



IJCSI

International Journal of Computer Science Issues

Volume 10, Issue 4, No 1, July 2013
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

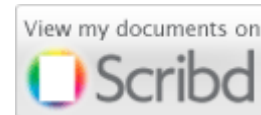
© IJCSI PUBLICATION
www.IJCSI.org

IJCSI proceedings are currently indexed by:



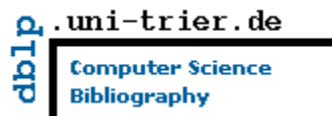
Cogprints

Google scholar



SciRate.com

CiteSeer^x beta



DOAJ DIRECTORY OF OPEN ACCESS JOURNALS



ProQuest

IJCSI Publicity Board 2013

Dr. Borislav D Dimitrov

Department of General Practice, Royal College of Surgeons in Ireland
Dublin, Ireland

Dr. Vishal Goyal

Department of Computer Science, Punjabi University
Patiala, India

Mr. Nehinbe Joshua

University of Essex
Colchester, Essex, UK

Mr. Vassilis Papataxiarhis

Department of Informatics and Telecommunications
National and Kapodistrian University of Athens, Athens, Greece

IJCSI Editorial Board 2013

Dr Tristan Vanrullen

Chief Editor

LPL, Laboratoire Parole et Langage - CNRS - Aix en Provence, France

LABRI, Laboratoire Bordelais de Recherche en Informatique - INRIA - Bordeaux, France

LEEE, Laboratoire d'Esthétique et Expérimentations de l'Espace - Université d'Auvergne, France

Dr Constantino Malagôn

Associate Professor

Nebrija University

Spain

Dr Lamia Fourati Chaari

Associate Professor

Multimedia and Informatics Higher Institute in SFAX

Tunisia

Dr Mokhtar Beldjehem

Professor

Sainte-Anne University

Halifax, NS, Canada

Dr Pascal Chatonnay

Assistant Professor

Maître de Conférences

Laboratoire d'Informatique de l'Université de Franche-Comté

Université de Franche-Comté

France

Dr Karim Mohammed Rezaul

Centre for Applied Internet Research (CAIR)

Glyndwr University

Wrexham, United Kingdom

Dr Yee-Ming Chen

Professor

Department of Industrial Engineering and Management

Yuan Ze University

Taiwan

Dr Gitesh K. Raikundalia

School of Engineering and Science,

Victoria University

Melbourne, Australia

Dr Vishal Goyal

Assistant Professor
Department of Computer Science
Punjabi University
Patiala, India

Dr Dalbir Singh

Faculty of Information Science And Technology
National University of Malaysia
Malaysia

Dr Natarajan Meghanathan

Assistant Professor
REU Program Director
Department of Computer Science
Jackson State University
Jackson, USA

Dr. Prabhat K. Mahanti

Professor
Computer Science Department,
University of New Brunswick
Saint John, N.B., E2L 4L5, Canada

Dr Navneet Agrawal

Assistant Professor
Department of ECE,
College of Technology & Engineering,
MPUAT, Udaipur 313001 Rajasthan, India

Dr Panagiotis Michailidis

Division of Computer Science and Mathematics,
University of Western Macedonia,
53100 Florina, Greece

Dr T. V. Prasad

Professor
Department of Computer Science and Engineering,
Lingaya's University
Faridabad, Haryana, India

Dr Saqib Rasool Chaudhry

Wireless Networks and Communication Centre
261 Michael Sterling Building
Brunel University West London, UK, UB8 3PH

Dr Shishir Kumar

Department of Computer Science and Engineering,
Jaypee University of Engineering & Technology
Raghogarh, MP, India

Dr P. K. Suri

Professor
Department of Computer Science & Applications,
Kurukshetra University,
Kurukshetra, India

Dr Paramjeet Singh

Associate Professor
GZS College of Engineering & Technology,
India

Dr Majid Bakhtiari

Faculty of Computer Science & Information System
University technology Malaysia
Skudai, 81310 Johore, Malaysia

Dr Shaveta Rani

Associate Professor
GZS College of Engineering & Technology,
India

Dr. Seema Verma

Associate Professor,
Department Of Electronics,
Banasthali University,
Rajasthan - 304022, India

Dr G. Ganesan

Professor
Department of Mathematics,
Adikavi Nannaya University,
Rajahmundry, A.P, India

Dr A. V. Senthil Kumar

Department of MCA,
Hindusthan College of Arts and Science,
Coimbatore, Tamilnadu, India

Dr Mashiur Rahman

Department of Life and Coordination-Complex Molecular Science,
Institute For Molecular Science, National Institute of Natural Sciences,
Miyodaiji, Okazaki, Japan

Dr Jyoteesh Malhotra

ECE Department,
Guru Nanak Dev University,
Jalandhar, Punjab, India

Dr R. Ponnusamy

Professor

Department of Computer Science & Engineering,
Aarupadai Veedu Institute of Technology,
Vinayaga Missions University, Chennai, Tamilnadu, India

Dr Nittaya Kerdprasop

Associate Professor

School of Computer Engineering,
Suranaree University of Technology, Thailand

Dr Manish Kumar Jindal

Department of Computer Science and Applications,
Panjab University Regional Centre, Muktsar, Punjab, India

Dr Deepak Garg

Computer Science and Engineering Department,
Thapar University, India

Dr P. V. S. Srinivas

Professor

Department of Computer Science and Engineering,
Geethanjali College of Engineering and Technology
Hyderabad, Andhra Pradesh, India

Dr Sara Moein

CMSSP Lab, Block A, 2nd Floor, Faculty of Engineering,
MultiMedia University, Malaysia

Dr Rajender Singh Chhillar

Professor

Department of Computer Science & Applications,
M. D. University, Haryana, India

EDITORIAL

In this forth edition of 2013, we bring forward issues from various dynamic computer science fields ranging from system performance, computer vision, artificial intelligence, software engineering, multimedia, pattern recognition, information retrieval, databases, security and networking among others.

Considering the growing interest of academics worldwide to publish in IJCSI, we invite universities and institutions to partner with us to further encourage open-access publications.

As always we thank all our reviewers for providing constructive comments on papers sent to them for review. This helps enormously in improving the quality of papers published in this issue.

Google Scholar reported a large amount of cited papers published in IJCSI. We will continue to encourage the readers, authors and reviewers and the computer science scientific community and interested authors to continue citing papers published by the journal.

Apart from availability of the full-texts from the journal website, all published papers are deposited in open-access repositories to make access easier and ensure continuous availability of its proceedings free of charge for all researchers.

We are pleased to present IJCSI Volume 10, Issue 4, No 1, July 2013 (IJCSI Vol. 10, Issue 4, No 1). The acceptance rate for this issue is 29.7%.



IJCSI Editorial Board
July 2013 Issue
ISSN (online): 1694-0784
ISSN (print): 1694-0814
© IJCSI Publications
www.IJCSI.org

IJCSI Reviewers Committee 2013

Mr. Markus Schatten, University of Zagreb, Faculty of Organization and Informatics, Croatia

Mr. Vassilis Papataxiarhis, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece

Dr Modestos Stavrakis, University of the Aegean, Greece

Dr Fadi KHALIL, LAAS -- CNRS Laboratory, France

Dr Dimitar Trajanov, Faculty of Electrical Engineering and Information technologies, ss. Cyril and Methodius Univesity - Skopje, Macedonia

Dr Jinping Yuan, College of Information System and Management,National Univ. of Defense Tech., China

Dr Alexis Lazanas, Ministry of Education, Greece

Dr Stavroula Mougiakakou, University of Bern, ARTORG Center for Biomedical Engineering Research, Switzerland

Dr Cyril de Runz, CReSTIC-SIC, IUT de Reims, University of Reims, France

Mr. Pramodkumar P. Gupta, Dept of Bioinformatics, Dr D Y Patil University, India

Dr Alireza Fereidunian, School of ECE, University of Tehran, Iran

Mr. Fred Viezens, Otto-Von-Guericke-University Magdeburg, Germany

Dr. Richard G. Bush, Lawrence Technological University, United States

Dr. Ola Osunkoya, Information Security Architect, USA

Mr. Kotsokostas N.Antonios, TEI Piraeus, Hellas

Prof Steven Totosy de Zepetnek, U of Halle-Wittenberg & Purdue U & National Sun Yat-sen U, Germany, USA, Taiwan

Mr. M Arif Siddiqui, Najran University, Saudi Arabia

Ms. Ilknur Icke, The Graduate Center, City University of New York, USA

Prof Miroslav Baca, Faculty of Organization and Informatics, University of Zagreb, Croatia

Dr. Elvia Ruiz Beltrán, Instituto Tecnológico de Aguascalientes, Mexico

Mr. Moustafa Banbouk, Engineer du Telecom, UAE

Mr. Kevin P. Monaghan, Wayne State University, Detroit, Michigan, USA

Ms. Moira Stephens, University of Sydney, Australia

Ms. Maryam Feily, National Advanced IPv6 Centre of Excellence (NAV6) , Universiti Sains Malaysia (USM), Malaysia

Dr. Constantine YIALOURIS, Informatics Laboratory Agricultural University of Athens, Greece

Mrs. Angeles Abella, U. de Montreal, Canada

Dr. Patrizio Arrigo, CNR ISMAC, italy

Mr. Anirban Mukhopadhyay, B.P.Poddar Institute of Management & Technology, India

Mr. Dinesh Kumar, DAV Institute of Engineering & Technology, India

Mr. Jorge L. Hernandez-Ardieta, INDRA SISTEMAS / University Carlos III of Madrid, Spain

Mr. AliReza Shahrestani, University of Malaya (UM), National Advanced IPv6 Centre of Excellence (NAv6), Malaysia

Mr. Blagoj Ristevski, Faculty of Administration and Information Systems Management - Bitola, Republic of Macedonia

Mr. Mauricio Egidio Cantão, Department of Computer Science / University of São Paulo, Brazil

Mr. Jules Ruis, Fractal Consultancy, The Netherlands

Mr. Mohammad Iftekhar Husain, University at Buffalo, USA

Dr. Deepak Laxmi Narasimha, Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia

Dr. Paola Di Maio, DMEM University of Strathclyde, UK

Dr. Bhanu Pratap Singh, Institute of Instrumentation Engineering, Kurukshetra University Kurukshetra, India

Mr. Sana Ullah, Inha University, South Korea

Mr. Cornelis Pieter Pieters, Condast, The Netherlands

Dr. Amogh Kavimandan, The MathWorks Inc., USA

Dr. Zhinan Zhou, Samsung Telecommunications America, USA

Mr. Alberto de Santos Sierra, Universidad Politécnica de Madrid, Spain

Dr. Md. Atiqur Rahman Ahad, Department of Applied Physics, Electronics & Communication Engineering (APECE), University of Dhaka, Bangladesh

Dr. Charalampos Bratsas, Lab of Medical Informatics, Medical Faculty, Aristotle University, Thessaloniki, Greece

Ms. Alexia Dini Kounoudes, Cyprus University of Technology, Cyprus

Dr. Jorge A. Ruiz-Vanoye, Universidad Juárez Autónoma de Tabasco, Mexico

Dr. Alejandro Fuentes Penna, Universidad Popular Autónoma del Estado de Puebla, México

Dr. Ocotlán Díaz-Parra, Universidad Juárez Autónoma de Tabasco, México

Mrs. Nantia Iakovidou, Aristotle University of Thessaloniki, Greece

Mr. Vinay Chopra, DAV Institute of Engineering & Technology, Jalandhar

Ms. Carmen Lastres, Universidad Politécnica de Madrid - Centre for Smart Environments, Spain

Dr. Sanja Lazarova-Molnar, United Arab Emirates University, UAE

Mr. Srikrishna Nudurumati, Imaging & Printing Group R&D Hub, Hewlett-Packard, India

Dr. Olivier Nocent, CReSTIC/SIC, University of Reims, France

Mr. Burak Cizmeci, Isik University, Turkey

Dr. Carlos Jaime Barrios Hernandez, LIG (Laboratory Of Informatics of Grenoble), France

Mr. Md. Rabiul Islam, Rajshahi university of Engineering & Technology (RUET), Bangladesh

Dr. LAKHOUA Mohamed Najeh, ISSAT - Laboratory of Analysis and Control of Systems, Tunisia

Dr. Alessandro Lavacchi, Department of Chemistry - University of Firenze, Italy

Mr. Mungwe, University of Oldenburg, Germany

Mr. Somnath Tagore, Dr D Y Patil University, India

Ms. Xueqin Wang, ATCS, USA

Dr. Borislav D Dimitrov, Department of General Practice, Royal College of Surgeons in Ireland, Dublin, Ireland

Dr. Fondjo Fotou Franklin, Langston University, USA

Dr. Vishal Goyal, Department of Computer Science, Punjabi University, Patiala, India

Mr. Thomas J. Clancy, ACM, United States

Dr. Ahmed Nabih Zaki Rashed, Dr. in Electronic Engineering, Faculty of Electronic Engineering, menouf 32951, Electronics and Electrical Communication Engineering Department, Menoufia university, EGYPT, EGYPT

Dr. Rushed Kanawati, LIPN, France

Mr. Koteswar Rao, K G Reddy College Of ENGG.&TECH,CHILKUR, RR DIST.,AP, India

Mr. M. Nagesh Kumar, Department of Electronics and Communication, J.S.S. research foundation, Mysore University, Mysore-6, India

Dr. Ibrahim Noha, Grenoble Informatics Laboratory, France

Mr. Muhammad Yasir Qadri, University of Essex, UK

Mr. Annadurai .P, KMCPGS, Lawspet, Pondicherry, India, (Aff. Pondicherry Univeristy, India

Mr. E Munivel , CEDTI (Govt. of India), India

Dr. Chitra Ganesh Desai, University of Pune, India

Mr. Syed, Analytical Services & Materials, Inc., USA

Mrs. Payal N. Raj, Veer South Gujarat University, India

Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal, India

Mr. Mahesh Goyani, S.P. University, India, India

Mr. Vinay Verma, Defence Avionics Research Establishment, DRDO, India

Dr. George A. Papakostas, Democritus University of Thrace, Greece

Mr. Abhijit Sanjiv Kulkarni, DARE, DRDO, India

Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius

Dr. B. Sivaselvan, Indian Institute of Information Technology, Design & Manufacturing, Kancheepuram, IIT Madras Campus, India

Dr. Partha Pratim Bhattacharya, Greater Kolkata College of Engineering and Management, West Bengal University of Technology, India

Mr. Manish Maheshwari, Makhanlal C University of Journalism & Communication, India

Dr. Siddhartha Kumar Khaitan, Iowa State University, USA

Dr. Mandhapati Raju, General Motors Inc, USA

Dr. M.Iqbal Saripan, Universiti Putra Malaysia, Malaysia

Mr. Ahmad Shukri Mohd Noor, University Malaysia Terengganu, Malaysia

Mr. Selvakuberan K, TATA Consultancy Services, India

Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India

Mr. Rakesh Kachroo, Tata Consultancy Services, India

Mr. Raman Kumar, National Institute of Technology, Jalandhar, Punjab., India

Mr. Nitesh Sureja, S.P.University, India

Dr. M. Emre Celebi, Louisiana State University, Shreveport, USA

Dr. Aung Kyaw Oo, Defence Services Academy, Myanmar

Mr. Sanjay P. Patel, Sankalchand Patel College of Engineering, Visnagar, Gujarat, India

Dr. Pascal Fallavollita, Queens University, Canada

Mr. Jitendra Agrawal, Rajiv Gandhi Technological University, Bhopal, MP, India

Mr. Ismael Rafael Ponce Medellín, Cenidet (Centro Nacional de Investigación y Desarrollo Tecnológico), Mexico

Mr. Shoukat Ullah, Govt. Post Graduate College Bannu, Pakistan

Dr. Vivian Augustine, Telecom Zimbabwe, Zimbabwe

Mrs. Mutalli Vatile, Offshore Business Philipines, Philipines

Mr. Pankaj Kumar, SAMA, India

Dr. Himanshu Aggarwal, Punjabi University, Patiala, India

Dr. Vauvert Guillaume, Europages, France

Prof Yee Ming Chen, Department of Industrial Engineering and Management, Yuan Ze University, Taiwan

Dr. Constantino Malagón, Nebrija University, Spain

Prof Kanwalvir Singh Dhindsa, B.B.S.B.Egg.College, Fatehgarh Sahib (Punjab), India

Mr. Angkoon Phinyomark, Prince of Singkla University, Thailand

Ms. Nital H. Mistry, Veer Narmad South Gujarat University, Surat, India

Dr. M.R.Sumalatha, Anna University, India

Mr. Somesh Kumar Dewangan, Disha Institute of Management and Technology, India

Mr. Raman Maini, Punjabi University, Patiala(Punjab)-147002, India

Dr. Abdelkader Outtagarts, Alcatel-Lucent Bell-Labs, France

Prof Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India

Mr. Prabu Mohandas, Anna University/Adhiyamaan College of Engineering, india

Dr. Manish Kumar Jindal, Panjab University Regional Centre, Muksar, India

Prof Mydhili K Nair, M S Ramaiah Institute of Technnology, Bangalore, India

Dr. C. Suresh Gnana Dhas, VelTech MultiTech Dr.Rangarajan Dr.Sagunthala Engineering College,Chennai,Tamilnadu, India

Prof Akash Rajak, Krishna Institute of Engineering and Technology, Ghaziabad, India

Mr. Ajay Kumar Shrivastava, Krishna Institute of Engineering & Technology, Ghaziabad, India

Dr. Vu Thanh Nguyen, University of Information Technology HoChiMinh City, VietNam

Prof Deo Prakash, SMVD University (A Technical University open on I.I.T. Pattern) Kakryal (J&K), India

Dr. Navneet Agrawal, Dept. of ECE, College of Technology & Engineering, MPUAT, Udaipur 313001 Rajasthan, India

Mr. Sufal Das, Sikkim Manipal Institute of Technology, India

Mr. Anil Kumar, Sikkim Manipal Institute of Technology, India

Dr. B. Prasanalakshmi, King Saud University, Saudi Arabia.

Dr. K D Verma, S.V. (P.G.) College, Aligarh, India

Mr. Mohd Nazri Ismail, System and Networking Department, University of Kuala Lumpur (UniKL), Malaysia

Dr. Nguyen Tuan Dang, University of Information Technology, Vietnam National University Ho Chi Minh city, Vietnam

Dr. Abdul Aziz, University of Central Punjab, Pakistan

Dr. P. Vasudeva Reddy, Andhra University, India

Mrs. Savvas A. Chatzichristofis, Democritus University of Thrace, Greece

Mr. Marcio Dorn, Federal University of Rio Grande do Sul - UFRGS Institute of Informatics, Brazil

Mr. Luca Mazzola, University of Lugano, Switzerland

Mr. Hafeez Ullah Amin, Kohat University of Science & Technology, Pakistan

Dr. Professor Vikram Singh, Ch. Devi Lal University, Sirsa (Haryana), India

Dr. Shahanawaj Ahamad, Department of Computer Science, King Saud University, Saudi Arabia

Dr. K. Duraiswamy, K. S. Rangasamy College of Technology, India

Prof. Dr Mazlina Esa, Universiti Teknologi Malaysia, Malaysia

Dr. P. Vasant, Power Control Optimization (Global), Malaysia

Dr. Taner Tuncer, Firat University, Turkey

Dr. Norrozila Sulaiman, University Malaysia Pahang, Malaysia

Prof. S K Gupta, BCET, Guradspur, India

Dr. Latha Parameswaran, Amrita Vishwa Vidyapeetham, India

Mr. M. Azath, Anna University, India

Dr. P. Suresh Varma, Adikavi Nannaya University, India

Prof. V. N. Kamalesh, JSS Academy of Technical Education, India

Dr. D Gunaseelan, Ibri College of Technology, Oman

Mr. Sanjay Kumar Anand, CDAC, India

Mr. Akshat Verma, CDAC, India

Mrs. Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia

Mr. Hasan Asil, Islamic Azad University Tabriz Branch (Azarshahr), Iran

Prof. Dr Sajal Kabiraj, Fr. C Rodrigues Institute of Management Studies (Affiliated to University of Mumbai, India), India

Mr. Syed Fawad Mustafa, GAC Center, Shandong University, China

Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA

Prof. Selvakani Kandeegan, Francis Xavier Engineering College, India

Mr. Tohid Sedghi, Urmia University, Iran

Dr. S. Sasikumar, PSNA College of Engg and Tech, Dindigul, India

Dr. Anupam Shukla, Indian Institute of Information Technology and Management Gwalior, India

Mr. Rahul Kala, Indian Institute of Information Technology and Management Gwalior, India

Dr. A V Nikolov, National University of Lesotho, Lesotho

Mr. Kamal Sarkar, Department of Computer Science and Engineering, Jadavpur University, India

Prof. Sattar J Aboud, Iraqi Council of Representatives, Iraq-Baghdad

Dr. Prasant Kumar Pattnaik, Department of CSE, KIST, India

Dr. Mohammed Amoon, King Saud University, Saudi Arabia

Dr. Tsvetanka Georgieva, Department of Information Technologies, St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria

Mr. Ujjal Marjit, University of Kalyani, West-Bengal, India

Dr. Prasant Kumar Pattnaik, KIST, Bhubaneswar, India, India

Dr. Guezouri Mustapha, Department of Electronics, Faculty of Electrical Engineering, University of Science and Technology (USTO), Oran, Algeria

Mr. Maniyar Shiraz Ahmed, Najran University, Najran, Saudi Arabia

Dr. Sreedhar Reddy, JNTU, SSIETW, Hyderabad, India

Mr. Bala Dhandayuthapani Veerasamy, Mekelle University, Ethiopia

Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia

Mr. Rajesh Prasad, LDC Institute of Technical Studies, Allahabad, India

Ms. Habib Izadkhah, Tabriz University, Iran

Dr. Lokesh Kumar Sharma, Chhattisgarh Swami Vivekanand Technical University Bhilai, India

Mr. Kuldeep Yadav, IIT Delhi, India

Dr. Naoufel Kraiem, Institut Supérieur d'Informatique, Tunisia

Prof. Frank Ortmeier, Otto-von-Guericke-Universität Magdeburg, Germany

Mr. Ashraf Aljammal, USM, Malaysia

Mrs. Amandeep Kaur, Department of Computer Science, Punjabi University, Patiala, Punjab, India

Mr. Babak Basharirad, University Technology of Malaysia, Malaysia

Mr. Avinash Singh, Kiet Ghaziabad, India

Dr. Miguel Vargas-Lombardo, Technological University of Panama, Panama

Dr. Tuncay Sevindik, Firat University, Turkey

Ms. Pavai Kandavelu, Anna University Chennai, India

Mr. Ravish Khichar, Global Institute of Technology, India

Mr. Ahsan Ali Zaidan Ansaef, Multimedia University, Cyberjaya, Malaysia

Dr. Awadhesh Kumar Sharma, Dept. of CSE, MMM Engg College, Gorakhpur-273010, UP, India

Mr. Qasim Siddique, FUIEMS, Pakistan

Dr. Le Hoang Thai, University of Science, Vietnam National University - Ho Chi Minh City, Vietnam

Dr. Saravanan C, NIT, Durgapur, India

Dr. Vijay Kumar Mago, DAV College, Jalandhar, India

Dr. Do Van Nhon, University of Information Technology, Vietnam

Dr. Georgios Kioumourtzis, Researcher, University of Patras, Greece

Mr. Amol D. Potgantwar, SITRC Nasik, India

Mr. Lesedi Melton Masisi, Council for Scientific and Industrial Research, South Africa

Dr. Karthik.S, Department of Computer Science & Engineering, SNS College of Technology, India

Mr. Nafiz Imtiaz Bin Hamid, Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT), Bangladesh

Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia

Dr. Abdul Kareem M. Radhi, Information Engineering - Nahrain University, Iraq

Dr. Manuj Darbari, BBDNITM, Institute of Technology, A-649, Indira Nagar, Lucknow 226016, India

Ms. Izerrouken, INP-IRIT, France

Mr. Nitin Ashokrao Naik, Dept. of Computer Science, Yeshwant Mahavidyalaya, Nanded, India

Mr. Nikhil Raj, National Institute of Technology, Kurukshetra, India

Prof. Maher Ben Jemaa, National School of Engineers of Sfax, Tunisia

Prof. Rajeshwar Singh, BRCM College of Engineering and Technology, Bahal Bhiwani, Haryana, India

Mr. Gaurav Kumar, Department of Computer Applications, Chitkara Institute of Engineering and Technology, Rajpura, Punjab, India

Mr. Ajeet Kumar Pandey, Indian Institute of Technology, Kharagpur, India

Mr. Rajiv Phougat, IBM Corporation, USA

Mrs. Aysha V, College of Applied Science Pattuvam affiliated with Kannur University, India

Dr. Debotosh Bhattacharjee, Department of Computer Science and Engineering, Jadavpur University, Kolkata-700032, India

Dr. Neelam Srivastava, Institute of engineering & Technology, Lucknow, India

Prof. Sweta Verma, Galgotia's College of Engineering & Technology, Greater Noida, India

Mr. Harminder Singh Bindra, MIMIT, INDIA

Mr. Tarun Kumar, U.P. Technical University/Radha Govinend Engg. College, India

Mr. Tirthraj Rai, Jawahar Lal Nehru University, New Delhi, India

Mr. Akhilesh Tiwari, Madhav Institute of Technology & Science, India

Mr. Dakshina Ranjan Kisku, Dr. B. C. Roy Engineering College, WBUT, India

Ms. Anu Suneja, Maharshi Markandeshwar University, Mullana, Haryana, India

Mr. Munish Kumar Jindal, Punjabi University Regional Centre, Jaito (Faridkot), India

Dr. Ashraf Bany Mohammed, Management Information Systems Department, Faculty of Administrative and Financial Sciences, Petra University, Jordan

Mrs. Jyoti Jain, R.G.P.V. Bhopal, India

Dr. Lamia Chaari, SFAX University, Tunisia

Mr. Akhter Raza Syed, Department of Computer Science, University of Karachi, Pakistan

Prof. Khubaib Ahmed Qureshi, Information Technology Department, HIMS, Hamdard University, Pakistan

Prof. Boubker Sbihi, Ecole des Sciences de L'Information, Morocco

Dr. S. M. Riazul Islam, Inha University, South Korea

Prof. Lokhande S.N., S.R.T.M.University, Nanded (MH), India

Dr. Vijay H Mankar, Dept. of Electronics, Govt. Polytechnic, Nagpur, India

Mr. Ojesanmi Olusegun, Ajayi Crowther University, Oyo, Nigeria

Ms. Mamta Juneja, RBIEBT, PTU, India

Prof. Chandra Mohan, John Bosco Engineering College, India

Dr. Bodhe Shrikant K., College of Engineering, Pandhapur, Maharashtra, INDIA

Dr. Sherif G. Aly, The American University in Cairo, Egypt

Mr. Sunil Kashibarao Nayak, Bahirji Smarak Mahavidyalaya, Basmathnagar Dist-Hingoli., India

Prof. Nikhil gondaliya, G H Patel College of Engg. & Technology, India

Mr. Nisheeth Joshi, Apaji Institute, Banasthali University, India

Mr. Nizar, National Engineering School of Monastir, Tunisia

Prof. R. Jagadeesh Kannan, RMK Engineering College, India

Prof. Rakesh.L, Vijetha Institute of Technology, Bangalore, India

Mr B. M. Patil, Indian Institute of Technology, Roorkee, Uttarakhand, India

Dr. Intisar A. M. Al Sayed, Associate prof./College of Science and IT/Al Isra University, Jordan

Mr. Thipendra Pal Singh, Sharda University, K.P. III, Greater Noida, Uttar Pradesh, India

Mrs. Rajalakshmi, JIITU, India

Mr. Shrikant Ardhapurkar, Indian Institute of Information Techonology, India

Ms. Hemalatha R, Osmania University, India

Mr. Hadi Saboohi, University of Malaya - Faculty of Computer Science and Information Technology, Malaysia

Mr. Sunil Kumar Grandhi, Maris Stella College, India

Prof. Shishir K. Shandilya, NRI Institute of Science & Technology, INDIA

Dr. Umesh Kumar Singh, Vikram University, Ujjain, India

Prof. Prasun Ghosal, Bengal Engineering and Science University, India

Dr. Nagarajan Velmurugan, SMVEC/Pondicherry University, India

Dr. R. Baskaran, Anna University, India

Dr. Wichian Sittiprapaporn, Mahasarakham University College of Music, Thailand

Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia

Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology, India

Mrs. Inderpreet Kaur, PTU, Jalandhar, India

Mr. Palaniyappan, K7 Virus Research Laboratory, India

Mr. Guanbo Zheng, University of Houston, main campus, USA

Mr. Arun Kumar Tripathi, Krishna Institute of Engg. and Tech-Ghaziabad, Affiliated to UPTU, India

Mr. Iqbaldeep Kaur, PTU / RBIEBT, India

Mr. Amit Choudhary, Maharaja Surajmal Institute, New Delhi, India

Mrs. Vasudha Bahl, Maharaja Agrasen Institute of Technology, Delhi, India

Dr. Ashish Avasthi, Uttar Pradesh Technical University, India

Dr. Manish Kumar, Uttar Pradesh Technical University, India

Prof. Vinay Uttamrao Kale, P.R.M. Institute of Technology & Research, Badnera, Amravati, Maharashtra, India

Mr. Suhas J Manangi, Microsoft, India

Mr. Shyamalendu Kandar, Haldia Institute of Technology, India

Ms. Anna Kuzio, Adam Mickiewicz University, School of English, Poland

Mr. Vikas Singla, Malout Institute of Management & Information Technology, Malout, Punjab, India, India

Dr. Dalbir Singh, Faculty of Information Science And Technology, National University of Malaysia, Malaysia

Dr. Saurabh Mukherjee, PIM, Jiwaji University, Gwalior, M.P, India

Mr. Senthilnathan T, Sri Krishna College of Engineering and Technology, India

Dr. Debojyoti Mitra, Sir Padampat Singhanian University, India

Prof. Rachit Garg, Department of Computer Science, L K College, India

Dr. Arun Kumar Gupta, M.S. College, Saharanpur, India

Dr. Todor Todorov, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria

Mrs. Manjula K A, Kannur University, India

Mrs. Sasikala R., K S R College of Technology, India

Prof. M. Saleem Babu, Department of Computer Science and Engineering, Vel Tech University, Chennai, India

Dr. Rajesh Kumar Tiwari, GLA Institute of Technology, India

Mr. Rakesh Kumar, Indian Institute of Technology Roorkee, India

Prof. Amit Verma, PTU/RBIEBT, India

Mr. Sohan Purohit, University of Massachusetts Lowell, USA

Mr. Anand Kumar, AMC Engineering College, Bangalore, India

Dr. Samir Abdelrahman, Computer Science Department, Cairo University, Egypt

Dr. Rama Prasad V Vaddella, Sree Vidyanikethan Engineering College, India

Dr. Manoj Wadhwa, Echelon Institute of Technology Faridabad, India

Mr. Zeashan Hameed Khan, Universit  de Grenoble, France

Mr. Arup Kumar Pal, Indian School of Mines, Dhanbad, India

Dr. Pouya, Islamic Azad University, Naein Branch, Iran

Prof. Jyoti Prakash Singh, Academy of Technology, India

Mr. Muraleedharan CV, Sree Chitra Tirunal Institute for Medical Sciences & Technology, India

Dr. E U Okike, University of Ibadan, Nigeria Kampala Int Univ Uganda, Nigeria

Dr. D. S. Rao, Chitkara University, India

Mr. Peyman Taher, Oklahoma State University, USA

Dr. S Srinivasan, PDM College of Engineering, India

Dr. Rafiqul Zaman Khan, Department of Computer Science, AMU, Aligarh, India

Ms. Meenakshi Kalia, Shobhit University, India

Mr. Muhammad Zakarya, Abdul Wali Khan University, Mardan, Pakistan, Pakistan

Dr. M Gobi, PSG college, India

Mr. Williamjeet Singh, Chitkara Institute of Engineering and Technology, India

Mr. G.Jeyakumar, Amrita School of Engineering, India

Mr. Osama Sohaib, University of Balochistan, Pakistan

Mr. Jude Hemanth, Karunya University, India

Mr. Nitin Rakesh, Jaypee University of Information Technology, India

Mr. Harmunish Taneja, Maharishi Markandeshwar University, Mullana, Ambala, Haryana, India

Dr. Sin-Ban Ho, Faculty of IT, Multimedia University, Malaysia

Dr. Mashiur Rahman, Institute for Molecular Science, Japan

Mrs. Doreen Hephzibah Miriam, Anna University, Chennai, India

Mr. Kosala Yapa Bandara, Dublin City University, Ireland.

