connections among perceived privacy, security, and trust towards the intention to use OSNs in the Vietnamese context.

Trust
One of the important determinants for users' intention in adopting online services is trust [21]. Krasnova et al, [16] measured the impact of users' trust with regards to privacy concerns. Their findings show that the ability to trust the OSN to protect their data impacted users intention to share their personal information and increased their attention to the type of information being disclosed. On the other hand Tufekci [23], discovered little to no relationship between users' online trust and information disclosure; users preferred to manage the viewers they wanted to share information with rather than restricting the type of information being disclosed. Numerous studies found a direct correlation between online security and users' willingness to make online purchases [17, 22]. These studies show that if a websites engaged in electronic commerce mismanaged consumers' personal details, such as by selling data to a third parties that resulted in spam being sent to consumers' Internet mailboxes, those consumers were unlikely to continue using the service and website reliability decreased correspondingly. These efforts propose that users trust of an OSN will affect their intention to use a particular online offering.

Without a doubt, it is worth exploring the connections among the three antecedents, privacy concerns, trust and intention in relation to the OSN context in Vietnam given the current and potential future growth in the use of online resources by Vietnamese users.

The following hypotheses were developed to inform the research:
H1: Vietnamese users' intention to use a site has a positive correlation to privacy protections offered by OSNs
H2: Vietnamese users' intention to use a site has a positive correlation to trust in OSNs.
H3: There is a positive correlation between perceived security trust and privacy for Vietnamese users.
H4: Perception of security in general has a positive correlation with Vietnamese users intention to use a site.

By studying the relationship among privacy concerns, security, and users' trust, this research aims to give a preliminary understanding of factors that could be applied to predict Vietnamese's intentions towards using a specific OSN environment.

3. RESEARCH METHODOLOGY

A quantitative research methodology was applied in this study to achieve the research objectives. Our research model was developed and modified based on Shin's [23] Social Networking Site (SNS) acceptance model, which was well validated by a survey of SNS users in South Korea and was employed to predict users' intention towards using SNS. The qualitative findings from the Krasnova group's research served as the basis for questionnaire items designed to develop a quantified understanding of Vietnamese security concerns.

Two hundred and ten respondents who used at least one OSN were recruited to answer the questionnaire including 5-point Likert scale questions that explored issues of privacy, security, trust, and intention. Next, an online survey was conducted through the Qualtrics platform available at https://asia.qualtrics.com/CP/Index.php to obtain additional data from the respondent sample. To resolve the hypotheses, IBM SPSS AMOS (Analysis of Moment Structures) software version 20 was employed to analyze the dataset with regard to the Structural Equation Modeling (SEM). According to Kline [24], Jöreskog [25], and Jöreskog and Sörbom [26], this SEM model allows both confirmatory and exploratory modeling, meaning they are suited to both theory testing and theory development and therefore are an appropriate model for this study.

To validate the SEM output, it is necessary to analyze the reliability and discriminant validity to examine the reliability of the scale and data validity as proposed by Straub et al. [27].

Average variance extracted (AVE) analysis was conducted to test the discriminant validity. According to Fornell and Larcker [28], the value of AVE for each construct should be at least 0.50. The results show AVE values of the constructs of 'privacy', 'security', 'trust', and 'intention' are 0.593, 0.643, 0.616, 0.689 respectively which proves validity of the data.

Based on the reliability of the scale, a Cronbach's alpha coefficient whose value is greater than 0.7 indicates a reliable and consistent scale, although sometimes the cut-off point of 0.5 is acceptable according to Hair [29]. In this study, the Cronbach's alpha coefficient of the constructs of 'privacy', 'security', 'trust', and 'intention' are 0.814, 0.821, 0.633, 0.803, respectively, therefore indicating that the scale is reliable by both benchmarks.

Before interpreting the output of the SEM model, the overall fitness of the model must be examined. There are primary indices of fit measures that can be utilized to make this determination, particularly Chi-square, normed Chi-square/df (CMIN/df), P-value of Chi-square test, Goodness of Fit Index (GFI), TuckerLewis Index (TLI), Comparative Fit Index (CFI), and Root Mean-Square Error of Approximation (RMSEA).

The model used is proven to fit the population data because of the following indicators: the Chi-square/df is smaller than 2 (Chi-square/df = 1.754 < 2) [31]; all of the GFI, TLI, and CFI indicators are approximately equal to 0.9 (GFI = 0.857; TLI =

0.882; CFI = 0.897) [32]; and the RMSEA is less than 0.08 [32].

4. RESULTS

The P-value and the sign of the Estimate of the Regression Weights shown in Table 1 demonstrate that 'privacy' has a positive significant impact on 'security' (P-value < 0.001), confirming in part H3; 'trust' has a positive significant impact on 'intention' (P-value (trust) = 0.002 < 0.05 confirming H2), security has a positive significant impact on 'intention' P-value

Table 1 Regression Weights

			Estimate	P-value
SECURITY	<---	PRIVACY	**.856**	*******
TRUST	<---	SECURITY	.047	.678
TRUST	<---	PRIVACY	-.055	.669
INTENTION	<---	TRUST	**.293**	**.002**
INTENTION	<---	PRIVACY	-.118	.315
INTENTION	<---	SECURITY	**.212**	**.050**

(*** is smaller than .001)

(security) = 0.05 < 0.1) confirming H4; and 'security' and 'privacy' have no significant correlation to 'trust', therefore not confirming H3; and finally that 'privacy' has no significant correlation to 'intention', therefore not confirming H1.

The Estimate of Standardized Regression Weights demonstrates that 'security' and 'privacy' have approximate impact on 'trust' [Estimate (security) = 0.068; Estimate (privacy) = -0.069)]. However, 'security' has a positive impact on 'trust' while 'privacy' has a negative impact. 'Trust' and 'security' have approximate positive impact on 'intention' [Estimate (trust) = 0.291; Estimate (security) = 0.299]. Conversely, 'privacy' has the negative impact on 'intention' [Estimate (privacy) = -0.149]. The final outcomes of the research are summarized in Table 2 below.

Table 2 Hypothesis Testing Outcome

	Relationship	Impact
H1	Users' concerns regarding privacy have a positive correlation to intention in using safe OSNs	Negative
H2	Users' trust has a positive correlation to intention in using safe OSNs	Positive
H3	Users' concerns regarding security have a positive correlation to trust in OSN	Positive
H3	Users' concerns regarding privacy have a positive correlation to security with regard to OSNs	Positive
H3	Users' concerns regarding privacy have a positive correlation to trust in OSNs	Negative
H4	Users' concerns regarding security have a positive correlation to intention in using safe OSNs	Positive

5. CONCLUSION AND LIMITATIONS

Although the growth rate of the Internet usage in Vietnam is ranking among the top in the world, the literature provides very little research, leading to a low understanding of the perceived privacy and security concerns of Vietnamese OSN users. This research was conducted to investigate the awareness and impact of Vietnamese security concerns in using OSNs by testing Vietnamese users' intention to use OSN sites and its relationship to users' concerns on privacy, trust and security. The findings of this research hopefully provide valuable insights for the implementation of enhanced OSN security measurements to avoid further threats to Vietnamese users.

Using Structural Equation Modeling (SEM), the findings show that Vietnamese OSN users are aware of their online privacy and that privacy concerns, in turn, affect their awareness of security threats they might encounter on OSNs. Nevertheless, privacy and security concerns do not appear to affect Vietnamese OSN users' trust. This is possibly due to Vietnamese users believing individuals should be responsible for their own online security and that a strong password should be sufficient enough to protect their personal information and online privacy [18]. As a result, trust and security have a significant impact on users' OSN intention because they believe the complexity of the password they create will prevent their personal information from getting into the hands of strangers who might use such information for improper purposes. Privacy concerns, on the other hand, perhaps do not correlate well with OSN users' intentions because Vietnam is a collectivist culture and it has been suggested that people from collectivist cultures are more comfortable with sharing their personal information and life experiences with their friends.

A limitation to this study is that the researchers designed questionnaires and worked with a conceptual research model in which each antecedent was expected to correlate with the other. Respondents, however, might have responded to survey questions as independent variables. Moreover, this research has only been based on quantitative results. In addition, the questionnaire was translated from English into Vietnamese, thus word meanings might have been slightly misinterpreted. It is recommended that in the future research, focus groups should be conducted in order to better understand possible links among those variables.

6. REFERENCES

[1] D.M Boyd and N.B. Ellison, "Social Network Sites: Definition, History, and Scholarship", **Journal of Computer-Mediated Communication**, Vol. 13, No. 1, 2007, pp 210-230

[2] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks", Proceedings of the 2005 ACM workshop on Privacy in the electronic society. 2005. ACM

[3] APCERT Annual Report 2011. 2011 January 23, 2012; Available from: http://www.apcert.org/documents/p df/ APCERT_Annual_Report_2011.pdf

[4] VNNIC. Report on Vietnam Internet Resources 2012. 2012; Available from: http://vnnic.vn/sites/default/files/ tailieu/ ReportOnVietNamInternetResources2012.pdf

[5] H. Krasnova, O. Günther, S. Spiekermann, K. Koroleva, "Privacy concerns and identity in online social networks" **Identity in the Information Society**, Vol. 2, No. 1, 2009. pp. 39-63.

[6] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen, "Security Issues in Online Social Networks", **IEEE Internet Computing**, Vol. 15, No. 4, 2011, pp. 56-63.

[7] C. Flavián and M. Guinalíu, "Consumer trust, perceived security and privacy policy: Three basic elements of loyalty to a web site loyalty", **Industrial Management & Data Systems**, Vol. 106 No. 5, 2006, pp. 601-620.

[8] J. Berman and P. Bruening, "Is Privacy Still Possible in the Twenty-first Century?" **Social Research**, Vol. 68, No. 1, 2001, pp. 306-318.

[9] E.V.D. Garde-Perik, P. Markopoulos, B.D. Ruyter, B.Eggen, and W. Ijsselsteijn, "Investigating Privacy Attitudes and Behavior in Relation to Personalization", **Social Science Computer Review**, Vol. 26, No. 1, 2008, pp. 20-43.

[10] D.J. Solove, "Conceptualizing Privacy", **California Law Review**, Vol. 90, No. 4, 2002, pp. 1087-1155.

[11] U. Hugl, "Reviewing person's value of privacy of online social networking", **Internet Research**, Vol. 21, No. 4, 2011, pp. 384-407.

[12] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: challenges and opportunities", **IEEE Network**, Vol. 24, No. 4, 2010, pp. 13-18.

[13] H.J. Smith and S.J. Milberg, "Information privacy: measuring individuals' concerns about organizational practices", **MIS Quarterly**, Vol. 20, No. 2, 1996, pp. 167-196.

[14] C. Dwyer, S. Hiltz, and K. Passerini, "Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace" in AMCIS 2007 Proceedings. 2007.

[15] B. Krishnamurthy and C.E. Wills, "Characterizing privacy in online social networks" in Proceedings of the first workshop on Online social networks. 2008. ACM.

[16] N.K. Malhotra, S.S. Kim, and J. Agarwal, "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model", Information Systems Research, Vol. 15, No. 4, 2004, pp. 336-355.

[17] N. Mohamed and I.H. Ahmad, "Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia", **Computers in Human Behavior**, Vol. 28, No. 6, 2012, pp. 2366-2375.

[18] P. Sharbaugh and P.T. Le Trang, "What's Mine Is Yours: An Exploratory Study of Online Personal Privacy in the Socialist Republic of Vietnam" in Unpublished paper presented at International Communication Association 2012: Phoenix.

[19] B. O'Neill, "Trust in the information society", **Computer Law & Security Review**, Vol. 28, No. 5, 2012, pp. 551-559.

[20] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision", **Decision Support Systems**, Vol. 43, No. 2, 2007, pp. 618-644.

[21] F. Belanger, J.S. Hiller, and W.J. Smith, "Trustworthiness in electronic commerce: the role of privacy, security, and site attributes", **The Journal of Strategic Information Systems**, Vol. 11, No. 3, 2002, pp. 245-270.

[22] Z. Tufekci, "Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites. Bulletin of Science", **Technology & Society**, Vol. 28, No. 1, pp. 20-36.

[23] D.-H. Shin, "The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption", **Interacting with Computers**, Vol. 22, No. 5, 2010, pp. 428-438.

[24] R.B. Kline, **Principles and practice of structural equation modeling**. Guilford Press, 2011.

[25] K.G. Jöreskog, **Testing structural equation models.** Sage Focus Editions, 1993.

[26] K.G. Jöreskog, and D. Sörbom, "Recent developments in structural equation modeling", **Journal of Marketing Research**, Vol. 19, November, 1982, pp. 404-416.

[27] D. Straub, M.-C. Boudreau, and D. Gefen, "Validation guidelines for IS positivist research", **Communications of the Association for Information Systems**, Vol. 13, No. 24, 2004, pp. 380-427.

[28] C. Fornell and D.F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error", **Journal of marketing research**, Vol. 18, No. 1, 1981, pp 39-50.

[29] J.F. Hair, **Multivariate Data Analysis**, New Jersey: Harlow Pearson Education, 2009.

Water Quantity Prediction Using Least Squares Support Vector Machines (LS-SVM) Method

Nian Zhang, Charles Williams
University of the District of Columbia, Department of Electrical and Computer Engineering
4200 Connecticut Ave. NW, Washington, DC, 20008, USA
nzhang@udc.edu, charles.williams4@udc.edu

and

Pradeep Behera
University of the District of Columbia, Department of Department of Engineering, Architecture and Aerospace Technology
4200 Connecticut Ave. NW, Washington, DC, 20008, USA
pbehera@udc.edu

ABSTRACT

The impact of reliable estimation of stream flows at highly urbanized areas and the associated receiving waters is very important for water resources analysis and design. We used the least squares support vector machine (LS-SVM) based algorithm to forecast the future streamflow discharge. A Gaussian Radial Basis Function (RBF) kernel framework was built on the data set to optimize the tuning parameters and to obtain the moderated output. The training process of LS-SVM was designed to select both kernel parameters and regularization constants. The USGS real-time water data were used as time series input. 50% of the data were used for training, and 50% were used for testing. The experimental results showed that the LS-SVM algorithm is a reliable and efficient method for streamflow prediction, which has an important impact to the water resource management field.

Keywords: Water Quantity Prediction, Least Squares Support vector Machine.