Mrs. Mitu Dhull, GNKITMS Yamuna Nagar Haryana, India

Dr. Chitra A.Dhawale, Professor, Symbiosis Institute of Computer Studies and Research, Pune (MS), India

Dr. Arun Sharma, GB Technical University, Noida, India

Mr. Naoufel Machta, Faculty of Science of Tunis, Tunisia

Dr. Utpal Biswas, University of Kalyani, India

Prof. Parma Nand, IIT Roorkee, India

Prof. Mahesh P K, Jnana Vikas Institute of Tevhnology, Bangalore, India

Dr. D.I. George Amalarethnam, Jamal Mohamed College, Bharathidasan University, India

Mr. Ishtiaq ahmad, University of Engineering & Technology, Taxila, Pakistan

Mrs. B.Sharmila, Sri Ramakrishna Engineering College, Coimbatore Anna University Coimbatore, India

Dr. Muhammad Wasif Nisar, COMSATS Institue of Information Technology, Pakistan

Mr. Prabu Dorairaj, EMC Corporation, India/USA

Mr. Neetesh Gupta, Technocrats Inst. of Technology, Bhopal, India

Dr. Ola Osunkoya, PRGX, USA

Ms. A. Lavanya, Manipal University, Karnataka, India

Dr. Jalal Laassiri, MIA-Laboratory, Faculty of Sciences Rabat, Morocco

Mr. Ganesan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India

Mr. V.Ramakrishnan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India

Prof. Vuda Sreenivasarao, St. Mary's college of Engg & Tech, India

Prof. Ashutosh Kumar Dubey, Assistant Professor, India

Dr. R.Ramesh, Anna University, India

Mr. Ali Khadair HMood, University of Malaya, Malaysia

Dr. Vimal Mishra, U.P. Technical Education, India

Mr. Ranjit Singh, Apeejay Institute of Management, Jalandhar, India

Mrs. D.Suganyadevi, SNR SONS College (Autonomous), India

Mr. Prasad S.Halgaonkar, MIT, Pune University, India

Mr. Vijay Kumar, College of Engg. and Technology, IFTM, Moradabad(U.P), India

Mr. Mehran Parchebafieh, Douran, Iran

Mr. Anand Sharma, MITS, Lakshmangarh, Sikar (Rajasthan), India

Mr. Amit Kumar, Jaypee University of Engineering and Technology, India

Prof. B.L.Shivakumar, SNR Sons College, Coimbatore, India

Mr. Mohammed Imran, JMI, India

Dr. R Bremananth, School of EEE, Information Engineering (Div.), Nanyang Technological University, Singapore

Prof. Vasavi Bande, Computer Science and Engineering, Hyderabad Institute of Technology and Management, India

Dr. S.R.Balasundaram, National Institute of Technology, India

Dr. Prasart Nuangchalerm, Mahasarakham University, Thailand

Dr. M Ayoub Khan, C-DAC, Ministry of Communications & IT., India

Dr. Jagdish Lal Raheja, Central Electronics Engineering Research Institute, India

Mr G. Appasami, Dept. of CSE, Dr. Pauls Engineering College, Anna University - Chennai, India

Mr Vimal Mishra, U.P. Technical Education, Allahabad, India

Mr. Amin Daneshmand Malayeri, Young Researchers Club, Islamic AZAD University, Malayer Branch, Iran

Dr. Arti Arya, PES School of Engineering, Bangalore (under VTU, Belgaum, Karnataka), India

Mr. Pawan Jindal, J.U.E.T. Guna, M.P., India

Dr. Soumen Mukherjee, RCC Institute of Information Technology, India

Dr. Hamid Mcheick, University of Qubec at Chicoutimi, Canada

Dr. Mokhled AlTarawneh, PhD computer engineering/ Faculty of engineering/ mutah university, Jordan

Prof. Santhosh.P.Mathew, Saintgits College of Engineering, Kottayam, India

Ms. Suman Lata, Rayat Bahara institute of engg. & Nanotechnology, Hoshiarpur, India

Dr. Shaikh Abdul Hannan, Vivekanand College, Aurangabad, India

Prof. PN Kumar, Amrita Vishwa Vidyapeetham, India

Dr. P. K. Suri, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India

Dr. Syed Akhter Hossain, Daffodil International University, Bangladesh

Mr. Sunil, Vignan College, India

Mr. Ajit Singh, TIT&S Bhiwani, Haryana, India

Mr. Nasim Qaisar, Federal Urdu University of Arts, Science and Technology, Pakistan

Ms. Rshma, Maharishi Markandeshwar University, India

Mr. Gaurav Kumar Leekha, M.M.University, Solan (Himachal Pradesh), India

Mr. Ordinar Tucker, Ministry of Finance Jamaica, Jamaica

Mr. Mohit Jain, Maharaja Surajmal Institute of Technology (Affiliated to Guru Gobind Singh Indraprastha University, New Delhi), India

Dr. Shaveta Rani, GZS College of Engineering & Technology, India

Dr. Paramjeet Singh, GZS College of Engineering & Technology, India

Dr. G R Sinha, SSCET, India

Mr. Chetan Sharma, TechMahindra India Ltd., India

Dr. Nabil Mohammed Ali Munassar, University of Science and Technology, Yemen

Prof. T Venkat Narayana Rao, Department of CSE, Hyderabad Institute of Technology and Management, India

Prof. Vasavi Bande, HITAM, Engineering College, India

Prof. S.P.Setty, Andhra University, India

Dr. C. Kiran Mai, J.N.T.University, Hyderabad/VNR Vignana Jyothi Institute of Engineering & Technology/, India

Ms. Bindiya Ahuja, Manav Rachna International University, India

Mrs. Deepa Bura, Manav Rachna International University, India

Mr. Vikas Gupta, CDLM Government Engineering College, Panniwala Mota, India

Dr Juan José Martínez Castillo, University of Yacambu, Venezuela

Mr Kunwar S. Vaisla, Department of Computer Science & Engineering, BCT Kumaon Engineering College, India

Mr. Abhishek Shukla, RKGIT, India

Prof. Manpreet Singh, M. M. Engg. College, M. M. University, Haryana, India

Mr. Syed Imran, University College Cork, Ireland

Dr. Intisar Al Said, Associate Prof/Al Isra University, Jordan

Dr. Namfon Assawamekin, University of the Thai Chamber of Commerce, Thailand

Dr. Shiv KUMar, Technocrat Institute of Technology-Bhopal (M.P.), India

Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran

Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran

Dr. Mohamed Ali Mahjoub, University of Monastir, Tunisia

Mr. Adis Medic, Infosys Ltd, Bosnia and Herzegovina

Mr Swarup Roy, Department of Information Technology, North Eastern Hill University, Umshing, Shillong 793022, Meghalaya, India

Prof. Jakimi, Faculty of Science and technology my ismail University, Morocco

Dr. R. Manicka Chezian, N G M College, Pollachi - 642 001, Tamilnadu, India

Dr. P.Dananjayan, Pondicherry Engineering College, India

Mr. Manik Sharma, Sewa Devi SD College Tarn Taran, India

Mr. Suresh Kallam, East China University of Technology, Nanchang, China

Dr. Mohammed Ali Hussain, Sai Madhavi Institute of Science & Technology, Rajahmundry, India

Mr. Vikas Gupta, Adesh Institute of Engineering & Technology, India

Dr. Anuraag Awasthi, JV Womens University, Jaipur, India

Dr. Mathura Prasad Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), Srinagar (Garhwal), India

Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia, Malaysia

Mr. Adnan Qureshi, University of Jinan, Shandong, P.R.China, P.R.China

Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India

Mr. Mueen Uddin, Universiti Teknologi Malaysia, Malaysia

Mr. Manoj Gupta, Apex Institute of Engineering & Technology, Jaipur (Affiliated to Rajasthan Technical University, Rajasthan), Indian

Mr. S. Albert Alexander, Kongu Engineering College, India

Dr. Shaidah Jusoh, Zarqa Private University, Jordan

Dr. Dushmanta Mallick, KMBB College of Engineering and Technology, India

Mr. Santhosh Krishna B.V, Hindustan University, India

Dr. Tariq Ahamad Ahanger, Kausar College Of Computer Sciences, India

Dr. Chi Lin, Dalian University of Technology, China

Prof. VIJENDRA BABU.D, ECE Department, Aarupadai Veedu Institute of Technology, Vinayaka Missions University, India

Mr. Raj Gaurang Tiwari, Gautam Budh Technical University, India

Mrs. Jeysree J, SRM University, India

Dr. C S Reddy, VIT University, India

Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Bio-Technology, Kharar, India

Mr. Muhammad Shuaib Qureshi, Iqra National University, Peshawar, Pakistan, Pakistan

Dr Pranam Paul, Narula Institute of Technology Agarpara. Kolkata: 700109; West Bengal, India

Dr. G. M. Nasira, Sasurie College of Engineering, (Affiliated to Anna University of Technology Coimbatore), India

Dr. Manasawee Kaenampornpan, Mahasarakham University, Thailand

Mrs. Iti Mathur, Banasthali University, India

Mr. Avanish Kumar Singh, RRIMT, NH-24, B.K.T., Lucknow, U.P., India

Mr. Velayutham Pavanasam, Adhiparasakthi Engineering College, Melmaruvathur, India

Dr. Panagiotis Michailidis, University of Western Macedonia, Greece

Mr. Amir Seyed Danesh, University of Malaya, Malaysia

Dr. Nadeem Mahmood, Department of computer science, university of Karachi, Pakistan

Dr. Terry Walcott, E-Promag Consultancy Group, United Kingdom

Mr. Farhat Amine, High Institute of Management of Tunis, Tunisia

Mr. Ali Waqar Azim, COMSATS Institute of Information Technology, Pakistan

Mr. Zeeshan Qamar, COMSATS Institute of Information Technology, Pakistan

Dr. Samsudin Wahab, MARA University of Technology, Malaysia

Mr. Ashikali M. Hasan, CelNet Security, India

Dr. Binod Kumar, Lakshmi Narayan College of Tech.(LNCT), India

Mr. B V A N S S Prabhakar Rao, Dept. of CSE, Miracle Educational Society Group of Institutions, Vizianagaram, India

Dr. T. Abdul Razak, Associate Professor of Computer Science, Jamal Mohamed College (Affiliated to Bharathidasan University, Tiruchirappalli), Tiruchirappalli-620020, India

Mr. Aurobindo Ogra, University of Johannesburg, South Africa

Mr. Essam Halim Houssein, Dept of CS - Faculty of Computers and Informatics, Benha - Egypt

Dr. Hanumanthappa. J, DoS in Computer Science, India

Mr. Rachit Mohan Garg, Jaypee University of Information Technology, India

Mr. Kamal Kad, Infosys Technologies, Australia

Mrs. Aditi Chawla, GNIT Group of Institutes, India

Dr. Kumardatt Ganrje, Pune University, India

Mr. Merugu Gopichand, JNTU/BVRIT, India

Mr. Rakesh Kumar, M.M. University, Mullana,Ambala, India

Mr. M. Sundar, IBM, India

Prof. Mayank Singh, J.P. Institute of Engineering & Technology, India

Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India

Mr. Khaleel Ahmad, S.V.S. University, India

Mr. Amin Zehtabian, Babol Noshirvani University of Technology / Tetra Electronic Company, Iran

Mr. Rahul Katarya, Department of Information Technology , Delhi Technological University, India

Dr. Vincent Ele Asor, University of Port Harcourt, Nigeria

Ms. Prayas Kad, Capgemini Australia Ltd, Australia

Mr. Alireza Jolfaei, Faculty and Research Center of Communication and Information Technology, IHU, Iran

Mr. Nitish Gupta, GGSIPU, India

Dr. Mohd Lazim Abdullah, University of Malaysia Terengganu, Malaysia

Ms. Suneet Kumar, Uttarakhand Technical University/Dehradun Institute of Technology, Dehradun, Uttarakhand, India

Mr. Rupesh Nasre., Indian Institute of Science, Bangalore., India.

Mrs. Dimpi Srivastava, Dept of Computer science, Information Technology and Computer Application, MIET, Meerut, India

Dr. Eva Volna, University of Ostrava, Czech Republic

Prof. Santosh Balkrishna Patil, S.S.G.M. College of Engineering, Shegaon, India

Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology Solan (HP), India

Mr. Ashwani Kumar, Jaypee University of Information Technology Solan(HP), India

Dr. Abbas Karimi, Faculty of Engineering, I.A.U. Arak Branch, Iran

Mr. Fahimuddin.Shaik, AITS, Rajampet, India

Mr. Vahid Majid Nezhad, Islamic Azad University, Iran

Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore-641014, Tamilnadu, India

Prof. D. P. Sharma, AMU, Ethiopia

Dr. Sukumar Senthilkumar, School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia

Mr. Sanjay Bhargava, Banasthali University, Jaipur, Rajasthan, India

Prof. Rajesh Deshmukh, Shri Shankaracharya Institute of Professional Management & Technology, India

Mr. Shervan Fekri Ershad, shiraz international university, Iran

Dr. Vladimir Urosevic, Ministry of Interior, Republic of Serbia

Mr. Ajit Singh, MDU Rohtak, India

Prof. Asha Ambhaikar, Rungta College of Engineering & Technology, Bilai, India

Dr. Saurabh Dutta, Dr. B. C. Roy Engineering College, Durgapur, India

Dr. Mokhled Altarawneh, Mutah University, Jordan

Mr. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar, India

Mr S. A. Ahsan rajon, Computer Science and Engineering Discipline, Khulna University, Bangladesh

Ms. Rezarta Mersini, University of Durres, Albania

Mrs. Deepika Joshi, Jaipuria Institute of Management Studies, India

Dr. Niraj Shakhakarmi, Prairie View A&M University, (Texas A&M University System), USA

Mrs. A. Valarmathi, Anna University, Trichy, India

Dr. K. Balamurugan, Institute of Road and Transport Technology, India

Prof. K S Sridharan, Sri Sathya Sai Institute of Higher Learning, India

Mr. Okumoku-Evrero Oniovosa, Delta State University, Abraka, Nigeria

Mr. Rajiv Chopra, GTBIT, Delhi, India

Mr. Harish Garg, Department of Mathematics, IIT Roorkee, India

Mr. Ganesh Davanam, Sree Vidyanikethan Engineering College, India

Mr. Bhavesh Shah, VIT, India

Dr. Suresh Kumar Bhardwaj, Manav Rachna International University, India

Dr. Muhammad Nawaz Khan, School of electrical engineering & Computer Science, Pakistan

Ms. Saranya, Bharathidasan University, India

Mr. Sumit Joshi, GRD-IMT, Dehradun, India

Dr. Mohammed M. Abu Shquier, Tabuk University, School of Computers and Information Technology, Kingdom of Saudi Arabia

Ms. Shalini Ramanathan, PSG College of Technology, India

Mr. S.Munisankaraiah, Geethanjali college of Engineering & Technology, Hyderabad, India

Dr. Satyanarayana, KL University, India

Mr. Sarin CR, Anna University, India

Mr. Sayed Shoaib Anwar, Mahatma Gandhi Mission College of Engineering, India

Mrs. Gunjan, JSSATE, Noida, India

Dr. Ramachandra V Pujeri, Anna University, India

Mrs. Antima Singh Puniya, Shobhit University, Meerut, India

Dr. Avdhesh Gupta, College of Engineering Roorkee, India

Ms. Shiva Prakash, Madan Mohan Malaviya Engg. College, Gorakhpur, India

Dr. Kristijan Kuk, School of Electrical Engineering and Computer Science Applied Studies, Belgrade, Serbia

Prof. Dinesh Vitthalrao Rojatar, Govt. College of Engineering, Chandrapur, India

Prof. Lalji Prasad, RGTU/TCET, Indore, india

Dr. A. John Sanjeev Kumar, Thiagarajar College of Engineering, Madurai, Tamilnadu, India

Mr. Harishbabu Kalidasu, Priyadarshini Institute of Technology and Science, Tenali, Guntur(DT), Andhra Pradesh, India

Prof. Vaitheeshwaran, Priyadarshini Indira Gandhi College of Engineering, India

Mrs. P.Salini, Pondicherry Engineering College, India

Mr. Vivek Bhambri, Desh Bhagat Institute of Management and Computer Sciences, Mandi Gobindgarh(Punjab), India

Mr. Slavko Zitnik, Faculty of Computer and Information Science Ljubljana, Slovenia

Ms. Sreenivasa Rao, CMJ University/Yodlee Infotech, India

Mr. Shihabudheen P M, TATA ELXSI LTD, India

Dr. Ahmed Moustafa Elmahalawy, Faculty of Electronics Engineering, Computer Science and Engineering, Egypt

Mr. Kamlesh Kumar, Kumaun University, Nainital, India

TABLE OF CONTENTS

1. Keystroke Dynamics Authentication: A Survey of Free-text Methods Arwa Alsultan and Kevin Warwick	1-10
2. A New Energy Consumption Algorithm with Active Sensor Selection Using GELS in Target Coverage WSN Ali Bagrezai, Seyed Vahab AL-Din Makki and Ali Shokouhi Rostami	11-18
3. Flexible boom numerical simulation and vibration characteristics analysis Hongbin Tang and Wu Ren	19-22
4. Adaptive approach for spam detection Sumant Sharma and Amit Arora	23-26
5. Adoption of ICT to Enhance Educational Development in Nigeria Odeniyi O.A, Ayinde A.Q, Adetunji A.B and Sarumi O.A	27-31
6. Model for determining the impact analysis of open source adoption in software development economics Ahmad Adnan and Muhammad Faisal Akram	33-43
7. Application of the Gradient Vector Flow Method for Treating Satellite Image Mohammed Ait Oussous, Nouredine Alaa and Youssef Ait Khouya	44-48
8. An Enhanced Application of Modified PSO for Association Rule Mining Bharathi.T and P.Krishnakumari	49-55
9. An Exploratory Study of Arabic Language Support in Software Project Management Tools Arwa Al-Amoudi, Hailah Almazrua, Hebah Al-Moaiqel, Noura Alomar and Sarah Al-Koblan	56-63
10. Web-Marketing in Social Networks Using MDA Approach Lamlili El Mazoui Nadori Yasser, Erramdani Mohamed, Arrassen Ibtissam, Esbai Redouane and Moussaoui Mimoun	64-73
11. Combining a drug therapy and oncolytic virotherapy to treat cancer: an optimal control approach Adil El Alami Laaroussi, Elhia Mohamed and Mostafa Rachik	74-80
12. Mixed Immunotherapy and Chemotherapy of Tumors: Optimal Control Approach Samira Zouhri, Smahane Saadi, Ilias Elmouki, Amine Hamdache and Mostafa Rachik	81-97
13. Task Migration in Cloud Computing using Combination of Yu and Post-Copy Techniques Elham Shamsinezhad, Asadollah Shahbahrami, Alireza Hedayati, Ahmad Khadem Zadeh and Hamid Banirostam	98-102
14. Assessment of offline Digital Signature Recognition Classification Techniques Dina Darwish	103-111

15. A Model for Optimizing Data Caching of Dual Mode Handheld Devices Mohamed Grida, Hasnaa Soliman and Mohamed Hassan	112-120
16. On the Limits of Perfect Security for Steganographic System Khan Farhan Rafat and Muhammad Sher	121-126
17. A Review of Fault Detection Techniques for Wireless Sensor Networks Saurabh Mittal and Rinkle Rani Aggarwal	127-138
18. New Efficient Technique for Compression of ECG Signal Nidhal K. El Abbadi and Abbas M. Al-Bakry	139-146
19. Performance Evaluation of Naive Bayes and Decision Stump Algorithms in Mining Students Educational Data Ayinde A.Q, Adetunji A.B, Bello M and Odeniyi O.A	147-151
20. Effectiveness of Online Job Recruitment System: Evidence from the University of the East Mary Grace G. Ventura and Rex P. Bringula	152-159
21. Evaluation of Different Query Expansion Techniques by using Different Similarity Measures in Arabic Documents Hayel Hussain Khafajeh and Nidal Abedalrahman Yousef	160-166
22. Design and Implementation of Feynman Gate in Quantum-dot Cellular Automata (QCA) Md. Anisur Rahman, Fatema Khatun, Angona Sarkar and Md. Fazlul Huq	167-170
23. Face Recognition Using SVM Based on LDA Anissa Bouzalmat, Jamal Kharroubi and Aرسالane Zarghili	171-179
24. Blind Fake Image detection Nidhal Khdhair El Abbadi, Adil Mohamad Hassan and Mudher Mohammed Al-Nwany	180-186
25. Iterative Decoding Termination Schemes for Turbo Code Performance Optimization In Mobile Wi-Max Environment Jagdish D. Kene and Kishor D. Kulat	187-191
26. Design a Secure Electronic Voting System Using Fingerprint Technique Sanjay Kumar and Manpreet Singh	192-199
27. The Effect of Changing the Speed and the Number of Nodes on Packet Delivery Ratio in MANET Imad I. Saada, Magdy Z. Rashad and Mohamed A. Abu Eloud	200-202
28. The development of algorithms for alleviating the problem of discontinuity in speech synthesis from text written in Albanian Adnan Maxhuni, Agni Dika, Avni Rexhepi and Dren Imeraj	203-207
29. A Fuzzy Based Feature Extraction Approach for Handwritten Characters Mahmood K Jasim, Anwar M Al-Saleh and Alaa Aljanaby	208-215
30. RBAC Architectural Design Issues in Institutions Collaborative Environment Muhammad Umar Aftab, Amna Nisar, Muhammad Asif, Adeel Ashraf and Burhan Gill	216-221

31. An Efficient Approach for Sky Detection Irfanullah, Kamal Haider, Qasim Sattar, Sadaqat Ur Rehman and Amjad Ali	222-226
32. Real Time Network Server Monitoring using Smartphone with Dynamic Load Balancing Dhuha Basheer Abdullah and Zeena Abdulgafar Thanoon	227-232
33. Heuristics for Routing and Spiral Run-time taskMapping in NoC-based Heterogeneous MPSOCs Mohammed Kamel Benhaoua, Abbou El Hassen Benyamina and Pierre Boulet	233-238
34. Broad View of Cryptographic Hash Functions Mohammad A. Alahmad and Imad Fakhri Alshaikhli	239-246
35. Heterogeneous Vehicle Routing Problem with profits Dynamic solving by Clustering Genetic Algorithm Sawsan Amous Kallel and Younes Boujelbene	247-253
36. An Investigation of the Incidences of Repetitive Strain Injury among computer Users in Nigeria Olabiyisi Olutunde, Akingboye Yusuff, Abayomi-Alli Adebayo, Izilien Fred and Adeleke Iyiola	254-261
37. Improving the methods of email classification based on words ontology Foruzan Kiamarzpour, Rouhollah Dianat, Mohammad Bahrani and Mehdi Sadeghzadeh	262-266
38. OpenMP - Automatic Parallelization Tools: An Empirical Comparative Evaluation Emna Kallel Laadhar and Yassine Aoudni	267-275
39. Ontology based data warehouses federation management system Naoual Mouhni and Abderrafiaa El Kalay	276-281
40. An Effective Web Service Ranking Algorithm based on Quality of Experience of Users Vandan Tewari, Nirmal Dagdee and Aruna Tiwari	282-288
41. Energy Efficient Topology Control Approach for Mobile Ad hoc Networks T.S.Asha and N.J.R.Muniraj	289-296
42. A High-Speed Residue-to-Binary Converter for Three-Moduli Set Amani Goniemat and Andraws Swidan	297-305
43. An Energy Efficient Data Redundancy Reduction Approach for Data Aggregation in WSN Bharathi M A, B P Vijayakumar and D H Manjaiah	306-311
44. An online collaborative framework for orthography system development Sook-Kuan Chin, Alvin W. Yeo and Nadianatra Musa	312-315
45. Optimal Transmission Power of target tracking with Quantized Measurement in WSN Amr Lotfy Elewa Mohamed and Osama El-Ghandour	316-322

46. Dimension Reduction in Intrusion Detection Features Using Discriminative Machine Learning Approach Karan Bajaj and Amit Arora	324-328
47. A Novel Design to Increase Trust in Cloud IaaS model Jitendra Kumar Seth and Satish Chandra	329-336
48. Continuous Improvement of Production system in Algerian Industry Mechenene Athmane and Aouag Hichem	337-344

Keystroke Dynamics Authentication: A Survey of Free-text Methods

Arwa Alsultan¹ and Kevin Warwick²

¹School of Systems Engineering, University of Reading
Reading, Berkshire, UK

²School of Systems Engineering, University of Reading
Reading, Berkshire, UK

Abstract

Current computer systems depend greatly on authentication methods in order to provide sufficient protection to the data handled by these systems. Rather than using the common username and password scheme which suffers from many security and usability limitations, we investigate in this paper the use of keystroke dynamics as a more useable authentication alternative. We focus on the research done on free-text keystroke systems and its ability to provide continual identity verification during the whole time that the user is using the system.

Keywords: *free-text, keystroke dynamics, authentication, identification, performance, survey.*

1. Introduction

The use of computer systems has proliferated at an unforeseen rate. They are now used in almost all aspects of our lives. This is a strong reason to protect them against illegal intrusions. However many computer systems use the simple username/password scheme for authentication, even though it suffers from the security-usability trade-off dilemma. Passwords can be guessed using different methods such as social engineering, spyware, dictionary attack and mere brute force attacks. These are all reasons for the user to employ extreme measures to safeguard his/her computer by using long and complex passwords which are unfriendly and hard to memorize. It is therefore ideal to use an alternative authentication method that can be low-cost yet provide ease of use and transparency to the user in addition to security robustness.

Keystroke dynamics is a behavior biometric scheme that provides sturdy system protection while maintaining a high level of usability. In particular, using free-text keystrokes provides real-time identity verification by continuously monitoring the keyboard's activities. This is a very important, yet frequently ignored, part of the authentication process since it is fairly simple to establish a level of confidence about the user's identity at log-in time. However there is no guarantee that the user who was successfully authenticated is the same person who is still using the system. There is always a chance that the system was left unattended which is a golden opportunity for the attacker who is physically close to the machine to have access to it and, for example, alter some documents or send an e-mail on behalf of the original user.

In this method of authentication, it is not obligatory to memorize any text such as a password or a passphrase;

instead authentication is conducted through finding the resemblance of the typing rhythm of a user, in a non-intrusive manner, regardless of the text typed.

One important fact in looking at research to date in free-text keystroke systems is that results from most studies are far from ideal, i.e. either the resulted accuracy is not satisfactory or it has a high accuracy level which was obtained under strictly controlled conditions, which is not at all representative of real-life situations. Thus, we aim in this paper to look at the various factors that might affect the authentication system performance in addition to covering the methods used for feature extraction and classification. Situations where free-text keystroke dynamics are best used are also discussed in this paper.

The rest of this paper proceeds as follows. Section two introduces keystroke dynamics theory and describes the differences between fixed-text and free-text systems. The third section lists some of the techniques followed for feature extraction while the section after that lists the methods used for classification. Performance measurement schemes are considered next. After that we list some of the factors affecting performance in free-text systems. A variety of applications that can benefit from free-text systems is given in the seventh section. Finally we discuss the level of protection that free-text systems can provide against some of the common security threats.

2. Keystroke Dynamics

Monitoring keystroke dynamics is considered to be an effortless behavioral based method for authenticating users which employs the person's typing patterns for validating his/her identity. As mentioned in [1], keystroke dynamics is "not what you type, but how you type." In this approach, the user types in text, as usual, without any kind of extra work to be done for authentication. Moreover, it only involves the user's own keyboard and no other external hardware. The original idea of using keystroke patterns for user identification purposes was originated from the idea of identifying the sender of Morse code on a telegraph machine, where operators have been able to identify the sender of a message by the rhythm, pace and syncopation of the received taps [2].

As early as 1980, researchers such as Gaines et al. [3] started to show interest in proving the hypothesis that typing patterns can be used as a mean of user

authentication. Experiments were conducted to find typing patterns that can be used effectively for authentication. Results from these tests showed that the similarity between typing samples from the same person is high with respect to the time delays it takes the user when typing one key or two successive keys. All of this early research though was only concerned with keystrokes generated by typing fixed words.

It wasn't until 1995 when Shepherd et al. [4] showed interest in continuous authentication. In 1997, the first organized attempt to use free-text keystroke system was conducted by Monroe and Rubin [1] where both fixed-text and free-text were used. The overall performance was not encouraging for free-text giving only 23% correct classification while fixed-text produced about 90% correct classification. This shows the complexity of using free-text systems compared with the fixed-text systems. Nevertheless, free-text systems have gone a long way since that experiment and much better results have been obtained using more sophisticated techniques.

There are two main phases that a user has to go through in order to be authorized by keystroke dynamic systems; namely: the enrolment phase and the log-in phase. The first phase has to do with collecting data about the user such as username and password in addition to capturing the user's typing behavior. The system gathers the keystroke times and extracts the timing features to create a template for each user's typing behavior. This template, also referred to as a user's profile, is stored in a database in correspondence to the user's other details.

The second phase takes place whenever the user needs to actually use the system. At that time, the system collects the user's keystroke times and then extracts the timing features in the same manner pursued in the enrolment phase. After that, the system performs feature matching with the user's template which is stored in the database. Next, based on the results of the matching process, one of two actions will take place: granting access to the user if the two sets of data are sufficiently similar or denying access to the user otherwise.

Two types of keystroke systems are used and discussed in the literature; they are: fixed-text and free-text keystroke systems. Fixed-text, also referred to as static, obliges the users to use only a predefined text to produce the typing samples. The predefined text varies in the research done in this area in the way that some have utilized the same shared password for all users [5] and others used different fixed text for each user such as using the user's name [6] or log-in IDs [7]. The main function of the fixed-text systems is applying it at log-in time in order to verify the user's identity at the beginning of the session only. This is done by forcing the user to retype their password a number of times at the enrolment phase in order to determine the user's typing rhythm for that specific password. This is considered a critical usability issue because of the amount of burden it adds on the user; still, the user needs to memorize the predefined text. Generally speaking, fixed-text keystrokes are mainly used for password hardening.

Free-text systems, also known as dynamic, don't restrict users to a particular text; on the contrary, they are given complete freedom to use any text of any length without any constraints. Unlike fixed-text, free-text systems will continue to collect the keystrokes, after successfully passing the log-in session, throughout the whole time that the user is logged-in for the reason of assuring the identity of the user during the full duration of that session. In free-text systems, the user's typing pattern is typically monitored during several days where he/she is performing regular typing tasks such as writing e-mails or typing word documents i.e. the enrolment phase is long yet transparent to the user. Even though, free-text and fixed-text systems are quite similar in the way that they both utilize the key press and release times to build a user behavior profile, they clearly differ in the way that the system is trained and applied.

All keystroke dynamics studies involve conducting five main experiment parts in the following order: recruiting participants, requesting a typing task to be done by the participants, collecting the keystrokes timing data, obtaining timing features from the raw keystroke data, training the classifier using part of the keystroke data and using the other part for testing the classifier [8]. We will go through the previous mentioned stages in order to compare and contrast what has been done in this area as reported in the current literature.

3. Feature Extraction and Profile Creation

The manner in which user data is collected in free-text keystroke systems is quite different from that of fixed-text systems in the way that a user is normally monitored for a period of time, a number of days for example. From all the typing data collected during this time, the system infers the typing pattern that the user typically follows which will be then stored as the user profile. The time it takes to type single letters or combinations of letters i.e. di-graphs, tri-graphs, even longer combinations is considered in free-text keystroke systems, yet there is a condition for including a particular letter or combination of letters in the template. It has to be typed often enough during the enrolment phase which will cause its mean and standard deviation to be statistically sound [9].

This implies that it is not necessary to include all letters and letter combinations, typed during the enrolment phase, in the template. Therefore, much research includes a pre-processing stage for removing noise from the data set. Extreme duration or latency values, i.e. very small or very large outliers, are discarded; for example: only the durations and the latencies of keys for which the standard deviation was below a predefined value were added to the user's template in [10, 1] while minimum and maximum values were fixed for the latencies that were used in [11].

Timing features are basically calculated using the press and release times of every key the user types and then processed in a specific way before being stored in the user's profile. Different methods were followed to carry out this part of the system, as shown in Table 1. Here we

focus on some of the common methods used for feature extraction and profile creation in free-text keystroke systems.

First we go through some of the simple feature extraction techniques found in the literature. Profiles in [1, 23] consisted of the mean latency and standard deviation of each di-graph in addition to the mean duration and standard deviation of each individual key. While the profiles in [11, 10] only included the latencies' means and standard deviations for di-graphs that have occurred a minimum number of times. On the other hand, the down-down duration time of di-graphs was used in [12, 13, 14]. This was extended in [15] to include more n-graphs including di-graphs, tri-graphs, and other longer n-graphs.

Although, di-graph and tri-graph time has been used in plenty of research, Sim and Janakiraman [16] concluded from their several experiments that using di-graphs/tri-graphs is not a good discriminative between users when the actual typed words are not taken into consideration. This is because the context of the text that a particular letter is included in regulates the manner in which it is typed [9] i.e. the letter 't' has different duration in the word 'sentence' and 'question'. Therefore, di-graphs/tri-graphs are more effective for keystroke dynamics when using context-specific n-graphs.

A more structured feature extraction was followed in some research where the timing features were extracted for only a set of key pairs which helped to increase the number of the di-graphs that can be found and compared in both the training and the testing samples. This increases the stability of its mean and standard deviation, in addition to reducing the required computation time. This was done in [17] by dividing all keystrokes into four attributes: left hand side keys, right hand side keys, spacebar and backspace bar; then, creating 16 diagraphs using these attribute combinations.

A keyboard grouping technique was introduced in [18] for classifying the keys based on their location on the keyboard, which was divided into 8 sections; two left and right halves and then each half divided into 4 lines representing the rows of the keyboard. For example WM is represented as Left 2- Right 4. Moreover, only a fixed set of letters and two letter combinations were used in [9, 19]; these sets were chosen based on each letters frequency in the English language. Letters including E, A, T ... etc. and di-graphs including: AT, TH, HE ... etc. are frequently found in English text, therefore, it is a good idea to use the mean and standard deviation of their duration and latencies in the user template which will increase its stability.

More complex features were also taken into consideration for the purpose of distinguishing users typing behavior. In addition to the usual key-press duration and di-graph's down-down (duration) and up-down (latency) times, other features were utilized in [19, 20, 21]; such as: typing speed, error rate, press-release ordering and the percentages of using special characters. Other features that capture the editing patterns of the user which includes

the usage of specific keys i.e. Home, End, Backspace, Delete, Insert, shortcut keys, arrow keys ... etc. were also used.

Another interesting feature was used in [22]; where all the commands executed in the first 10 minutes were collected. Although, it might seem irrelevant on first glimpse, an attacker is more likely to hurry to execute as many commands as he can on the victim's machine during the first few minutes. This shows an obvious change in the users habits which can be used to detect illegal intrusion.

4. Methods

After extracting the users' typing features and creating their profile templates has been completed, the classification process is performed to find the similarities and differences between the user's template stored at the enrolment phase and the sample provided during the session the system is being used. Similar to fixed-text systems, many methods have been used for classification in free-text keystroke systems; ranging from simple statistical methods to more complex pattern recognition and neural network algorithms. Moreover, an even more sophisticated combination of methods was used in some cases. This section highlights the major classification approaches used in the current literature. Please refer to Table1 for more details.

Simple statistical methods were used as a classification mean for typing behavior in several free-text keystroke systems studies. A variety of distance techniques have been used; Euclidean distance [18], weighted Euclidean distance [23], scaled Manhattan distance [9] and Bhattacharyya distance [24] were all utilized to find the level of similarity between samples. In addition, other statistical techniques were also used; decision trees were used in [22] while Kolmogorov-Smirnov Test (KS-test) was used in [13, 17].

One of the most cited free-text studies was that conducted by Gunetti and Picardi (P&G) [15] which depended on two measures, the first of which was the relative measure which was used to find the degree of disorder between the two samples. The second was the absolute measure which was used to calculate the absolute distance between the two samples. In both the relative and absolute measures only n-graphs occurring in both typing samples were considered. Even though the results were very good, the computational costs required to identify users was expensive because it needed to compare the test sample with all users' templates in the database which obviously makes it less scalable. Hu et al. [25] attempted to solve the scalability issue of P&G's method using the k-nearest classifier. In this approach, training samples were divided into clusters such that, every test sample was compared only with the samples of those users in the same cluster. Results for this modification revealed accuracy which compared well with that of P&G. Computation speed, on the other hand, proved to be 66.7% better.

A number of extensions have been carried out on P&G's method by Davoudi and Kabir. In [12] they combined P&G's method with a distance-calculating method that used histogram-based density estimation for each di-graph in order to find the probability density function of the di-graph's duration time. While in [26], they modified the relative distance in the P&G method by choosing the di-graph with the highest difference in duration between the

two samples to compute the difference of its positions first. After that, it was removed from the two timing vectors, and then, the new vectors were sorted again. They also applied one further modification to P&G's method in [14] by adding a weight factor to the digraphs when computing the relative distance. This weight was defined as the ratio of the number of occurrences of this digraph and its standard deviation.

Table 1: Chronological list of free-text keystroke systems.

<i>Study</i>	<i>Features</i>	<i>Method</i>	<i>Subjects</i>	<i>Samples</i>	<i>Performance</i>
Monrose & Rubin [1]	Di-graph latency, key duration	Euclidian distance, probability score, weighted di-graph probability	31	-	23% accuracy
Gunetti & Ruffo [22]	Di-graph latency, executed commands	Decision tree	10	-	90% accuracy
Dowland et al. [11]	Di-graph latency	Mean, Standard deviation	4	-	50% accuracy
Gunetti & Picardi [15]	N-graph duration	Relative distance, absolute distance	205	765	0.005% FAR, 5% FRR
Villani et al. [20]	Di-graph latency & duration, key duration, typing speed, percentage of special characters, editing patterns	Euclidian distance, k-nearest neighbour	118	2360	99.8% - 44.2% accuracy
Curtin et al. [19]	Di-graph latency & duration, key duration, typing speed, percentage of special characters, editing patterns	Euclidian distance, k-nearest neighbour	30	-	100% - 97% accuracy
Filho & Freire [30]	Di-graph latency	Simplified Markov chain model	15	150	41.6% - 12.7% EER
Janakiraman & Sim [24]	Di-graph latency, key duration	Bhattacharyya distance	22	-	100% - 70% accuracy
Buch et al. [34]	Di-graph latency & duration, percentage of special characters	Euclidian distance	36	650	100% - 98% accuracy
Hu et al. [25]	N-graph duration	Relative distance, absolute distance, k-nearest neighbour	36	36554	0.045% FAR, 0.005% FRR
Hempstalk et al. [21]	Di-graph latency, key duration, typing speed, error rate, P-R ordering	One-class classification	10	150	11.3% FAR, 20.4% FRR
Ahmed et al. [29]	Di-graph latency	Neural network	22	-	0.015% FAR, 4.82% FRR
Davoudi & Kabir [12]	Di-graph duration	Relative distance, absolute distance, histogram-based density estimation	21	315	0.015% FAR, 0.0025% FRR
Pilsung et al. [13]	Di-graph duration	Kolmogorov-smirnov Test	-	-	0.17% EER
Samura & Nishimura [23]	Di-graph latency & duration, key duration	Weighted Euclidian distance	112	-	67.5% - 81.2% accuracy
Bours & Barghouthi, [10]	Di-graph latency, key duration	Distance measure	25	-	79 – 348 keystrokes
Davoudi & Kabir [26]	Di-graph duration	Modified relative distance	21	315	0.08% FAR, 18.8% FRR
Davoudi & Kabir [14]	Di-graph duration	Weighted relative distance	21	315	0.07% FAR, 15.2% FRR
Park et al. [17]	Key-pair duration	Kolmogorov-smirnov Test	35	-	0.089% EER
Messerman et al. [38]	N-graph duration	Normalized relative distance	55	-	2.20% FAR, 1.84% FRR
Singh & Arya [18]	Key-pair latency	Euclidian distance	20	-	0.02% FAR, 0.04% FRR
Chantan et al. [28]	Di-graph duration	Bayes classifier	-	-	0% EER
Bakelman et al. [27]	Di-graph duration	K-nearest neighbour	20	200	4% EER
Bours [9]	Di-graph latency, key duration	Scaled Manhattan distance	25	-	182 keystrokes

Pattern recognition methods were also exploited in order to be used as a classification method for free-text keystroke authentication. For example: K-nearest neighbor was used in [27] and Bayes classifier was used in [28].

Ahmed et al. [29] used a feed forward multi-layer perceptron neural network system for the purpose of classifying users. Two neural networks were used; a behavior-modeling network and a detection network. The first used the di-graph's first and second keys press and release times to find the elapse time it took a user to press two successive keys. The second neural network used the di-graph's times and the matching output from the behavior-modeling network to estimate which user's typing patterns it represented.

5. Performance

Unfortunately, not only keystroke systems but all biometric authentication systems sometimes suffer from mistakes in the authentication decision. This is due to a number of reasons that has to do not only with the efficiency of the technique but also with the user himself or with his surroundings. First of all it is possible, yet not likely, that an imposter is mistakenly identified as the legitimate user if by chance the two persons typing patterns are close enough to the extent that the classification method fails to distinguish between them. Conversely, when one of the legitimate user's fingers slips off the keyboard and causes the typing pattern to change slightly, the user may not be successfully authenticated. Thus, it is important to have some metrics to exactly measure the error rate which will help to identify the performance level that can be expected and tolerated by that system's users.

A very simple way to measure the error rate was used in earlier studies; using the Accuracy measure which is the percentage of successfully authenticated attempts compared to the total number of completed attempts. This technique was adapted in [1, 27, 22].

The most frequently used error rates for inferring the performance of an authentication system are: the False Accept Rate (FAR), also referred to as the Imposter Pass Rate (IPR) and the False Reject Rate (FRR), also called the False Alarm Rate (FAR). The FAR is the percentage of impostors who have successfully gained access to the system while the FRR is the percentage of legitimate users who have been denied access to the system. These two error rates were used by the majority of free-text keystroke systems including [15, 21, 18].

Clearly, there is a trade-off between the FAR and FRR which can be controlled according to the level of security strictness required. FAR is required to be as low as possible in strictly secure applications while there is a compromise of having a higher FRR. Meanwhile, a higher FAR is acceptable in systems where security is not the major aim yet system usability has higher priority.

The other commonly used error rate is the Equal Error Rate (EER), also referred to as Cross-over Error Rate (CER), which is the value where FAR and FRR are equal. It was used in many methods such as [17, 30, 27] where lower EER values indicate a more secure system.

Due to the fact that free-text keystroke authentication is a continuous process, another metric which defines exactly how much time, in number of keystrokes, did it take the system to discover that an imposter had had access to the system has been proposed in some studies. This aims to detect the impostor as fast as possible, incorporating as few keystrokes as possible. This implies that an attacker would be detected before he can do more harm to the system. A penalty-reward technique was introduced in [9, 10] where a user was initially given the highest trust level prior to the user being successfully authenticated via a static authentication procedure. During the typing session, the user obtained a reward which he received in the form of an increase of his trust level when he typed in a manner sufficiently close to his typing template. Likewise, he obtained a penalty in the form of a decrease of his trust level when he typed in a manner far from his typing template. The system then locks-out a user if his/her trust level falls below a pre-determined threshold.

6. Factors Affecting Performance

There are many different performance measures used to determine the error rate in free-text keystroke systems, it is therefore often difficult to compare studies. This is also due to not having any form of standardization in the data collection process in these different experiments. Even though, the error rate in study A is lower than the error rate in study B, that does not necessarily mean that the method adapted in A is better than that used in B. Different factors may have a positive or a negative impact on the authentication process regardless of the actual method's functionality. Standardization of such factors requires information exchange amongst researchers which would offer an improved comparing mechanism between different algorithms. There are a lot of different factors to be considered in free-text keystroke systems; a detailed list of these factors is provided in this section.

Nevertheless, there are some solutions that can be used to standardize the factors involved. The first solution is using a widely available automated program for collecting data. A broad range of software is available commercially; for example: BehavioSec and KeystrokeID. Another solution involves the use of standardized databases which has been formerly created and published for the purpose of keystroke dynamics research. A list of some of the databases available online can be found in [31]. Using these solutions could not only standardize the data collection method, it could also decrease a duplication of effort among researchers.

6.1 Environment Controlling

There are two basic categories in the way experiments have been conducted in free-text keystroke studies.

Experiments have either been conducted in a controlled environment or in an uncontrolled one. In a controlled environment, users are asked to type on a specific machine which has built-in software for recording the keystrokes. Thus, the same external conditions are consistent for all users. The issue with this kind of arrangement is that it may not have the same characteristics as those encountered in realistic situations, therefore, the response may not be representative of a user's typical typing patterns.

In uncontrolled environments, on the other hand, users are asked to either download a program on their machines to collect their keystrokes [24, 30] or to use an online data collection form [15, 25]. This indicates that the data is collected wherever and whenever it is convenient to the user. Although, this method provides a realistic representation of normal circumstances for the user, each user's surroundings can be very different, which makes the data harder to analyze. This might be the reason for inconsistencies in the keystroke data provided by the users. A lot research done using free-text keystroke systems has so far been conducted in uncontrolled environments due to the desire to imitate the lifelike conditions of a real authentication system [15, 12, 28].

6.2 Keyboard Type

Using different brands of computer has a big impact on the user's typing pattern since the keyboard of different brands differs in key size and spacing between keys which is clearly a reason that users may type differently than normal [31]. Furthermore, different keyboards have different key pressing sensitivity levels which consequently may affect the timing data collected from the users. Using a laptop keyboard adds another variation which can also affect the typing behavior; because laptops provide the freedom of movement, users may use it in different positions such as on a bed or on a table.

Villani et al. [20] investigated the case of using different keyboards in free-text keystroke systems. One of their experiments was conducted using a desktop keyboard and another was performed on a lab top keyboard. A significant finding was produced in this study which can be summarized as: the system has a good chance of accurately identifying a user as long as he uses the same type of keyboard for training and testing. It is therefore important that researchers attempt to stick to using the same keyboard in order to maintain the same level of consistency throughout the data collecting process [19].

6.3 Entry Mode

Because free-text keystroke systems are used for long text, it makes more sense to allow the users to enter whatever text they prefer. Having said that, studies conducted have actually used two different methods for text entry in the experiments conducted for free-text authentication. The first technique allowed the users to type completely free text as they desired, such as: typing an e-mail or typing a report for work or an essay for school [15, 16]. The second approach required the users to

type a specific long text from an article, in which the users needed to copy specific text into a section specified for text entry [19, 25].

In the research conducted by Villani et al. [20], participants were asked to be a part of several experiments with different conditions. One of these tests incorporated a copy-task in which the participants were asked to copy a predefined long text. Another included a free-text input where users were free to type-in arbitrary text. In this study, it was found that the accuracy of correctly authenticating a user decreased considerably when the user used different input modes in the training and testing phases. Moreover, it was also shown that the accuracy in free-text typing mode was higher than that in the copying-task mode. This can be explained by the frequent pauses that a user has to perform in order to look at the text during the copy-task which might cause the collected data to be inconsistent.

6.4 Text Length

One area that keystroke systems lack in is the amount of information that can be obtained. The only data that can be collected while the user is actually typing is the time each key is pressed and released, from which only little information can be inferred, including the time interval between each two consecutive keys and the duration time for each key press. In addition the data is often not stable since it changes based on the environment surrounding the experiment or based on the state of mind of the user at the time. As a result, to reduce the effect of such instabilities, much research has shown more interest in using short free-text [e.g. 28, 1, 18]. Realistically though, it is not enough to use short texts to analyze keystrokes since it does not offer an adequate amount of information to distinguish between users. Consequentially, using longer sample texts is considered a better alternative [11, 15, 16].

Moreover, Curtin et al. [19] provided evidence that using long-texts increases the chance of having more repetitions of the same di-graph in the training and testing samples which will, consequently, increase the stability of its mean and standard deviation significantly. The only problem with using long texts systems is that the training phase unavoidably needs more time. In their experiment, Curtin et al. investigated the accuracy of identifying users when typing long-texts under the condition that training and testing texts were different in length. The accuracy from different text/same length experiments was better than that from different text/different length experiments. Therefore, improving authentication accuracy can be achieved via standardization of the feature measurements i.e. the text size in this case.

6.5 User's Experience

The user's health and state of mind are a very crucial part of the authentication process using keystroke dynamics. The user's typing skills and level of comfort while using a keyboard are additional characteristics that have a clear impact on the user's typing behavior. The more skillful the user is, the more stable his/her fingers are located on

the keyboard and the more familiar he is with the position of each character on the keyboard. This will result in a more consistent typing pattern all through.

Samura and Nishimura [23] conducted a study that examined keystroke dynamics for long free-texts. The experiment participants were divided into three groups based on their typing speed, specifically the number of letters typed in a 5 minute period. This study indicated that the best recognition accuracy was obtained from the group which typed fastest.

6.6 Monitoring Mode

A free-text keystroke system is a continual process of identity verification which is taking place during the course of the whole time a user is using the system. This can be done in either a continuous manner or a periodic manner. Continuous authenticating is done, in real time, every time a key is clicked on the keyboard [9]. Although this method provides strong imposter detection, it is computationally expensive. Periodic authentication, alternatively, is repeated every time a certain text is entered [24]. This is a less strict method, security wise, yet it is computationally cheaper. Moreover, waiting until a specific text is entered may cause the system to wait for long periods of time if this particular text does not occur frequently enough in the typed text; which will represent a security threat for the system.

A periodic verification scheme that included the use of interruptions was utilized in [27]. In this research, the identity of the user was only verified after text breaks e.g. user leaving the PC for a coffee break. The system only captured the first burst of input after each pause in order to analyze it. The method does though reduce the frequency of authentication checks which is a key reason for reducing the false alarm rate in addition to decreasing the computational cost.

6.7 Words Choice

As mentioned, some free-text systems depend on periodic authentication where the authentication process is actually performed every time a particular text is entered. It is clear that choosing a specific piece of text is crucial for training and testing the system. It might be thought that using familiar English words may realize more consistent typing patterns. However this has been shown to be wrong by Janakiraman and Sim [24].

In their research, Janakiraman and Sim introduced a new "goodness" measure which was suggested to be used to calculate the universality, accuracy and expectancy of a word used for free-text keystroke authentication. Universality is a measure to identify if a word is one of the words commonly used by users or not. Accuracy measures how unique a word is. Lastly, expectancy is used to calculate the average number of keystrokes typed before that word actually appears in the text. Unexpectedly, using the goodness measure, the result of this experiment revealed that non-English words, such as: 'tmr' which is an abbreviation of 'tomorrow' used in

online chats, are better than English words for identification and verification purposes.

6.8 Number of Training Samples

When considering the training phase in fixed-text keystroke systems, it is hard to ignore the time required for training the system by retyping the password again and again. This is not an issue in free-text keystroke systems where the user's data is collected while performing daily tasks. This implies that the free-text method is more practical in real life situations and easier to use since it causes less burden for the user. For example, 15 samples were collected from the participants over a two weeks period in [27]; each sample was 400 characters long of whatever the user needed to type at the time. This demonstrates that even though the samples were long, they were collected transparently to the user.

From the experiment results conducted by Gunetti et al. [32] it was found that the accuracy of the system generally escalated when the number of samples in the user's profile was increased. Meanwhile an effective mechanism for profile enhancement was suggested in [18] where the user's profile was expanded, during the typing session, by adding new key-pairs timing data attained from text entered by the user after being authenticated.

7. Applications

Although more than a quarter of a century has passed since keystroke authentication was first researched, it has not yet been applied much in the security field. In addition to the security that keystroke authentication systems can provide by locking-up the workstation when an imposter is detected at any point of time during which the system is used, a wide variety of other applications can also benefit from such authentication schemes. The applications, listed in this section, are examples of some situations where free-text keystroke authentication is more applicable than fixed-text systems.

7.1 Different Languages Authentication

Most of the work done on keystroke dynamics has concentrated on using the same language for training and testing the system. Gunetti et al. [32] though gave empirical proof that free-text typing patterns could be used to authenticate the user even when the test samples were written in different languages to that of the samples in the user's profile. Evidently, this only works when the two languages share a significant number of di-graphs. So, languages like English and Italian which have largely the same alphabet can be used for this kind of authentication but English and Arabic, for example, cannot be used because they have a completely different set of letters.

The data used in this study was provided by Italian speakers each of whom provided two samples typed in Italian and another two samples typed in English. From the experimental results, about 10% mistakes in identification occurred when the test sample was in a

language different to that of the user's profile. Better performance was obtained when the user's profile contained samples in both languages. The error rate was even smaller when the test sample was in the same language as that which dominated the samples in the user's profile. By experimenting with different combinations of template samples and test samples, it was clear that samples provided by the same person while typed in different languages were more similar than samples provided by different persons while typed in the same language. An average performance of 1.61% FRR and 3.23% FAR was achieved in total. Thus, keystroke authentication for different language texts, is possible, though more difficult than the case where all samples are in the same language.

7.2 Old Profile Authentication

Most of the studies conducted in the free-text keystroke authentication field have had only a few months gap between the time the training samples were collected and the time in which the test samples were gathered. Gunetti et al., however, showed in [33] that a typing profile could still be used to identify a user, even though, it has been created a long time before the test samples are provided and investigated. Their original experiment involved 30 participants whom were asked to provide 15 samples each. The samples consisted of whatever the users choose to type. One and a half years later, the same 30 volunteers were asked again to provide another two free-text samples. It was discovered that even after such a long period of time their keystroke dynamics system was still able to identify users with an average accuracy of a 1.67% FAR and a 11.67% FRR.

7.3 Intrusion Detection

The continual authentication scheme that the free-text method provides is a very effective intrusion detection method. It is mainly used to notice any warnings with regards to irregularities in the typing patterns of a specific user. Moreover, free-text keystroke systems are used for active monitoring of the system which can aid in finding any intrusion quickly and reliably. One important issue that has to be addressed here is the generation of false alarms in continual keystroke-based authentication systems. It might cause frequent and rapid system halts with much annoyance for the users when they falsely occur. Therefore, Gunetti et al. [32] suggested using it combined with other authentication methods in order to reduce the false alarm rate.

7.4 Online Marketing

Free-text keystroke systems can also be utilized for identifying users over the internet. This is done by capturing a user's typing patterns on their first visit to the website and then it can be used to identify returning users [32]. This data can be used to determine user preferences and interests which can be employed for marketing purposes. This approach, on the other hand, has many

privacy issues regarding the amount of information that users are happy to hand-in to the websites they visit.

7.5 Cybercrime Investigating

User tracking through typing patterns can also help in cybercrime and investigating illegal electronic movements of anonymous users. Using free-text keystroke schemes was suggested for network forensics in [29] through attacker profiling which is conducted by collecting his/her typing patterns when surfing websites on the internet. This profile, collected for each user, can be used as a digital fingerprint gathered from the cybercrime scenes. This is considered as passive fingerprinting because it can be created without the knowledge of the attacker which can be extremely beneficial in fraud or identity theft cases where attackers are completely oblivious that they are being monitored. The issue with such a digital fingerprint is that it must be built progressively which requires a lot of internet service providers to collaborate and work together in facing such threats.

7.6 Identification and Authentication

Keystroke dynamics systems are used for two different purposes. Firstly: identification, which is a way of determining the user's identity when no data is available about their identity beforehand. In this method, a test sample is matched with all the templates stored in a database. The system assigns the user to the identity of the person whose template is the most similar to the test sample. The second purpose is authentication which is used to verify the identity of the user. The user supplies his identity and the system takes on the responsibility of making sure that the user is who he/she claims to be. The test sample in this case is only compared with the user's template in the database.

The complexity of performing identification is clearly higher than that of authentication since it includes comparing the test sample with all available templates which may be a very large undertaking in large scale systems. Identification also requires a larger amount of data i.e. longer text. From the definition of both methods, fixed-text keystrokes system is used mostly for authentication since it employs a password that is considered a mean for providing the user's identity. Free-text keystrokes system, on the other hand, is used for both identification and authentication [e.g. 34, 15].

7.7 User's Emotion Detection

Since free-text keystroke systems gather a lot of data from the user during the whole time he/she is using the computer, this data can also be used to infer the emotional state that the user is going through during the typing process. This has been employed in [35] to determine what the user is feeling during every day free typing. Feelings like frustration, focus, anger, stress, relaxation, excitement and tiredness were derived from the user's typing behavior. Extracting the emotional state that the user is going through in a particular period of time that the user is using the system has many benefits for intelligent

computers. It helps the system make the right decisions regarding the best interaction method to practice with the user. The issue with using keystrokes for user emotion detection is that it can cause an invasion into the user's typing experience. For example, in [35], the user was required to determine his emotion every 10 minutes in order to train the system to identify his emotions automatically.

8. Security Issues

In this section we discuss the security level that free-text keystroke authentication systems provide. A list of the most common threats is provided here along with the degree of safety that free-text keystroke systems deliver against these dangers [36].

1. Shoulder surfing and user mimicking: is an attack in which the attacker monitors the victim typing, during the typing process, in order to try imitating his/her typing behavior. Even though there is little possibility of an attacker successfully mimicking a user typing pattern in fixed-text keystrokes, it is even harder to do so in free-text systems. Since it requires the attacker to observe the user's behavior for the whole time the user is logged-in, it is very rare that an attacker can actually imitate all the aspects of the user's typing behavior.

2. Spyware: is software downloaded into the victims' computer without their consent which is used to record information about them. Spyware is perhaps the biggest threat to keystroke dynamic authentication systems because it can record exactly the time each key is pressed and released. This can be used by the attacker to simulate the legitimate user's typing behavior. Nevertheless, it is still a hard task for the attacker to undertake in the case of the huge amount of data that free-text systems need to analyze.

3. Social engineering: is manipulating the user in order to obtain his/her private information. Tricking the victim to reveal his typing pattern is though not possible using telephone calls or face to face meetings. Yet, phishing e-mails can be used to trick the user to type some text which can be used to extract the victim's typing patterns. But even then, the attacker has to get hold of a sufficient amount of keystrokes to be able to actually simulate the victim's free-text typing patterns.

4. Guessing: is trying to guess the way that a victim types. There are simply too many different ways that a user might normally follow when typing. Therefore, guessing the typing behavior of another person is almost impossible in free-text keystroke dynamics.

9. Conclusions

Free-text keystroke dynamics is a non-intrusive method, since it only uses the behavioral data that users convey during regular typing tasks. In addition to that, it is relatively inexpensive; the only required hardware is the keyboard. However the most important benefit that free-text keystroke systems provide is that the typing patterns

can still be used for authenticating users even after the authentication phase has passed. In addition, free-text authentication provides a valuable balance between security and usability which is highly desirable in the businesses world.

One concern about free-text keystrokes is that it tends to be instable in the sense that it might be influenced by the user state or by environmental conditions. Indeed some level of instability might occur without any obvious cause. Therefore, free-text authentication is probably best used as a part of a multi-factor authentication scheme [28, 37] that provides a higher level of security.

Generally, it is obvious that keystroke dynamics works more accurately for fixed-text compared with free-text. Therefore, it might be a good practice for free-text tests to take into consideration the actual words that the user is typing, in addition to the key hold time the di-graph's duration and latency times.

Moreover, determining the best method to follow to achieve the best authentication accuracy is not a straightforward task. Due to the variation of conditions that might be affecting the study participants, environment or procedure, the comparison between two or more methods is not always accurate. Therefore, a standardization mechanism has to be established to assure that factors affecting performance are in agreement in all the studies and hence can be properly compared.

Lastly, it is clear that the idea of using keystroke dynamics is not only restricted to the traditional keyboard, it can be conveyed to many other mechanisms like ATM machines and cell phones, which will then provide better every day protection for the standard user.

References

- [1] F. Monrose and A. Rubin, "Authentication via Keystroke Dynamics", in the Fourth ACM Conference on Computer and Communication Security, 1997.
- [2] M. Karnan, M. Akila and N. Krishnaraj, "Biometric Personal Authentication Using Keystroke Dynamics: a Review", Applied Soft Computing, Vol. 11, No. 2, 2011, pp. 1565–1573.
- [3] R. Gaines, W. Lisowski, S. Press and N. Shpiro, "Authentication by Keystroke Timing: some Preliminary Results", Rand Corporation, Rep. R-256-NSF, 1980.
- [4] S. J. Shepherd, "Continuous Authentication by Analysis of Keyboard Typing Characteristics", in European Convention on Security and Detection, 1995, pp. 111-114.
- [5] R. Giot, M. El-Abed, B. Hemery and C. Rosenberger, "Unconstrained Keystroke Dynamics Authentication with Shared Secret", Computers and Security, Vol. 30, 2011, pp. 427-445.
- [6] J. Garcia, "Personal Identification Apparatus", U.S. Patent 4621334, 1986.
- [7] M. S. Obaidat and B. Sadoun, "Verification of Computer Users Using Keystroke Dynamics", in IEEE Transactions on Systems, Man and Cybernetics - Part B: cybernetics, 1997, Vol. 27.
- [8] K. S. Killourhy, "A Scientific Understanding of Keystroke Dynamics", PhD thesis, Computer Science Department, Carnegie Mellon University, Pittsburgh, US, 2012.

- [9] P. Bours, "Continuous Keystroke Dynamics: a Different Perspective Towards Biometric Evaluation", Information Security Technical Report, Vol. 17, 2012, pp. 36-43.
- [10] P. Bours and H. Barghouti, "Continuous Authentication Using Biometric Keystroke Dynamics", in The Norwegian Information Security Conference, 2009.
- [11] P. S. Dowland, H. Singh and S. M. Furnell, "A Preliminary Investigation of User Authentication Using Continuous Keystroke Analysis", in the IFIP 8th Annual Working Conference on Information Security Management and Small Systems Security, 2001.
- [12] H. Davoudi and E. Kabir, "A New Distance Measure for Free Text Keystroke Authentication", in the 14th International CSI Computer Conference, 2009, pp. 570-575.
- [13] K. Pilsung, P. Joosong, P. Sunghoon, Y. Joonha and C. Sungzoon, "Keystroke Dynamics Analysis Based on Long and Free Text", in Fall Korean Industrial Engineering Conference, 2009.
- [14] H. Davoudi and E. Kabir, "Modification of the Relative Distance for Free Text Keystroke Authentication", in the 5th International Symposium on Telecommunications, 2010.
- [15] D. Gunetti and C. Picardi, "Keystroke Analysis of Free Text", ACM Transactions on Information System Security, Vol.8, No.3, 2005, pp. 312-347.
- [16] T. Sim and R. Janakiraman, "Are Digraphs Good for Free-text Keystroke Dynamics?", in the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp.1-6.
- [17] S. Park, J. Park and S. Cho, "User Authentication Based on Keystroke Analysis of Long Free Texts with a Reduced Number of Features", in the Second International Conference on Communication Systems, Networks and Applications, 2010.
- [18] S. Sing and K.V. Arya, "Key Classification: a New Approach in Free Text Keystroke Authentication System", in Third Pacific-Asia Conference on Circuits, Communications and System, 2011, pp. 1-5.
- [19] M. Curtin, C. Tappert, M. Villani, G. Ngo, J. Simone, H. St. Fort, and S.-H. Cha, "Keystroke Biometric Recognition on Long-text Input: a Feasibility Study", in IMECS, 2006.
- [20] M. Villani, C. Tappert, G. Ngo, J. Simone, H. St. Fort and S. Cha, "Keystroke Biometric Recognition Studies on Long-text Input Under Ideal and Application-oriented Conditions", in the IEEE Computer Society Workshop on Biometrics, 2006.
- [21] K. Hempstalk, E. Frank and I. H. Witten, "One-class Classification by Combining Density and Class Probability Estimation", in the European Conference on Machine and Learning and Principles and Practice of Knowledge Discovery in Database, 2005, pp.505-519.
- [22] D. Gunetti and G. Ruffo, "Intrusion Detection through Behavioral Data", in the Third International Symposium on Advances, 1999, pp. 383-394.
- [23] T. Samura and H. Nishimura, "Keystroke Timing Analysis for Individual Identification in Japanese Free Text Typing", in ICROS-SICE International Joint Conference, 2009.
- [24] R. Janakiraman and T. Sim, "Keystroke Dynamics in a General Setting," in Advances in Biometrics, 2007, Vol. 4642, pp. 584-593.
- [25] J. Hu, D. Gingrich and A. Sentosa, "A K-nearest Neighbor Approach for User Authentication through Biometric Keystroke Dynamics", in IEEE International Conference on Communications, 2008, pp. 1556-1560.
- [26] H. Davoudi, E. Kabir, "User Authentication Based on Free Text Keystroke Patterns", in the 3rd Joint Congress on Fuzzy and Intelligent Systems, 2010.
- [27] N. Bakelman, J. V. Monaco, S. H. Cha, and C. C. Tappert, "Continual Keystroke Biometric Authentication on Short Bursts of Keyboard Input", in Pace University CSIS Research Day, 2012.
- [28] C. Chantan, S. Sinthupinyo and T. Rungkasiri, "Improving Accuracy of Authentication Process via Short Free Text Using Bayesian Network", International Journal of Computer Science Issues, Vol. 9, No. 3, 2012.
- [29] A. A. E. Ahmed, I. Traore and A. Almulhem, "Digital Fingerprinting Based on Keystroke Dynamics", in the Second International Symposium on Human Aspects of Information Security and Assurance, 2008.
- [30] J. R. M. Filho and E. O. Freire, "On the Equalization of Keystroke Timing Histogram", Pattern Recognition Letters, Vol. 27, No.13, 2006, pp. 1440-1446.
- [31] S. P. Banerjee and D. L. Woodard, "Biometric Authentication and Identification Using Keystroke Dynamics: a Survey", Journal of Pattern Recognition Research, Vol. 7, 2012, pp. 116-139.
- [32] D. Gunetti, C. Picardi, and G. Ruffo, "Keystroke Analysis of Different Languages: a Case Study", in the Advances in Intelligent Data Analysis, 2005, Vol. 3646, pp. 133-144,
- [33] D. Gunetti, C. Picardi, and G. Ruffo, "Dealing with Different Languages and Old Profiles in Keystroke Analysis of Free Text", in the Advances in Artificial Intelligence, 2005, Vol. 3673, pp. 347-358.
- [34] T. Buch, A. Cotoranu, E. Jeskey, F. Tihon and M. Villani, "An Enhanced Keystroke Biometric System and Associated Studies", in Pace University CSIS Research Day, 2008.
- [35] C. Epp, M. Lippold and R. L. Mandryk, "Identifying Emotional States Using Keystroke Dynamics", in the Conference on Human Factors in Computing Systems, 2011, pp. 715-724.
- [36] D. Shanmugapriya and G. Padmavathi, "A Survey of Biometric Keystroke Dynamics: Approaches, Security and Challenges", International Journal of Computer Science and Information Security, Vol. 5, No. 1, 2009.
- [37] O. S. Adeoye, "Evaluating the Performance of Two-factor Authentication Solution in the Banking Sector", International Journal of Computer Science Issues, Vol. 9, No. 2, 2012.
- [38] A. Messerman, T. Mustafic, S. A. Camtepe and S. Albayrak, "Continuous and Non-intrusive Identity Verification in Real-time Environments Based on Free-text Keystroke Dynamics," in the International Joint Conference on Biometrics, 2011, pp.1-8.