1. INTRODUCTION

Land development activities inevitably change watershed conditions, primarily due to an increase in the impervious area through paving, construction, drainage systems and removal or alteration of vegetation which results in water quantity and quality problems for local receiving bodies. The examples of water quantity problems include flooding and stream bank erosion, while examples of water quality problems include pollution loading and receiving water impairments. The impact of this type of activity is more pronounced for highly urbanized areas and the associated receiving waters such as the Potomac River and Anacostia River within the Chesapeake Bay Area watershed. In addition, it has been recognized that climate change can have severe impacts on our streams and rivers due to extreme weather events such as frequent flooding. In this regard, reliable estimation of streamflow at various locations is very important from the water resources management viewpoint. Engineers, water resources professionals, and regulatory authorities need this streamflow information for planning, analysis, design, and operation & maintenance of water resources systems (e.g., water supply systems, dams, and hydraulic structures). Currently USGS provides the streamflow data at various locations in the form of gage height and discharge volume at specific locations, and we used this input to design a reliable prediction model.

The study area is focused on the Potomac River watershed, as shown in Fig. 1. The Potomac River is one of the least dam-regulated large river systems in the eastern United States [1]. The Potomac River has the highest level of nitrogen and the third highest level of phosphorus loading of all the major rivers in the Chesapeake Bay watershed. These nutrients can limit the growth of submerged aquatic vegetation, cause low oxygen conditions and create dead zones.

Approximately 90% of DC area drinking water comes from Potomac River. The Washington Aqueduct is located directly adjacent to the Potomac River. It produces drinking water for approximately one million citizens living, working, or visiting the District of Columbia, Arlington County, Virginia, and the City of Falls Church, Virginia, and its service area [2]. In the last three decades, many areas in the watershed have seen their population more than double. A growing population alters and stresses the natural state of an area's land and water. The Potomac watershed is expected to add more than 1 million people to its population over the next 20 years [3]. The most densely populated area in the watershed is the Middle Potomac, including Washington, DC, which is home to 3.72 million or about 70% of the watershed's population. In the next 20 years, the population of the Potomac watershed is expected to grow 10% each decade, adding 1 million inhabitants to reach a population of 6.25 million. The Potomac River delivers the largest amount of sediment to the Chesapeake Bay each year which can limit the growth of submerged aquatic vegetation and affect populations of all fish, shellfish and birds that depend on this vegetation as a source of food or shelter.
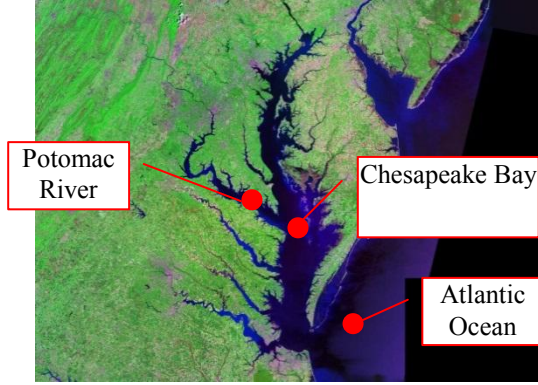
Fig. 1. Satellite landsat photo of Potomac River watershed. The Potomac river is a key entry point to the Chesapeake Bay for millions living in or visiting metropolitan Washington.

Given the existing flow conditions of Potomac River, there is need to analyze the flow conditions at specific locations for future flow, specifically streamflow rate, and a reliable estimate under changing climactic conditions.

To resolve the above problems, it is extremely important to investigate state-of-the-art computational intelligence with the potential for higher rates for urban runoff forecast. Based on the fact that support vector machine has very successfully applications on the time series prediction problems [17], and because time series prediction is a generalized form of runoff quantity prediction, we expect this method will also work the best for the runoff prediction problem.

This paper is organized as follows. In Section II, the principle scheme and the method of Least Squares Support Vector Machines (LS-SVM) are illustrated. In Section III, the function estimation tool is introduced. The water data are briefly introduced. The experimental results of LS-SVM predictions on the water data are demonstrated. In Section IV, the conclusions are given.

2. METHOD

Support Vector Machines (SVMs) are a powerful kernel based statistical learning methodology for the solving problems of nonlinear classification, pattern recognition and function estimation [4]. Least Squares Support Vector Machines (LS-SVM) are an advanced version of the standard SVMs which incorporates unsupervised learning and recurrent networks. Recent developments of LS-SVM are especially relevant to the fields of time series prediction, kernel spectral clustering, and data visualization [5]-[14]. The preliminary results show that the LS-SVM modeling method is promising for time series prediction, thus we want to study the present a current LS-SVM toolbox run through Matlab to implement a number of LS-SVM algorithms.

Support Vector Machines are a new and potential data classification and regression instrument. The basic idea of SVM is based on Mercer core expansion theorem which maps sample space to a high dimension or even unlimited dimension feature space (Hilbert space) via nonlinear mapping φ. And it will boil algorithm which searches for optimal linear regression hyper plane down to a convex programming problem of solution of a convex restriction condition. And it will also obtain overall situation optimum solution so as to use the method of linear learning machine in feature space to solve the problem of high-degree nonlinear regression in sample space [15].

The principles of SVM can be summarized by Fig. 2 as follows:
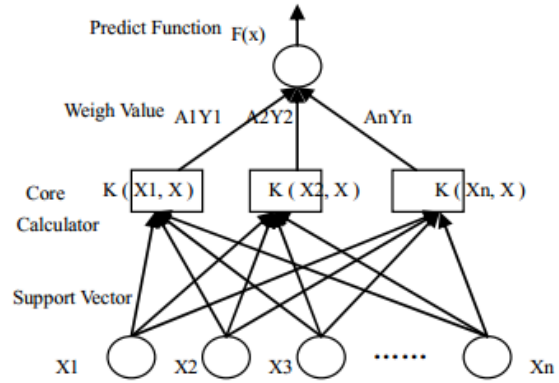
Fig 2. Principle scheme of Support Vector Machine

In Fig. 2, N input support vectors are in the first layer and the second layer is nonlinear operation of N support vectors, that is, the core operation. For nonlinear problems, assume sample to be n-dimension vector, then in one certain domain, N samples and their values can be expressed as:

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N) \in R^n \times R \qquad (1)$$

Firstly, a nonlinear mapping $\psi(\cdot)$ is used to map samples from former space R^n to feature space:

$$\psi(x) = (\emptyset(x_1), \emptyset(x_2), \dots \emptyset(x_N)) \qquad (2)$$

Then, in this high-dimension feature space, optimal decision function:

$$y(x) = w\emptyset(x) + b \qquad (3)$$

is established. In this function, w is a weighed value vector and b is a threshold value. In this way, nonlinear prediction function is transformed to linear prediction function in high-dimension feature space. As development and improvement of classical SVM, Least Squares Support Vector Machine (LSSVM) defines a cost function which is different from classical SVM and changes its inequation restriction to equation restriction. As a result, the solution process becomes a solution of a group of equations which greatly accelerates the solution speed [16]. In Least Squares Support Vector Machines, the problem of optimization is described as follows:

$$\min_{w,b,\varepsilon} L(w, b, \varepsilon) = \frac{1}{2}\|w\|^2 + \frac{c}{2}\sum_{i=1}^{l}\varepsilon_i^2$$
$$(4)$$

Such that: $y_i = w^t\emptyset(x_i) + b + \varepsilon_i$ (i=1,2,…,l)

The extreme point of Q is a saddle point, and

differentiating Q can provide the formulas as follows, using Lagrangian multiplier method to solve the formulas:

$$\frac{\partial Q}{\partial w} = w - \sum_{i=1}^{l} \propto_i \emptyset(x_i) = 0 \qquad (5)$$

$$\frac{\partial Q}{\partial b} = -\sum_{i=1}^{l} \propto_i = 0$$

$$\frac{\partial Q}{\partial \propto} = w^T - \emptyset(x_i) + b + \varepsilon_i - y_i = 0$$

$$\frac{\partial Q}{\partial \varepsilon_i} = C\varepsilon_i - \propto_i = 0$$

From formulas above:

$$\frac{1}{2}\sum_{i=1}^{l} \propto_i \emptyset(x_i) \sum_{j=1}^{l} \propto_j \emptyset(x_j) + \frac{1}{2C}\sum_{i=1}^{l} \propto_i^2 +$$
$$b\sum_{i=1}^{l} \propto_i = \sum_{i=1}^{l} \propto_i y_i \qquad (6)$$

The formula above can be expressed in matrix form:

$$\begin{bmatrix} 0 & e^T \\ e & \Omega + C^{-1}I \end{bmatrix}(l+1)(l+1)\begin{bmatrix} b \\ \propto \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \qquad (7)$$

In this equation,
$$e = [1, \dots, 1]_x^T$$
$$\Omega_{ij} = K(x_i, x_j) = \emptyset(x_i)^T\emptyset(x_j) \qquad (8)$$

Formula (7) is a linear equation set corresponding to the optimization problem and can provide us with α and b. Thus, the prediction output decision function is:

$$\bar{y}(x) = \sum_{i=1}^{l} \propto_i K(x_i x) + b \qquad (9)$$

where $K(x_i, x)$ is the core function.

We are ultimately using the LS-SVM method to calculate and predict the USGS water data, specifically using time-series data prediction. After loading the data into Matlab, we first build the training and testing sets from the data. Next we cross-validate based upon a feed-forward simulation on the validation set using a feed-forwardly trained model. This will supply us with the tuning parameters: γ (gamma) which is the regularization parameter and $\sigma2$ (sigma squared) or the squared bandwidth. The tuning parameters were found by using a combination of coupled simulated annealing (CSA) and a standard simplex method. The CSA finds good starting values and these values were passed to the simplex method in order to fine tune the result. One of the parameters, γ is the regularization parameter, determining the trade-off between the training error minimization and smoothness. The other parameter, σ represents the squared bandwidth. Once the parameters are calculated, we are able to plot the function estimation or use the predict function to predict future values of the data. By using only a subset of the total data available, we can compare the predictions against real values to see how accurate the prediction is.

3. EXPERIMENTAL RESULTS

Function Estimation
Before plotting our USGS water data set, we first test the LS-SVM toolbox with sample data within Matlab to examine its functionality. First is the function estimation tool, which can be implemented with only a few lines of code:

```
X = linspace(-1,1,50)';
Y = (15*(X.^2-1).^2 .*X.^4).*exp(-X)+
    normrnd(0,0.1,length(X),1);
type = 'function estimation';
[gam,sig2] = tunelssvm({X,Y,type,[],[],'RBF_kernel'},
    'simplex','leaveoneoutlssvm',{'mse'});
[alpha,b] = trainlssvm({X,Y,type,gam,sig2,'RBF_kernel'});
plotlssvm({X,Y,type,gam,sig2,'RBF_kernel'},{alpha,b});
```

In this case, we generate sample data sets X and Y, with X being a linearly spaced set of fifty values from -1 to 1, and Y being a quasi-random exponential function. After setting LS-SVM processing type to 'function estimation,' we then tune the coefficients using the tunelssvm command, passing it the values for X, Y, type, and other settings for the computation. The tunelssvm function outputs the two tuning hyper-parameters, gamma and sigma squared. After the algorithm is done tuning, the program generates the following output and specifically values for the hyper-parameters, as shown in Fig. 3.

Fig. 3. Sample output generated from tunelssvm function

After the tuning shown above, the trainlssvm command is called to generate alpha and b, which are then both input into the plotlssvm function to generate the following plot representing the LS-SVM estimation of the data set, as shown in Fig. 4.
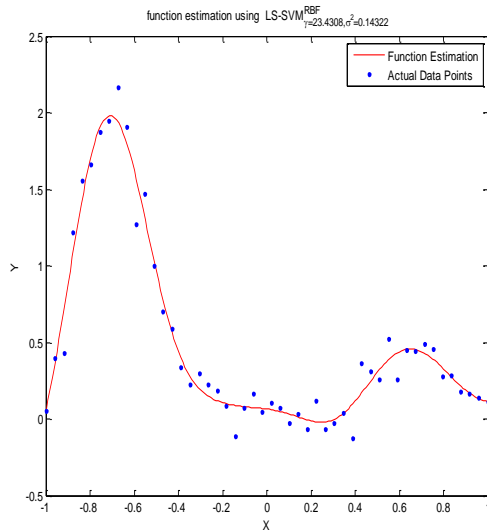
Fig. 4 Plot of Y vs. X, where blue dots represent actual data points and the red line is the prediction

This is a useful tool when trying to fit a line to a random set of data, as the LS-SVM algorithm is very good at classification and regression.