Arwa Alsultan is pursuing a PhD degree in Computer Science from the School of Systems Engineering at the University of Reading, Reading, Berkshire, UK. She completed her Master's degree in Computer Science from the Computer and Information Science College at the King Saud University, Riyadh, SA in 2010. She works as a lecturer at the IT Department in the Computer and Information Science College at the King Saud University, Riyadh, SA.

Kevin Warwick is Professor of Cybernetics at the University of Reading. His research interests are in Artificial Intelligence, Robotics, Biomedical Engineering and Control Systems. He has D.Sc. degrees from both Imperial College London and the Czech Academy of Sciences. He has published over 500 research papers and is perhaps best known for his experimentation with implant technology.

A New Energy Consumption Algorithm with Active Sensor Selection Using GELS in Target Coverage WSN

Ali Bagrezai¹, Seyed Vahab AL-Din Makki², Ali Shokouhi Rostami³

¹Department of Communication, Kermanshah Science and Research branch, Islamic Azad university, Kermanshah, Iran

²Electrical Department, Engineering Faculty, Razi University, Kermanshah, Iran.

³Department of Computer, Islamic Azad University, Behshahr Branch, Iran

Abstract

In wireless sensor network, due to impossibility of replacing battery, the problem of energy and network lifetime is one of the important parameters. In asymmetric sensor networks, due to limited range of normal sensors it is not possible to communicate directly with central station by these sensors. In noted network, manager nodes are used which have more energy, processing power and broader telecommunication range. Connectivity and sending information to central station are done through them. The optimal selection and considering the energy of intermediate nodes to select and transmit data and also increasing network lifetime is one of the most important parts of wireless network design. In this paper, a gravitational force algorithm is used to solve the problem that is a power aware Selection algorithm in sensor network.

Keywords: *asymmetric sensor network, point coverage, the network energy, gravitation, velocity, Newton's law*

1. Introduction

Recent technology developments in micro-electro-mechanical systems and in integrated circuits led to development of small sensors with high processing power of information and low power consumption. These sensors have numerous applications such as multimedia, medical, surveillance, military telecommunications and home applications. Pocket PC, pager and cell phones are among them. A set of these sensors make a powerful network as wireless sensor network that is able to sample from local values ,process and send them to other sensors and finally to main observer (user).

Service quality is a versatile combination with multiple meanings and is one of network designer's goals .in order to achieve such design, many mechanisms are designed and evaluated [1,2].

The main challenge in wireless and mobile systems design Originate from two main sources of these systems i.e. telecommunication bandwidth and energy. To solve these limitations it is needed to Design telecommunication

techniques to increase bandwidth needed for each user and design powerful protocol for efficient use of energy. Designs will be different Depending on expected capabilities of system and in various applications.

For example, in many applications, the optimal number of nodes and consumed energy in executive rounds, and maximizing network lifetime are basic requirements of network. (In networks classification, Time interval of network activity is divided into certain parts so that each interval just after choosing selected category is activated in size of that time and other nodes of network will turn off. This part is called a round).

Clustering is a solution for this. To collect and aggregate data in a sensor network, the nodes can be organized in small groups called cluster. Each cluster contains a central node called cluster heads and some member nodes. A two-level hierarchy of cluster heads (in high level) and member nodes (in low level) is constructed through clustering [3].

As replacing battery in many applications is not appropriate, low energy consumption is one of the basic needs in these networks and lifetime of each sensor can be effectively increased by optimizing energy consumption [4]. Schemes that are efficient In terms of power have applied more in these networks. These schemes are being investigated in all layers of network in two aspects of hardware designing and algorithm and protocol designing. One way to reduce energy consumption is to decrease the number of sensors in sensing area to ensure identification of each target in the area. If the network is scalable, the Algorithms to decrease number of sensors can be efficiently implemented [5].

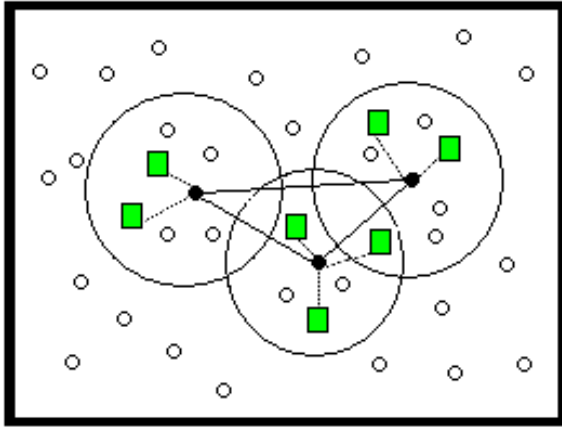


Fig. 1. Point coverage [6]

In point coverage, the aim is to produce coverage in set of points. A set of sensors is shown in figure 1 that are randomly arranged to cover a set of targets (square nodes which are green). The connected black nodes produce a set of active sensors that are result from a timing mechanism [6]. Point coverage scenarios have many applications. In this scenario, a number of targets with certain position are considered that should be controlled. A large number of sensors are randomly distributed Very close to targets. These sensors send collected information to central processing node. Based on this method, each target should be controlled by at least one sensor at any moment, assuming that each sensor is able to control all targets in its sensing rang. One way to reduce energy consumption is to decrease number of active sensors in Coverage area. A method for increasing sensor network lifetime through saving energy is to divide a set of sensors into several separated sets. This classification should be in a way that each set covers all targets completely. These separated sets are activated consecutively so that only one set is active at any moment [7].

2. Related works

Many researches have been investigated in the field of power aware algorithm and optimization of power consumption. Carburnar and et al presented a way to save energy consumption by detecting position of sensors and decreasing their overlapping [8]. In [6], a method is presented to save energy consumption. Based on this method, each area of sensors limit is divided into two sets. Only a set of them are active at any time and they will be activated alternatively. Based on the method presented in [7], at first, nodes are active or inactive distributive to obtain considered coverage range and remain unchanged. In the networks that the nodes distributed statistically, there is problem of heterogeneous distribution of energy in

nodes. In fact, as the sensor is closer to target, its energy consumption is more. So, connectivity and network coverage will not be ensured completely [9]. In [10], a method is presented to achieve a scalable coverage. This method is used to enhance energy efficiency when there is high computational complexity and slag. In method proposed in [11], at first the sensors are randomly distributed then a self-healing algorithm is used to produce a complete coverage. Energy optimization is obtained based on energy optimizer algorithm. Numerical simulation verifies lower energy consumption in this network in comparison with initial randomly distributed network. In [12], it is noted that one way to reduce energy consumption is to decrease energy consumption in boundaries of covered regions. In [13], a method is presented to reduce the number of sensors and energy consumption based on biological algorithms. The advantage of this method compared to other methods is uniform distribution of sensors. In [14], a method is presented to increase network lifetime which is based on maximizing the number of sensor classes. In this method, a node is allowed to be a member of more than one group which will increase network lifetime. In [14], a relation is presented for sinks velocity, energy optimization and reduction of data packets Failure possibility. In another research, in paper [15] Using data transmission in multiple paths the network is resistant to node Failure. Here, each node determines its next hop based on a node which has the highest residual energy. In [16], a method is presented for clustering. This method is based on maximum delay and wasted energy by intermediate nodes and cluster size. This algorithm is based on creating a spanning tree whose root is node of cluster head. In another research, in [17], unlike other papers about sensing range of node, a disk with fixed radius around the node is not considered. In this paper, a scheme is presented to reduce network power consumption by establishing cooperation between nodes. In article [18], Data Compression Problem in wireless networks is formulated according to energy. In this model, a percentage of each sensor data doesn't send; based on data correlation, therefore, result in reduced energy consumption. It is also shown that the greedy method is the most optimal method in terms of energy consumption. In the paper [19], a number of intermediate nodes are distributed in the network to save energy. This paper aims to obtain distribution of intermediate nodes in network So that the network lifetime increases. In [20], hierarchical clustering is used in wireless networks to achieve lower energy consumption. Considering the cluster head, energy consumption needed for communication of each node with processing center will decrease. In many studies including references [5-8], clustering methods are presented to reduce the number of clusters. In [21], methods are proposed to decrease power consumption and increase

lifetime of target coverage network. In this paper, the advantage of using an algorithm based on greedy protocol is presented. In [22], to select relay stage, two algorithms based on clustering -the shortest distance and greedy algorithm are compared. In presented greedy algorithm, Classes of nodes that do not have manager node are merged with other classes of nodes. This process continues until all classes will obtain manager nodes. Although this method has low computational and telecommunication complexity, it is not an optimal method to decrease network energy consumption. Another method presented in [22] is the shortest distance algorithm. In [23], a method is presented to increase network lifetime based on gravitational algorithm. This algorithm aims to increase network lifetime by optimizing and decreasing energy consumption and increasing productivity of monitoring network. In this paper, each node that is active in current round finds its shortest distance to closest manager node. The shortest distance Selection decreases network energy consumption.

3. Energy model

This algorithm is based on timing protocol of activity duration networks. Timing protocol is one of the grouping protocols which are placed in sensor networks. It uses two-step mechanism (initiative and executive) and works on the basis of data communication in shape of single-hop or multi-hop Including some super nodes and relay and monitoring sensor. In this protocol, group selection is done by using size function designed in the protocol.

In initiative phase, some nodes which is called "sensor" send their propagation messages to their neighbors. In second phase (executive) which is known as stable phase, data reception or transmission is done from sensor nodes to relay nodes and from relay nodes to destination. Figure 2 shows the plan of protocol operation.

Some nodes of super nodes transmit data carefully, like LEACH algorithm [15], According to the plan The energy is saved by grouping in remaining time of inactive nodes. In grouping protocols, energy consumption is constant in whole network due to periodic circulation of active sensors. Hence, we used this feature in our paper.

As shown in Figure 2 each round includes two phases: initiative phase and executive phase. The initiative phase includes two parts. The former is devoted to monitoring sensors selection. The latter is for relay sensors selection. It is obvious that using super nodes increases network lifetime.

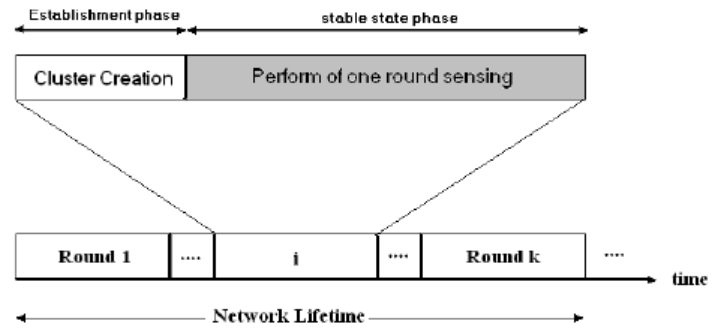


Fig. 2. Timeline of proposed protocol Performance

3.1 Energy Model

The energy model is considered for transmitting and receiving one of data in accordance with LEACH energy model. Assume that the distance between a transmitter and a receiver is d in energy model mentioned above. If d is more than d_0 , the multi-path model (with less path coefficient 4) is used; otherwise open space model (with less path coefficient 2) is used.

$$E_{Tx}(l, d) = E_{Tx-elec}(l) + E_{Tx-amp}(l, d) = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2 & d < d_0 \\ lE_{elec} + l\epsilon_{fs}d^4 & d \geq d_0 \end{cases} \quad (1)$$

E_{elect} Is required energy to activate the electrical circuit
 ϵ_{mp} and ϵ_{fs} are activation energies for power amplifiers in multi-path and open space modes, respectively. Its general form is represented:

(1) With constant coefficients p and q

(2) In receiver case

$$E_{Tx}(l, d) = p + qd^\alpha \quad (2)$$

The consumed energy is received with one of data sizes (3).

$$E_{Rx}(l) = E_{Rx-elec}(l) = lE_{elec} = p \quad (3)$$

In presented asymmetrical networks, it is assumed that initial energy of super nodes is several times greater than initial energy of normal sensors. The consumption energy of a relay and monitoring node are denoted by E_{s1} and E_{c1} in each round respectively.

3.2 Sensor Network Protocol Design

The problem is emphasized on how to design a protocol to increase network lifetime and decrease energy consumption in available nodes. The benchmarks are trying to use more from usual energy of sensors.

In covering networks, the physical positions of nodes and times of using them should be considered in designed protocol. The times of using sensor and also the distance between selected node (in fact in relay path) and super nodes have crucial role for energy consumption of that group. Therefore, we should seek for a relation between these two parameters and their energy consumption. At first, we state the problem and considered situations. Then, similar parameters that include timing algorithm based on the super nodes (for point coverage) will be explained below.

Our network contains N sensors named S_1 to S_N . We have M super nodes named S_{u1} to S_{uM} ($M < N$). The proposed timing algorithm is divided into time intervals with certain rounds and identical intervals T_r .

Selected group is only active during time of T_r and other nodes are off during a round. during round a, T_r can be computed by considered grouping time, the groups of energy estimate physical parameters of lifetime and types of normal sensors is used in network.

3.3 dominating Provisions network

The provisions of network are listed below:

There are K targets with defined positions in network composed of sensor nodes and super nodes. In considered scenario, sensor nodes and super nodes are randomly distributed. This plan of sensor nodes activities must be guaranteed according to following conditions after running algorithm for network lifetime:

- Targets T_{a1} to T_{ak} must be covered.
- There are nodes S_1 to S_N which perform monitoring task and are deployed randomly.
- The super nodes S_{u1} to S_{uM} are deployed.
- A set of nodes C_1 to C_j should be selected. Each C_j is set of active nodes and is generated by protocol in each round.
- Each set of C_j is necessary and sufficient to cover k targets.

In fact, the objective is to divide sensor nodes into active and inactive groups. Active sensors must be able to

communicate and cover. The objective is to use this algorithm for maximizing the groups, reducing energy consumption and increasing network lifetime. In each executive round, it should be checked whether a node is active as a sensor node or a relay node.

- ✓ Each normal sensor has initial E_i and high processing power. Common sensors Dissimilar to super nodes have higher energy, greater lifetime and higher processing power.
- ✓ All super nodes are connected to each other by a path between two super nodes.
- ✓ Each active sensor exists in one of C_j groups and connected to a super node by relay nodes. Each sensor is connected to one of super nodes through Data transmission path.
- ✓ Sensor nodes possess initial energy E_i , communication range R_c , and sensing range R_s ($R_c \geq R_s$).
- ✓ This selection must be located and distributed. Decision making is done for using data in neighboring node with fixed multi-hop distance.

Definition 1:

In defined point coverage, it should be said that when the Euclidean distance between nodes and target is Less than or equal R_s , the target is covered.

Definition 2:

Sensors can connect to each other or super nodes if the Euclidean distance is less than R_s .

Definition 3:

Network lifetime is defined as time interval in which all k targets will be covered by a set of active sensor nodes that are connected to super nodes.

3.4 Sensor Nodes Selection Algorithm

As indicated before, designed grouping algorithm which are executed at the beginning of each performance round, includes two sections. The first section is selected active nodes. The second section is attributed to data collection from nodes and data transmission through relay nodes.

In the first section, one of C_j groups is formed in a way that must be satisfied in above provisions. When this group is active, all other nodes are inactive (Sleep Mode) and consume little energy.

They should be evaluated in next phase. This evaluation is done by considering a series of physical factors of sensors during a round.

3.5 System Specifications

The network is offered a squared environment. There are T_{a_k} targets in the environment that are covered with connection of covering network.

$Tars_n$ Includes all targets in sensing domain S_n . They are covered by nodes. The number of targets is located in sensing range of node S_1 which is shown by m_1 .

The initial energy of common sensors is E_i and initial energy of super nodes is three times greater than E_i . The energy consumed in each round is called Es_1 and the consumed energy of a relay in each round is called Ec_1 .

The first section include sensor node selection, checking size function for evaluation and selecting active monitoring nodes that are w time units (the Second is the time unit here). The waiting time of node S_n is computed by a function measuring physical parameters of sensor S_n . Waiting time is stated as a multiple coefficient for total time of a round by using the parameters of a node: remaining energy, initial energy and number of targets seen in the range of a sensor.

A sensor decides to sleep or awaken after passing remaining time.

If $E_n < E_{s_1} + E_{c_1}$ (E_n is remaining energy of sensor node S_n) then the node cannot be converted to a sensor node.

So waiting time is not computed and t_n is waiting time of node n which is equivalent to w. It means that the node is not a sensor.

Otherwise, when $E_n > E_{s_1} + E_{c_1}$, t_n is computed and inspected. When t_n is finished $Tars_n \neq \phi$, S_n introduces itself as a sensor node and joins active nodes in the group.

Then, new selected node shows the position of two-hop neighboring nodes. If there is a node such as S_j at the end of the round that $Tars_n \neq \phi$ and $E_n < E_{s_1} + E_{c_1}$, the node sends the “no coverage” message to super node. It means that network lifetime is terminated. At this time, a message containing “no completed coverage” is sent to super nodes and the network sends this message to final monitoring destination.

4. Gravitational force

In 1995 for the first time vadoris and Tsang [24] proposed GLS algorithm to search and solve NP-complete problems

and in 2004 Barry Webster [25] presented it as a robust algorithm and named it GELS.

The idea of this algorithm is based on gravitational force principle that causes objects are attracted to each other in the nature, So that the heavier object has more gravitational force and imposes it on other objects and attracts the objects with lower weight toward itself. However, the Distance of two objects is very effective on size of this force; consider two objects with same weight and different Distance compared to object with less weight, the Object which has less distance to low weight object can impose more gravitational force on it. In GELS, Newton's law of gravity formula between two objects is:

$$F = \frac{Gm_1m_2}{r^2} \quad (4)$$

Where m_1 and m_2 are mass of first object and second object respectively. G Equals to gravitational constant value 6.672, R is the radius parameter and the distance between two objects.

Also GELS imitates this process of nature to search through a search space. So that search space, world and objects in this world are possible responses to search. Each object has a weight, the weight of each object is the performance or the search criteria, in which the best response has maximum weight and none of objects, cannot hold a zero weight [26- 28].

In this way, the possible responses in search space based on the criteria that depends on type of problem are divided into categories that each category is known as a dimension of problem response and a value called initial velocity is considered for each dimension of problem response which will be explained below.

GELS include a vector whose size specifies the number of response dimensions. The values of this vector represent relative velocity in each dimension. The algorithm starts with an initial response, initial velocity vector and movement direction. For each dimension in velocity vector, a random number between one and maximum speed is selected and it is the value of each element in each dimension. The initial response is generated by user or randomly as current response.

For each dimension in initial velocity vector, according to initial velocity vector of response dimensions, a direction is selected to move which is equals to response dimension that has maximum initial velocity in initial velocity vector.

The algorithm consists of a pointer object that can move in search space and the weight considered for object pointer is fixed in all calculations and the object always refers to a response with maximum weight. The algorithm is completed with occurrence of one of two conditions: All

components of initial velocity vector are zero, or number of algorithm iterations reaches its maximum.

In used Newton's formula, by replacing two mass in numerator of equation and replacing with difference between cost of candidate response and current response, the gravitational force between two objects is calculated using the following equation:

$$f = \frac{G(CU - CA)}{R^2} \quad (5).$$

Where CU and CA are cost of current response and candidate response respectively. This formula has a positive value if cost of current response is greater than cost of candidate response and has a negative value if cost of candidate response is larger. Then value of this force, positive or negative, is added to velocity vector in status of current path. If this action causes the value of velocity parameter exceeds the maximum setting, it takes a maximum value. If the update results in negative value, it takes zero.

The available Parameters in GELS:

Maximum Velocity: The maximum value that can be allocated to each element of initial velocity vector and this parameter prevents from getting too big.

Radius: the Radius which is used in the formula to calculate the gravitational force.

Iteration: Defines the maximum number of algorithm iterations which Ensures that the algorithm is terminated [26].

5. The proposed algorithm

In new proposed method, gravitational emulation local search algorithm (GELS) is used as a strategy to select optimal sensors. The choice is done for monitoring in Point coverage wireless sensor network. The goal of this algorithm is to increase network lifetime by optimization and reducing power consumption and increasing monitoring network efficiency. At first For each executive rounds in network, a number of sensors are activated to monitor, they Will lose some of their energy For each activation and These sensors must be chosen in a way To ensure that these sensors will cover all points needed for monitoring and Also there will be a distance from these nodes to the sink which has a cost.

To solve the problem of optimal sensor selection for monitoring in Point coverage wireless sensor network ,at first we consider three distance matrix, initial velocity matrix and time matrix which distance matrix, initial velocity matrix are randomly produced. In velocity matrix,

an initial velocity will be given to each available sensor in the network which is considered as a mass and then in later stages the speed will change besides the time matrix is obtained from following equation Based on distance matrix and velocity matrix:

$$T = \frac{\sqrt{(Y_B - Y_A)^2 + (X_B - X_A)^2}}{V_{in A,B}} \quad (6)$$

Then After creation three mentioned matrices, Sensors will be placed randomly within an array and For each executive rounds in network, the sensors which cover limitations of problem and have greater mass and speed and shorter distance compared to targets of network and also have less frequent than other sensors and they Have been selected with this condition that they perform Monitoring in network, will be activated. At this time, a solution has created for problem and then the Suitability of solution will be calculated and will be recognized as a mass of that solution. According to law of gravity, the best solution must be the largest mass.

Next solution will be created from current solution based on problem limitation that the suitability of this solution is calculated and will be known as mass of solution and In the case of optimization to current solution, created Solution is chosen as current solution I.e. the problem will go toward optimization.

If created solution is not optimal compared to current solution, The algorithm does not consider created solution and then makes another solution from current solution and examines it And this action is repeated until the algorithm return the optimal solution. The algorithm is completed with occurrence one of two conditions:

All components of initial velocity vector are zero, or the number of algorithm iterations reaches its maximum.

6. Simulation results

To evaluate the performance of proposed algorithm, the software C# is used. Sensor network is simulated in various modes. Selection methods to simulate and compare interface distance Selection, include a method based on greedy algorithm for distance selection, a method based on clustering -the shortest distance Selection and the method proposed in this paper. Table 1 is used for above simulations and proposed algorithm is compared with sources [22] and [23] and [29] and [30].

It should be noted that in proposed algorithm the run time of each round of proposed algorithm due to algorithm simplicity and low slag Is very low and very substantial

compared to other algorithms that are simulated ,in addition it reduces energy consumption And lifetime that can be seen in the simulation.

Table 1: The values used in simulation

Parameter	Value
Network Size	500 * 500 m
SNodes Location	Random
Nodes Location	Random
Nodes Initial Energy	0.1 J
SuperNode Initial Energy	0.5 J
Communication Range	90 m
Sensing Range	60 m
Number of Nodes	300
Number of SNodes	25
Number of Target	20
Eelec	50 nJ/bit

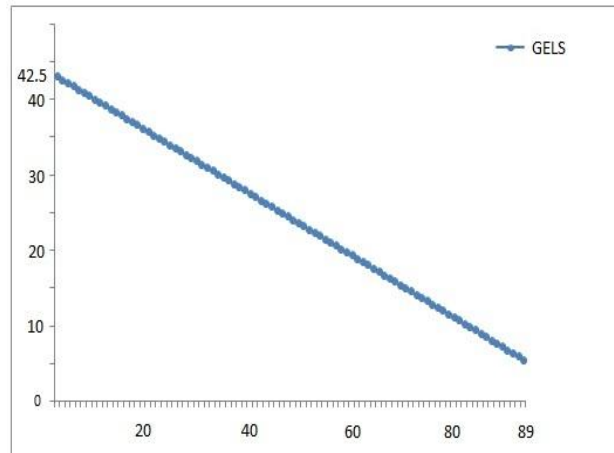


Fig. 5. Results of the gravity algorithm for 300 Normal sensor, 25 Manager Sensor and 20 o target

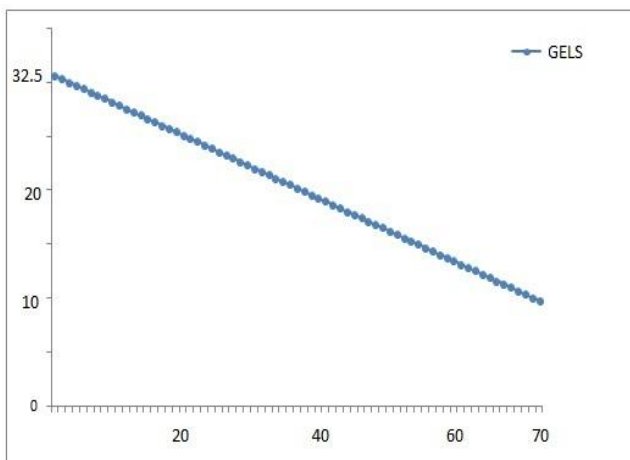


Fig. 3. Results of gravity algorithms for 200 Normal sensors, 25 Manager Sensor and 20 targets

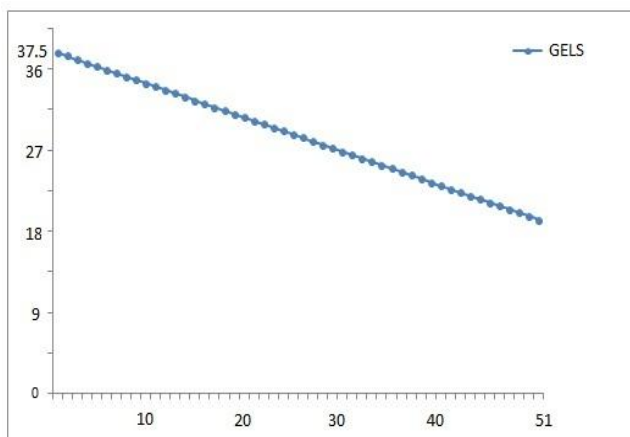


Fig. 4. Results of the gravity algorithm for 300 Normal sensors, 15 Manager Sensor and 30 targets

7. Conclusions

In this paper, a gravitational emulation local search algorithm is used to solve Optimal sensor selection for monitoring in point coverage wireless sensor network .as Well as a New Method is proposed To Calculation suitability And evaluate presented Solutions To solve Optimal sensor selection problem for monitoring in point coverage wireless sensor network. The advantages of this algorithm are speed, low run time, increase lifetime of network by optimization and reduce energy consumption and increase monitoring network efficiency. The results shows improvement and superiority of proposed algorithms compared to sources [22] and [23] and [29] and [30]. This improvement is more apparent in large-scale systems.

References

- [1] E.H. Callaway, "Wireless Sensor Networks, Architectures and Protocols, Aurebach Publications," vol 1, pp. 20-28, 2004.
- [2] M. Cardei, D. MacCallum, X. Cheng, M. Min, X. Jia, D. Li, and D.-Z. Du, "Wireless Sensor Networks with Energy Efficient Organization", Journal of Interconnection Networks, Vol 3, No 3-4, pp 213-229, Dec 2002.
- [3] G. Gupta and M. Younis, Load-Balanced Clustering in Wireless Sensor Networks, IEEE International conference on communications, Anchorage, Alaska, 2003.
- [4] M. Cardei M. T. Thai, Y. Li , W. Wu, "Energy-Efficient Target Coverage in Wireless Sensor Networks," IEEE INFOCOM, 2005.
- [5] J. Carle, D. Simplot, " Energy efficient area monitoring by sensor networks," IEEE Computer 37 (2) pp 40-46, 2004.
- [6] S. Slijepcevic and M. Potkonjak, "Power efficient organization of wireless sensor networks," Proc. IEEE

- International Conference on Communications, pp. 472–476, 2001.
- [7] D. Tian and N. D. Georganas, “A node scheduling scheme for energy conservation in large wireless sensor networks,” *Wireless Comm. Mob. Comput.*, vol. 3, pp. 271–290, 2003.
- [8] B. Carbutar, A. Grama and J. Vitek. “Distributed and Dynamic Voronoi Overlays for Coverage Detection and Distributed Hash Tables in Ad-hoc Networks,” (ICPADS 2004), Newport Beach, CA, pp. 549-559, July 2004.
- [9] C. Chen & J. Ma, “Designing Energy-Efficient Wireless Sensor Networks with Mobile Sinks,” *ACM International Workshop on (WSNA)*, pp. 343–349, 2006.
- [10] J. Lu, J. Wang, and T. Suda, “Scalable Coverage Maintenance for Dense Wireless Sensor Networks,” *EURASIP Journal on Wireless Communications and Networking* Volume, 2007.
- [11] S. Commuri, M. Watfa, “Coverage Strategies in 3D Wireless Sensor Networks”, *International Journal of Distributed Sensor Networks*, pp. :333-353 ,October 2006.
- [12] M. Watfa, S. Commuri, “Boundary Coverage & Coverage Boundary Problems in Wireless Sensor Networks”, *International Journal of Sensor Networks, IJSNET*, Volume: 2, Number: 3/4, Pages:273-283, 2007.
- [13] Y. Xu and X. Yao, “A GA Approach to the Optimal Placement of Sensors in Wireless Sensor Networks with Obstacles and Preferences,” *IEEE CCNC*, 2006.
- [14] A. S. Rostami, M. R. Tanhatalab, H. M. Berneti, “Novel Algorithm of Energy-Aware in Asymmetric Wireless Sensor Networks Routing for In-Point Coverage”, *CICN*, pp 308-313 ,IEEE 2010.
- [15] J. Kim, D. Kim, “ Energy-efficient Data Gathering Techniques Using Multiple Paths for Providing Resilience to Node Failures in Wireless Sensor Networks,” *JOURNAL OF COMMUNICATIONS*, VOL. 1, NO. 3, JUNE 2006.
- [16] B. Aoun, R. Boutaba, “Clustering in WSN with Latency and Energy Consumption Constraints,” *Journal of Network and Systems Management*, Vol. 14, No. 3, September 2006.
- [17] A. S. Rostami, M. R. Tanhatalab, H. M. Berneti, “Decreasing the Energy Consumption by a Distribution Classification New Algorithm in Choosing the Best Sensor Node in Wireless Sensor Network with Point Coverage”, *CICN*, pp 269-274, IEEE 2010.
- [18] S. J. Baek, G. d. Veciana, X. Su, “Minimizing Energy Consumption in Large-Scale Sensor Networks Through Distributed Data Compression and Hierarchical Aggregation,” *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, VOL. 22, NO. 6, AUGUST 2004.
- [19] Y. T. Hou, Y. Sh. Hanif, D. Sherali, S. F. Midkiff, “On Energy Provisioning and Relay Node Placement for Wireless Sensor Networks,” *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, VOL. 4, NO. 5, 2005.
- [20] S. Bandyopadhyay, E. J. Coyle, “ An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks,” *IEEE*, 2003.
- [21] M. Cardei, M. T. Thai, Y. Li, “Energy -Efficient Target Coverage in Wireless Sensor Networks”, *IEEE*, 2005.
- [22] W. Awada and M. Cardei, “Energy Efficient data Gatering in heterogeneous Wireless Sensor networks,” *IEEE, WiMob*, 2006.
- [23] A. S. Rostami, H. M. Bernety and A. R. Hosseinabadi, “A Novel and Optimized Algorithm to Select Monitoring Sensors by GSA”, *ICCIA*, 829–834, 2011.
- [24] C. Voudouris and E. Tsang, “Guided Local Search,” *Department of Computer Science, University of Essex, UK*, August 1995.
- [25] B. Webster, “Solving Combinatorial Optimization Problems Using a New Algorithm Based on Gravitational Attraction, Ph.D., thesis, Melbourne,” *Florida Institute of Technology*, 2004.
- [26] A. R. Hosseinabadi, A. B. Farahabadi, M. S. Rostami, A. F. Lateran, “Presentation of a New and Beneficial Method Through Problem Solving Timing of Open Shop by Random Algorithm Gravitational Emulation Local Search,” *International Journal of Computer Science Issues*, Vol. 10, Issue 1, 745-752, 2013.
- [27] A. R. Hosseinabadi, M. Yazdanpanah and A. S. Rostami, “A New Search Algorithm for Solving Symmetric Traveling Salesman Problem Based on Gravity”, *World Applied Sciences Journal* 16 (10): 1387-1392, 2012.
- [28] A. R. Hosseinabadi, M. R. Ghaleh, S. E. Hashemi, “Application of Modified Gravitational Search Algorithm to Solve the Problem of Teaching Hidden Markov Model”, *International Journal of Computer Science Issues*, Vol. 10, Issue 3, 2013, 1-8.
- [29] Zh. Liu, “Maximizing Network Lifetime for Target Coverage Problem in Heterogeneous Wireless Sensor Networks”, *MSN 2007, LNCS 4864*, pp. 457–468, 2007.
- [30] Sh. Rostami, M.H. Tanhatalab, H. M. Bernety, S. E. Naghibi, “Decreasing the Energy Consumption by a New Algorithm in Choosing the Best Sensor Node in Wireless Sensor Network with Point Coverage”, *IEEE CICN*, 2010.