Next we used the LS-SVM command to calculate error bars based on a sample training data set. After loading a randomized sinc function into Y, and a linearly spaced vector into X, we implement the following code to both calculate the LS-SVM regression and error bars, a representation in the error of each predicted value. The command 'bay_error' bar is sent the variables involved after LS-SVM processing and returns a figure that graphically depicts the accuracy of the prediction:
[Yp,alpha,b,gam,sig2] = lssvm(X,Y,type);
sig2e = bay_errorbar({X,Y,type, gam, sig2},'figure');

After executing the above code, two plots are generated, The first is the plot of the actual data points (blue dots) compared to the function estimated to predict those points (red line), as shown in Fig. 5.
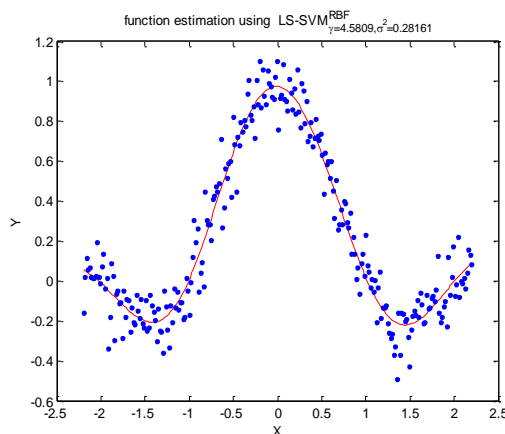
Fig. 5 Another LS-SVM function estimation with randomized sinc function

Fig. 6 shows the error bars, or the algorithm's confidence in its own predictions. The black line covered by black '+' signs is the predicted function and data, while the red dotted lines represent the 95% error bar; basically the area in which 95% of the data resides.
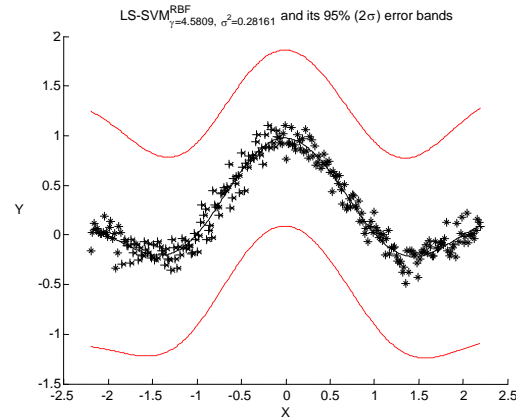
Fig. 6 Data points and prediction (black) with 95% error bars (red)

Time Series Prediction
We used a sample set of about 35,000 data points, all taken at a regular time intervals. We examined both gage height and discharge. The discharge is the volume of water flowing past a certain point in a water-flow. For example, the amount of cubic feet passing through a drain per second is a measure of discharge. Gage height is simply the height of water at a certain point, like the level of the Potomac River measured at Key Bridge. Initially we are only looking to input one of these variables into the LS-SVM algorithm, but in the future it would probably prove to increase prediction accuracy to include the use of both variables at once, the more data input into the system often translates into better results. Fig. 7is a plot of the discharge vs. time.
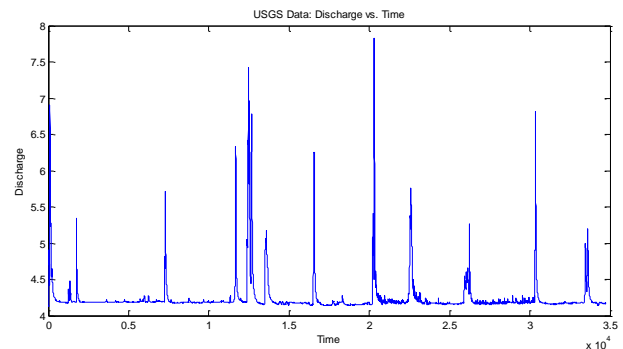
Fig. 7 Plot of entire discharge data set vs. time

These discharge values vary significantly over time- the baseline is at around 4 on the Y-axis, with peaks reaching 8, with very little repetition to the pattern, making it more difficult to predict future values. The gage height is an even more varied set, as seen below.

254
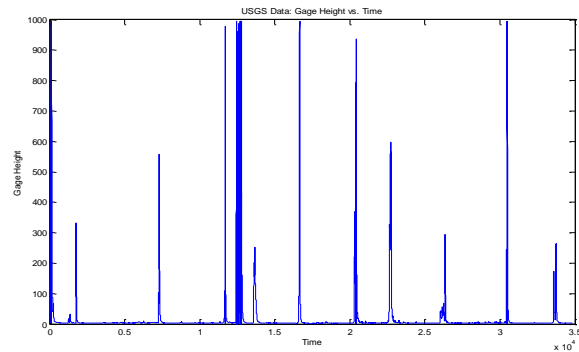
Fig. 8 Plot of entire gage height data set vs. time

The gage height plot contains peaks in similar timeframes as the discharge plot, likely due to large rainfall events or local flooding, as shown in Fig. 9.
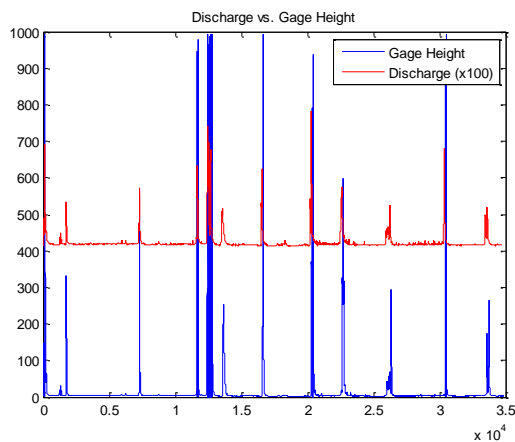
Fig. 9 USGS gage height vs. discharge (scaled) for side-by-side comparison

When looking at both gage height and discharge on the same plot as shown above (with discharge multiplied by 100 so it is visible on the same scale), you can see that they closely correlate with each other. Again this is likely due to the patterns in local weather, specifically precipitation. Sending only one of these variables to the LS-SVM function will produce good predictions, but if we can go further and implement a two input algorithm, that will analyze both discharge and gage height at the same time, this will definitely increase the accuracy of predictions.

To test LS-SVM predictions on the water data we selected a random portion of the discharge data, 500 data points from the original sample of about 35,000. The LS-SVM algorithm is known to be very resource efficient, meaning it can process large amounts of data without using too much processor or memory power. By even this algorithm would take a very long time to process more than a few thousand data points.

The following plot was processed in a similar way to the examples above, but instead of using a random function for Y, we utilize a random selection of points from the discharge dataset. In this case X is just the time axis, as in

all time-series datasets, this data was taken at regular intervals so the values used for X increment one by one.

After loading the data, we tune the hyper-parameters gamma and sigma squared with the tunelssvm command, which generates the following values after 10 iterations, as shown in Fig. 10.

Iteration	Func-count	min f(x)	log(gamma)	log(sig2)	Procedure
1	3	7.406416e-004	9.2756	-7.2357	initial
2	5	6.985583e-004	9.2756	-8.4357	reflect
3	7	5.531121e-004	9.8756	-8.1357	contract inside
4	11	5.133046e-004	9.5756	-7.6857	shrink
5	13	4.963790e-004	9.8006	-7.7232	contract outside
6	17	4.845395e-004	9.8381	-7.9295	shrink
7	18	4.845395e-004	9.8381	-7.9295	reflect
8	22	4.788851e-004	9.8193	-7.8264	shrink
9	23	4.788851e-004	9.8193	-7.8264	reflect
10	27	4.785728e-004	9.8287	-7.8779	shrink

optimisation terminated sucessfully (MaxFunEvals criterion)

Simplex results:
X=18559.201058   0.000379, F(X)=4.785728e-004

Obtained hyper-parameters: [gamma sig2]: 18559.2011   0.000379018892
Start Plotting...finished
>>

Fig. 10 Output generated from tunelssvm command operating on USGS water data

Once the hyper-parameters are tuned, we just assigned alpha and b with the trainlssvm command and finally use the predict function to make a prediction for the next set of values. The final plot generated is shown in Fig. 11, with real USGS discharge datapoints shown in blue while the LS-SVM prediction is the red line.
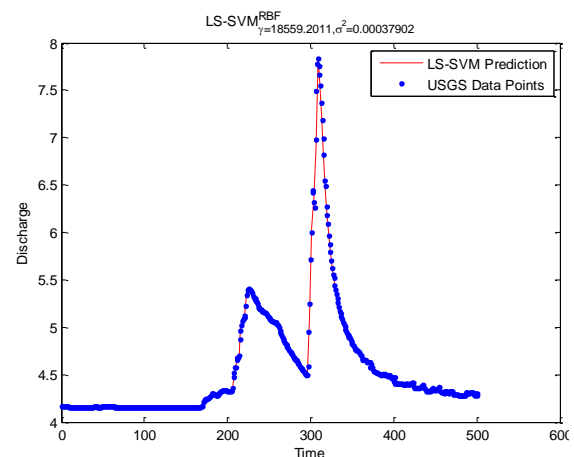
Fig. 11 USGS discharge data set (blue dots) and LSSVM prediction (red)

4.   CONCLUSIONS

In this paper, the least squares support vector machine (LS-SVM) based algorithm to forecast the future streamflow discharge. A Gaussian Radial Basis Function (RBF) kernel framework was built on the data set to optimize the tuning parameters and to obtain the moderated output. The training process of LS-SVM was designed to select both kernel parameters and regularization constants. The USGS real-time water data were used as time series input. 50% of

the data were used for training, and 50% were used for testing.

The experimental results demonstrated that the proposed LS-SVM based predictive model and the training algorithm ensure an accurate prediction of LS-SVM, and by association any natural measurable system. In addition, this provides an excellent prediction method for the time series data, and if correctly implemented can be an invaluable tool in predicting natural weather events. Even outside of storm-water, this algorithm could be very useful to engineers who wish to develop a resource efficient prediction model for any quantifiable data set, i.e. solar radiation, global warming, glacier melting, and more.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1]   Potomac Conservancy, State of the Nation's River, Potomac Watershed. 2007. Available: http://www.potomac.org.

[2]   US Army Corps of Engineers, http://washingtonaqueduct.nab.usace.army.mil/.

[3]   A Snapshot of Potomac Watershed Health, Potomac Conservancy, http://www.potomac.org/site/snapshot-watershed/index.php.

[4]   Support Vector Machines Toolbox, http://www.esat.kuleuven.be/sista/lssvmlab/.

[5]   Z. Liu, X. Wang, L. Cui, X. Lian, J. Xu, "Research on Water Bloom Prediction Based on Least Squares Support Vector Machine," **2009 WRI World Congress on Computer Science and Information Engineering**, vol.5, pp. 764 - 768, 2009.

[6]   Y. Xiang, L. Jiang, "Water Quality Prediction Using LS-SVM and Particle Swarm Optimization," **Second International Workshop on Knowledge Discovery and Data Mining**, 2009. pp. 900- 904, 2009.

[7]   X. Wang, J Lv, D. Xie, "A hybrid approach of support vector machine with particle swarm optimization for water quality prediction," **International Conference on Computer Science and Education (ICCSE)**, pp. 1158-1163, 2010.

[8]   W. Liu, K. Chen; L. Liu, "Prediction model of water consumption using least square support vector machines optimized by hybrid intelligent algorithm," **2011 Second International Conference on Mechanic Automation and Control Engineering (MACE)**, pp. 3298- 3300, 2011.

[9]   L. Yu, X. Wang, Q. Ming, H. Mu, Y. Li, "Application and comparison of several modeling methods in spectral based water quality analysis," **2011 30th Chinese Control Conference (CCC)**, pp. 5227- 5230, 2011.

[10]   L. Liang, F. Xie, "Applied research on wastewater treatment based on least squares support vector machine," **2011 International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE)**, pp. 4825- 4827, 2011.

[11]   X. Zhang, S. Wang, Y. Zhao, "Application of support vector machine and least squares vector machine to freight volume forecast," **2011 International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE)**, pp. 104- 107, 2011.

[12]   R.J. Liao, J.P. Bian, L.J. Yang, S. Grzybowski, Y.Y. Wang, J. Li, "Forecasting dissolved gases content in power transformer oil based on weakening buffer operator and least square support vector machine–Markov," **Generation, Transmission & Distribution**, IET, vol. 6, no. 2, pp. 142- 151, 2012.

[13]   L. Hou, Q. Yang, J. An, "An Improved LSSVM Regression Algorithm," **International Conference on Computational Intelligence and Natural Computing**, vol. 2, pp. 138- 140, 2009.

[14]   X. Zhang, Y. Zhao, S. Wang, "Reliability prediction of engine systems using least square support vector machine," **2011 International Conference on Electronics, Communications and Control (ICECC)**, pp. 3856- 3859, 2011.

[15]   T. A. Stolarski. **System for wear prediction in lubricated sliding contacts**. Lubrication Science, 1996, 8 (4): 315 -351.

[16]   Suykens J A K,Vandewalle J . Least squares support vector machine classifiers. **Neural Processing Letter**, 1999, 9(3):293-300.

[17]   N.I. Sapankevych, Ravi Sankar, "Time Series Prediction Using Support Vector Machines: A Survey," **IEEE Computational Intelligence Magazine**, vol. 4, no. 2, pp. 24-38, 2009.

Case Study: Implementation of IT Governance in a major industry located in Brazil´s Central Region (Issues and Results).

Altino José Mentzingen de Moraes

altino.moraes@gmail.com / +55 (61) 9844-1844

AR - ODL Universities IT Support Co.

CEP: 73.252-200 / DF - Brazil

ABSTRACT

This paper presents a Case Study (Issues and Answers), applied in a major industry located in Brazil´s Central Region, related to rise up checkpoints linked to IT Governance.

As one of its result, is presented a graph of how the IT Governance was approached in researched Industry, in which, it is possible to see which areas this method of working was active and in which this was not applied.

Also a Timeline describes the steps that were undertaken for the implementation of this Case Study, which, can be reused for the resubmission – in a future moment – in the same industry, to evaluate if there was a change of scenery, or in other Organization with the same characteristics of this studied one.

Keywords: Case Study; Survey Results; IT Governance; IT Management; Strategic Planning; Frameworks (Models) of IT Governance; Assessment Maturity Level; M2A3-ITGov Model.

1. INTRODUCTION

It is essential, for the success of the core business of the Corporate, to control how are being done the use of IT Resources and also how are being developed the IT Management actions. Under this context, the IT Governance is inserted and is responsible by its effectiveness.

The question is how is possible, for IT Governance, keep track of the implementation of IT Procedures to guarantee that these are always focused on meeting the Strategic Guidelines that were pre-planned for the activities for this Area.

One of the possible ways, to audit the level of the integration between the IT Governance and the processes of Strategic Planning, is to establish Goals and Targets those can assess the degree of effectiveness of IT actions to meet the expectations of the Strategic Plan of the Organization.

Many ways can be found and can be used as a support to this purpose (to get Goals and Targets in its actual position). As an example, the IT Governance could perform daily meetings with all IT Teams to retrieve the updated status of its tasks and trying to check how these are being conducted but, maybe in the end, this procedure could not show the real status and, in practice, could be impossible be done (since it would involve many Persons and Technicians in several points).

Another way, and that seems more effective, could be the implementation of a Model of Action of IT Governance with the application of this Model to retrieve – in regular manner – this kind of status which could load the Decision Process in order to evaluate whether the Organization can continue with the same initiatives or it needs to move to new ones.

Exercising this another way, this work presents the results achieved and the perceived issues with the implementation of the Guidelines and Techniques related to IT Governance – in the form of a Case Study – in a major industry located in Brazil´s Central Region.

Guidelines and Techniques, mentioned in the previous paragraph and applied in this Case Study, were compiled by researches, studies and surveys conducted by the author of this work since 2007. This work continues those researches, studies and surveys conducted by MORAES ([1], [2], [3], [4], [5], [6], [7], [8], [9] e [10]) during these last years.

During his academic career, he has further enhanced its Implementation Models proposed, with the deepening of concepts built by the various experiences related to implementation of IT Governance in different Organizations and Business Niches. These were divulged in several papers published by the author in previous congresses sponsored by IIIS, as well as, in events held in Brazil.