Flexible boom numerical simulation and vibration characteristics analysis

Hongbin Tang¹, Wu Ren²

¹College of Automobile and Mechanical Engineering, Changsha University of Science and Technology, Changsha 410004, China;

²School of Mechanical and Electrical Engineering, Central South University, Changsha, 410083, China,

Abstract

Truck mounted concrete pump boom is typical flexible structure. Due to the increasing of boom length, the flexible characteristics play important role in the dynamics behavior. Three simulation models are built. The first is rigid model. The second is flexible model with four booms flexible used modal reduction technique. The last is added with boom hydraulic cylinder equivalent virtual spring damper with the second one. Then the trajectory, hydraulic cylinder force of the flexible booms are analyzed. Finally a test is done on a boom test rig to validate the proposed approach is correct. It provides method to vibration control, trajectory prediction for such flexible structure.

Keywords: flexible boom; multi-body dynamics; tip displacement; vibration

1. Introduction

Truck mounted concrete pump boom is typical multi-body system consisting of pumping boom, joint, link. Until to now the long boom becomes more and more popular (vertical elongation more than 40 m), especially in recent years 66 m, 72 m, 80 m, 86 m and even 101 m's boom has appeared. So influence of flexibility, large deformation and non linear characteristics must be taken into account[1].

By far many methods have been developed for flexible multi-body dynamics of truck mounted concrete pump boom system: Oliver Lenord did research on a 4 booms concrete pump test rig, then he established three models of the interdisciplinary system and obtained a simplified linear damping to simulate the model[2]. Liu jie, Dai li established a boom model and done numerical computation about it using rigid-flexible multibody dynamics theory[3]. F. Resta, F. Ripamonti, G. Cazzulani establish a nonlinear flexible multi-body test rig and the modal observation method and disturbance estimate strategy were adopted to study the boom and hydraulic device. Finally flexible boom and pump vibration suppression and control are researched[4-5].

Among the above experts, different boom and model influence to the system are not studied. So in the first part of this paper a rigid model of a truck mounted concrete pump boom was established, then modal reduction and virtual spring damper method are used to build flexible simulation models on which kinematics and dynamic response was studied. Finally an experiment was done on a test rig to prove the established model and proposed method are correct.

2. Flexible multibody boom test rig model methods

2.1 Virtual spring damper method

The boom system is driven and controlled by hydraulic cylinder. It is equivalent to certain stiffness and damping system by virtual spring damper method as shown in Fig. 1. Driving force and simulation position can be written[6-7]:

$$F_{cyl} = k \cdot (y_0(t) - y_{cyl}(t)) + c \cdot (y_0(t) - y_{cyl}(t))' \quad (1)$$

$$y = \int_{t_0}^t -\frac{F_{cyl} - k(y_0(t) - y_{cyl}(t))}{c} dt \quad (2)$$

$$c = \frac{\pi \eta l d^2}{(D-d)^2} \left[3 + \frac{3d}{4D-d} \right] \quad (3)$$

In Eq.1 $y_0(t)$ is the initial position of the cylinder. $y_{cyl}(t)$ is the terminated position. t is the time. The damping c of each cylinder is obtained by Eq.3, they are 1.08 N•s/mm, 0.9 N•s/mm, 0.2 N•s/mm, 0.02 N•s/mm, η is the kinematics viscosity, l is the piston length, d is the piston Diameter. D is the cylinder inner diameter. The stiffness k of each cylinder is 40 kN/mm, 30 kN/mm, 20 kN/mm, 6 kN/mm.

2.2 Modal reduced method

Modal reduced method is also called FMBD theory (finite element multi-body dynamics) [8]. Flexible body point i in the three dimensional coordinate can be expressed as:

$$\bar{r}_i = \bar{r} + A \bar{S}_i n = \bar{r} + A(\bar{S}_i^0 n + \bar{u}_i n) \quad (4)$$

In Eq. 4 A is the transfer matrix. \bar{S}_i^0 is the undeformed position vector, \bar{u}_i is the deformation vector. The translation and rotation modal matrix of point i is:

$$\Psi^i = [\Psi_i^{tT}, \Psi_i^{rT}] \quad (5)$$

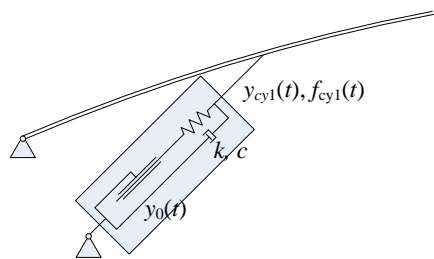


Fig. 1 Physical model of the hydraulic cylinder

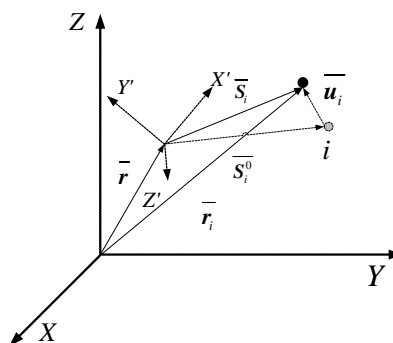


Fig. 2 Flexible body coordinate system
 flexible-rigid physical model are established in Fig. 3 above.

The translation modal matrix and modal coordinate is written:

$$u_i' = \Psi_i' \bar{a} \quad (6)$$

In Eq. 6 \bar{a} is the modal position vector, applying Eq. 6 to Eq. 4 and Eq. 7 we can get:

$$\bar{r}_i = \bar{r} + A(\bar{S}_i^{0'} + \Psi_i' \bar{a}) \quad (7)$$

According to the modal information Ψ the overall displacement are as follow:

$$[u] = \sum a_i [\phi]_i \quad (8)$$

In Eq. 8, a_i is the modal participation factor; $[\phi]_i$ is the structure mode.

2.3 Physical model

Three assumptions are needed in modeling and simulation of boom system:

- (1) Only considering posture transform in planar, ignoring the impact of spatial torque.
- (2) Translational joint and fix joint are used to connect booms with links, translational joint are used to simulate hydraulic cylinder motion.
- (3) Small boom motion speed. The topology model of the concrete pump boom test rig is shown in Fig. 3.

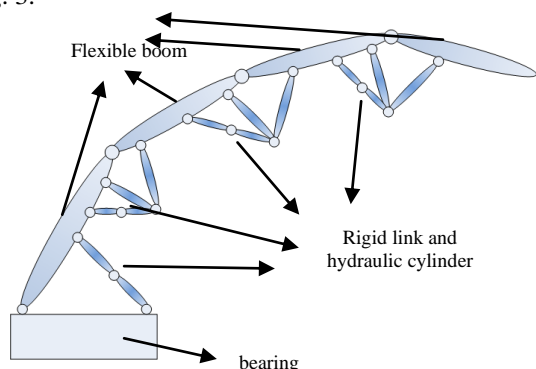


Fig. 3 Boom topology system

3.Simulation and experimental analysis

According to the above assumptions a truck mounted concrete pump boom rigid model and

Three simulation models are established using Recurdyn software. All the parameters of the booms are shown in Table 1:

Table 1 Main parameters of each boom

	boom1	boom2	boom3	boom4
length(mm)	3753	2830	2875	3205
Mass(kg)	88.4	37.9	32.0	18.3

The motion stress, joint force, tip trajectory of each model are studied in Recurdyn as follow.

3.1 Motion stress simulation

In Fig. 4 the maximum stress of rigid flexible model is between the 2nd and 3rd boom during simulation. Due to the flexible feature the booms stop gradually after motion stop. Under allowable operating speed the stress is 135 MPa, less than Q345B boom material's allowable stress 230 Mpa.

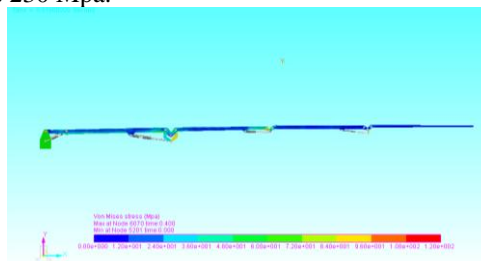


Fig. 4 Flexible model stress contour

3.2 Tip displacement simulation analysis

The tip displacement of three simulation models and test data are in Fig. 5. It can be seen the rigid model has the smallest boom tip displacement which is about 145 mm, the other two flexible model's tip displacement are 330mm and 709 mm. The test data is 460mm. Due to the simulated cylinder stiffness error the data of the 3rd model is 349 mm larger than the test value.

Fig. 6 is different cylinder Equivalent spring damping influence to the tip displacement of the 3rd model. We can see the 2nd and 3rd cylinder play more important role to the total displacement.

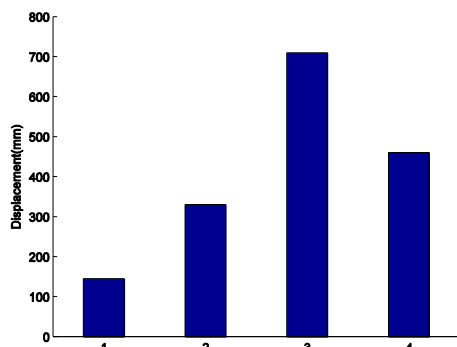


Fig. 5 Different models tip displacement
 1-The 1st model; 2-The 2nd model; 3-The 3rd model;
 4-test data

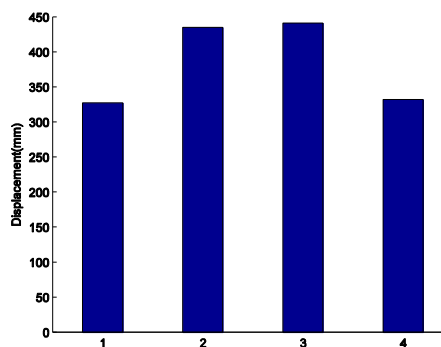


Fig 6 Different boom spring damper influence
 1- The 1st boom; 2-the 2nd boom;
 3- The 3rd boom; 4- the 4th boom;

3.3 Cylinder force simulation research

The three model cylinder force are in Fig. 7 it can be seen the maximum force between the second and third cylinder in the rigid model is about 52.8 kN, while the other two flexible model are 98.6kN and 90.1 kN.

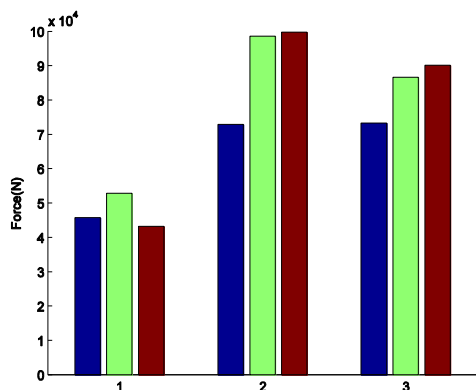


Fig. 7 Different models hydraulic cylinder force

1-rigid model; 2- Flexible rigid model; 3- Flexible rigid model with cylinder mass damping mode

In Fig. 7 the blue column represents the first cylinder force; while the green and brown are the 2nd and 3rd



Fig. 8 flexible boom test rig

cylinder force. The two flexible model forces are nearly two times of the rigid one. Therefore the flexible deformation should be taken full account in hydraulic cylinder design process

3.4 Experimental study and analysis of results

In order to verify the proposed method and simulation result, an experiment is done on a 13 m truck mounted concrete pump boom test rig (Fig. 8). Dewesoft multi channel signal acquisition instrument (Fig. 9), three axis acceleration sensor, tilt sensor, strain rosette are used. The measured boom stress, simulation value and MSC / Nastran value are in Table 2. It is well proved the established model is appropriately. The different model vibration displacement and cylinder force are discussed above.

Table. 2 Maximum stress comparisons between Simulation and test

	By Nastran	Simulation	Test value
stress [Mpa]	138.6	135.0	118.3



Fig. 9 24-channel multi-signal test System

4. Conclusion

(1) A 13 m rigid boom model was established. Then modal reduction method is adopted to build a flexible body model. Next spring damping method is used to simulate the hydraulic cylinder.

(2) Boom motion stress is calculated of different model. It can be found the stress of flexible model is closed to the actual structure. The hydraulic cylinder force, tip displacement of the rigid model is smaller than the flexible model. The 2nd and 3rd cylinder have the most significant influence on the boom tip displacement.

(3) An experiment is done on the boom test rig to verify the simulation result proposed above. And it provides theoretical basis to vibration control, trajectory prediction and life assessment for such structure.

Acknowledgements

A Project Supported by National Natural Science Foundation of China (No.51075138) and Scientific Research Fund of Hunan Provincial Education Department (No.09C073) and Key Laboratory of Lightweight and Reliability Technology for Engin

Vehicle, College of Hunan Province (No.2012kfjj08)

References

- [1] National natural science foundation of engineering and materials science department of mechanical engineering disciplines development strategy report (2011-2020) [M]. Beijing: Science Press, 2010: 24-46.
- [2] Lenord O, Fang S, Franitza D, et al. Numerical linearisation method to efficiently optimise the oscillation damping of an interdisciplinary system model[J]. *Multibody System Dynamics*, 2003, 10(2): 201-217.
- [3] LIU Jie, Dai Li, Zhao Li-juan, et al. Modeling and simulation of flexible multi-body dynamics of concrete pump truck arm[J]. *Chinese Journal of Mechanical Engineering*, 2007, 43(11): 131-135.
- [4] Cazzulani G, Ghielmetti C, Giberti H, et al. A test rig and numerical model for investigating truck mounted concrete pumps[J]. *Automation in Construction*, 2011, 20(8): 1133-1142.
- [5] Cazzulani G, Resta F, Ripamonti F. A Feedback and Feedforward Vibration Control for a Concrete Placing Boom[J]. *Journal of Vibration and Acoustics*, 2011, 133(5): 021-028.
- [6] Shabana A A. *Dynamics of multibody systems*[M]. London: Cambridge Univ Pr, 2005: 190-200.
- [7] Nissing D. A vibration damped flexible robot: Identification and parameter optimization[C]. *Proceedings of the American Control Conference*. Chicago: IEEE Press, 2000: 1715-1719.
- [8] Yoon J W, Park T W, Lee S H, et al. Synthetic analysis of flexible multibody system including a very flexible body[J]. *Journal of mechanical science and technology*, 2009, 23(4): 942-945.

Adaptive Approach for Spam Detection

Sumant Sharma¹, Amit Arora²

¹ Student, Department of Computer Science, Chitkara University, Himachal Pradesh.

² Assistant Professor, Department of Computer Science, Chitkara University, Himachal Pradesh.

Abstract

Spam has emerged as a major problem in recent years. The most widely recognized form of spam, is email spam. The accounts which contain spam messages must waste time deleting annoying and possibly offensive message. In this paper, we present a variety of machine learning algorithms to identify spam in e-mail accounts. We design classifier model to automatically determine spam in the accounts so that time of account holder can be saved and utilized on other work.

The dataset we used for our project is named as SPAMBASE dataset download from UCI Machine Learning Repository. We used the labeling data in conjunction with machine learning techniques provided by WEKA tool kit, to train a computer to recognize spam instances automatically. The accuracy of 94.28 is shown by the Random committee through the experiment.

Keywords: Machine Learning, NumericToBinary Filter, Spam Detection, Weka.

1. Introduction

Spam is the use of electronic messaging systems to send unsolicited bulk messages, especially advertising, indiscriminately [1].

While the most widely recognized form of spam is e-mail spam. Now a days spam messages are sent to large number of users via internet. The Spammers are targeting the users having accounts on e-mail sites like Gmail, Hotmail, I Cloud and social networking sites like Facebook and Twitter. An account holder waste time deleting annoying and possibly offensive message, It also cause delay to deliver important e-mails to the account due to large amount of spam traffic in between host and the e-mail servers. Hence, filtering the spams from bulk of data is very challenging task. One of the popular method for spam detection is Bayesian spam filtering (Thomas Bayes) which is a statistical technique for e-mail filtering. In this technique a naïve Bayes classifier is used to identify spam e-mail. Bayesian classifier work by correlating the use of tokens with spam and non-spam e-mails and then using Bayesian inference to calculate a probability that an e mail is spam or not [8].

Another approach is content based spam filtering in which a word or a phrase of an email is analyzed by using machine learning algorithm. It matches the content, if the content is found then numeric value is given to the e-mail. A threshold value is set and after crossing a threshold, that e-mail is considered as spam [4].

Another approach uses Support Vector Machine algorithm for content base filtering. This algorithm gives a remarkable performance on text classification also. Support Vector Machine algorithm give very high performance when applied on large benchmark dataset. An equivalent performance was also evaluated by applying the Relaxed Online SVM (ROSVM) on same dataset to detect E-mail spam, Blog spam and SP-LOG [6].

Another approach detects spam by using the text clustering based on vector space model. The disjoint clusters are computed for all Spam or Non-Spam mails by using the spherical k-means algorithm. For each centroid vectors, label is assigned by calculating the number of spams in the cluster. When new mail arrives in the account, the cosine is calculated between the new mail vector and centroid vector. Finally, the label of the most relevant cluster is assigned to the new mail [7].

In this paper SPAMBASE dataset is used to classify e-mail as spam or non-spam e-mails. Spam base dataset is multivariate dataset contains data from a single email account. This data is used to apply various machine learning algorithm to classify the spams present in that. For the various machine learning algorithm **WEKA** tool is used. WEKA is open source software made in java. It provides collection of algorithms used for data analysis and predictive modeling. After applying the algorithm, percentage of precision, recall accuracy, score and Correctly Classified Instances at ten Fold Cross-validation is calculated. The classifier which is having high accuracy and correctly classified instances.

2. Experiment

In this section, we describe the experiment done in order to detect spam. How we apply the machine learning algorithm and which algorithm is giving the maximum number of correctly classified instances and accuracy. We will draw our attention on detecting maximum number of spams from the spam-base dataset. In particular we look for the following-

- [1] **Bayes Net (BN)**
- [2] **Logic Boost (LB)**
- [3] **Random Tree (RT)**
- [4] **JRip (JR)**
- [5] **J48 (J48)**
- [6] **Multilayer Perceptron (MP)**
- [7] **Kstar (KS)**
- [8] **Random Forest (RF)**
- [9] **Random Committee (RC)**

Bayes network model is a probabilistic graphical model that represents a set of random variable and their conditional dependencies via a directed graph. For example a Bayesian network could represent the probabilistic relationships between diseases and symptoms [9].

Random forests are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees. The algorithm for inducing a random forest was developed by Leo Breiman and Adele Cutler [10].

J48, In machine learning are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification analysis. The basic j48 takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier [11].

Multilayer Perceptron (MLP) is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate outputs. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network [12].

To do all the experiments, we used WEKA (Waikato Environment for Knowledge Analysis), the open source software to make and test the classifier. WEKA is a popular suit of machine learning algorithms for data mining tasks. It is widely used for developing new machine learning models. WEKA has a highly customizable interface and very easy to use, which enabled us to run a large number of experiments.

For all the experiments, we used a Spam-base dataset. This dataset was collected from UCI Machine Learning Repository for the research purpose. Dataset is multivariate having 4601 instances and 55 attributes, consisting of tagged emails from a single email account. We have a training set of labeled spam mails to train the classifiers. Testing of the classifier was performed on a testing set, the performance was measured by evaluating the accuracy, precision, recall and score for various classifier.

2.1 Preprocessing Data

The data which is available in the spam-base data set is in numeric form. The fifty five attributes in the dataset represent relative frequencies of various salient words and characters in emails. We wish to convert these to Boolean values for the experiment. The attribute will take a value 1 if the word or character is present in the email and 0 if it is not present in the email. To do this we apply a Numeric to Binary filter present in the WEKA tool. This filter will convert all the numeric values to the binary

2.2 Making Classifier

The converted data set is used to train the classifier to detect spam from regular email by checking the number of occurrences of each word for all the spam and non-spam emails. A variety of algorithm is given into the WEKA tool that can be used. We apply the algorithm by choosing "TEN Fold Cross Validation".

3. Examining Results

Now, we examine results of classifier/model produced by the weka. Which basically tells Five things which are correctly classified instances (CCI), True positive value (TP), False positive value (FP), True negative value (TN), False negative value (FN). For each classifier all these five things are distant. All these five values were used to calculate Accuracy, Recall Precision and Score for a particular classifier.

TABLE-1

Algorithm	CCI	TP	FP	TN	FN
NB	88.54	2596	192	1478	335
BN	88.56	2596	192	1479	334
NBU	88.54	2596	192	1478	335
LOGISTIC	92.95	2654	134	1623	190
MLP	93.28	2630	158	1662	151
SGD	93.28	2655	133	1637	176
SMO	93.21	2659	129	1630	183
VP	92.56	2615	173	1644	169
KSTAR	93.56	2665	123	1640	173
DT	91.71	2666	122	1554	259
RIP	92.32	2634	1544	1614	199
PATR	93.06	2631	157	1651	162
DS	77.2	2041	747	1511	302
J48	92.34	2618	170	1631	182
RF	93.89	2673	115	1647	166
RT	91.54	2586	202	1626	187
BAGGING	92.93	2650	138	1626	187
LOGICBOOST	89.76	2589	199	1541	272
MCC	92.95	2654	134	1623	190
RS	92.37	2667	121	1583	230
CVR	92.15	2637	151	1603	210
FC	92.34	2618	170	1631	182
RC	94.28	2680	108	1658	155

As we can see in the above table each individual classifier is having five values. These values are used for further calculations.

4. Indentation And Equations

Now, Precision, Recall, Accuracy and Score are calculated for each of the individual classifier. To calculate we need to have the values from Figure 1. From the figure rather than calculating values for all the classifiers we took some of the classifiers based on the percentage of correctly classified instances (CCI). After that the future calculation is done by using the formulas describe below-

$$\text{Precision (PPV)} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall (TPR)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Accuracy (ACC)} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN}) + (\text{FP} + \text{TN})$$

$$\text{Score(F1)} = 2 * \text{TP} / (2 * \text{TP} + \text{FP} + \text{FN})$$

TABLE-2

ALGOS	CCI	PPV	TPR	ACC	F1
BAYSNET	88.56	93.113	88.6	88.56	90.8
LOGICBOOST	89.76	92.862	90.4	89.7	91.66
RANDOMTREE	91.71	92.754	93.2	91.54	93
JRIP	92.32	94.476	92.9	92.32	93.71
J48	92.34	93.902	93.5	92.34	93.7
MULTILAYER PERSAPTRON	93.28	94.332	94.5	93.28	94.45
KSTAR	93.56	95.588	93.9	93.56	94.73

RANDOM FOREST	93.89	95.875	94.1	93.89	95
RANDOM COMMITTEE	94.37	96.126	94.5	94.28	95.32

Table- 2 describes all the algorithms with their calculated values of percision, recall, accuracy and score. All the values are calculated manually by using the formulas describe above. Here percentage values is taken for all the algorithms. As we can see the accuracy of the algorithm is lies between 88.56 to 94.5.

Figure-1

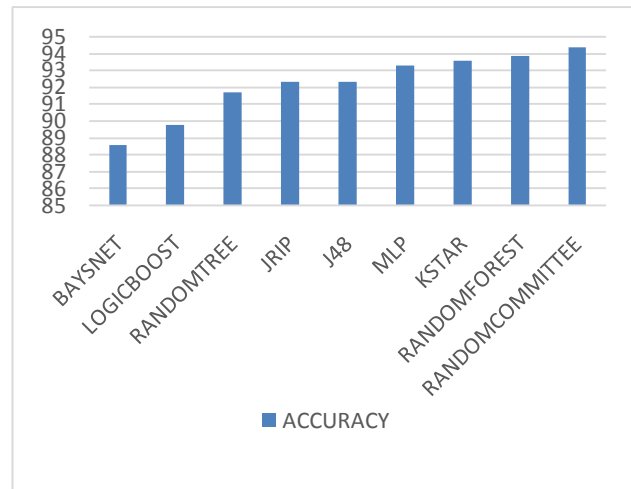


Figure-1 shows Accuracy of 9 algorithm on spambase dataset.

Figure-2

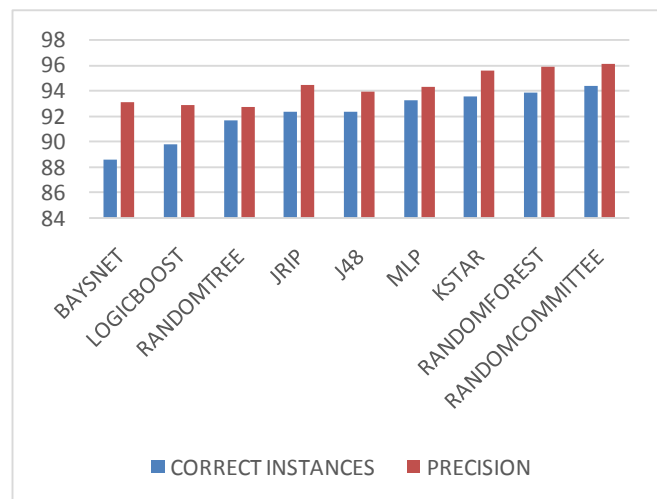


Figure-2 shows Correctly Classified Instances and Percision after applying the classifier on spam base dataset in percentage.

Figure- 3

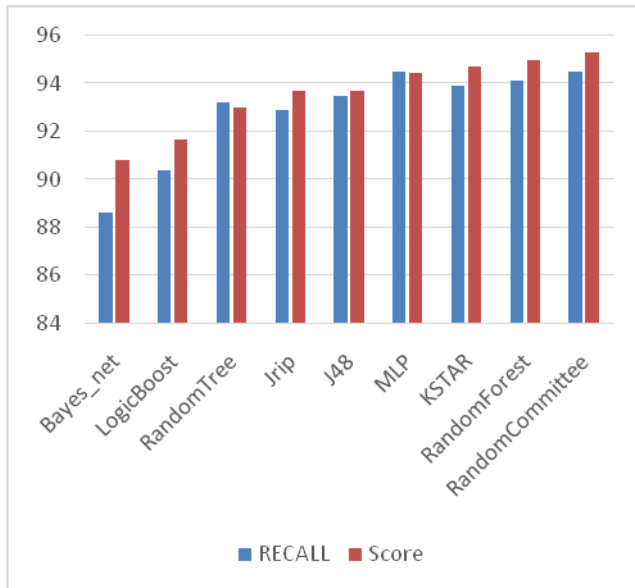


Figure- 3 shows the Recall and Score value in percentage.

5. Conclusions

In this paper we apply various machine learning algorithms to classify the spams from e-mail. We compare 24 algorithms (Table 1) on spam base dataset using 55 spam base-attributes. We focus on the accuracy and performance of the algorithm to classify the spam/non-spam e-mails from tagged emails of a single account. As an initial experiment to analyze accuracy and performance 10 classifiers are used as described in the Table-2 above. The results of experiencing on Spambase datasets show better performance of the proposed method. It shows 94.28 accuracy and Score value of 95.32 for Random Committee, which is high in all other methods

Acknowledgments

I would like to express a deep sense of gratitude and thanks profusely to Mr. Amit Arora, Assistant professor, Department of Computer Science and Engineering, COSET who was the thesis Supervisor. Without the wise counsel and able guidance, it would have been impossible to complete the research work in this manner and on time.

References

- [1] Guzella T. S., Caminhas W. M. 2009 "A review of machine learning approaches to spam filtering". 36(7), 10206-10222.
- [2] Koprinska I., Poon J., Clarck J., Chan J. "Learning to classify e-mail". Info. S. 177: 2167-2187, 2007.
- [3] Kuncheva L. "Combining Pattern Classifiers, Methods and Algorithms, Wiley Inter Science, 2005".
- [4] S. Mason. New Law Designed to Limit Amount of Spam in e-mail.

- [5] Jongsub Moon, Taeshik Shon, Jungtaek Seo, Jongho Kim, Jungwoo. "An Approach for Spam E-mail Detection with Support Vector Machine".
- [6] D.Sculley and Gabriel m.wachman. "Relaxed Online SVMs for Spam Filtering".
- [7] Spam detection using text clustering Sasaki, M.; Dept. of Computer & Inf. Sci., Ibaraki University; Shinnou, H.
- [8] M. Sahami, S. Dumais, D. Heckerman, E. Horvitz (1998). "A Bayesian approach to filtering junk e-mail".
- [9] Pearl, Judea (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press. ISBN 0-521-77362-8.
- [10] Breiman, Leo (2001), "Random Forests". Machine Learning 45 (1): 5–32. 10.1023/a: 1010933404324
- [11] Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995.
- [12] Rumelhart, David E., Geoffrey E. Hinton "Learning Internal Representations & Error Propagation".

ADOPTION OF ICT TO ENHANCE EDUCATIONAL DEVELOPMENT IN NIGERIA

Odeniyi O.A¹, Ayinde A.Q², Dr Adetunji A.B³, Sarumi O.A⁴

1. Computer Science Department, Osun State College of Technology,
Esa-Oke, 234035/South West, Nigeria

2 Computer Science and Engineering Department, LAUTECH,
Ogbomoso, 23402/South West, Nigeria

3. Computer Science and Engineering Department, LAUTECH
Ogbomoso, 23402/South West, Nigeria

4. Computer Science Department, Osun State College of Technology,
Esa-Oke, 234035/South West, Nigeria.

Abstract

Education has always played a central role in human development while today the world accepts universal primary education as an achievable goal. Formal schooling for everyone is a relatively recent phenomenon, even when it was less formalized or standardized scientific and technical curiosity help moved mankind from agricultural to the industrial .Now into the knowledge economy know as the ICT. ICTs has been prime as the "ultimate saviour" for reaching the millennium development goals but they are more of an enabler. They have the potential to help reach both, but not without reliance on more traditional learning system and technologies. This research related the relationship between an education-centered approach with ICTs, adding tutorial and the efficiency of the educational system in the classroom and beyond.

Keywords: Education, ICT, Rationale, School, Traditional Learning, Technology

1. Introduction

At personal level, education helps individuals more beyond subsistence agriculture, and helps them compete against their peers. However, in today's globalized world, the competition is not just with people of the same village or region but across continents. A century ago, improving transportation was a driving force globalization. Now, information and communications technology (ICT) is a major factor in spurring positive educational development but it has been seen as symmetry, that is, it has the potential to be the great equalizer and democratizer for

those who had been left outside its scope, or who had failed to harness its resources.