In this Case Study development were applied ways of addressing the management of IT Governance and also were implemented Frameworks (Models) of Governance – which includes – charts and functional descriptions of the functions of the Manager in charge of each Department (besides the inclusion of a definition of their profiles, roles and responsibilities) and of its Technicians as well.

This same implementation of this IT Governance Model, subject of this paper, has covered 12 (twelve) Fields of Action of IT, where were tested, whether or not they were getting success in keeping alignment and the commitment with Strategic Planning.

In the final of this Case Study, some new approaches were created and other tasks – those already existed – were reviewed. Both, the new approaches created and the existent tasks reviewed, were aggregated to the initial IT Governance Model.

Among other conclusions on how industry has been driving IT Governance management processes, this Case Study results also show suggestions on organizational and functional structure patterns towards this field. Those patterns were created based on the reality noticed during this Case Study application in the field of practice.

It is possible also to understand, as a extra conclusion, that data and information presented in this paper can be useful and can help to implement an effective IT Governance Model in any kind of Corporation – despite of the User-Corporation size – by the academic manner as this work was built (assembly and simulation) and by the methodological structure exposed for the treatment regarding the integration of the IT Governance with the processes of Strategic Planning.

2. PROBLEM STATEMENT

The main question that this paper intends to answer is how initiatives of IT - Information Technology can align with the initiatives and expectations of Corporate Vision with the goal of adding operational value for core-business of the Organization.

3. CLASSIFICATION OF SAMPLE

According SEBRAE [11] the classification of Brazilian Industry, in terms of their size (Micro Enterprise, Small Enterprise, Medium Enterprise and Large Enterprise) within its Annual Gross

Operating Revenues and Numbers of Employees, is presented in Table 1.

Table 1 - Enterprise Classification

Range	Annual Gross Operating Revenue	Number of Employees
Micro Enterprise	Less than or equal to R$ 2,4 millions of Brazilian Rreais	Until 19 Employees
Small Enterprise	Greater than R$ 2,4 millions of Brazilian reais and less than or equal to R$ 16 millions of Brazilian reais	From 20 to 99 Employees
Medium Enterprise	Greater than R$ 16 millions of Brazilian reais and less than or equal to R$ 300 millions of Brazilian reais	From 100 to 499 Employees
Large Enterprise	Greater than R$ 300 millions of Brazilian reais	More than 499 Employees

According to this classification, the researched Industry qualifies as a Large Enterprise, because in 2013, achieved an Annual Gross Operating Revenue of over R$ 870.000.000,00 (eight hundred and seventy millions of Brazilian reais) and had as Number of Employees a value around of 3,000 distributed throughout Brazil.

4. CASE STUDY DEVELOPMENT

The research had, as steps to be implemented, the below presented:

1. Survey of IT Systems;

   Rmk.: In this step, were also considered new projects in Implantation Phase.

2. Ranking of IT Systems;

3. Knowledge Capture about IT Systems Operation;

4. Attribution of Significance Value for IT Systems;

5. IT Systems Value Add for Business;

6. Research of Technologic Solutions applied to IT Systems.

The Annex A provides a Timeline, of 8 (eight) weeks equal to 2 (two) months without gaps, which was used as basis for the development of the above activities, under the operational aspect, which generated specific results – that when were analyzed – subsidized the conclusions presented in this work.

According to the processes listed, this research began with the identification of the Systems of IT - Information Technology found in the Organization. The total raised was 117 (one hundred and seventeen) and some EXCEL spreadsheets also were recorded (because, for the researched Industry by their complexity, these were also considered as "Systems").

After this step, these Systems were evaluated how these would fit in accordance with the 12 (twelve) Fields of Action of IT defined by MORAES [6] the author of this paper. The Table 2 shows the quantification (some Systems may be counted in more than one area or none) found for the Systems of IT - Information

Technology and the distribution of these among the 12 (twelve) Fields of Action of IT defined by MORAES [6].

Table 2 - General Total of System distributed by the 12 (twelve) Fields of Action of IT

Fields of Action		Total of IT Systems
Code	Name	
AD	AUDITING	9
CN	COMPLIANCE	3
DV	DEVELOPMENT	34
KL	KNOWLEDGE	2
MG	MANAGEMENT	27
PL	PLANNING	44
PR	PRODUCTION	68
PJ	PROJECT	17
QL	QUALITY	28
RQ	REQUIREMENT	78
SC	SECURITY	44
TS	TESTING	39

The Maturity Levels presented by MORAES [6] Are: A = Match (which means that the procedures of this Field of Action of IT correspond to meet all the expectations of the Corporation), B = Match with Restrictions (which means that the procedures of this Field of Action of IT correspond to meet some all the expectations but not all) and C = No Match (which means that the procedures of this Field of Action of IT correspond to meet none of the expectations of the Corporation).

In the sequence of processes applied for the implementation of this research, the identification of the Maturity Level – achieved for each Systems of IT - Information Technology – was performed. This is shown in Table 3 and, for purposes of presentation as an Executive Summary, it is organized as an extract results compiled in the end of the final application of MORAES [6] with the balancing of the Significance Value (which can change the score originally assigned).

Table 3 - General Total of System distributed by the 12 (twelve) Fields of Action of IT and with its Maturity Level assigned

Fields of Action		Total of IT Systems	Maturity Level (A/B/C)
Code	Name		
AD	AUDITING	9	C
CN	COMPLIANCE	3	A
DV	DEVELOPMENT	34	B
KL	KNOWLEDGE	2	C
MG	MANAGEMENT	27	B
PL	PLANNING	44	A
PR	PRODUCTION	68	A
PJ	PROJECT	17	A
QL	QUALITY	28	A
RQ	REQUIREMENT	78	C
SC	SECURITY	44	B
TS	TESTING	39	C

In Annex B is presented a Graphic Vision that associates, for each Systems of IT - Information Technology, in which Business Niche each one is positioned and also shows other information that can facilitate the comprehension of the details of the survey done and its results.
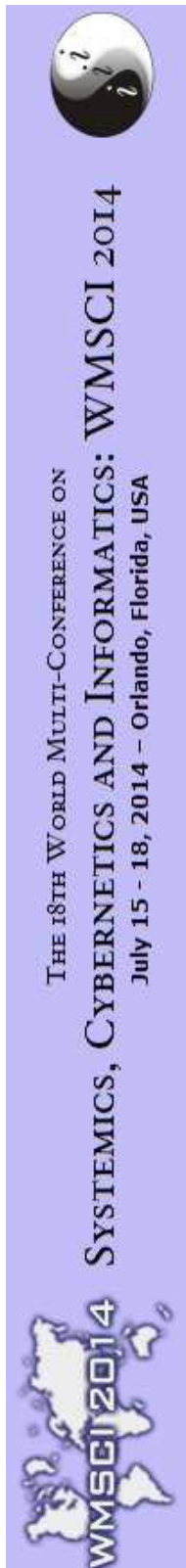
5. CONCLUSION

The below conclusions, listed as Points of Attention, directs consideration in evaluating solutions to problems that can also be

observed in Organizations with the same size and field of business.

- Due to globalization and need for high agility to remain competitive, the researched Organization has invested heavily in Business and Administrative Management and less on Systems of IT - Information Technology in Production Area;

- As differential in the market, and to enable competition on equal terms in the concurrent environment, the researched Organization also invested heavily in Technological Innovation , both in Industrial Machinery as in the Products (Software) and Equipment (Hardware) to meet more effectively their applications of Systems of IT - Information Technology;

- The training of employees, in the researched Organization, has focused on the improvement of procedures aimed at improving Quality Management (Customer Service) and Production Control (Parameter Indicators), just for that purpose (under a Program of Continuous Quality Improvement) increase market share by attracting and retaining new Customers, by the perception of them about this differential in its Services and Products.

6. REFERENCES

[1]MORAES, Altino. J. M.. **Cenário sobre Governança de Tecnologia da Informação em grandes organizações no Brasil.** Knowledge and IT Management Master Dissertation. MGCT - Mestrado em Gestão do Conhecimento e da Tecnologia da Informação. UCB - Universidade Católica de Brasília. Programa de Pós-Graduação *Lato-Sensu*. Brasília, 2007.

[2]MORAES, Altino J. M.; COSTA, Ivanir; FERNEDA, E.; SPINOLA, Mauro M.; SANTOS, R. R.. **Proposta de Modelos de Governança aplicáveis à TI - Tecnologia de Informação**. XXVIII ENEGEP - National Production Engineering Meeting, 2008, Rio de Janeiro / RJ - Brazil. Proceedings of XXVIII ENEGEP - National Production Engineering Meeting, Rio de Janeiro / RJ - Brazil: ABEPRO - Brazilian Association of Production Engineering, 2008.

[3]MORAES, Altino J. M.; COSTA, Ivanir. **Proposta de alinhamento das atividades da TI, na produção de aplicações, com as determinações geradas pelo Planejamento Estratégico da Organização**. XXIX ENEGEP - National Production Engineering Meeting, 2009, Salvador / BA - Brazil. Proceedings of XXIX ENEGEP - National Production Engineering Meeting. Rio de Janeiro / RJ - Brazil: ABEPRO - Brazilian Association of Production Engineering, 2009.

[4]MORAES, Altino J. M.; COSTA, Ivanir; FERNEDA, E.; SPINOLA, Mauro M.. **Practical approach for implementation of Governance Process in IT: Information Technology Areas**. Nan Si Shi; Gilbert Silvius. (Org.). Enterprise IT Governance, Business value and Performance Measurement. Hershey / Pennsylvania (USA): IGI Global, v. p. 19-40, 2010.

[5]MORAES, Altino J. M.. **Adequação das iniciativas de melhoria dos Processos de Produção dos Serviços de TI - Tecnologia de Informação com políticas de preservação do Meio-Ambiente**. XXX ENEGEP - National Production Engineering Meeting, São Carlos / SP - Brazil. Anais do XXX ENEGEP - National Production Engineering Meeting. Rio de Janeiro / RJ - Brazil: ABEPRO - Brazilian Association of Production Engineering, 2010.

[6]MORAES, Altino J.M.. **Proposta do M2A3-GovTI - Modelo de Maturidade para Análise do Alinhamento das Atividades relacionadas à Governança da TI em conformidade com as expectativas de resultado planejadas pela Organização**. Production Engineering Doctorate Thesis. PPGEP - Programa de Pós-Graduação *Stricto-Sensu* em Engenharia de Produção. Universidade Paulista (UNIP), 406 p., São Paulo, 2011.

[7]MORAES, Altino J. M.. **Proposal of guidelines for corporations conciliate the needs for the improvement of IT Services Production with the issues relating to Clean Production**.. 2nd. ICEME, Orlando / FL - USA. 2nd. ICEME. 2011 Proceedings. Winter Garden / FL - USA: IIIS, 2011.

[8]MORAES, Altino J. M.. **Plan-for-Gov[IT] - Planning for Governance of IT Method: use of the techniques of text retrieval for mapping the expected support needs from it area to serve of the corporation score-business expectations**. 3rd. ICEME. Orlando / FL - USA. 3rd. ICEME 2012 Proceedings. Winter Garden / FL - USA: IIIS, 2012.

[9]MORAES, Altino J. M.. **Proposal of Result Indicators for monitoring of operational initiatives of Information Technology (IT) by the IT Governance - through 12 (twelve) Operational Areas of IT - in order to check how these are related to Corporate's Strategic Planning**.17th. WMSCI. Orlando / FL - USA. 17th. WMSCI 2013 Proceedings. Winter Garden / FL - USA: IIIS, 2013.

[10]MORAES, Altino J. M.. **Monitoramento do cumprimento pela TI - Tecnologia da Informação das metas estratégicas estabelecidas pelo Planejamento Corporativo via a auditoria executada pela Governança da TI em suas Áreas de Atuação por meio de Indicadores de Resultado**. XXXIII ENEGEP - National Production Engineering Meeting, 2013, Salvador / BA - Brazil. Proceedings of XXXIII ENEGEP - National Production Engineering Meeting. Rio de Janeiro / RJ - Brazil: ABEPRO - Brazilian Association of Production Engineering, 2013.

[11]SEBRAE accessed "**http://www.sebrae.com.br/uf/goias/ indicadores-das-mpe/classificacao-empresarial/critérios-e-conceitos-para-classificacao-de-empresas/criterios-e-conceitos-para-classificacao-de-empresas**" in 14/01/31.
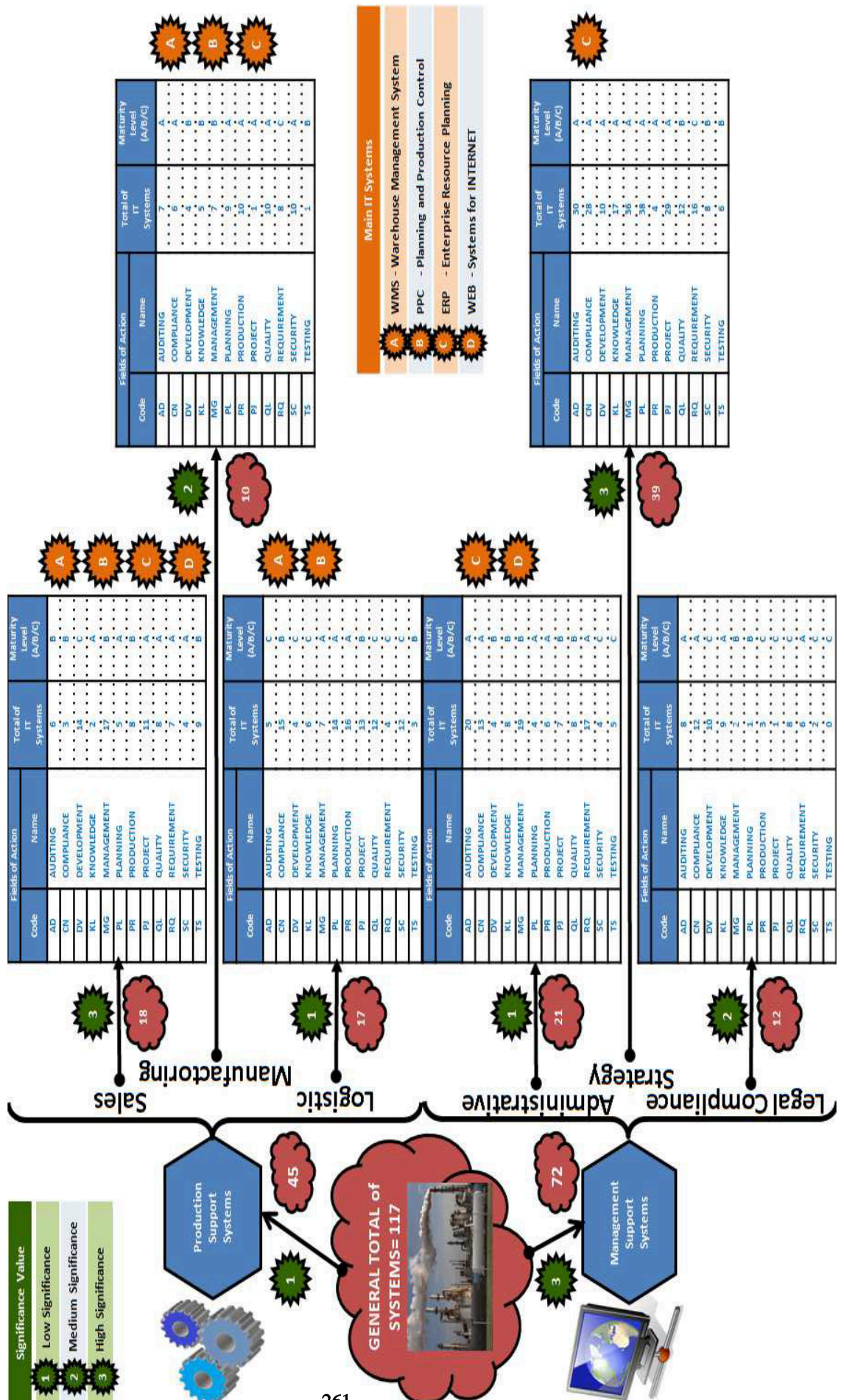
Annex A

The 18th World Multi-Conference on
Systemics, Cybernetics and Informatics: WMSCI 2014
July 15 - 18, 2014 – Orlando, Florida, USA

WMSCI 2014

Research Timeline

2013/08/28 - V:05.a

DESCRIPTION of ACTIVITIES	WEEKS
	1st. · 2nd. · 3rd. · 4th. · 5th. · 6th. · 7th. · 8th. (M T W T F)
1. Survey of IT Systems	
1.1. Study of Existent Documentation	
1.2. Interview with Final-Users	
1.3. Compilation of Results	
2. Ranking of IT Systems	
2.1. Prospection of Types of Software Application	
2.2. Prospection of # of Final-Users of Applications	
2.3. Prospection of # of BU Users of Applications	
2.4. Compilation of Results	
3. Knowledge Capture about IT Systems Operation	
3.1. Overview of Features	
3.2. Real Operation Participation with Evaluated Company Final-Users	
3.3. Compliance Grade with Requirements	
3.4. Quality Evaluation	
3.5. Compilation of Results	
4. Attribution of Significance Value for IT Systems	
4.1. Definition of Significance Value by Evaluated Company Managers	
4.2. Reranking of IT System using Significance Value	
4.3. Reranking Validation with Evaluated Company Managers	
4.4. Compilation of Results	
5. IT Systems Value Add for Business	
5.1. Submission of General Questionaire	
5.2. Submission of Financial Questionaire in Financial Area	
5.3. Submission of Production Questionaire in Production Area	
5.4. Submission of Administrative Questionaire in Administrative Area	
5.5. Compilation of Results	
6. Research of Technologic Solutions applied to IT Systems	
6.1. INTERNET Generation Resources	
6.2. Data Base Facilities	
6.3. Computer Interconnection	
6.4. Web Service	
6.5. Languages and Components	
6.6. Technician Expertise	
6.7. Compilation of Results	
FINAL COMPILATION OF RESULTS	

Annex B

THE 18TH WORLD MULTI-CONFERENCE ON
SYSTEMICS, CYBERNETICS AND INFORMATICS: WMSCI 2014
July 15 - 18, 2014 – Orlando, Florida, USA

WMSCI 2014

Main IT Systems

- A WMS - Warehouse Management System
- B PPC - Planning and Production Control
- C ERP - Enterprise Resource Planning
- D WEB - Systems for INTERNET

Significance Value

- 1 Low Significance
- 2 Medium Significance
- 3 High Significance

Fields of Action (common to all tables):

Code	Name
AD	AUDITING
CN	COMPLIANCE
DV	DEVELOPMENT
KL	KNOWLEDGE
MG	MANAGEMENT
PL	PLANNING
PR	PRODUCTION
PJ	PROJECT
QL	QUALITY
RQ	REQUIREMENT
SC	SECURITY
TS	TESTING

Each Fields of Action table has columns: Code | Name | Total of IT Systems | Maturity Level (A/B/C)

Departments: Sales, Manufacturing, Logistic, Administrative, Strategy, Legal Compliance

Production Support Systems = 45
Management Support Systems = 72

GENERAL TOTAL of SYSTEMS = 117

Implementation of Massive Agent Model Using Repast HPC and GPU

Shinsaku Segawa, Shofuku Kin, Hidenori Kawamura, Keiji Suzuki
Hokkaido University, Sapporo, Hokkaido, Japan

ABSTRACT

Agent Based Model (ABM) is efficient for analysis of various social mechanisms. Recently, there are many studies on massive agent model to explain more complex social phenomena.

Then, we aim for implementation of large scale simulation model using Repast HPC toolkit, a platform for massive agent model. In this article, we build "Schelling Segregation Model" for spatial model using geospatial data provided OpenStreetMap, an open source project creating a free editable map. In this model, agents are located continuous space, not grid in original. When an agent is "unhappy" and migrates to new location, it costs agents some simulation time depending on distance between old location and new one.

This article reports simulation results using Japanese cities and verification result about execution time.

Keywords: Multi Agent Simulation, Social Simulation, High Performance Computing

1. INTRODUCTION

Recently, agent based models (ABMs) become larger, high performance computing systems makes it possible to perform simulations at huge scales only a few years ago [1]. ABMs are efficient for the simulation of objects whose mathematical modeling is difficult [2].

ABMs are seen as attractive approach to reproducing and analyzing diverse social systems including autonomous and heterogeneous decision making entities, i.e., humans [3]. Researchers are possible to explore how through the interaction of many individuals more emergent phenomena arise. Moreover, ABMs allows for practitioners to build models of complex social phenomenon by simulating the interactions of the many actors in such systems. Thus gaining insights that will lead to greater understanding and, in some cases, better management of the behavior of complex social systems [4].

It is reported that in some cases simulation results are significantly different depending on simulation scale [2]. Massive ABMs platform is required for building model of large social systems.

In this article, it is introduced to promote large ABM tool.

This article introduces about implementation of agent model on Repast HPC, ABMs toolkit for large scale.

This article's aim is to detect bottleneck of simulation through implementation of simple model that is extension of Schelling segregation model using geo-spatial data.

The rest of this article is organized as follows. Sec.2 shows detail of Repast HPC framework, Sec.3 introduces related works for building models on Repast HPC and some toolkit for large scale ABMs. Sec.4 shows the implementation of the model. Sec.5 describes an analysis of simulation. In Sec6, we conclude this paper with future works.

2. Repast HPC

Repast High Performance Computing (Repast HPC) [5] is an agent based modeling and simulation framework for high performance distributed computing platforms written in C++ and using MPI for parallel operations. Most of the other open source ABM software does not support parallelization (See Sec.3.). Repast HPC is designed for parallel environments where many processes are running in parallel and where the agents themselves are distributed across processes. In addition of being a parallel platform, the fact that Repast HPC supports parallelizing the simulation world at prominent modeling methodologies (shared grid, shared network and shared continuous space) independently or in combination, makes it excellent software for high performance computing.

The functions specific to the HPC environment deal with the theoretical challenge of sharing agent information across multiple processes. Repast HPC deals with this through a two-stage process. In the first, agents are 'requested' from other processes; the process that manages the agent will answer a request by providing the other process with a copy of the original agent. In the second, these copies are updated with the borrowed agents' current information. Fig.1 shows an example of sharing agent information across multiple processes. The effect is that an agent on a given process can derive information from and react to the current state of agents on any other process. In the simulation presented here the 'request' for other agents is explicitly scheduled, Repast HPC also provides a way for spatially located and moving agents to interact with agents near themselves in simulation space but across a process boundary; when this functionality is used both the request and synchronization are handled automatically. Fig.2 shows an example**.**
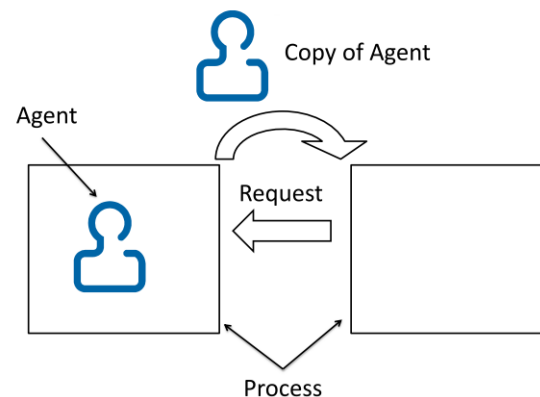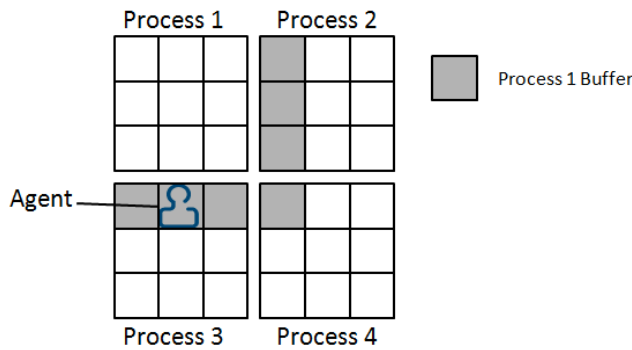

Fig. 1 Requesting Agents Across Processes

Fig.2 Synchronized Buffer

3. Related Work

In this section, ABM tool kit and the study case about the large scale model using Repast HPC is introduced.

3.1 ABM Toolkit

Many ABM simulation frameworks have been developed and shared among research communities. There are some reports that review ABM tools [6] [7]. Swarm [8] is the first ABM software, launched in 1994 at the Santa Fe Institute. Swarm supports hierarchical modeling approaches whereby agents can be composed of swarms of other agents in nested structures. Swarm provides object oriented libraries of reusable components for building models and analyzing, displaying, and controlling experiments on those models. MASON [9] is a single-process discrete-event simulation core and visualization library written in Java, designed to be flexible enough to be used for a wide range of simple simulations, but with a special emphasis on swarm multi-agent simulations of many (up to millions) of agents. It provides a toolkit to execute high speed simulations with many agents in a single process. ZASE (Zillions of Agents-based Simulation Environment) [Yamamoto 08] is a scalable platform for multi-agent simulations potentially using billions of agents. The purpose of ZASE is to develop ABM applications on multiple general computers such as PCs or workstations where they are connected with a high performance but generally available network such as gigabit Ethernet. ZASE aims at performing even larger-scale simulations by managing many agents in each process and by combining the processes hierarchically. SOARS (Spot Oriented Agent Role Simulator) [10] aims to support a PC grid. It provides a common platform for education and real world applications, and it is easy to develop simulation models with its GUI and also to visualize its simulation results.

3.2 Study Case Using Repast HPC

Only a few examples of Repast HPC are available because it is a new toolkit for ABM, first version has been released in 2010.

It is expected that more examples of multi agent simulation on Repast HPC are given in future.

In [1], an example model called "Triangles simulation" in Repast HPC is shown and run at scales up to billions of agents. And practical issues of performance, network analysis, file output and data visualization are illustrated with example. It is reported that execution time scales almost linearly with the number of cores for mainly independent agents, whereas speed-up drops significantly in case of highly interdependent agents.

Recently, ABMs are seen as attractive approach to analysis Computational Social Science (CSS). In [11], models of configurations of social agents at a massive scale are introduced. The paper reports implementation of Cellular Automata (CA) for spatial model, to use urban evacuation analysis.

4. Implementation

In this article, to satisfy three requirements, 1) interaction between agents, 2) agent's movement, 3) agent placement using geographic information, "spatial model" that is extension of Schelling Segregation Model using geospatial information is implemented.

Segregation model, proposed by Thomas Schelling [12], attempts at understanding the phenomenon of residential segregation by considering it as an aggregated result of the decisions of residents in choosing their housings. Residents are represented explicitly as agents in the model. Residents having certain similarities are classified as belonging to the same class. In reality, criteria used to classify resident may be educational level, religion, annual income, skin color, political point of view, etc. Same class residents are represented by same color agents. When the simulation runs, we see groups of nearby same color residents appear. We call these groups emerging structures because they are not explicitly represented in the model as agents. As soon as these emerging structures appear in the simulation, isolated residents tend to be attracted to them. Isolated residents move to join group of other residents similar to them. The forming of such emerging structures is the result of residents' decision in choosing places where they live in a city. Vice-versa, these emerging structures have certain feedback influences on the behaviors of the residents. Fig.3 shows an example of Segregation Model. In this figure, circle and cross represent same class residents.

In the original model, agents are located on 8×8 grid. In this model, each cell represents "agent's residents". In the suggested model, residents are placed along the road, which is continuous space extracted from GIS. It is based on an assumption that typical housing is located along the road.

In spatial model, geospatial data used as GIS is provided by OpenStreetMap (OSM) [13]. OSM is an open source project for creating a free editable map, and geographic information can be obtained easily. OSM provides a file named "planet.osm", an xml format file. This file includes in one file; all the nodes, ways and relations made up by OSM community members.

In this model, road information is extracted from planet.osm and used to make residents. Information about road is explained by "node" elements and "way" elements in osm file. A node defines a single geospatial point using a latitude and longitude. In many case, a node represents a crosspoint. Each node stores its node-ID, longitude and latitude. A way is an ordered list of between 2 and 2000 nodes. A way element describes road, e.g. street, railway and highway. Ways can also represent areas, such as buildings or forests. In this case, the first and last node will be the same - a "closed way". Fig.4 describes each element. Road information is can be obtained by extracting "node" and "way" elements from osm file.