Learning occurs in four different streams in the 20th century:

(i) Formal learning takes place in schools and higher education institutions providing systematic education

(ii) Non-formal learning occurs outside the formal education system but is nevertheless an organised event with specific target groups or clients and learning objectives. This includes continuing education, adult education, professional training and literacy programmes.

(iii) Informal learning is the individual acquisition of skills, knowledge and attitudes from the everyday experience and environment.

(iv) Virtual collaborative learning occurs both as an independent learning structure or within networks made up of learners with similar learning needs in the case of initiatives like the African virtual university and as an integrated strategy added on to

The first section is used to describe the current trends in education and different learning streams. The second section is factor influencing the adoption of ICT in educational domain. The third section covers why ICTs was adopted for the student and how it will be deploy to meet their needs. The fourth section covers the conclusion.

2. Literature Review

In this literature review we highlight the factors affecting the ICT adoption process and their numerous impacts. As our study is focusing on education and the impact of ICT to enhance education was also being explored.

Four attributes from Roger's model [7] are relative advantage, compatibility, complexity, image and cost which will be adopted in this framework to test the impacts on ICT adoption process.

Besides focusing on particular ICT factors, the organizational and environmental factors also may impact on the process of ICT adoption in an organization. According to Rashid and Al-Qirim [1], the organizational factors collectively impact on the resources of the business in relation to adoption of ICT innovation [2]. However, the process of ICT adoption could be quite difficult for a firm because of its requirements. The willingness for adoption of ICT is, usually, associated with organizational readiness where organization must adapt with a large investment and firms may not have sufficient financial resources to support the high investment in hardware and software technology that is required [3, 4]. Therefore we would expect that organizational awareness, encouragement and readiness might influence technological innovation. Damanpour [5] found that environments with high uncertainties would have positive influence on the relationship between organizational structures and organizational innovation. Apart from that, competitor also could be one of the important external factors considered in ICT adoption. ICT adoption decision would be influenced by the relative advantage gained by LSPs compared to their competitors. If there is no relative advantage gained by LSPs, ICT might not be adopted [6].

3. Adoption of ICT in Nigeria Schools

There is an immense opportunity for the utilization of ICTs in the education sector in Nigeria, for learners, teachers and the administrators. Enormous opportunities are present in the use of ICTs in education and we can also weigh the opportunities against the constraints in their education systems. Education choices have to be made first in terms of objectives, methodologies and roles of teachers and students before decisions on appropriate technologies can be made. No technology will fix bad educational philosophy and practice. Learning objective should be aligned with learning technologies.

There are four rationales for the use of computer in schools:

- (i) Social rationale – this is the demystification of the importance of computers at school level;
- (ii) Vocational rationale – the need to prepare learners for employment through providing computer competencies, including educational programmes;
- (iii) Pedagogical rationale – the use of computers to improve on the delivery of education and as an aid in the teaching and learning process;
- (iv) Catalytic rationale – use of the computer in the overall performance of schools, integrating functions of teaching and learning management and administration.

Whereas the developed nations are placing emphasis on the development of e-commerce, the worldwide web and the internet, the developing world is still grappling with

attaining access to ICT equipment and gathering hands-on experience in operating ICT devices. Affordability should be addressed, as it is the developing world.

Four attributes from Roger's model [7] are relative advantage, compatibility, complexity, image and cost which will be adopted in this framework to test the impacts on ICT adoption process.

Besides focusing on particular ICT factors, the organizational and environmental factors also may impact on the process of ICT adoption in an organization.

The capacity of ICTs to achieve development goals will not be effectively leveraged without content that is responsive to user needs and local conditions. If ICTs are to be used for education purposes, the content should be relevant to the curricular requirements of the education system and the needs of the students, teachers and administrators.

Generally, the following should be considered if ICTs should serve as an added value in education:

- (a) Educational policy
Establishing the best policy framework is essential if the development potential of ICTs is to be fully realised. Effective introduction of ICTs requires policies and planning at various levels. At the national level, when leaders recognise the benefits of ICT it is easier to allocate funds for their support. Incentive can be provided to increase the involvement of the private sector. The government should show support and commitment and be involved in the democratisation process of access to ICTs. The regulatory framework should be adaptable to allow faster expansion of ICT use and enterprise growth.
Hence, the following education policy initiatives are important conditions and facilitators of ICT based learning:
 - (1) Government awareness of the importance of ICTs for national education. This calls for understanding that:

- (i) ICTs are virtually important to the development of the economy and to participation in the global information society, with a corresponding need to develop appropriate skills and
- (ii) ICT based learning and distance education can play a crucial role in broadening access to education for the whole society and reaching the education for all goals.

- (2) A strategy plan or policy. this must be based on an analysis of needs and priorities for the use of ICT to improve education – the school net Africa value chain being a possible basis to begin with. Some key elements and concrete steps of such a strategic plan include:

- (i) ICT skills integration in national curricula – in teacher training institutions and in the classroom.

- (ii) Equipping schools with appropriate technologies such as computers
- (iii) Teacher and administrator training on ICT
- (iv) Initiatives and programmes which invite and attract private sector involvement

(3) The presence of the local participation and initiative and serious considerations regarding the self sustainability of projects. Financially, skills wise (e.g through train the trainers); equipping teachers to use these technologies in teaching; long term maintenance of such technologies; support system to ensure longevity to get optimal use from resources committed to such technologies. In order to develop sustainable ICT based education in developing countries like Nigeria, models of financing need to be devised which provide for transition from initial dependency on the international donor agencies to self-sustaining and continuing institutions.

A policy on ICT in education is just one fact of preparing the population and workforce in Nigeria to be productive, competitive and at ease in the information economy. Education is fundamental to achieving development – education not just in terms of skills acquisition, but in the application and implementation of skills to meet needs of the communities to bring about development.

(b) CAPACITY BUILDING

There is need for low cost solutions, government support and increased involvement of the telecommunication providers in narrowing the gap between the capacities of urban and rural schools. Technology infrastructure needs upgrading, more so with poor telecommunications infrastructure in Nigeria. Maintaining permanent qualified technical support is costly and trade off should be made. Teachers in the schools should be trained to cut across teaching as well as providing technical support.

The success of ICTs will to a large extent be based on the cooperation between the technical and educational infrastructure. The quality of manpower and teachers on the ICT programmes has a great impact on the success of ICT learning. There is a need for dedicated human resources with several roles. A full time IT teacher for technical backstopping would be appropriate for each school.

Basic literacy is of crucial importance for the development of ICT in Nigeria. To deploy ICT for educational development in Nigeria, we should develop a critical mass of trained personnel at various levels. Developing human capacity in ICT will be supported by the promotion of relevant educational curricula and the creation of new educational facilities with emphasis on ICT skills development. Teachers often do not know what

they can do with technology, and the tendency is to use ICT simply to automate traditional teaching methods. Teachers need to get critical guidelines and upgrading of their skills for effective use of ICT.

(c) PARTNERSHIPS

Partnerships between the various stakeholders in ICT programmes will be central to the full adoption of ICTs. Issues raised with partners resolve around the need for strong coordination and support from the government, active partnerships between the government ,NGOs, the private sector, donors and telecommunication providers; private – public sector partnerships for long-term sustainability; and complementary donor funding for such programmes. Partnerships need to be flexible, strength related, privileged and here clearly defined goals of importance are the rural communities in Nigeria. This group of people is often overlooked but forms the strongest basis for the sustainability and successful implementation of ICT projects.

(d) ACCESS TO ICTs

Access to ICT is a fundamental requirement and the solution to the problems of the digital divide and other development divides. If the desired end-goals of empowerment and opportunities are considered, access leads to information, which can lead to knowledge, leading to empowerment and opportunities. Of course, it is not linear and one requires complementary capabilities, especially to interpret information into usable knowledge. In fact, knowledge is an interpreted extension of information that captures relevance and context, and it is tightly coupled with opportunities. The digital divide, however defined, is a stark divide and a challenge for development and technology professionals. It is actually a manifestation of other underlying divides, spanning economic, social, geographic, gender and other divides. Attempting to address the digital divide as a cause instead of a symptom of other divides has led to many failure of ICT drives development projects.

The digital divide can be considered at four levels: awareness, availability, accessility and affordability.

- (i) Awareness: relates to knowing what can be done with ICT; people must also be open to using ICT (attitudes)
- (ii) Availability: ICT must be offered within reasonable proximity, with appropriate hardware/software.
- (iii) Accessibility: relates to the ability to use the ICT (spanning literacy, e-literacy, language, interfaces, etc)
- (iv) Affordability: all ICT usage together should, ideally, be only a few percent of one's income (under 10% maximum); this covers life-cycle costs (termed total costs

of ownership – TCO), spanning hardware, software, connectivity, education, etc.

Reducing the divide requires improvements across all the dimensions of ICT (dubbed the 4c framework): computing, connectivity, content and human capacity.

(i) Computing: personal computers (pcs) are prohibitively expensive for most people, and shared access (e.g schools, community centres or cybercafés) becomes inevitable. PCs today are very difficult to use, and even ‘experts’ spend a lot of time maintaining their machines, worrying about upgrades, security, compatibility of hardware etc as a complementary (but not substantive) technology, non-pc devices are an important option e.g mobile phones.

(ii) Connectivity: while mobile telephony is improving worldwide (witness in Africa it is now twice the number of landlines), it remains expensive, limited in rural areas, and poor at providing data connectivity. Many areas are now grappling with limited connected options, such as dial up. Instead, broadband connectivity can be affordable, even in rural areas, with the right network and business models.

(iii) Content: meaningful content is lacking in many languages, and most content is not locally relevant. Today’s system tends to make people passive consumers of information, instead of enabling the generation of local information. In addition, rich content demands multimedia (useful to overcome literacy issues), which in turn, requires broadband connectivity.

(iv) Human capacity: users need to be aware, literate and innovative to harness the power of ICT. They also should be empowered to use ICT, societal and governmentally.

Of course, ICT usage does not occur in a vacuum, rather within social and cultural norms that also shape the divide. In addition, ICT usage is based on policy and business models, especially regulation. In the long run, ICT must provide value and be sustainable from both a user and a provider perspective. As the Markle foundation’s report on national strategies of ‘ICT for development’ (2003) states, ‘digital divides are not just the result of economic differences in access to technologies (Have-Not’s), but also in cultural capacity and political will to apply these technologies for development impact (Do’s Do-Not’s)’. However affordability is certainly a limiting factor, since we have seen that many people could have access of some form of ICT but do not (e.g. mobile telephony footprint extends to over 80% of developing country populations, but the actual usage rates are much lower).

(e) Investment

There are many components of ICT for education that require investment. These include hardware (components),

software, training/education, supporting infrastructure such as electricity and connectivity.

Upgrade and install some new applications which needed more computing horsepower. Specifying the right PC is a key decision in better enabling access to ICTs in education. One method of calculating what specification to buy and what the appropriate refresh cycle should be is to calculate the ‘equivalent annual cost’ (EAC) of different specifications and different refresh cycles. (EAC is a fairly common financial method to compare capital investments with different lifetimes as an annuity that has the same present value and life as the underlying cost flow). In determining total cost of ownership for PCs, governments should also take into account all the cost components of a PC’s lifecycle including PC deployment, PC usage, PC support and PC retirement costs. Calculating the helps in choosing the right system and refresh cycle for PCs in education.

Consolidating blackened systems to reduce the numbers of servers required unit cost of ICT. Again, this is a practise that many corporations have used to reduce ICT unit costs with typically e-mail servers being a frequent target for quick cost reduction. The ever-improving price/performance ratio of communications, particularly the role of wireless as a disruptive technology, is also an enabler to consolidate multiple data centres into fewer larger data centres. Also as blade technology becomes more mainstream, computer density in computer rooms can be significantly increased, again lowering cost.

4. Conclusion

Education using ICT is especially difficult as it involves specialised knowledge for both ICT and education. Many electronic/distance educational models failed because existing providers thought it was enough to digitise and put their current material on the web. In fact, a number of reputed schools and universities failed or saw enormous setbacks in such efforts. To succeed, new content and ideally, new methods of instruction are key. Students learn and retain far more by doing than by ‘taking in’. This also relates to the failure of many syllabi or curricula in being relevant for either rural areas or the modern (global) economy.

ICT is not yet a stage where it can substitute for human. It is best used to enhance and extend humans’ capabilities. One major challenge is meeting minimum skill sets for all students while allowing those who were able to progress rapidly. The catch is that some can progress rapidly not only due to skills but also due to advantages of family support, infrastructure, etc. Many people acknowledge that

curriculum based rich content is the biggest gap or issue in increased adoption of ICT in education.

Armed with the appropriate ICT tools and the appropriate standards, teachers can and should be encouraged to develop curriculum focuses content which can be re-used by other teachers. Teach to the future programme has trained more than two million teachers worldwide and these kinds of programmes rapidly improve teacher competency in using and handling ICT infrastructures in the classroom. A disruptive innovative that could transform the development and proliferation of content is the establishment of a Nasper-like solution based on peer-to-peer networks which could enable easy sharing of teacher prepared content between schools and across borders.

Another tactic to improve access to ICTs in Nigeria education is the increasingly popular concept of public-private sector organisations come relationship to fund, develop and operate ICT solutions and services. As we move into an era of third generation corporate social responsibility, corporation to drive large-scale changes which produce win-win outcomes. While developing strategies which address thus, we also need to continue to drive pilots and leverage learning from rapid solution prototyping.

Finally, Nigeria government should try to create various circles of innovation through coordinated strategies on subsidize broadband deployment, PC purchase programmes, digital literacy programmes and on-line e-services provisioning while each of these components have value in isolation, network effect in education can only be achieved through the co-management and co-evolution of strategies which co-evolve the 4Cs of ICT- computing, connectivity, Content and (human) capacity.

References

- [1]. Rashid, M.A.a.A.-Q., N. A. (2001)., *E.Commerce Technology Adoption Framework by New Zealand Small to Medium Enterprises*. Research Letters Information Mathematical Science, 2001. **2**(1): p 63
- [2]. Lin, C.-Y. and Y.-H. Ho, *Technological Innovation for China's Logistics Industry*. Journal of Technology Management Innovation, 2007. **2**(4).
- [3]. Lai, et al., *Information Technology Adoption in Hong Kong's Logistic Industry*. Transportation Journal, 2005.**44**(4): p. 1-10.
- [4]. Tang, L.-L. and W.-C. Tsai, *RFID adoption Model for*

Taiwan's Logistic Service Providers. 2009.

- [5]. Damanpour, F., *Organizational innovation: a metaanalysis of effects of determinants and moderators*. Academy of Management Journal, 1991. **34**(3): p. 555-590.
- [6]. Galliers, R. D. & Sutherland, and A. R., *Information Systems Management and Strategy Formulation: Applying and Extending The Stages of Growth'*
- [7]. Rogers, E.M., *Diffusions of Innovations*. 4th. Ed. New York:, 1995.

O.A Odeniyi obtained his B.Tech (Computer Science) from LAUTECH (1996). He held Post Graduate Diploma (PGD) in Education from National Teachers' Institute Kaduna (2006). He is presently a research student at the Department of Computer Science and Engineering, LAUTECH. He is a Lecturer 1 at Osun State College of Technology Esa Oke. He is a member of Computer Professionals Registration Council of Nigeria (CPN) and Nigeria Computer Society (NCS). His research areas are soft computing, ICT, Data mining and Database.

A.Q Ayinde obtained his B.Tech (Computer Science) from LAUTECH (2008). He is a research student in the Department of Computer Science and Engineering, Ladoko Akintola University of Technology (LAUTECH) and a Lecturer in the Department of Computer Science at Osun State College of Technology, Esa-Oke. His research areas include Data Mining, ICT, Soft Computing and Database. He is a member of the following professional bodies Redhat Linux (RHCE Certified) Microsoft (MCITP Certified) and Information and Technology in Infrastructure Library (ITIL Certified).

A.B Adetunji obtained her B.Sc (Computer Science) from University of Ibadan (1988), M.Sc (Computer Science) from Obafemi Awolowo University Ile- lfe (2002) and obtained her PhD in Computer Science from LAUTECH (2010). She is a Senior Lecturer, presently working in Computer Science and Engineering Department, LAUTECH. She is a member of Computer Professionals Registration Council of Nigeria (CPN). Her research areas include Database, Artificial Intelligent and Data Mining.

O.A Sarumi obtained his B.Sc (Computer Economics) from OAU Ile-lfe (2001). He is a research student in the Department of Computer Science and Engineering, Obafemi Awolowo University Ile-lfe (O.A.U) and he is a lecturer II in the Department of Computer Science at Osun State College of Technology Esa Oke. He is a member of Computer Professionals Registration Council of Nigeria (CPN) and Nigeria Computer Society (NCS). His research areas are in soft computing, ICT and Database.

Model for determining the impact analysis of open source adoption in software development economics

Muhammad Faisal Akram

*Department Of Computer Science
University Of Agriculture, Faisalabad,
Pakistan*

Ahmad Adnan

*Department Of Computer Science
University Of Agriculture, Faisalabad,
Pakistan*

Abstract

Paper presents the open source software adoption and also shows the effect of open source software reuse on some software development economic factors. Research paper also discussed and identified some of the open source adoption factors which were checked for relationship with software economic factors. Cost, time, productivity and software quality are the software economic factors which were discussed in research. An enhanced research model was proposed after conducting the interviews with different project, quality managers and senior developers from IT industry of Pakistan. On the basis of interview, different hypotheses were formulated between open source software adoption factors and software development economics. How different software organizations will achieve economic gain in terms of software development productivity, to decrease the time and cost to development and improve quality of the software. The results of the paper show the most important and most critical open source software factors. Those factors are developer's experience and OSS adoption which affects the software economics including cost, time, quality and productivity during the open source software development. The results shows that cost and time will be reduced if organizations hire more experienced

developer and similarly, quality and productivity also improved.

Keywords: Open Source Software, Open Source Projects, Software Development, Quality, Productivity.

1. Introduction

Open source is a software development approach to design the soft wares, development of a software, and free distribution of software (source code), offering free access to the source code of that software.

To make source code of a software available to any one at no cost with the permission of changing

The code according to any one's requirement is open source approach. The term "open source" also used as a single term that fits for the following valuable conditions: (1) non-time delimited, complete software for which source code is openly available for (re)distribution without any cost to the user at any time, (2) forces minimal, non-restrictive licensing environments, and (3) is itself either based on technologies which are non-proprietary or on the technologies which are proprietary that conform to (1) and (2) (Thomas and Hunt, 2004). The use of OSS in the field of computer science education has been highlighted in recently years (Attwell, 2005). Developing OSS could also assist learners in their future career paths, suggested by (Cusumano, 2004). The OSS components reuse of saleable software is rising and in many cases a lawful

choice in commercial application (Stone, 2002).

As the source code for open source software is available, so any one can revolutionize this code, fix errors or bugs and get better the developed software from the giving source code. This means any one can change the open source software according to his/her requirements. This is the great facility in open source soft wares. It is also easy to decide what to reuse and where to reuse and when to reuse open source software. In the last few years this way of building software, the open source process, has gained disrepute just as the products of this process have gained market share across key segments of the information economy. Moral values of open source are divides into two different strands:

1. Open Source Ethics as an Ethical School.
2. Open Source Ethics as a Professional Body of Rules.

There are many advantages of reuse which are following- reduced time to market, compact cost for development of product, increased the quality of product and increased the customer satisfaction (Madanmohan and De, 2004). Software is a digital product that can be copied an infinite number of times at zero cost and is thus purely non-rival in economic terms. There are some findings for the important Project starting points which are motivation, community, software development support, licensing and size (Gacek and Arief, 2004). These starting points must keep in mind while starting software development. First of all motivation for all the team, then for which community the software is going to develop, support for that software and at the end the licensing issues and the size of the software .Open source software is being developed in that projects where every one can contribute his best knowledge to improve the software, support and fix from errors. Such projects are called open source projects. Every one can share his/her ideas in these projects (Ruffin and Elbert, 2004).

OSS software communities are virtual work groups consisting of different members with different skills in software development. These members work in temporary, cultural diverse, geographically dispersed, electronically communicating work groups (Maass, 2004). Based on role of every user, open source communities, are generally organized as presented below (Gabriel and Goldman, 2002). To take advantage of reuse, the development team must be aware of the realities of reused products (Samuel, 2006). In this existing research there are 6 variables which are OSS reuse factors and these variables are:

- 1- Developer's experience
- 2- Degree of OSS reuse
- 3- OSS adoption maturity level
- 4- OSS reuse experience and skill
- 5- OSS select criteria
- 6- Difference of small and large organization

The correlation of these variables was checked with the different software development economic factors. There is no research on the developer's experience till today (Sumila, 2006).

Thus the main and the core objectives of this research are to correlate different independent open source adoption factors (described above) with different economic factors. Research can analyze that there is any effect of Developer's experience of development has any positive impact on the economic factors of software like cost, quality, productivity etc. so the objectives of the research are to investigate that

- Is developer's experience has any effect to software economic factors cost, time, security, quality and productivity etc?
- Is there any relationship between degree of OSS components reuse adoption and software economic factors?
- Is OSS adoption maturity level has any effect to cost, time, quality and productivity of software?

- Is there any positive impact of OSS “select criteria” on the software economic factors?
- Is there any difference between the organization of small and large organizations?
- Is there any correlation or relationship between the OSS adoption factors and software development economic factors?

2. Methodology

Pakistani IT companies are the population for this research. The organizations which were targeted in the research are of different sizes. These organizations are developing the software for running the software in their organization. Some of the organizations from these organizations are developing software for the external market including the local and international market. The questionnaire was designed to conduct the survey.

Purpose of conducted research was to identify the relationship between the OSS reuse factors and software economic factors and to find out the effects of OSS reuse on software economic factors. Previous research targeted on some of the OSS adoption factors and some of the software economic factors. There is no research on the developer’s experience till today (Sumila, 2006). All the previous research missed an important point as OSS adoption factor. No one can describe this point in any research which is “Developer’s experience of development”.

My research focuses on OSS adoption factors which are discussed earlier along with one extra variable which is very important to be consider. “Developer’s experience” is very important OSS adoption factor. This factor has a great effect on the software economic factors like cost, time, productivity and quality of open source software. There are also some new software economic factors which were not described in earlier research. These missing

factors are described in this research and explained in detail below.

If a developer is more experienced in development then he can easily reuse the open source software. Experience developer knows very well that how to reuse an open source component and where it has to be reused because he has experience of making different components. He/she also knows that what to reuse? If a developer has less experienced as compared to first and he has only experience of reusing components then he will be not be so important to the organization as first developer. First developer deeply understands the complexity of software and also understands the complexity of integrating the components and merging them. He knows the pros and cons of these techniques so he can better handle the situation of reusing components in a better manner. While a developer only with some experience of reusing open source not able to understand the complexity of integrating components so much. He did not deeply know about the software complexity due to his only experience in reusing.

These are the important questions to conduct this research. Research based on these questions.

Is there any effect of developer’s experience to software economic factors cost, time, security, quality and productivity etc?

Is there any effect of OSS adoption maturity level to cost, time, quality and productivity?

Is there any effect of degree of OSS components reuse adoption on software economic factors? This section defines the methodology used for the research. First of all in the research the surveys of software companies were conducted. For population we can say that software IT companies were chosen for the study. These IT companies are of different sizes. And these are developing different kind of software and also focus on open source adoption reuse at different level. To conduct the research on this topic first a detailed study and literature was carried out. In

the detailed study the focus was on the reusing the different components, development of open source software, quality of the software, productivity of the software and economic factors like security, time and cost issues etc.

To collect the data for research conclusion and to design the survey questionnaire, interviews of different top management were conducted on open source reuse. Top management includes the project managers, quality managers, product managers and senior developers. By this interviews, a questionnaire was designed to take feed backs from different organizations on open source adoption.

From this population, sampling technique was applied to take random samples. Samples are not self selected but these are random. The top management was involved in this research. As above discussed more experienced persons have well known about the complexity of the software developed. Project managers and product managers are directly involved in every phase of the software development. Interview was taken by these people to clarify the objectives of OSS adoption and the reasons why they adopt OSS. Some objectives of the Interview also include the things like:

How the process of OSS adoption can be done by the organization?

What is the importance of developer's over all experience and how much experience is required to be fully and deeply understand the theory for OSS adoption?

What skills and experience of reuse of OSS is required?

Results were collected and research used the questionnaire as follows:

What are the reasons that your organization use the OSS components reuse?

How and from where your organization did you find OSS components to reuse in your project according to need?

During project in which phase or stage your management decides to use open source components for reuse? What is the reason to use OSS in your project?

These are the important questions from different interviewee for research.

3. Research Hypothesis

The main objectives which are considered for the research to be conducted are the OSS adoption factors, developer's experience of development, OSS maturity level, OSS reuse experience, OSS select criteria and the difference between small and large size organizations. All the hypothesis including previous research and the new hypothesis are describe below as follows:

3.1 Developer's experience and skill

This is the point on which no one conducted research. This is the main point of this research. This is ultimately the very important point to be studied. If a developer is more experienced in development then he can easily reuse the open source software. Experience developer knows very well that how to reuse an open source component and where it has to be reused because he has experience of making different components. He/she also knows that what to reuse? The hypotheses for this point are the following:

M1 (a): Developer's experience helps to reduce the cost of software development.

M1 (b): Developer's experience helps to reduce the time for software development.

M1 (c): Developer's experience is positively related to software development productivity.

M1 (d): Developer's experience is positively related to software product quality.

This point is the most important as my research was on this point.

3.2 Degree of OSS adoption

There are many benefits of reuse software that are lower cost of software development, shorter time to market, and in most cases best quality of software (Samuel, 2006).

Because of the claims of high-level OSS adoption, the success stories from literature

review, and the results obtained from our initial study on the subject, we propose the following hypotheses:

- M2 (a): Degree of OSS adoption helps to reduce the cost of software development.
- M2 (b): The degree of OSS adoption helps to reduce software development time.
- M2 (c): The degree of OSS adoption and software development productivity has strong positive relationship.
- M2 (d): The degree of OSS adoption and software product quality are positively related to each other.

3.3 Software Reuse maturity level

Maturity level of OSS reuse is an important and critical point in the consideration of this research. This is important while in the development of open source projects. The other consideration includes that how much the developed software is mature to reuse. Components reuse is a powerful tool that can help significantly if employed correctly (Samuel, 2006).

Regarding maturity level there are the following hypotheses.

- M3 (a): Reuse maturity level for OSS reduces the cost of software development.
- M3 (b): Reuse maturity level for OSS helps to reduce the time of software development.
- M3 (c): There is positive relationship between maturity level and software development productivity.
- M3 (d): Reuse maturity level is positively related to software product quality.

3.4 Reuse experience and skill

A number of researchers have explored the reasons why systematic reuse succeeds or fails (Basili, 1996).

The hypotheses for this factor are the following:

- M4 (a): Reuse experience and skill for open source software reduce the cost of software development.
- M4 (b): Reuse experience and skill for open source software helps to reduce the time for software development.
- M4 (c): Reuse experience and skill is positively related to software development productivity.
- M4 (d): Reuse experience and skill is positively related to software product quality.

3.5 OSS reuse selects criteria

For the selection of OSS reuse adoption all the organizations have deep eye on the components for the reuse in their projects.

- M5 (a): OSS reuse selection criteria for open source software reduce the cost of software development.
- M5 (b): OSS reuse selection criteria for open source software reduce the time of software development.
- M5 (c): OSS reuse selection criteria and software development productivity are positively related.
- M5 (d): OSS reuse selection criteria software product quality is positively related.

This research was conducted in the IT industry of Pakistan. All the variables were conducted through the survey. Questionnaire was designed and distinguished to all the software developers to get their feed backs.

4. Results and Discussion

The main concern of the analysis is to determine the effects of OSS components reuse on the cost of software development, software quality and software productivity. So, the purpose of research is to find the impact of open source reuse on software development economic factors. Thus, from the help of

research we are able to define the relationship between the different variables and between small to medium and large sized organizations (see Fig. 4.1).

Variables

In this section all the variables were defined which were used during the research. There are independent variables, dependent variables and control variable. The independent variables are actually the objectives of our research. We also called that these are the OSS adoption factors.

Independent Variables

Developer's Experience of development

Degree of OSS adoption

OSS adoption maturity level

OSS adoption skill and experience

OSS adoption selection criteria

Dependent Variables

Software cost

Time for software development

Productivity

Quality of software

Control Variable

Company size

4.1. M1 model

Hypotheses M1 (a, b, c and d) were tested and Correlation Coefficient were determined to examine whether there is a positive relationship between the developer's experience and software development economic factors (cost, time, productivity, and quality of the software). Pearson correlation test (see Table 4.1) was conducted. Overall analysis in Table 4.1 strongly supports the 3 hypotheses M1 (a) with a correlation ($r = -0.104$) M1, (b) with a correlation ($r = -0.371$), M1(c) with correlation value ($r = 0.459$) and M1 (d) with a statistically significant strong correlation ($r = 0.477$). From the analysis it is shown that there is a strong relationship between developer's experience and dependent variables (cost, time, productivity and quality). Correlation value of hypotheses M1a and M1b is negative which means that if developer's experience increases the cost and time to the development of

software decreases which strongly supports the hypotheses M1a and M1b. Hypotheses M1 (a) is supported by the analysis in Table 4.1 but the cost of the development not so much decreases as values shown in Table 4.1.

Dependent variables are also called the software economic factors. This research can test the impact of open source software on these software development economic factors. The impact of every independent variable which are five defined above can be checked for every dependent variable or software economic factors. Company size is the control variable. Company size can be control to check the over-all impacts of open source software development economic factors.

In terms of developer's experience, Table 4.1 shows significant statistical strong relationship. Cost of software development decrease very little as developer's experience increases. Time to software development decreases as developer's experience increases which shows positive sign in the development of software. The productivity and quality of software is positively impacted by developer experience, so, productivity and quality will increase if experience of developer goes up and up.

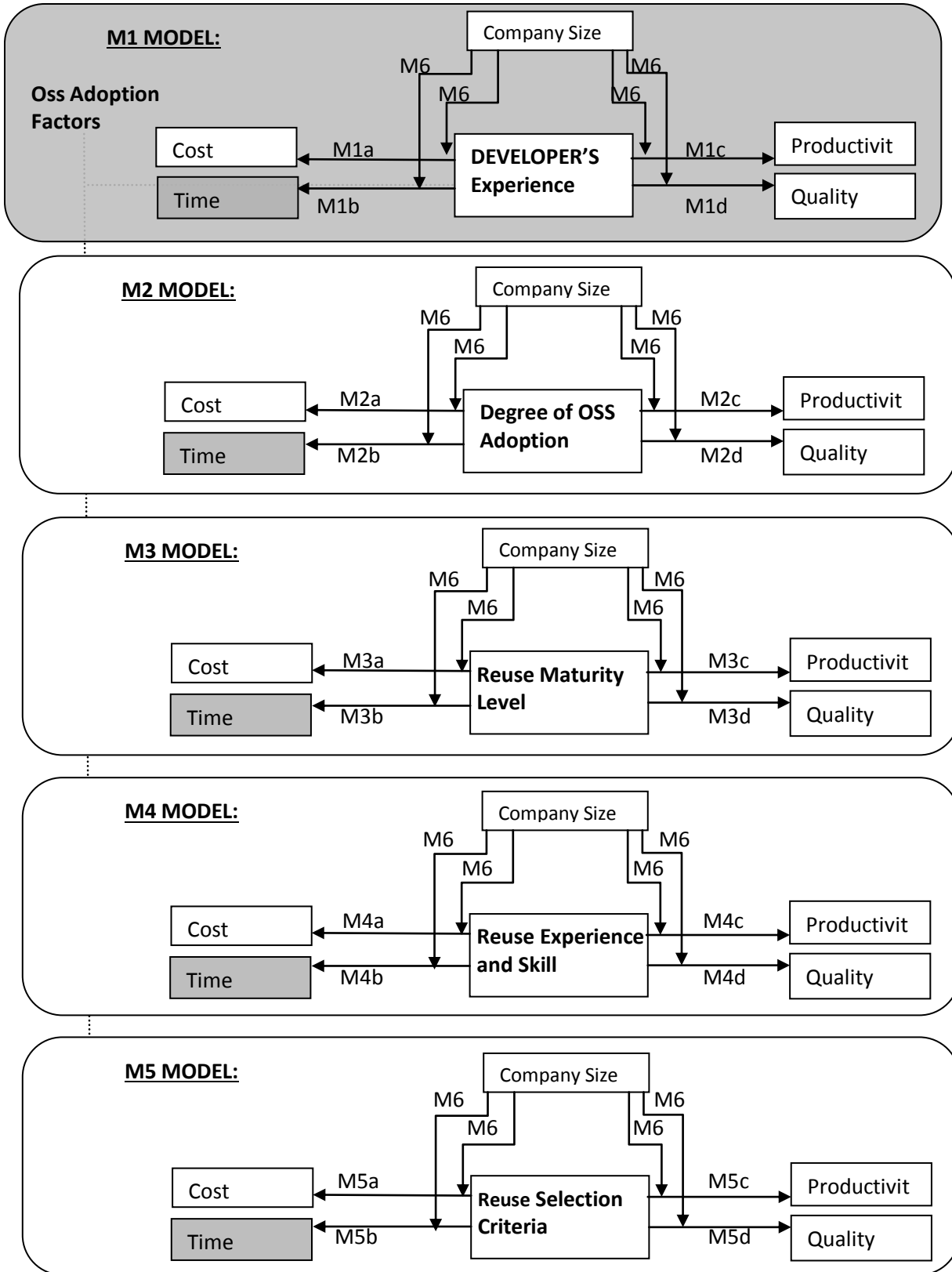


Fig 4.1 – Model for determining s/w development economics effect

4.2. M2 model

Hypotheses M2 (a, b, c and d) were examined and the relationship between the variables were determined. Pearson Correlation was applied to test the hypotheses.