Fig.5 Conversion map data to 100 × 100 grid
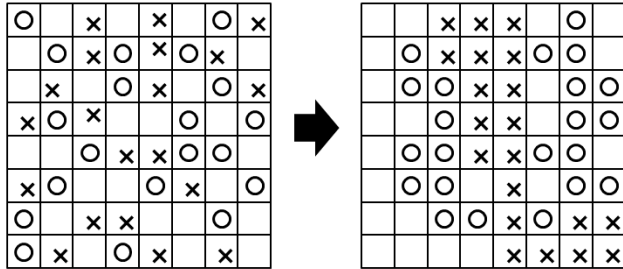
Fig.3 Schelling Segregation Model

Fig.4 Elements in osm file

Fig.6 Definition of neighbor agent on suggested model

Fig.7 Spatial model
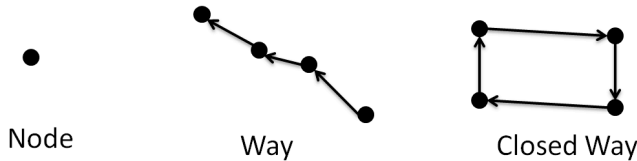
5. Simulation and Analysis

In this article, simulation is processed on Japanese cities to test the influence of the geographic information on the simulation. Table shows a list of cities simulated in this article. Four rows, left, bottom, right and top represent longitude and latitude of extracted city. Number of residents depends on total road distance. In the simulation, agents are separated into two groups. Each group consists of 500,000 agents as enough number of large-scale agents. Agents are given their residential randomly when simulation starts. If an agent is in an "unhappy" situation, it starts migration to empty chosen randomly. The simulation continues until agents are happy, and this represents the simulation time distance between new residence and old one. In this simulation, it cost agent 1 simulation time per 5km as real world.

In the original model, an agent refers neighbor agents and decides migration. In the original model, "neighbor" is defined 4cells (Von Neumann Neighborhood in Cellular Automata). In the spatial model, new search method is needed to search continuous space. In this article, as referred Fig.5, city map is converted to 100 × 100 grids and "neighbor" is defined agents on same cells. Fig.6 shows an example. A small example of simulation is shown in Fig.7.

The part of city data is obtained from planet.osm via Osmosis, a command line Java application for processing OSM data. This can extract a rectangle. An example of command extracting the part of data from osm file is described below.

```
$ bzcat planet.osm.bz2 | osmosis --rx - --bb left=141.2035
   bottom=42.9469 right=141.4754 top=43.1413 --wx sapporo-city.osm
```

In this article, ratio that an agent hopes that neighbor agents are same group is called "preference". Simulation is processed on 3 preferences, 30%, 40% and 50% to investigate the relation between preference and execution steps. On each simulation, all agents have same preferences. The graph of the simulation steps for number of "unhappy" agents is shown Fig.8. Simulation cost many steps as preferences increases. Fig.9 shows the graph of the simulation steps for average of moving day of agents. We can see that average of moving day increases as map size is larger.
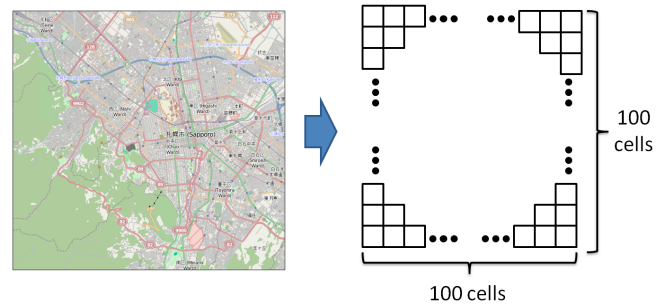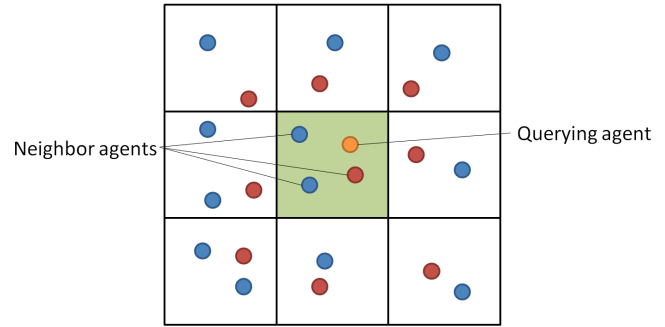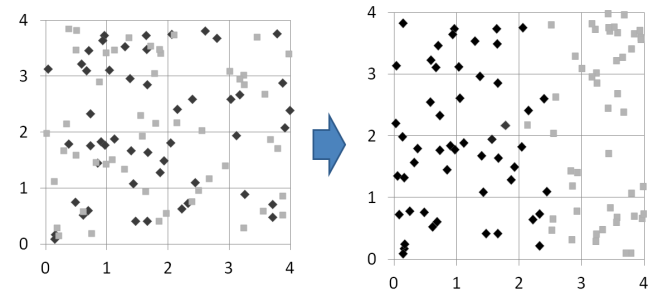
Execution time of simulation is discussed as below. To detect bottleneck of simulation, Simulation step is mainly divided into two phases, initial phase and running phase. Initial phase is a phase initializing simulation, creating agent objects and locating agents into grid or space. Running phase is a main part of the simulation, iteration of searching neighbor agents and migration to other location. Fig.10 and Fig.11 show transition of execution time. Each graph shows execution time increases as number of agents increases. Initial phase is more divided in detail because it costs longer time than run phase. It is shown that a part for giving agents coordinates and locating on grid or space takes most execution time. In this part, Repast HPC API is used. The API is repast::Grid, functions for discrete grids and continuous spaces. It is expected that execution is more efficient by using C++ standard library instead of API.

6. Conclusion

In this article, we verified about large scale ABM using epast HPC which is ABM toolkit. We constructed segregation model which is added the geospatial information in the real world to the Schelling Segregation Model, and in an agent's migration, we added the concept of move cost according to distance. In order to check influence that a geospatial situation has on a simulation, we simulated 1 million agents using the map data which is the city in Japan. We also divided the simulation into two processing parts, and then measured execution time of each division to specify the bottleneck of processing time that is caused by large-scale simulation.

City	Sapporo	Sendai	Niigata	Yokohama	Kyoto
Left	141.2035	140.6595	138.9273	139.5094	135.6496
Bottom	42.9469	38.2192	37.7020	35.3741	34.8859
Right	141.4754	141.0131	139.3873	139.7226	135.8823
Top	43.1413	38.3571	38.0440	35.5136	35.0744
Area()	479.5	474.7	1542.5	300.5	445.9
Number of Residences	1,698,750	1,445,642	1,904,392	1,421,162	1,935,092

Table. 1 List of Simulated Cites

(a) preference 30% (b) preference 40% (c) preference50%
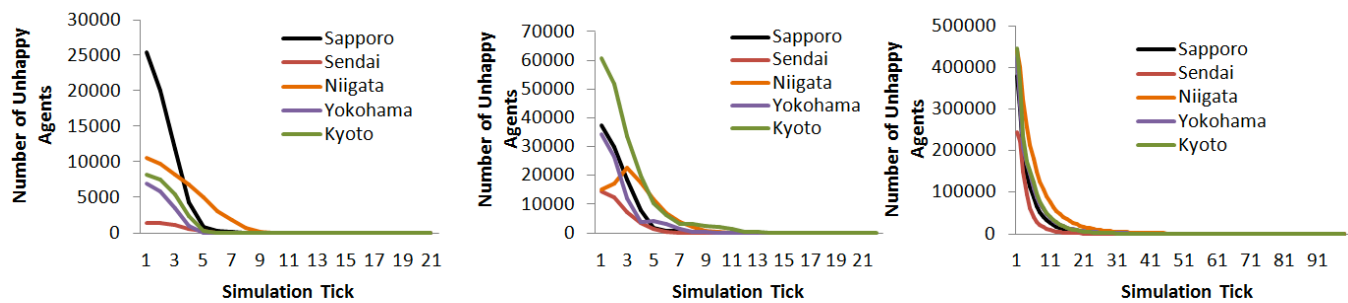
Fig.8 Transition of "unhappy" agents

(a)preference30% (b) preference40% (c) preference50%
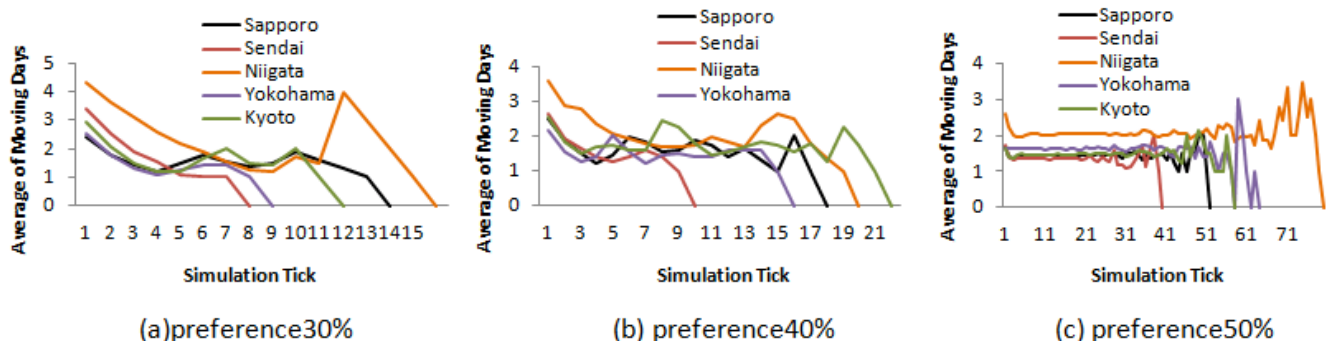
Fig.9 Transition of average of moving day
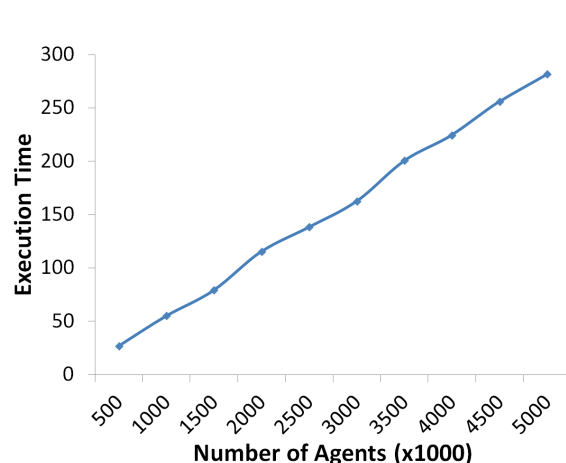
Fig.10 Transition of initial phase
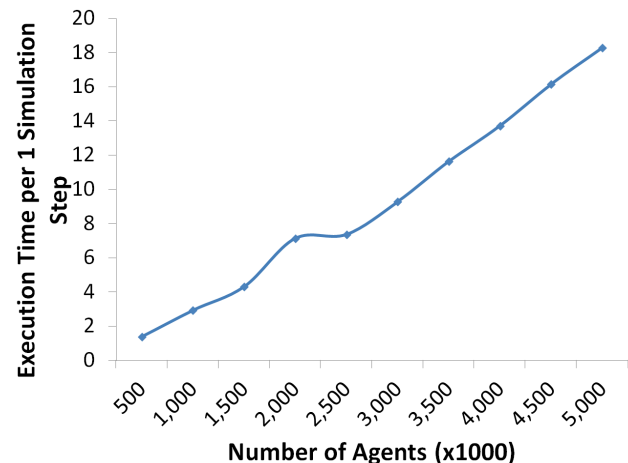
Fig.11 Transition of run phase

265

7. REFERENCES

[1] John T. Murphy, Computational Social Science and High Performance Computing, A Case Study of a Simple Model at Large Scales, The Computational Social Science Society of the Americas, (2011)

[2] Gaku Yamamoto, Hideki Tai, and Hideyuki Mizuta, A Platform for Massive Agent-Based Simulation and Its Evaluation, Massively Multi-Agent Technology, pp. 1 - 12, (2008)

[3] Yuu Nakajima, Shohei Yamane, and Hiromitsu Hattori, Multi-model Based Simulation Platform for Urban Traffic Simulation, Principles and Practice of Multi-Agent Systems, pp. 228 - 241, (2012)

[4] A.J.Heppenstall, A.T.Crooks, L.M.See & M.Batty, The integration of agent-based modelling and geographical information for geospatial simulation, Agent- based models of geographical systems, pp. 219 - 252, (2012)

[5] http://sourceforge.net/projects/repast/files/Repast \\HPC/Repast HPC 1.0.1/repasthpc-1.0.1.tar.gz/\\download

[6] Steven F. Railsback, Agent-based Simulation Platforms - Review and Development Recommendations, Fig.10 Transition of initial phase   Society for Computer Simulation International, Vol. 82, Issue 9, pp. 609 - 623, (2006)

[7] CM Macal, MJ North, Agent-based modeling and simulation, WSC '09 Winter Simulation Conference, pp. 96 -98, (2009)

[8] Nelson Minar, Roger Burkhart, Chris Langton, The swarm simulation system, A toolkit for building multi-agent simulations, pp. 96-06-42, (1996)

[9] Sean Luke, Claudio Cioffi-revilla, Liviu Panait, Keith Sullivan, MASON, A Multiagent Simulation Environment, Simulation, Vol. 81, No. 7, (2005)

[10] Hideki Tanuma, Hiroshi Deguchi, Tetsuo Shimizu, SOARS, Spot Oriented Agent Role Simulator- Design and Implementation, Agent-Based Social Systems, Vol.1, pp. 1-15, (2005)

[11] Zia, K. Riener, A. Farrahi, K. Ferscha, A New Opportunity to Urban Evacuation Analysis - Very Large Scale Simulations of Social Agent Systems in Repast HPC, Principles of Advanced and Distributed Simulation (PADS), 2012 ACM/IEEE/SCS 26th Workshop on, pp. 233 - 242, (2012)

[12] Thomas C. Schelling, Dynamic Models of Segregation, The Journal of Mathematical Sociology, Volume 1, Issue 2, (1971)

[13] http://www.openstreetmap.org/

Transport System for Predicting Occupancy of Parking Lots on the Highway Network

Martin SROTYR, Michal JERABEK, Zdenek LOKAJ, Tomas ZELINKA, Petr KUMPOST
Faculty of Transportation Sciences, Czech Technical University in Prague
Konviktska 293/20, 110 00 Prague 1, Czech Republic
srotymar@fd.cvut.cz, jerabem1@fd.cvut.cz, lokajzde@fd.cvut.cz, zelintom@fd.cvut.cz, kumpopet@fd.cvut.cz

ABSTRACT

More and more often occurs problem with the lack of parking places for trucks over 3.5t, which is related to the increase in the intensity of freight transport. Consequently this can lead to very dangerous situations with parking of trucks because drivers do not have information about the occupancy of the nearest parking. The paper presents an intelligent transportation system, which provides information about the predicted occupancy of parking lots to truck drivers. Providing this information leads to optimizing the use of existing parking areas on the highway network and also makes it easier for drivers to deciding on a suitable location for parking, which ultimately contributes to the fluency and safety of traffic.