Table 4.2 shows the significance relationship between the degree of OSS adoption and software economic factors (cost, time, productivity and quality). The analytical analysis shows that overall there is a positive relationship between the variables.

Relationship between Degree of OSS adoption and cost is negative with a correlation ($r = -0.048$) which means that cost will decrease if we will increase the degree of OSS adoption. There is a negative relationship between degree of OSS adoption and time; the second hypothesis is also negatively related which also supports the hypothesis M2 (b) with correlation ($r = -0.234$) and which means that time of development will decrease if level of degree of OSS adoption will increase gradually.

The third and fourth hypotheses which are M2 (c) with a correlation ($r = 0.476$) and M2 (d) with correlation ($r = 0.277$) are positively related. **Table 4.1** Correlation between **Developer's Experience** and cost, time, productivity and product quality

	Budget/Cost	Time	Productivity	Quality
Developer's Experience	-0.104	-0.371**	0.495**	0.477**

** . Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed)

Table 4.2 Correlation between the OSS adoption factors and software economic factors

	Budget/Cost	Time	Productivity	Quality
OSS adoption	-0.048	-0.234	0.476**	0.277

** . Correlation is significant at the 0.01 level (2-tailed).

related to the degree of OSS adoption. This means as degree of OSS adoption will increase then the productivity will increase as well as quality of the software will improve.

Therefore, degree of OSS adoption affects the dependent variables to decrease the cost and time and increase the productivity and quality of the software.

What is important here is that the developers are highly motivated to reuse OSS components judging from the results in Table 4.2. The correlations are equally stronger on productivity and quality when it comes to degree of OSS adoption.

4.3. M3 model

This model is about maturity level of OSS and software economic factors. This model also supports the four hypotheses. This model also helps to reduce the cost of development and reduce time to development, increase the productivity and improve the quality of the software.

First and second hypothesis has negative relationship with cost with correlation ($r = -0.172$) which means cost will decrease if OSS component will be more mature and time will decrease for development which has correlation ($r = -0.211$). Third and fourth hypotheses has the positive relationship and has the correlation ($r = 0.276$ and 0.211) respectively.

4.4. M4 model

M4 model is the model of reuse skill and experience. This model determines the relationship of OSS skill and reuse and

software economic factors (cost, time, productivity and quality). This model not supports the hypothesis M3 (a) which has correlation ($r = 0.068$) which means that cost will increase somehow if skill and experience of reuse increases. The second hypothesis has the negative value of correlation ($r = -0.059$) which is benefit and means that as reuse skill and experience increase the time of development decreases. Third and fourth hypotheses also supports and has the positive correlation which are ($r = 0.142$ and $r = 0.211$) respectively.

Table 4.3

Correlation between the **maturity level** and software economic factors (cost/budget, time, productivity and quality)

	Budget/Cost	Time	Productivity	Quality
Maturity Level	-0.172	-0.211	0.276	0.211

Table 4.4

Correlation between the reuse **skill and experience** and software economic factors (cost/budget, time, productivity and quality)

	Budget/Cost	Time	Productivity	Quality
Skill & Experience	0.068	-0.059	0.142	0.211

Table 4.5

Correlation between the **OSS selection criteria** and software economic factors (cost/budget, time, productivity and quality)

	Budget/Cost	Time	Productivity	Quality
OSS Selection Criteria	0.068	0.106	0.144	0.133

4.5. M5 model

M5 model is the model about OSS selection criteria and software economic factors (cost, time, productivity and quality). This model has

not so much impact on different economic factors. The hypotheses in this model are not significantly strong. The Table 4.5 shows that there is no significant relationship between the

variables. The results from the table 4.5 show that productivity and quality of the software increases which is very low to negligible. The values of correlation ($r = 0.144$ and $r = 0.133$) are too low to ignore. Similarly, the cost and time also has correlation values ($r = 0.068$ and $r = 0.106$) which are also too low to negligible. So this model has not so much impact on the software development economics.

5- Conclusion

All the results and hypotheses were discussed in detail. Results show the importance of all the variables. In the previous research one model was missed and previous research declared that the most important thing is the OSS reuse factor. In this research the missed model was included in this research. This model is the "Developer's Experience" in the development of open source software. So there is positive impact of developer's experience to the different software economic factors. This is the most important factor and has more impact on the economic factor of software development process.

The cost of software development process will decrease too little due to most experienced developer. The time decreases if the developer is most experienced. The productivity and quality of software increased if open source software is developed by the most experienced developer as compared to less experienced developer.

From this research it is clear that all IT organizations can achieve economic benefit in terms of reduced cost to software development, reduced time for development, productivity and most important the quality of the software if organizations hire most experienced developers.

This model (developer's experience) has great effect on increasing the productivity of the software and to improve the quality of software developed by most experience developer. This model also influences the time of development. Time will be reduced by the help of this model

which is shown from the results in the Table 4.1.

The second important variable in the research is the degree of OSS adoption by the organizations. Research results also shows that an organization can also get some economic benefits in terms of development productivity and more importantly the software quality if it implements OSS components reuse adoption in a systematic way (Samuel, 2006).

This model also helps to reduce the cost of the software as shown in Table 4.2. Cost reduction by reusing OSS components is benefit for maintenance and security of the software.

The research did not show any significance importance to the OSS selection criteria. This is the less important in the research. However, model (Maturity Level) also shows some significance to reduce the cost of development and to reduce the time to the development if the OSS components are more mature. This model also helps to improve the productivity and quality of the software of most mature OSS components.

Finally, the impact of developer's experience and OSS components reuse on software development cost and time need to be studied separately. Development cost and time has many dimensions (developer's cost, testing cost, Analysis time, Design time, development time and testing time etc), therefore, need to be more exploration on cost and time. Cost to development and time to development are two major factors in software development economics. It has been established that (proprietary) software reuse in general contributes a lot to reduction in time to market as well as cost reduction.

References

- [1] Boehm, B. 1999. "Managing software productivity and reuse", Computer, vol. 32, no. 9, pp. 111-113.
- [2] C. Gacek and B. Arief. pp.83-100.
- [3] Dedrick, J. and J. West., 2003. Why firms adopt open source platforms: a grounded theory of innovation and

- standards adoption. *MIS Quarterly Special Issue on Standard Making: A Critical Research Frontier for Information Systems*, Seattle, December, pp. 236–257.
- [4] El Emam, K., 2005. *The ROI from Software Quality*. CRC Press, ISBN 0849332982.
- [5] Gacek, C. and B. Arief., 2004. The many meanings of open source. *IEEE Software* 21 (1), 34–40.
- [6] Gacek and B. Ariel. pp. 197-203.
- [7] Gaffney, J.E. and T.A.Durek., 1989. Software reuse – key to enhanced productivity: some quantitative models. *Information and Software Technology* 31 (5), 258–267.
- [8] Koltun, P. and A.Hudson., 1991. A reuse maturity model. In: *Fourth Annual Workshop on Software Reuse*. Herndon, VA.
- [9] Kshetri, N., 2004. Economics of Linux adoption in developing countries.
- [10] Lawrie, T. and C. Jones 2002. Goal-Diversity in the Design of Dependable Computer-Based Systems. In *Proceedings of the Open Source Software Development Workshop, Newcastle upon Tyne, UK, February 25-26, 2002*, ed. C. Gacek and B. Ariel. pp.130-154.
- [11] Lerner, J. and J. Tirole., 2002. Some simple economics of open source. *Journal of Industrial Economics* 50 (2), p. 197, 38p.
- [12] Madanmohan, T.R. and R. De., 2004. Open source reuse in commercial firms. *Software, IEEE* 21 (6), 62–69.
- [13] Nasr, R., 2006. Open source software: The use of open source GIS software and its impact on organizations. *Software, IEEE* 21 (1), 74–81.
- [14] Ruffin, M. and C. Ebert., 2004. Using open source software in product development: a primer. *IEEE Software* 21 (1), 82–86.
- [15] Samuel, A. and J.A.Bailetti., 2006. Measuring change in software product line – an experience report. *International Journal of Product Development (IJPD)* 3 (2).
- [16] Samuel, A., 2006. The impact of firm size on knowledge reuse and exploration during software product development: an empirical study. In: *2006 IEEE International Conference on Information Reuse and Integration (IEEE IRI-2006)*, Waikoloa, Hawaii, USA, September 16–18, pp. 160–165.
- [17] Thomas, D. and A. Hunt., 2004. Open source ecosystems. *Software, IEEE* 21 (4), 89–91.
- [18] Wang, H. and C.Wang., 2001. Open source software adoption: a status report. *Software, IEEE* 18 (2), 90–95. Tony

Application of the Gradient Vector Flow Method for Treating Satellite Image

Mohammed AIT OUSSOUS¹, NourEddine ALAA² and Youssef AIT KHOUYA³

¹ *University Cadi Ayyad, UFR Metrology Automatic and System Analysis
Faculty of Science and Technology of Marrakech*

² *University Cadi Ayyad, Laboratory LAMAI, BP 549 Marrakech Morocco.*

³ *University Cadi Ayyad, UFR Metrology Automatic and System Analysis
Faculty of Science and Technology of Marrakech*

Abstract

In this work we propose a new approach to delineate regions in Satellite Image. This method is particularly applicable to the limitation of insoluble habitats. We are interested in detecting the boundaries of the villages near Marrakech from their satellite images, as a study case: Ouahat Sidi Brahim. Our approach has three stages: The first, using k nearest neighbour method for classification of the satellite image. The second, using a mean filter to regularize this classification. In the last we use the Gradient Vector Flow to detect the edge of the village. Our method gives satisfactory and encouraging results to detect the edge of the village and can be used as a tool to control the spread of slums.

Keywords: *Classification, Satellite Image, k-Nearest Neighbour (k-NN), median filter, active contour, Gradient Vector Flow.*

1. Introduction

Mathematics is heavily involved in developing science. These interactions revitalize and strengthen the field. The field of image analysis is clearly the future, therefore, mathematicians should be involved in image processing as the most important and exciting scientific discoveries of all time are in the field of computer vision. Research in image processing is useful and interesting, and the results are usually relevant. The image processing is the set of methods and techniques operating on them in order to make this operation possible, simpler, more efficient and more enjoyable, improve the visual image and extract information deemed relevant. The reader interested in this vast subject may consult the following references for more details [1], [2], [3], [4], [8].

Here we focus particularly on edge detection of the village from their satellite image to limit the spread of slums which is a district precarious illegal building formed by a poor suburb in or near a big city. This form of urban sprawl is growing almost always occupied government land without authorization, and therefore these parts become places of crime by excellence.

We are interested in the region of Marrakech is the first tourist destination in Morocco, appreciated for its quality of life and its landscapes, its rich heritage and historical conditions of easy access. It is perceived as a destination of confidence.

In this paper we propose a method based on remote sensing and Gradient Vector Flow to control the spread of slums in the regions of Marrakech. We took as an example Ouahat Sidi Brahim village located 20Km from the center of Marrakech.

This article is organized as follows: first time we discussed the classification method and more specifically the method of k-NN, then to regulate our classification we apply a median filter followed by a thresholding to eliminate isolated points. Edge detection of the village is obtained by applying the GVF method, this is the subject of Section 3. In the last section we present experimental results and discussion

2. Classification and treatment of satellite image

2.1 Principle of k-NN method

The classification is a method of data analysis that aims to separate the image into several classes of interest, ie to aggregate data in homogeneous subsets [1], [6], [7], which share common features. It may be supervised or automatic (ie unsupervised). In the supervised case, it requires a subset of data with known classes, called learning base in order to classify new data. Concerning our problem, we must determine two classes, the class of pixels buildings and the fields. For this we adopted the method of k-NN [1], [5], which is a supervised learning algorithm where the result of new instance query is classified based on majority of k-NN category. The purpose of this algorithm is to classify a new object based on attributes and training samples. The classifiers do not use any model to fit and only based on memory. Given a query point, we find k number of gold objects (training points) closest to the query point. The classification is using majority vote among the classification of the k objects. Any ties can be broken at random. The k-NN algorithm use the neighborhood classification as the prediction value of the new query instance. The algorithm of k-NN is as follows [1], [5]:

For each pixel (x,y) in the image:

- Calculate distances between every pixel and vector based $B_{\{i\}}$ of the database: $d((x,y), B_{\{i\}})$
- Search the k vectors of the training set closest to the pixel, ie, those with the k smallest distances.
- Assigns the pixel to the class most represented among these k vectors.

2.2 Post treatment

After classification, usually we get isolated pixels, poorly sorted and contains areas of small discontinuities. To improve the mapping, we must rectify by reallocating these isolated pixels. In our case to regularize the classification we used a mean filter followed by an adaptive thresholding.

The principle of the filter medium is very simple: a pixel is replaced by the average of itself and its neighbors. In the definition of neighborhood that the filters will vary.

Because of the variety of image content filtering (size M x N) we apply an adaptive thresholding to free ourselves of local contrast and allow good detection of edges in all regions of the image. We choose a threshold of the form:

$$t = m + \beta s$$

$$m = \sum u(x, y) / MN$$

$$s = \sqrt{\sum (u(x, y) - m)^2 / MN}$$

m and s are respectively the mean and standard deviation of the image u obtained in first step and we set β equal to 0.6. The thresholding is to compare the gray level of each pixel (x,y) of the image with a global fixed threshold t. In our case we use the new value of pixel thresholding given by the following expression:

$$U(x, y) = 255 \text{ if } u(x, y) \geq t \text{ and } U(x, y) = 0 \text{ if } u(x, y) < t$$

Where U is a new image obtained after post treatment.

3. THE DETECTION OF THE CONTOUR BY THE GVF METHOD

3.1 Gradient Vector Flow Field

In [9], [10] Xu and Prince proposed the gradient vector flow to resolve the problem of initialization and poor convergence of the classical model of snakes [y]. The energy function to be minimized by a snake can be described as follows:

$$E = \int_0^1 \left(\alpha \left| \frac{d}{ds} v(s) \right|^2 + \beta \left| \frac{d^2}{ds^2} v(s) \right|^2 + \nabla f(v(s)) \right) ds \quad (1)$$

Where, $\{v(s) = (x(s), y(s)), s \in (0,1)\}$ is a snake curve, and α and β are two parameters which satisfy $\alpha + \beta = 1$. The first two terms in (1) are internal energies. The latter term represents an external energy in this case f is a function of rising edge detector, which depends on the image. If a snake realizes the minimum of the energy E then the Euler equation is satisfied:

$$\alpha \frac{d^2 v}{ds^2}(s) - \beta \frac{d^4 v}{ds^4}(s) - \nabla E_{ext} = 0 \quad (2)$$

where, $\frac{d^2 v}{ds^2}(s)$ and $\frac{d^4 v}{ds^4}(s)$ are the second and fourth derivatives of $v(s)$ with respect to s , respectively, and

$$E_{ext} = -f$$

The gradient vector flow is defined as the vector field $v(x, y) = [u(x, y), v(x, y)]$ which minimizes the energy function:

$$E_{gvf} = \iint \left(\mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |v - \nabla f|^2 \right) \quad (3)$$

where, u_x and u_y are the first order partial derivatives of $u(x,y)$, and v_x and v_y are the first order partial derivatives of $v(x,y)$, respectively. It is easily shown that if $|\nabla f|$ is small. The function of the energy used to get all the minimum value when $v = \nabla f$. The solution to (3), which is the GVF field, can be found by solving the following Euler equations:

$$\mu \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0 \quad (4)$$

$$\mu \nabla^2 v - (v - f_y)(f_x^2 + f_y^2) = 0 \quad (5)$$

3.2 Numerical Implementation

Equations (4) and (5) can be solved by treating and as functions of time and solving

$$\frac{\partial u}{\partial t} - \mu \nabla^2 u + (u - f_x)(f_x^2 + f_y^2) = 0 \quad (6)$$

$$\frac{\partial v}{\partial t} - \mu \nabla^2 v + (v - f_y)(f_x^2 + f_y^2) = 0 \quad (7)$$

for the numerical implementation we use the following discretization

$$u_{i,j}^{t+dt} = u_{i,j}^t + \frac{\mu dt}{|N_r|} \sum_{(k,l) \in N_r} (u_{k,l}^t - u_{i,j}^t) + dt(u_{i,j}^t - dx f_{i,j}^2 + dy f_{i,j}^2) \quad (8)$$

$$v_{i,j}^{t+dt} = v_{i,j}^t + \frac{\mu dt}{|N_r|} \sum_{(k,l) \in N_r} (v_{k,l}^t - v_{i,j}^t) + dt(v_{i,j}^t - dx f_{i,j}^2 + dy f_{i,j}^2) \quad (9)$$

When Sobel gradient operator is used to calculate (f_x, f_y) .

4 APPLICATION

Images used are of size (512 × 512) pixels obtained from google maps, with a scale equal to (1/(20000)). The algorithm of k-NN is tested on these images and the parameter k is set at 3. Figure (2) shows the obtained

results. The calculation time is proportional to the size of the training set, which is logical since the algorithm calculates the distance between each new object and each element of the training set. To improve the classification obtained by taking into account the spatial information provided by neighboring pixels in the image, thresholding on the average yields good results as shown in Figure (3).

We apply the method of GVF to each separate region in our image, taking the initial contour is a circle outline that surrounds this region, will evolve until it reaches the edge of the village as it is represented in Figure (4).



Fig. 1. Image of Ouahat Sidi Brahim from

5 Conclusion

In this article, we are concerned with the problem of slums in limited regions of Marrakech from satellite images of high resolution. Our approach has two steps: the first is the classification of satellite images by the technique of k-NN, the second. Our method gives satisfactory results and encouraging, and this, by comparing the contours detected at different times which are intended to confirm the spread of slums around. For detection of the contour we applied the method of GVF, the initial contour is a circle, will evolve until it reaches the edge of the village as it is represented in Figures (4).

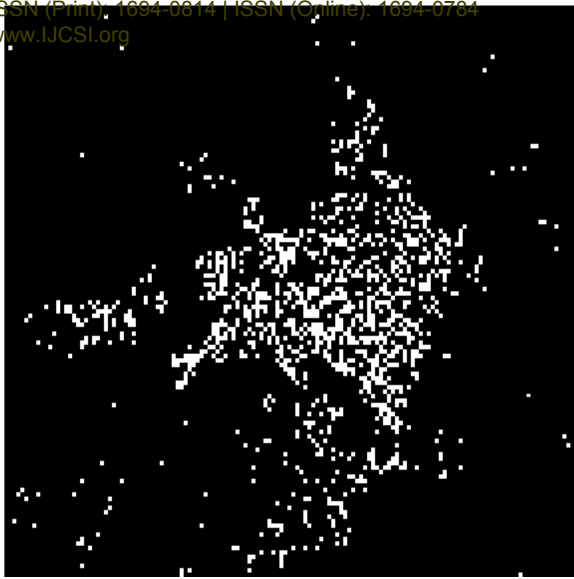


Fig. 2. Image after classification by k-NN



Fig. 3 Image after post treatment



Fig. 4 Final contour of the village Ouahat Sidi Brahim

Acknowledgment

The authors wish to thank Professor Michel Pierre of the ENS Cachan France for his valuable discussion.

REFERENCES

- [1] Bahram, Javidi, Image recognition and classification: Algorithms, Systems, and Applications, Marcel Deckkeerr 2002 .
- [2] Chan, F. T. and J. Jackie Shen , Image Processing and Analysis - Variational, PDE, Wavelet, and Stochastic Methods. Society of Industrial and Applied Mathematics, 2005.
- [3] Morris , T., Computer Vision and Image Processing. Palgrave Macmil, 2004.
- [4] Ian.Duda R.O., Hart P.E., Stork D.G. Pattern classification (2ed., Wiley), 2000.
- [5] R. A, Schowengerdt, Remote sensing : models and methods for image processing. Academic Press, Second edition.R. Nicole, "The Last Word on Decision Theory," J. Computer Vision, submitted for publication. (Pending publication) , 1997.
- [6] O. Zammit, Détection de zones brûlées après un feu de forêt à partir d'une seule image satellitaire Spot 5 par techniques SVM. Université de Nice-Sophia Antipolis , 2008.
- [7] C.Samson, Contribution à la classification d'image satellitaires par approche variationnelle et équations aux dérivées partielles. Thèse de doctorat de l'Université de Nice-Sophia Antipolis , 2000.
- [8] Ünsalan Cem , Kim L. Boyer, Multispectral Satellite Image Understanding From Land Classification to Building and Road Detection Springer 2011.
- [9] C. Xu and J.L. Prince, "Gradient Vector Flow: A New External Force for Snakes," Proc. IEEE Conf. on Comp. Vis. Patt. Recog. (CVPR), Los Alamitos: Comp. Soc. Press, pp. 66-71, June 1997
- [10] C. Xu and J. L. Prince, "Gradient Vector Flow Deformable Models", Handbook of Medical Imaging, edited by Isaac Bankman, Academic Press, September, 2000.

First Author Mohammed AIT OUSSOUS had his diploma DESA in telecommunications and networks in 2008 from the University of Cadi Ayyad, Morocco. He is actually working as a qualifying high school professor and prepares his Ph.D. in image processing.

Second Author Pr. NourEddine ALAA received his Master of Science and his Ph.D. degrees from the University of Nancy France respectively in 1986 and 1989. In 1996, he received the HDR in Applied Mathematics from the University of Cadi Ayyad, Morocco. He is currently Professor of modeling and scientific computing at the Faculty of Sciences and Technology of Marrakech. His research is geared towards non-linear mathematical models and their analysis and digital processing applications.

Third Author received in 2008 his Master in Telecommunications and Computer Networks from the University of Cadi Ayyad, Morocco. He is currently a Ph.D. Student at the Faculty of Sciences and Technology Marrakech. His research interests include Image Segmentation, Restoration and Classification..

An Enhanced Application of Modified PSO for Association Rule Mining

Bharathi.T¹ and Dr.P.Krishnakumari²

¹Computer Science Department, RVS Arts and Science College, Coimbatore, Tamilnadu-641402 India.

² Director of MCA, RVS Arts and Science College, Coimbatore, Tamilnadu-641402 India

Abstract

In data mining, association rule learning is a well-liked and well explored technique for finding out interesting relatives in large databases along with variables. It analyzes and present strong rules discovered in databases by means of diverse measures of interestingness. As all know two important parameters, minimal support and confidence, are forever deciding by the decision-maker him/herself or in the course of trial-and-error; and thus, the previous algorithms be deficient in both objectiveness and competence. As a result, the main purpose of proposed work is to recommend an improved algorithm that can provide feasible threshold values for minimal support and confidence. Earliest particle swarm optimization algorithm investigates for the optimum fitness value of each particle and next discovers equivalent support and confidence as minimal threshold values subsequent to the data are distorted into binary values. To improve the feasibility of the work the modified PSO (Particle Swarm Optimization) algorithm has a number of swarm population size, the number of highest generation, and three predetermined parameters will be determined C_w, C_p, C_g . In each generation, the particle's position value in all measurements will be reserved or be updated by its pbest value or be updated by the gbest value or restored by generating a new random number.

Keywords: Data Mining, Particle Swarm Optimization, Modified PSO, Minimal Support and Confidence, Association Rule Learning, Swarm intelligence.

1. Introduction

Data mining is an emerging technique to address the problem of reconstructing data into useful knowledge information from the user who can mine the results which they really want. That the rules are generated according to the knowledge by data mining algorithms in which one of most problematic steps in an association rule is discovery process knowledge validation. To solve this problem association rules have been widely used in many application domains for finding patterns in data and to generate the association rules. The pattern reveals

combinations of events that occur at the same time based on the interesting associations and/or correlation relationships among a large set of data items. The attributes value conditions will be shown by association rule that occur frequently together in a given dataset. Apart from the antecedent (the "if" part) and the consequent (the "then" part), an association rule has two numbers that express the degree of uncertainty about the rule. First is Support: It is simply the number of transactions that include all items in the antecedent and consequent parts of the rule. Second is Confidence: It is the ratio of the number of transactions that include all items in the consequent as well as the antecedent to the number of transactions that include all items in the antecedent. The objective of data mining is to find out significant associations along with items such that the occurrence of various items in a transaction will entail the occurrence of some other items.

To accomplish this principle, suggested quite a lot of mining algorithms depends on the perception of large item sets to discover association rules in the transaction data mining process into two stages. In the first stage, candidate item sets were produced and calculated through scanning the transaction data. The value is called as minimum support if the amount of an item set to emerge in the transactions be larger than a pre-defined threshold value then the item set is considered as a large item set. At the second stage the association rule, rules is generated from the first stage's result of large item sets. For each large item set all feasible association permutations were formed and the value is called as minimum confidence the output will be as association rules, when individuals calculated with confidence values larger than a predefined threshold. Modified PSO has been developed based on the standard PSO [16] in which each particle is implied as a positive integer number. In this work, modified PSO algorithm is exploited to resolve rule mining problem which can manage with the dataset enclosing both the discrete and continuous variables. This advance is significantly diverse from other methods which had only joined data mining

and PSO together. Most of their attempts had utilized to compact with the development of PSO as an optimization technique to resolve the data mining problems, for instance classification and clustering. To get better performance of the proposed algorithm planned to integrate the local search scheme to perform globally best solution obtained in each generation.

The main contribution of work is as follows:

1. First the input data from the database are transformed into binary data type
2. A novel technique of application that uses the concept of the particle swarm optimization approach.
3. Applying the modified PSO for a feasible solution of support and confidence.
4. At last Mining the association rule

The remainder of this work is described as follows: In section 2 the related work is dealt and in section 3 the proposed work is explained. In section 4 the experimental results and discussion is explained. In section 5 the conclusion of the work is described.

2. Previous Research

2.1 Association rule mining

Association rule mining is the most important and well researched data mining techniques were first introduced by J. Han [1]. In this method are used to extract the correlation, frequent interaction patterns, frequent item set association or casual relationship between the set of items in the transaction database or repositories. Association rule mining methods used in several areas such as telecommunication networks and risk management, inventory control etc. It is to find out the association rules that satisfy the predefined minimum support and confidence from a given database. Sometimes the association rules are very large to validate the results of each and every generation of rules. Several strategies have been proposed to reduce the number of association rules, such as generating only “interesting” rules, generating only “non redundant” rules, or generating only those rules by satisfying the certain number of criteria such as coverage, lift or strength and leverage.

2.2 Weighted Clustering based Particle Swarm Optimization (PSO) For MANET

Clustering is a method of organizing things into meaningful groups with respect to their similarities or grouping of the data. Clustered results the elements in a group are similar to each other but are different from other groups. The objective of clustering is to identify the groups in such a way that the identified groups are excluded so that any instance belongs to a single group. It

is very similar to a graph partitioning problem in the PSO. Optimally portioning a graph is an NP-hard problem with respect to certain parameters. In this method the set of cluster-heads is called the dominating sets S of the graph. Due to the mobility of the network, the nodes can go outside the transmission range of their cluster-head and move into another cluster thus changing their neighborhood. The PSO optimization result changes the number of clusters and the number of nodes in a cluster but this does not result in a change of the dominant set at all.

Clustering of nodes in MANETs is one of the biggest challenges. Finding the optimal number of clusters that cover the entire network becomes essential and an active area of research. Although, several authors have proposed different techniques to find the optimal number of clusters, none of them addresses all the parameters of a mobile ad hoc network. Clustering has numeral advantages in MANETs. The performance of the system can be improved by allowing the reuse of system resources. It can optimally manage the network topology by dividing the task among specified nodes called cluster-heads, which is very useful for network management and routing [2].

The clustering algorithm must be distributed, since every node in the network has only local knowledge and communicates outside its group only through its cluster-head as in the case of cluster-based routing. The algorithm should be robust as the network size increases or decreases; it should be able to adapt to all the changes. The clusters should be reasonably efficient, i.e. the selected cluster heads should cover a large number of nodes as much as possible. In this work, they propose a Comprehensive Learning Particle Swarm Optimization (CLPSO) based clustering algorithm to find the optimal number of clusters for mobile ad hoc networks. Particle swarm optimization is a stochastic search technique. It has simple parameters that need to be tuned during the execution of the algorithm. It has been an efficient and effective technique to solve complex optimization problems. Each particle contains the IDs of all mobile nodes of the network. The algorithm takes a set of parameters of MANETs into consideration such as mobility of nodes, transmission power, battery power and moving speed of the nodes. It is a weighted clustering algorithm in which each of these parameters is assigned a weight such that the sum of all the weights is equal to one.

2.3 Optimization based methods for frequent item set

Genetic algorithms (GA) Optimization techniques have also been applied in ARM [3]. In this Genetic algorithm based system assigns weighted values to the importance of individual items. These weighted values based items apply to the fitness function of heuristic genetic algorithms to estimate the value of different association rules. These

genetic algorithms can generate a suitable threshold value for association rule mining. In addition, Saggari et al. Proposed an approach concentrating on optimizing the rules generated using genetic algorithms. The approach predicts the more number negative attributes in the association rule [4] another study, a genetic algorithm was employed to mine the association rule oriented to the dataset in a manufacturing information system (MIS). According to the test results, the conclusion drawn stated that the genetic algorithm had a considerably higher efficiency [5].

In another study, an ant colony based algorithm was also employed in data mining under multi-dimensional constraints. The computational results showed that the ant colony based algorithm could provide more condensed rules than the Apriori method. In addition, the computation time was also reduced [6]. In addition, this method was integrated with the clustering method to provide more precise rules [7]. First, the dataset is clustered with the self-organizing map (SOM) network and the association rules in each cluster are then mined by an ACS-based association rule mining system. The results show that the new mining framework can provide better rules.

2.4 Particle swarm optimization (PSO) for applications

After the PSO algorithm was proposed in 1995, besides the above mentioned modifications, many different kinds of applications have been developed. PSO is applied to learn neural networks and it can classify XOR problem precisely. The results have shown that PSO can learn simple neural networks. Moreover, PSO was also utilized to develop the weight and structure of a neural network in 1998. It is more efficient than traditional training algorithms. Applications of PSO are gradually increasing, like in the medical treatment of human tremors of diseases such as Parkinson's disease, and in industrial automation of computer-aided design and manufacturing (CAD/CAM) [8]. In addition, PSO has been applied in clustering analysis. Cohen and Castro [9] presented a modified PSO that featured self organization of the updating rule for clustering analysis. In their PSO, it is not necessary to calculate fitness value. The results show that it is better than the K-means method. Kuo et al. [10] Proposed a PSKO which combined PSO-clustering with K-means. The PSKO was evaluated in four data sets, and compared with the performance of K-means clustering, PSO-clustering and hybrid PSO. The experimental results show that the PSKO algorithms outperform other algorithms. In addition, Kuo and Lin [11] further used binary PSO to solve a clustering analysis problem and applied it to an order clustering problem. Chen and Ye proposed a PSO based clustering algorithm [12], which they called PSO-clustering. This method used minimal

target function in PSO to automatically search for the data group center in multi-dimensional space. Compared with traditional clustering algorithms, PSO-clustering requires fewer parameter settings and avoids local optimal solutions.

2.5 An Optimized Particle Filter based on PSO Algorithm

Optimized PSO-UPF [13] was proposed for nonlinear dynamic systems. Based on the concept of re-sampling step where the particles with larger weights should be re-sampled more time. PSO-UPF optimization filter based algorithm the after calculation of weight values for particles, some particles will join in the refining process which means particles with higher weight values are moving to the region. The proposed PSO-UPF algorithm was compared with other filtering algorithms and variances of PSO-UPF are lower than other filtering algorithms. Major important feature in the particle optimization based filters is that the random measurement of the weight values and is recursively updated. Here three major operations are performed such as sampling, weight computation, and re-sampling. In sampling step one generates a set of new particles that represents the support of the random measure and with weight computation; one calculates the weight values of the particles. Re-sampling is an important operation because without this step proposed system will get poor results. Re-sampling step can be performed in two ways one replicates the particles that have larger weights and removes the ones with negligible weights.