Keywords: Electronic Toll, ITS, Parking, Prediction, Telematics.

1. INTRODUCTION

With the lack of parking spaces for heavy goods vehicles face the entire road network in the Czech Republic. The worst situation is on highways, where we often encounter dangerous situations parked truck right onto the ramp to gas stations or their exits. This situation is similar throughout Europe; the importance of this issue is demonstrated by the Action Plan for the Deployment of Intelligent Transport Systems in Europe, which contains measures "Development of appropriate measures including best practice guidelines on secure parking places for trucks and commercial vehicles and instructions for parking and reservation telematics systems". The high traffic intensity raises a number of transportation problems related to the lack of parking spaces, and therefore is accurate and accessible information about available parking spots very valuable in everyday use and also very important for the planning of efficient and safe transport.

The project, which partial result describes this article, is focused on creating a telematics system that will be based on the input data of the toll system to predict the occupancy of individual parking spaces to provide information to optimize the usage of existing parking areas on the highway network. The outputs of the model will be forwarded to appropriate channels for drivers who thereby greatly facilitate decisions about the appropriate place for parking and thus contribute to the overall fluency and traffic safety as a whole.

Reasons for system of increasing the usage of parking facilities for trucks on the highway network in the Czech Republic using prediction models are especially:

- increasing the intensity of freight transport on the highway network in the Czech Republic, for the reason that road freight transport is still cheaper than the railway, resulting in the observed lack of capacity
- 30-40% of road accidents are caused by driver fatigue, due to the extending of driving and ignoring the mandatory rest breaks

- priorities of the EU, increasing parking capacity on highways handles projects as CONNECT, Easyway, Intelligent truck parking, etc.
- inefficient use of parking lots
- uneven distribution of load of parking lots
- driving round highways and finding a parking space, which means the formation of undesirable emissions, congestion on local roads, noise and hazards caused by parking in places that are not intended to (emergency stopping lanes, ramps, etc.)
- lack of information on capacity and current vacancies in the parking areas
- determining the need for construction of new parking lots
- increased demand for parking due to the Just-in-Time
- reduction of investment in transport infrastructure construction, which increases the demand for optimizing the use of existing infrastructure

The expected benefits of implementing an information system to increase parking utilization capacity are divided by system participants, because for each group are different.

The Benefits of implementing the system from the perspective of truck drivers

- knowledge of the current traffic situation before entering to the parking lot
- reduce the risk of stressful situations for drivers due to pressure on compliance with statutory breaks and thus enhance the security (reducing the risk of accidents due to fatigue during driving overtime)
- increased comfort while driving
- higher probability of finding a free parking place - eliminates the need for finding alternative places
- more effective route planning associated with savings of distance traveled and time

The Benefits of implementing the system from the perspective of car park operators

- increase the satisfaction and comfort of drivers
- optimized parking and better use of existing parking capacity
- elimination of the problem of exceeding the capacity of parking areas
- an increase in revenues from products and services offered on the parking lots due to optimized parking
- optimization planning - supplying, staffing shifts, etc.

The Benefits of implementing the system from the perspective of the State

- reduce the externalities arising from traffic accidents, ie. disposal costs, damages, costs for emergency operations, the cost of treating injuries, reducing state revenues due to the loss of the taxpayer (in the case of death)
- reduce the externalities arising due to leaving the highway network (emissions, noise, damage to roads)
- financial savings for the construction and maintenance of highway infrastructure, which arise due to fewer

required newly built parking lots and also because there is no need to look for alternative parking outside the toll network

- achieving the objectives of the European "TEN-T" program by using telematics systems
- use of the toll system as a source of current traffic data (composition and characteristics of traffic flow) for the prediction model
- increase traffic safety
- reduce accidents
- increase driver satisfaction
- increase the usage of parking areas, and thus the potential increase in tax revenues (the driver will use more parking at us and there will be motivated to refuel, buy goods and services)

The Benefits of implementing the system from the perspective of roads administrator

- effective planning of maintenance of highways infrastructure
- effective planning the construction of parking areas on the highway network
- reduction of required maintenance and the associated costs on roads close to the highway network
- reducing the number of traffic restrictions associated with the construction and maintenance of new parking areas

2. ARCHITECTURE OF THE INTELLIGENT TRANSPORT SYSTEM FOR PREDICTING OCCUPANCY

Architecture of the telematics system for predicting the occupancy of parking spaces on highways and expressways in the Czech Republic connects the database, prediction model and distribution channels for the providing information to end users.
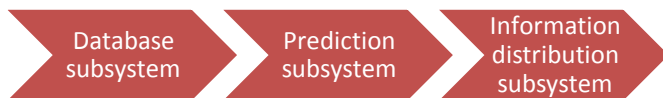
Fig. 1 Architecture of the system for predicting occupancy

Architecture of the telematics system consists of three basic subsystems:

- database subsystem - data primarily consists of transactional data from electronic toll collection system, which are provided on-line (or at intervals close to on-line transmission) to prediction subsystem;
- prediction subsystem – the core of this subsystem consists of a predictive model that calculates by defined algorithms the probable occupancy of parking spaces on the highway network based on the data
- Information distribution subsystem – this subsystem uses the results of prediction model that processes and distributes intelligible information to end users.

A. Database Subsystem

Subsystem of database combines data from data sources, both historical data and current data. This data are stored in the operational database in which the data are preprocessed so those with them are able to run a prediction model to predict the occupancy of parking

lots on the highway network in the Czech Republic. Especially it is the calculation of driving time each detected vehicle and calculates intensity in each toll section.

Data base draws data from the following sources.

- the primary source are transactional data from electronic toll collection system, namely: a) historical data, which are recorded to operational database one-off; b) actual data, which are sent to the production database on-line c) telematics data (speed and composition of traffic flow), which are sent to the production database in batches;
- secondary sources are a) data from detection sensors located at the entrance and exit of the selected parking lot, used to calibrate the prediction model, these data are sent to the production database on-line as an individual transactions and batch for a defined time intervals, b) data from manual vehicle census at selected parking lots serving the calibration and validation of the prediction model; c) data from automatic vehicle census at selected parking lots from mobile surveillance vehicle of toll collection system; d) other traffic data, e.g. data from traffic sensors on highways etc.
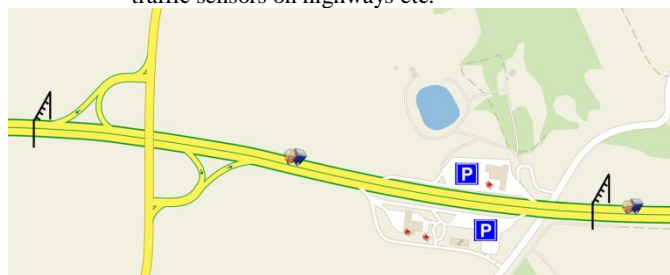

Fig. 2 Simplified diagram of the highway network with the location of parking areas, toll gates, entrance and exit ramps

B. Prediction Subsystem

The prediction subsystem is the core of the entire system. It is connected to the subsystem of database over which implements the defined prediction algorithms.

The predictions are calculated in systems derived on the basis of fully formalized statistical methodology. At the core of the whole approach is the fact that value which is needed to predict, i.e. the number of free parking spaces, is directly unobservable. The prediction of this value should be constructed on the basis of correct algorithms derived from models for time series of observable variables. Basically, it is an estimate of the status-type model, which we deal with in a few steps. First, we construct the proxy variable for the latent number of vehicles parked on the parking lot at a time from the observed data on the passage of vehicles through toll gates. For the proxy variable we build a statistical model completely describing its dynamic and probabilistic behavior. Unknown parameters of the model of a particular class we estimate from the historical data from a passage through the toll gates. After more statistically demanding estimate the unknown parameters of the data (identification), the model has been fixed and used for routine, already simply available, predictions with parameters fixed at estimated values. With extended operation will be appropriate in the context of the entire system service (e.g. annually) to "re-learn" model, i.e. to re-estimate the parameters, or modify (or choose an entirely new) class model. According to our previous tests the stability of the estimated parameters using several months of data is significant and re-learning is not a critical requirement. From the estimated model we derive using the theory of Markov chains (Resnick, 1992 Klebaner, 2005) method for calculating the prediction of the number of vehicles parked on the parking lot for a set of necessary horizons. For fast and efficient implementation of predictive calculations in real traffic, we use a method based on Monte Carlo simulations. Prediction of the

number of free parking spaces is obtained as the difference between extended parking capacity and prediction of parked vehicles, and for each of the required horizons. Prediction of the number of free parking spaces is using the subsystem distribution information transmitted to end users. Solution of the prediction system has been described in [1], [2] and [3].

The prediction subsystem is implemented in the first version of the telematics system as a script in the statistical software "R". This script implements algorithms for the minimum and maximum prediction horizons from 5 minutes to 150 minutes with a time step of 5 minutes.

Prediction process is performed at regular intervals and the results of the process are recorded in the table in the operational database where they are available to end users through distribution channels.

For each parking lot is available the information about the GPS position, the maximum capacity, detail of gas station or other information. This information is static and changes only when the layout changes such as reconstruction. For each parking lot are regularly at 10 minute intervals calculated prediction of the state of occupancy in several prediction horizon. In the functional sample we predict for the four parking lots on the highway D5. The prediction subsystem additionally maintains the entire history of the calculations of predicted occupancy for the all monitored parking lots. Historical data are another source of verification and validation of the prediction model.
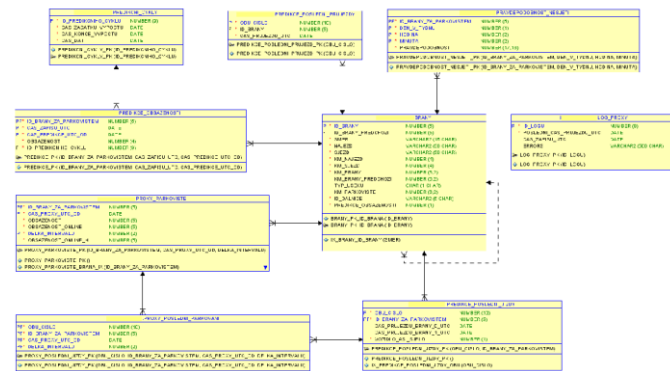
Fig. 3 Relational schema of the production database

Table of occupancy prediction for the parking lot contains the following parameters:

- ID of parking lot
- date and time of update
- prediction horizons – from a computational reasons the prediction is calculated for several horizons (windows), not continuously for each time point. The maximum prediction horizon is 150 minutes. Horizons are divided in 5 minutes and counting since the last update time,
- predicted occupancy status – for each forecast horizon is on the basis of the developed model calculated the predicted occupancy. Aggregate channel then assign a specified color according to the following table (Tab.1) to predicted occupancy.

Occupancy [%]	color
<0,60)	green
<60,80)	yellow
<80,+∞)	red

Tab. 1 Occupancy intervals and their color for drivers

The last interval is really up to infinity (better expressed greater than 100%) because the drivers are able to park in parking lots more vehicles than there are parking (marked) spaces.

C. Information Distribution Subsystem

Subsystem of distribution of information is from the end user perspective crucial component of the entire system. Distribution channels are the only interface between the telematics system and end-user, i.e. truck drivers. Selecting the appropriate information channel therefore fundamentally affects the efficiency of the entire system and the acceptability of the system by the drivers. Nowadays there are more and more options to display and distribute information.

For intelligent parking system is necessary to resolve the distribution of information channels that are suitable for this type of system. Distribution channel must be selected properly to be effective for distributing information and also for end users, truck drivers, acceptable.

3. PROTOTYPE OF A MOBILE APPLICATION

The developed prototype of mobile application communicates with the communication layer of prediction system and displays the acquired prediction of the parking lots in the driving direction, as shown in the following figure.

The application is adjusted so that it can be tested without the presence on the D5 highway, thanks to manually input the current position on the highway D5 and average vehicle speeds, which are input parameters for prediction of occupancy parking lots along the route.

Fig. 4 An example of the mobile application in the vehicle

Data exchange between applications and production database is based on XML document.

Fig. 5 An example of the request and response of mobile application
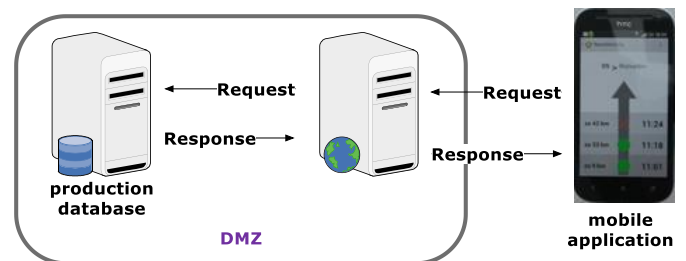
For example request for list of highways:
```
<?xml version="1.0" encoding="UTF-8"?>
<S:Envelope xmlns:S="http://schemas.xmlsoap.org/soap/envelope/">
  <S:Header/>
  <S:Body>
   <Request method="getHighwayList"/>
  </S:Body>
</S:Envelope>
```

Response with list of highways:
```
<?xml version="1.0" encoding="UTF-8"?>
<S:Envelope xmlns:S="http://schemas.xmlsoap.org/soap/envelope/">
  <S:Header/>
  <S:Body>
   <Response method="getHighwayList">
     <HighwayList>
       < Highway>
         < HighwayName>D5</ HighwayName>
         <HighwayFrom>Praha</HighwayFrom>
         <HighwayTo>Rozvadov</HighwayTo>
         <HighwayLength>151</HighwayLength>
       </Highway>
     </HighwayList>
   </Response>
  </S:Body>
</S:Envelope>
```

4. Key Performance Indicators

System key performance indicators dedicated to quantify occupancy prediction can be alienated into two categories – the operational and the functional category.

The operational performance indicators indicate the characteristics on the operation of the system. The functional performance indicators indicate characteristics of system functions provisioning with regard to the purpose of predicting occupancy.

The operating category includes the following indicators of performance:

- failure time
- reliability of operation
- failure rate (instability)
- activation time of availability

The functional category includes the following indicators of performance:

- rate of erroneous prediction of occupancy
- duration of response
- stability of serviceability
- loss rates
- safety

Most of those indicators have been described in [4], however, for this specific application for the prediction of occupancy of parking is important to quantify and define quality of prediction. For this purpose, it has been defined key performance indicators "Rate of Erroneous Prediction of Occupancy".

A. Rate of Erroneous Prediction of Occupancy

The rate of erroneous prediction of occupancy reflects the ability of telematics system to don't serve wrong prediction of occupancy to a maximum allowable limit that can be defined as the probability

$$P\left(\left(r_{e,\langle 0,T\rangle} - R_{\langle 0,T\rangle}\right) \le \varepsilon_e\right) > \gamma_e \ , \tag{1}$$

where the difference between the number of wrong prediction of occupancy $r_{e,\langle 0,T\rangle}$ and tolerated maximum number of wrong prediction of occupancy $R_{\langle 0,T\rangle}$ do not exceed the value ε_e on the probability level γ_e within time interval $\langle 0, T\rangle$.

5. Conclusions

The proposed telematics system, designed to provide information for truck drivers about the predicted occupancy for the nearest parking, was presented. For the verification of the quality of the prediction model data obtained by manual measurement in the parking lots were exclusively used. All the other tested methods like Doppler based or DSRC based mobile units have not been acceptable due to their insufficient for this application system performance. The quality of the prediction has been evaluated based on in paper defined performance indicator "Rate of Erroneous Prediction of Occupancy". The present phase of the project is being focused on the final validation and verification of the prediction model.

6. Acknowledgment

7. References

[1] M. Brabec, O. Konar, I. Kasanicky, E. Pelikan, M. Maly, Z. Lokaj, T. Zelinka: Forecasting system for truck parking based on statistical modeling of indirect, easily available data, **Proceedings of the 11th European Transport Congress**. Faculty of Transportation, Czech Technical University in Prague, Praha, pp. 26-28, 2013, ISBN 978-80-01-05416-1

[2] M. Brabec: Statistical metodology for truck parking modeling and prediction. TRANSBALTICA 2013, Vilnius, Lihuania, 2013, ISBN 978-609-457-470-2, ISSN 2029-2384

[3] M. Brabec, O. Konar, I. Kasanicky, E. Pelikan, M. Maly: Statistical modeling of truck parking lots occupancy, MOSATT 2013, Zlatá Idka, Slovakia

[4] T. Zelinka, M. Jerabek, Z. Lokaj: Authentication and Data Security in ITS Telecommunications Solutions. In **Proceedings of the 17th World Multi-Conference on Systemics, Cybernetics and Informatics**. Orlando: IIIS - International Institute of Informatics and Systemics, 2013, p. 295-300. ISBN 978-1-936338-87-0.

NEW OPTOELECTRONIC TECHNOLOGY SIMPLIFIED FOR ORGANIC LIGHT EMITTING DIODE (OLED)

Andre F. S. Guedes[1], Vilmar P. Guedes[2], Simone Tartari[2], Mônica L. Souza[2] and Idaulo J. Cunha[3]

[1]Image Processing Laboratory, SENAI College of Technology - 8450, Assis Brasil Avenue - Zip Code 91140000, Porto Alegre, Rio Grande do Sul, Brazil.

[2]Nanotechnology Laboratory, Intellectos - 301, Sao Miguel Avenue - Zip Code 95560000, Torres, Rio Grande do Sul, Brazil.

[3]Department of Production Engineering, UFSC, Campus Trindade, Zip Code 88040-970, Florianopolis, Santa Catarina, Brazil.

ABSTRACT

The development of Organic Light Emitting Diode (OLED), using an optically transparent substrate material and organic semiconductor materials, has been widely utilized by the electronic industry when producing new technological products. The OLED are the base Poly (3,4-ethylenedioxythiophene), PEDOT, and Polyaniline, PANI, were deposited in Indium Tin Oxide, ITO, and characterized by UV-Visible Spectroscopy (UV-Vis), Optical Parameters (OP) and Scanning Electron Microscopy (SEM). In addition, the thin film obtained by the deposition of PANI, prepared in perchloric acid solution, was identified through PANI-X1. The result obtained by UV-Vis has demonstrated that the Quartz/ITO/PEDOT/PANI-X1 layer does not have displacement of absorption for wavelengths greaters after spin-coating and electrodeposition. Thus, the spectral irradiance of the OLED informed the irradiance of 100 W/m^2, and this result, compared with the standard Light Emitting Diode (LED), has indicated that the OLED has higher irradiance. After 1000 hours of electrical OLED tests, the appearance of nanoparticles visible for images by SEM, to the migration process of organic semiconductor materials, was present, then. Still, similar to the phenomenon of electromigration observed in connections and interconnections of microelectronic devices, the results have revealed a new mechanism of migration, which raises the passage of electric current in OLED.

Keywords: organic materials, optoelectronic technology, OLED, organic electronic technology.

INTRODUCTION

In 1977, a new chapter in the evolution of organic semiconductor materials began, when Hideki Shirakawa, from the University of Tsukuba (Japan), Alan Macdiarmird, from the University of Pennsylvania, and Alan J. Heeger from the University of Santa Barbara (The U.S.), demonstrated the existence of conductive properties for doped Polyacetylene (intrinsically an insulator), which assured them the Nobel Prize in Chemistry, in 2000 [1][2]. Research studies, realized to synthesize and characterize the components of this new class of materials, have been currently passing through a continuous process of technological advance in search of new conductive polymers. In addition, the greatest interest in studying organic semiconductor materials has been present in its potential applications, among which are: lightweight batteries, gas sensors, electrochromic devices, capacitors, electrochemical cells, organic solar cells, organic light emitting diodes (OLED), shielding of electromagnetic radiation, artificial muscles, passivation of integrated circuits MOSFET, pn junction, satellites weighing less than 0.2 kg, and others [3].

Thin films made of organic semiconductor materials have drawn the attention of research groups, due to its enormous potential for its application in various industries as well as the impact their results can give to technological development. Moreover, the characterization and application of new techniques for the deposition of organic semiconductor materials has been the main focus of a large number of studies, being, thus, essential in reducing production costs. Furthermore, OLEDs can be utilized for large and small areas of flat panel flexible self-luminous displays, in many consumer products. Organic light emitting diodes are electronic devices made

by placing a thin film of an electroluminescent organic material between two conductors with different work functions. When an electrical voltage is applied, electrons and holes are injected into the electroluminescent material, and, when they are recombined, light is emitted [4][5]. This research aims to develop thin films electroluminescent organic material for OLED, by using the Electrodeposition technique, and characterizing the layers and devices utilizing UV-Visible Spectroscopy (UV-Vis), Optical Parameters (OP) and Scanning Electron Microscopy (SEM) techniques.

EXPERIMENTAL DETAILS

The OLEDs, developed in this research, have been utilizing optically transparent material, covered on surface with Indium Tin Oxide (ITO), as a substrate. The ITO has high conductivity and transmittance in the visible region of the electromagnetic spectrum, which enables their utilization, for instance, in organic solar cells, organic gas sensors, organic transistors and electrochromic devices [6].

The grass substrate, based on quartz, and utilized for the spin-coating of Poly (3,4-ethylenedioxythiophene), PEDOT, with thickness of 200 nm, was covered by a layer of ITO, with a thickness of 400 nm. The PEDOT layer was deposited through the spin-coating technique, adding from 100 to 100 µL at 750 rpm for ten seconds in each deposition. On these layers, an electroluminescent material, corresponding to the active layer, has been deposited. This layer was deposited through the Electrodeposition of PANI-X1 solution prepared with perchloric acid (HClO$_4$), and applying voltage of 10,0 Volts for 5 minutes, resulted in an active layer of PANI-X1, with thickness between 180 nm and 220 nm. The electrodeposition of PANI in acid solvent on the ITO layer originated the active layer of PANI-X1. The metal contact, utilized on the last layer of this device, is made of aluminum, with thickness of 400 nm, and it was deposited through the sputtering system.

The developed OLED device has presented a configuration layered of Quartz/ITO/PEDOT/ PANI-X1/Al, as represented schematically in figure 1.
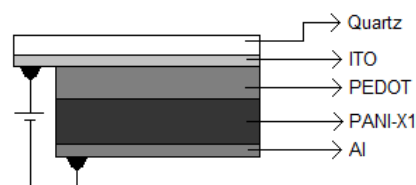
Figure 1. Configuration of OLED.

DISCUSSION

The analysis of a sample through the ultraviolet-visible spectroscopy has been the result in a spectrum of light, obtained by a graph of wavelength, or frequency versus the intensity of absorption (absorbance or transmittance) [7] [8]. The absorption spectrum in the spectral region of 200-1100 nm, of the OLED, has been shown in Figure 2. This result has indicated that there is no increase in the absorbance with the spin-coating technique and electrodeposition of organic materials.
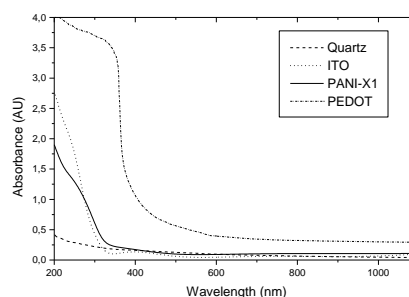
Figure 2. UV-Vis Spectroscopy of OLED.

The spectral irradiance of the OLED has been shown in Figure 3. The irradiance of this OLED is of 100 W/m^2. This optical parameter has indicated that the OLED operates in a wavelength region of 450-650 nm. This result, compared with the standard Light Emitting Diode (LED), has indicated that the OLED has higher irradiance.
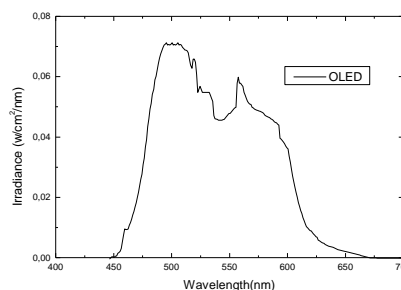

Figure 3. Spectral Irradiance of OLED

The microscopic analysis performed in OLED by Scanning Electron Microscopy, has allowed us to observe the induction on the surface of the samples before and after the application of voltage. Figure 4 has shown the micrograph of OLED before applying voltage. It may be observed that the surface of the sample has some homogeneous aspect, plane and without changes.
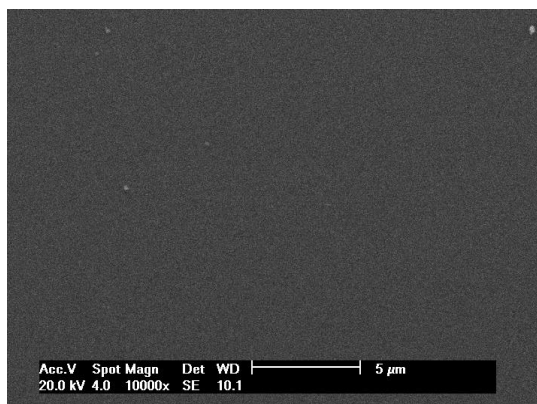
Figure 4. SEM of OLED before applying voltage.

Figure 5 has shown the micrograph of OLED after application of 5.0 volts during 1000 hours. In this micrograph, it may be observed the surface of the sample in some irregular aspect, with holes and cracks. In addition, the formation of nanoparticles due to the migration of the organic semiconductor material can be observed. Furthermore, an increase of time, when utilizing the device, and an increase in the application of voltage, may be observed as well as the existence of an increase in the formation of nanoparticles. Therefore, these nanoparticles can contribute to reducing the lifetime of the OLED, since the accumulation of organic semiconductor material may cause short circuit in the device.
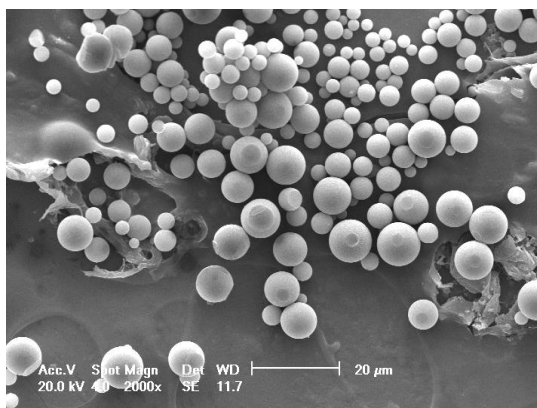
Figure 5. SEM of OLED after applying voltage.

CONCLUSIONS

The results presented in this research are that PEDOT/PANI-X1 can be utilized as an active layer of OLED. These organic semiconductor materials can be deposited, spin-coating and electrodeposition, respectively. The absorption spectrum in the spectral region of 200-1100 nm of the OLED, has indicated that there is no increase in absorbance with the utilization of these deposition techniques. The spectral irradiance of the OLED informed the irradiance of 100 W/m^2. This result, compared with the standard Light Emitting Diode, (LED) has indicated that the OLED has higher irradiance. The new phenomenon of migration, observed in OLED, demonstrated similarities with that observed in microelectronic devices. In OLED devices this new migration mechanism that rises with the passage of electric current can influence the lifetime reduction of devices.

ACKNOWLEDGMENTS

REFERENCES

[1] A. G. MacDiarmid, *Current applied physics* 1, p.269-279 (2001).

[2] H. Shirakawa et alii, *Chem. Soc.,* **99**, 578 (1977).

[3] H. Varela et alii, *Química Nova.* n. 2, v. 25, p.287-299 (2002).

[4] A. F. S. Guedes, E. M. Costa, V. P. Guedes, Formation of Nanoparticles in Organic Materials used in Semiconductors Optoelectronic Devices. *ECS Trans.* **31**, 231-236 (2010).

[5] R. C. Jaeger, Introduction to microelectronic fabrication. Addison-Wesley Publishing Co., Massachusetts, **5**, (1989) 230p.

[6] S. Yun et alli. *Noncontact characterization of sheet resistance of indium–tin-oxide thin films by using a near-field microwave microprobe. Thin Solid Films* **515**, 1354-1357 (2006).

[7] D. Voet, J. G. Voet, *Biochemistry,* 2nd ed., John Wiley & Sons, Inc., New York (1995).

[8] J. R. Dyer, *Aplicações da Espectroscopia de Absorção dos Compostos Orgânicos.* Edgard Blucher Ltda, São Paulo (1969).

AUTHORS INDEX
Volume I