2.6 Quantum Evolutionary based optimization Algorithm

Quantum Evolutionary Algorithm (QEA) is an optimization algorithm proposed by [14]. Here the QEA algorithm is performed by combining both the quantum based computing and Particle Swarm Optimization (PSO). By combining both these methods improves the performance of the system and solve optimization problems it is named as PSEQEA. PSEQEA is the algorithm used to solve multi-objective Optimization (MO) problems and single optimization problems. In this method non trivial points are used to evaluate the performance of the system to detect the Pareto optimal points and the shape of the Pareto optimal points by using both Fixed Weighted Aggregation method and Adaptive Weighted Aggregation method. The global optimization problem is still becoming a major problem in multidimensional functions. Quantum based concept with optimization techniques solves the optimization problems in function or multidimensional data. Hota et al proposed Quantum-behaved particle swarm optimization (QPSO)

[15] algorithm for global optimization of multi-dimensional functions. In this research, a modified and improved QPSO using fitness weighted recombination operator along with a fitness proportionate selection mechanism proposed to improve or solve optimization problems in the data or multidimensional functions. The experimental results are tested with different benchmark functions and compared with PSO and QPSO.

3. PSO and Modified PSO Based Association Rule Mining

3.1 Encoding

According to the definition of association rule mining, the intersection of the association rule of item set X to item set Y ($X \rightarrow Y$) must be empty. Items which appear in the item set X do not appear on item set Y, and vice versa. Hence, both the front and back partition points must be given for the purpose of chromosome encoding. The item set before the front partition point is called "item set X," while that between the front partition and back partition points is called "item set Y." The chromosome encoding approach in this study is "string encoding." Each value represents a different item name, which means that item 1 is encoded as "1" and item 2 is encoded as "2." The representative value of each item is encoded into a string type chromosome by the corresponding order.

3.2 Fitness value calculation

The fitness value is utilized to evaluate the importance of each particle. The fitness value of each particle comes from the fitness function.

$$Fitness(k) = confidence(k) \times \log(support(k) \times length(k) + 1)$$

Fitness (k) is the fitness value of association rule type k. Confidence (k) is the confidence of association rule type k. Support (k) is the actual support of association rule type k. Length (k) is the length of association rule type k. The objective of this fitness function is maximization. The larger the particle support and confidence, the greater the strength of the association, meaning that it is an important association rule. In the equation above, support, confidence and item set length must be calculated before calculating fitness value. This study uses the binary type data search method. This method first arranges the original data into a two-dimensional matrix where rows represent data records and columns represent product items.

3.3 Population generation

In order to apply the evolution process of the PSO algorithm, it is necessary to first generate the initial population. Here we select particles which have larger fitness values as the population. The particles in this population are called initial particles.

3.4 Search the best particle

First, the particle with the maximum fitness value in the population is selected as the "gbest." The initial velocity of the particle is set to be $v_0 = 0$, while the initial position is $x_0 = (2,5,1,3,4)$. The particle's initial "pbest" is its initial position, and it is updated as shown below:

$$v_{id}^{new} = v_{id}^{old} + c_1 rand() (pbest - x_{id}) + c_2 rand() (gbest - x_{id}), x_{id}^{new} = x_{id}^{old} + v_{id}^{new}$$

Since the values calculated by these two equations may not always be an integer or fall in the range (1, 5), we designed a method to constrain the search. The constrained method is to calculate the distance between the particle's new position and all the possible particles inside the constrained range before the particle's position is updated. Definitely, the particle with the smallest distance will be selected and treated as the particle's new position. In the distance measuring function, we use traditional "Euclidean distance" as shown below:

$$dist(x^n, y^m) = \sqrt{\sum_1^d (x_i^n - y_i^m)^2}$$

Where x^n the position of the particle at nth update and y^m is the possible particle number m in the constrained range. In addition, d is the dimension of the search space. The nearest possible particle is selected to be the target particle's new position. This method can prevent a particle from falling beyond the search space when its position is updated.

3.5 Termination condition

To complete particle evolution, the design of a termination condition is necessary. The evolution terminates when the fitness values of all particles are the same. In other words, the positions of all particles are fixed. Another termination condition occurs after 100 iterations and the evolution of the particle swarm is completed.

Finally, after the best particle is found, its support and confidence are recommended as the value of minimal support and minimal confidence. These parameters are employed for association rule mining to extract valuable information.

3.6 Item set weight

Based on the item weight $w(i)$, the weight of an item set, denoted as $w(is)$, can be derived from the weights of its enclosing items. One simple way is to calculate the average value of the item weights, denoted as:

$$w(is) = \frac{\sum_{k=1}^{|is|} w(i_k)}{|is|}$$

3.7 Transaction weight

Transaction weight is a type of item set weight. It is a value attached to each of the transactions. Usually the higher a transaction weight, the more it contributes to the mining result. In a supermarket scenario, the weight can be the "significance" of a customer who made a certain transaction.

3.8 Weighted support

Weighted support WSP of an item set. A set of transactions T respects a rule R in the form $A \rightarrow B$, where A and B are non-empty sub-item sets of the item space I and they share no item in common. Its weighted support is the fraction of the weight of the transactions that contains both A and B relative to the weight of all transactions. This can be formulated as:

$$wsp(AB) = \frac{\sum_{k=1}^{|w_{sr}| \& (A \cup B) \subseteq t_k} weight(t_k)}{\sum_{k=1}^{|w_{sr}|} weight(t_k)}$$

By this means, weighted support is modeled to quantify the actual quota of an item set in the transaction space in weighted association rule mining scenario. The weighted support of an item set can be defined as the product of the total weight of the item set (sum of the weights of its items) and the weight of the fraction of transactions that the item set occurs in. The goal of the weighted association rule mining is then changed to determining all rules that are above a user specified minimum weighted support threshold holding a minimum user specified confidence. In order to calculate weighted support of an item set, we need a method to evaluate transaction weight.

The transaction weight (t_k) can be derived from weights of the items presented in the transaction. One may formulate it easily as the average weight of the items presented in the transaction. Note that $WS_t(t_k)$ denotes the inner-transaction space for the k_{th} transaction in transaction space WS_T .

$$weight = \frac{\sum_{i=1}^{|w_{sr}(t_k)|} weight(item(k))}{|w_{sr}(t_k)|}$$

This value is used to calculate the weighted support of a potentially significant itemset described above. The item set is then validated as significant if its weighted support is above the pre-defined minimum weighted support.

The Modified PSO the number of swarm population size, the number of maximum generation, and three predetermined parameters will be determined. In every generation, the particle's position value in each dimension will be kept or be updated by its pbest value or be updated by the gbest value or replaced by generating a new random number according to the procedure depicted (1). In this equation, $i = 1, 2, \dots, m$, where m is the swarm population. $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$, where x_{id} is the position value of the i th particle with respect to the d^{th} dimension ($d = 1, 2, \dots, D$) of the feature space. C_w, C_p and C_g are three predetermined positive constant with $C_w < C_p < C_g$. $P_i = (p_{i1}, p_{i2}, \dots, p_{id})$ denotes the best solution achieved so far by itself (pbest), and the best solution achieved so far by the whole swarm (gbest) is represented by $G_i = (g_{i1}, g_{i2}, \dots, g_{id})$. X represents the new values for the particle in every dimension which are randomly generated from random function $rand()$, where the random number is between 0. The update strategy for particles' position value in Modified PSO is presented below.

Step 1: Initialize the swarm size (m), the maximum generation (max_{Gen}), the maximum fitness value (max_{Fit}), C_w, C_p and C_g .

Step 2: In every iteration, a random number R that is in the range of 0 and 1 will be randomly generated for each dimension.

Step 3: Perform the comparison strategy where:
 if ($0 \leq R < C_w$), then $\{x_{id} = x_{id}\}$;
 Else if ($C_w \leq R < C_p$), then $\{x_{id} = p_{id}\}$;
 Else if ($C_p \leq R < C_g$), then $\{x_{id} = g_{id}\}$;
 Else if ($C_g \leq R \leq 1$), then $\{x_{id} = new(x_{id})\}$;

Step 4: This process will be repeated until the termination condition is satisfied.

$$x_{id}^t = \begin{cases} x_{id}^{t-1} & \text{if } rand() \in [0, C_w) \\ p_{id}^{t-1} & \text{if } rand() \in [C_w, C_p) \\ g_{id}^{t-1} & \text{if } rand() \in [C_p, C_g) \\ x & \text{if } rand() \in [C_g, 1) \end{cases}$$

Step 5: Proceed the step 3 until convergence met.

Step 6: Find the best solution.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this module we measure the performance of the system in terms of the association rule mining accuracy and rule quality. The mining accuracy and result of the rule quality is measured by using the following equations: Data will be divided into two parts: training data and testing data. Training data is used to generate a model according to the given rules in the target problem, and later the model will be used on the testing data to obtain the validation accuracy. How well the rule will perform in the testing phase will depend on the reliability of the mining accuracy measurement. The standard mining accuracy rate can be written as:

$$\text{Standard mining accuracy rate} = \frac{TP+TN}{TP+TN+FP+FN}$$

The quality of the resulting rule is evaluated according to the rule-evaluation function

$$\text{Quality} = \text{sensitivity} \times \text{specificity} \frac{TP}{TP+FN} \times \frac{TN}{TN+FP}$$

True positive (TP): the numbers of examples that covered by the rule that have the class predicted by the rule. False positive (FP): the numbers of examples covered by the rule that have a class different from the class predicted by the rule. True negative (TN): the number of examples that are not covered by the rule that have a class different from the class predicted by the rule. False negative (FN): the number of examples that are not covered by the rule that have the class predicted by the rule.

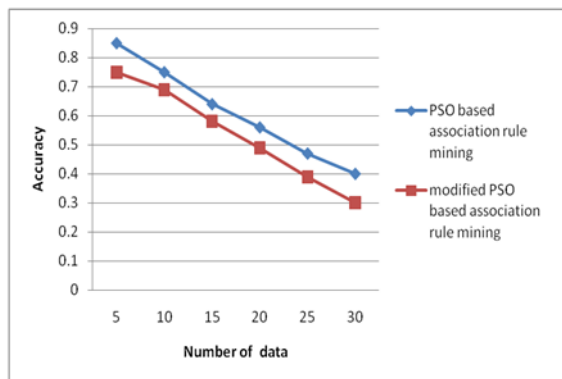


Fig 1: Comparison graph of Accuracy

This graph(Fig 1) shows the accuracy rate of existing and proposed system based on two parameters of accuracy and the number of datasets. From the graph we can see that,

when the number of number of datasets is advanced the accuracy also developed in proposed system but when the number of number of datasets is improved the accuracy is reduced somewhat in existing system than the proposed system. From this graph we can say that the accuracy of proposed system is increased which will be the best one.

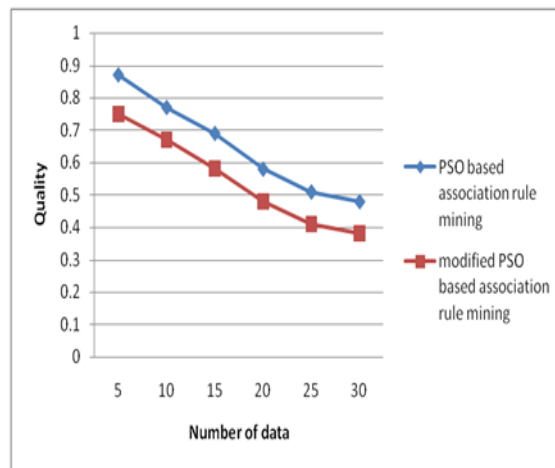


Fig 2: Comparison graph of Quality

This graph(Fig 2) shows the accuracy rate of existing and proposed system based on two parameters of quality and the number of datasets. From the graph we can see that, when the number of number of datasets is advanced the quality also developed in proposed system but when the number of number of datasets is improved the quality is reduced somewhat in existing system than the proposed system. From this graph we can say that the quality of proposed system is increased which will be the best one.

5. CONCLUSION

An important investigate that gets place in the area of data mining is the process of the extracting the exact required information based on the query. Thus the effective information can be retrieved based on the efficient association rules. By focusing on the problem of the generating the association rules, in this paper an effective approach of quantum particle swarm optimization is proposed in this paper. Here we have evaluated the proposed approach with the existing concept of association rule mining. From the experimental result shows that this approach provides an efficient association rule mining application for the information searching from the large databases. On applying this application of quantum particle swarm optimization approach to real time applications, could able to retrieve information effectively such as retrieving the transaction behavior, etc. This approach can be further enhanced by applying the other

association rule techniques that can outperform the proposed approach.

References

- [1]. J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, New York, 2000.
- [2]. Waseem Shahzad, Farrukh Aslam Khan, and Abdul Basit Siddiqui ,”Weighted Clustering using Comprehensive Learning Particle Swarm Optimization for Mobile Ad Hoc Networks”, *International Journal of Future Generation Communication and Networking*, Vol. 3, No. 1, March , 2010.
- [3]. S.S. Gun, Application of genetic algorithm and weighted itemset for association rule mining, Master Thesis, *Department of Industrial Engineering and Management*, Yuan-Chi University, 2002.
- [4]. M. Saggarr, A.K. Agrawal, A. Lad, Optimization of association rule mining using improved genetic algorithms, in: Proceeding of the *IEEE International Conference on Systems Man and Cybernetics*, vol. 4, 2004, pp. 3725–3729.
- [5]. C. Li, M. Yang, Association rule data mining in manufacturing information system based on genetic algorithms, in: Proceeding of the 3rd *International Conference on Computational Electromagnetics and Its Applications*, 2004, pp. 153–156.
- [6]. R.J. Kuo, C.W. Shih, Association rule mining through the ant colony system for National Health Insurance Research Database in Taiwan, *Computers and Mathematics with Applications* 54 (11–12) (2007) 1303–1318.
- [7]. R.J. Kuo, S.Y. Lin, C.W. Shih, Discovering association rules through ant colony system for medical database in Taiwan,” to appear, *International Journal of Expert Systems with Applications* 33 (November (3)) (2007).
- [8]. R.C. Eberhart, Y. Shi, Particle swarm optimization: developments, applications and resources, in: Proceedings of the *IEEE Congress on Evolutionary Computation*, 2001, pp. 81–86.
- [9]. S.C.M. Cohen, L.N.D. Castro, Data clustering with particle swarms, *IEEE Congress on Evolutionary Computations (2006)* 1792–1798.
- [10]. R.J. Kuo, M.J. Wang, T.W. Huang, Application of clustering analysis to reduce SMT setup time—a case study on an industrial PC manufacturer in Taiwan, in: Proceedings of *International Conference on Enterprise Information Systems*, Milan, Italy, 2008.
- [11]. R.J. Kuo, F.J. Lin, Application of particle swarm optimization to reduce SMT setup time for industrial PC manufacturer in Taiwan, *International Journal of Innovative Computing, Information, and Control*, in press.
- [12]. C.Y. Chen, F. Ye, Particle swarm optimization algorithm and its application to clustering analysis, in: Proceedings of the *IEEE International Conference on Networking, Sensing and Control*, Taipei, Taiwan, 2004, pp. 21–23.
- [13]. Hai Zhao, Chunhe Song, Dan Liu ,”A optimized particle filter based on PSO algorithm “,Inst. of Inf. & Technol., Northeastern Univ., Shenyang, China , Dec. 2009 pp 13-14.
- [14]. Hossain, M.A. Hashem, M.M.A.; Ali, M.M. , “ Quantum Evolutionary Algorithm based on Particle Swarm theory in multiobjective problems “,*Computer and Information Technology (ICCIT)*, 2010 13th International Conference Dec. 2010,pp 23-25 .
- [15]. Hota, A.R. An improved quantum-behaved Particle Swarm Optimization using fitness-weighted preferential recombination “,Dept. of Math., Indian Inst. of Technol., Kharagpur, India, Dec. 2010709 – 714, pp 15-17.
- [16]. J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the 1995 *IEEE International Conference on Neural Networks*. Part 4 (of 6) Perth, Aust, 1995, pp942–1948

Bharathi.T completed her Under graduation degree Bachelor of Science B.Sc (Computer Science) in Pioneer college of arts and science affiliated under Bharathiyar University in the duration of 2003-2006 ,She completed her Master of Science(Computer Science) and M.phil In Govt College of Arts and Science affiliated Under Bharathiyar University in the duration 2007-2009 and 2010.Currently she is doing her Ph.D under the Research area of Data Mining. She Successfully submitted one National level conference paper in her academics.

Dr P.Krishnakumari completed her Under graduation degree Bachelor of Science B.Sc (Computer Technology) in PSG College of Technology, Coimbatore in the duration of 1989-1991 ,She completed her Master of Science(Computer Science) in Avinashilingam University in the year of 1992-1994 and M.phil in Bharathiyar University in the year of 2010. She successfully completed her Ph.D in the field of Data Mining. She had a 12 years of working experience as a Lecturer in Ramakrishna Women’s Arts and Science College, Coimbatore.Currently, she is working as a Director of MCA in RVS Arts and Science College for past 1 year. She successfully submitted 10 National level Journal paper in her career.Her Research area is in the field of Data mining.

An Exploratory Study of Arabic Language Support in Software Project Management Tools

Arwa Al-Amoudi¹, Hailah AlMazrua¹, Hebah Al-Moaiqel¹, Noura AlOmar¹ and Sarah Al-Koblan¹

¹ Software Engineering Department, King Saud University
Riyadh, Saudi Arabia
{aalamoudi, halmazrou, halmoaiqel, nnalomar, salkoblan} @KSU.EDU.SA

Abstract

Due to the increasing demand for software in managing software development projects, the need of Arabic project management tools to be used by Arabic native speakers is emerging. Moreover, since there is hundreds of effective English project management tools, some of these tools were Arabized to fulfill the needs of Arabic users. However, they lack applying some basic rules in Arabic language. This study addresses the most important consideration elements that must be followed when Arabizing software. Also, a systematic review on three project management tools, which are TeamworkPM, ProjectLibre and Microsoft Project, is conducted based on the proposed guidelines. This study concludes with some recommendations and suggestions, which can be used to improve the software project management tools in particular, and the wide range of software applications that provide Arabic language interfaces in general.

Keywords: *Project management, Arabization, Tools, Localization, Arabic language.*

1. Introduction

The Arabic language is one of the oldest languages in the world. Currently, it is the fifth most used language worldwide. It is spoken by a significant percentage of the world's population; approximately more than 290 million speakers speak Arabic as their native language [1]. As a result, Arabic is becoming an important language on the Internet due to the growing number of Arabic speaking online users seeking Arabic content and applications.

Recent figures from the Internet World Statistics show that there are 90 million Internet users from the Arab world [2], 2.5 million of these users are from Saudi Arabia [3], which is known to have one of the strongest economies in Gulf Council Countries (GCC) and is considered one of the largest markets of the world's major projects [4], due to its richness in natural resources in general and oil in particular. For the purpose of using its revenues to boost local economy, Saudi Arabia is investing heavily in large-scale development projects [5]. Therefore, the need for project management (PM) expertise is highly needed along with a project management tool that support the growing demand for an Arabic PM application that servers the Arabic speaker.

Software developers often face several challenges when attempting to Arabize any application, such as character sets and Internet standards. As a result, several global leading companies like IBM and Microsoft are trying to increase the customers from the Arab world by adopting the Arabic language in their applications. The goal of this paper is to examine three Arabized project management tools which are TeamworkPM, ProjectLibre and Microsoft Project from four angles. These angles are language, layout, culture and word segmentation.

In this paper, we first start by proposing guidelines to be followed when Arabizing tools. Section 3 follows it by reviewing these guidelines in the chosen tools. Next, in Section 4, we provide a discussion of the current issues and the possible approaches to solve them. Finally in Section 5, we conclude the paper.

2. Proposed Guidelines

The most important aspects that have to be considered in Arabic tools and their absence affects the customers' interaction with the software negatively are language, Right-to-Left (RTL) direction, culture and word segmentation.

2.1 Language

In order to localize a tool to a specific language, the first step is to understand their language rules and apply them to the interface that is being designed to fulfill the needs of the target audience in the targeted context. This study focuses on the Arabic language since it's a way of communication between Arabic and non-Arabic Muslims [7]. The following sections illustrate Arabic language from three angles which are the grammatical structure, the way in which the words are correctly spelled and the process of translating the words from the source language to Arabic language without changing their meanings [6].

2.1.1 Grammar and Spelling

To customize software to Arabic language, the grammatical structure and spelling rules of it must be adopted for the purpose of making the sentences clear, logical and understandable. Therefore, detection function should be executed in all applications that support Arabic language. For this reason, avoiding grammar and spelling mistakes is essential to gain the trust of the users. This can be viewed from two angles which are ensuring that there are no mistakes in the translated interface as well as providing a grammar-checker and a spell-checker in the software. For the spelling checker, having a reference word list (dictionary), an error model and a language model in the Arabic supporting tools is important [9].

2.1.2 Translation

Albalooshi and Aljaroodi [7] define the process of translation in the context of interfaces as "Communicating the meaning of some data and information from the original language into another intended language". Translating software to another language isn't an easy as it sounds because it can affect the usability of its interface negatively or positively. Online users can find intended content in their native language via translation process. Thus, if the translation is done successfully, it can increase the number of users of the software worldwide. There are two ways to translate a system into Arabic language, either by using the translators that are available as online tools and do the task automatically or by hiring a specialist who is an Arabic native speaker. The latter one is better, as it can provide meanings that are more precise [7] since there are tools that were translated from a specific language to Arabic language inaccurately. Moreover, changing the language of the system's interface involves translating the words written on the icons, hyperlinks, dialog boxes and menus.

Terminologies related to a specific area must be translated into the Arabic language in an understandable and a scientific way. Additionally, the terminologies shouldn't be written in its source language such as keeping the English terminologies as they are. Furthermore, the absence of the abbreviations in Arabic language adds more work on the translators, as they need to specify the complete words and provides a meaningful sentence [8]. Moreover, there are many dialects in the Arabic language and they differ based on the country, thus the Arabization must not be affected by these dialects. Ideally, Modern Standard Arabic (MSA) which is derived from Holy Qur'an has to be adopted. Another Arabic language consideration is numerals and Hindi digit shapes. Numbers in Arabic language are read from left to right. There are two possibilities for writing

numbers in Arabic language: "Arabic digits"¹ (same as Latin numbers) and "Hindi shapes"², these two possibilities must be configured by the user of the system [10].

2.2 Layout

The interface design and the layout, which affects the look and feel of the project management tools, are influenced entirely by the writing direction. The interface layout should be aligned from right to left since the Arabic script is written in that direction. This means that the placement of the side menu will change to be on the right [8] [11] [12]. Consequently, the text alignment and the navigation should be from right to left. Additionally, if there is a logo it should be at the top-right whereas the scroll bar should be at the left side [13].

2.3 Culture

Culture is considered as one of the major aspects that affects any website, software or new technology. It has many factors like: environment, weather, religion, gender and economy. These factors are taken under consideration in producing or releasing a new version of such tool to a foreign country like European or Arabian countries due to increasing sells or spotlight reputation of the company. Hence, this section will highlight with more details on the vacations days, currency and Hijri calendar.

2.3.1 Working and Weekend Days [8] [14]

Official vacations days are categorized in two types, special ceremonies and official weekend days. For the former type, Islamic nations have two special ceremonies, which are "Eid-alfetar" and "Eid-aladha". The latter type is the official weekend days, which differ from one country to the other. For example, Yemen has a Thursday-Friday weekend, while Saudi Arabia, The United Arab Emirates and Qatar have a Friday-Saturday weekend, and finally both Lebanon and Tunisia have a Saturday-Sunday weekend. In a nutshell, you can infer that within the same region, each country has different occasions.

2.3.2 Currency [14]

American dollar is generally the most used currency worldwide. However, in most countries, project managers are requested to run their projects by using local currencies. For instance, almost all Saudi Arabian running projects are applying Saudi Riyal as the prime currency. For this reason,

¹ Arabic digits: (1, 2, 3 ...).

² Hindi shapes: (1• 2• 3 ...).

this paper looks up for how TeamworkPM, ProjectLibre and Microsoft Project tools are supporting the local currency.

2.3.3 Hijri Calendar [14]

The Islamic culture has the Islamic calendar (Hijri calendar) that is purely a lunar calendar. It contains twelve months that are based on the phases and stages of the moon. The Hijri calendar is mostly used in the Islamic countries and in Saudi Arabia's governmental organizations and institutions. Said countries use the Hijri calendar in their work such as adopting it in projects' deadlines, paying employees' salaries, specifying vacations' days and the dates of releasing new products or decisions. Therefore, software and tools are required to provide both Hijri and Gregorian calendars to make all processes easier and suitable for the Islamic regions and culture. Moreover, since the Hijri calendar is strictly based on lunar cycles as mentioned before, which means that each year is eleven-days short of a solar year. Hence the start of each month will be different from one year to the next. Therefore, adding a feature of editing the dates (start of the month) should be applied.

2.4 Word Segmentation

The cursive nature of Arabic writings is one of the difficulties facing the developers of text recognition systems. Moreover, there are applications that support Arabic language; however, the words are segmented letter by letter and this is a major error that needs to be eliminated in Arabic applications. As a rule, there are different graphical forms (glyph) for the same Arabic character and it varies based on the location of this character. Some characters can be located at the beginning of the word (first letter) and connected with the second character (left side), or not connected (table 1.a). Some characters can be located in the middle of the word and linked to another character on either side, or both sides, or not linked (table 1.b). Some characters can be placed at the end of the word and linked to another character on right side, or not linked (table 1.c) [10]. All of these glyph types of the Arabic character must be considered in the tools that support Arabic language.

Table 1: Different Arabic Characters' Positions in a Word

<i>a</i>	<i>b</i>	<i>c</i>
صالح	الوليد	محمد
أسماء	الوليد	عدي
-	الاميراطور	-

In addition, several research studies addressed the segmentation issue and proposed different methods to split the words into characters to allow the processing of the language such as editing the words. [15] Presents nine techniques of segmenting the Arabic word and the pros and cons for each one. Another approach was discussed in [16],

where its algorithm relies on extracting the segments from pre-specified points.

3. Project Management Tools' Review

Three project management tools have been selected to be evaluated according to the pre-mentioned consideration elements. Two of these tools are desktop-based ones, which are ProjectLibre and Microsoft (MS) Project and the remaining tool named TeamworkPM is a web-based one.

3.1 Language

3.1.1 Grammar and Spelling

Several spelling errors in the Arabic language have been noticed in the chosen tools. Firstly, TeamworkPM doesn't provide the spelling detection and correction features. For example, a basic spelling rule hasn't applied when we attempt to write this sentence as a title for a new project "موقع ألكتروني تجاري". In this example, there are two obvious mistakes which are the (أ) in the second word and the repeated letter (ج) in the last word. Furthermore, there are many spelling errors in its Arabic interface such as the word "ملاحظه" in the features tab of creating new project dialog box while the correct writing of this word is "ملاحظة". Secondly, ProjectLibre's interface doesn't include clear spelling mistakes; however, it doesn't detect and correct the typed spelling errors. Lastly, MS Project has a feature which allows the addition of new correctly spelled words to its dictionary and it detects/corrects the errors accordingly. Moreover, there are no apparent spelling mistakes in its interface. Turning to the grammar side of the Arabic language in these tools. All of the three mentioned programs don't support the detection and correction of the grammatical mistakes.

3.1.2 Translation

Four translation elements are addressed here which are the correct meaning of the translated words, English terminologies, abbreviations and Arabic numbers. The following table summarizes the translation errors in TeamworkPM that were noted by us (Wrongly translated words) along with its accurate meaning in English (Literal translation), the source English words (Intended English word) and the precise Arabic translation (The correct translation).

Table 2: TeamworkPM's Translation Mistakes


Tool Name	Wrongly Translated Word	Literal Translation	Intended English Word	The Correct Translation
TeamworkPM	منزل	House	Home	الصفحة الرئيسية
	بلدي تفاصيل	Description country	My details	معلوماتي الشخصية
	أنت تسير على هذا الحب	You are in love	You are going to love this	سوف يعجبك هذا البرنامج
	لون الموضوع	Subject color	Color theme	لون العرض
	الذين و عندما	They & then	Who & When	من ومتى

In contrary, the Arabic translation in ProjectLibre and MS Project is almost accurate and leads to the correct meaning. As for the project management's terminologies and abbreviations, it is noticeable that the terminologies are translated in an understandable way and reflect the actual meaning in all of the three tools. Furthermore, all of them have considered the transformation of English abbreviations into complete sentences particularly ProjectLibre tool in which it provides accurate Arabic sentences along with their English abbreviations. What is more, the Arabic digits appear appropriately in TeamworkPM, ProjectLibre and MS Project whereas the Hindi shapes are supported only by MS Project.

3.2 Layout

As depicted in table 3, 4 and 5, layout is reviewed in this section by considering five points which are text alignment, side menu placement, scroll bar position, logo location and navigation.

Table 3: Layout Evaluations (TeamworkPM)

TeamworkPM	
Text Alignment	 <p>The text alignment is from LTR while it should be from RTL.</p>





Menu Place	 <p>The menu is in the left while it should be at the right.</p>
Scroll Bar	<p>There is no scroll bar since it depends on the scroll bar of the browser.</p>
Logo	 <p>The logo is on the top left while it should be on the top right.</p>
Navigation	 <p>When you see the navigation you think that the opened page is : "جامعة الملك سعود" While it's : "المشروع الأول" The correct navigation is: جامعة الملك سعود << المشروع الأول</p>

Table 4: Layout Evaluations (ProjectLibre)

ProjectLibre	
Text Alignment	 <p>The text alignment is from LTR while it should be from RTL.</p>

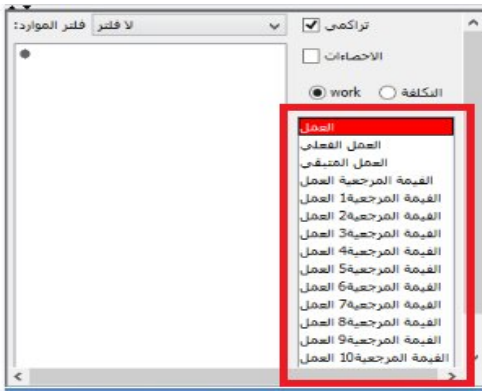
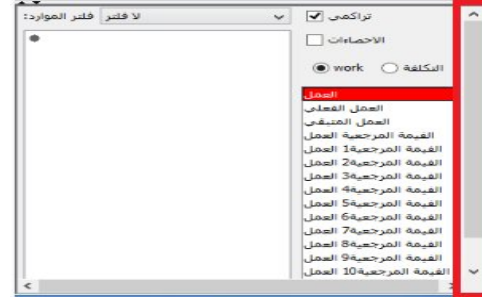
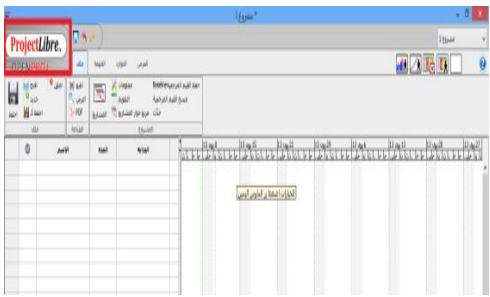


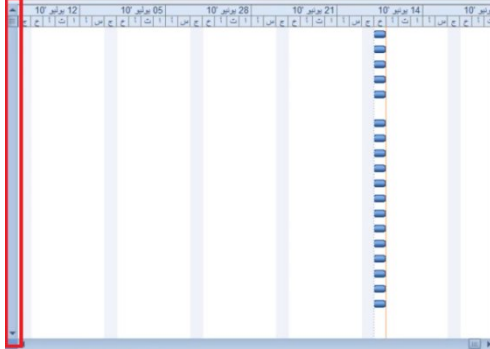

<p>Menu Place</p>	 <p>The menu is on the right.</p>
<p>Scroll Bar</p>	 <p>The scroll bar is on the right while it should be on the left.</p>
<p>Logo</p>	 <p>The logo is on the top left while it should be on the top right.</p>
<p>Navigation</p>	<p>Didn't support the navigation criteria.</p>

Table 5: Layout Evaluations (MS Project)

<p>MS project</p>	
<p>Text Alignment</p>	 <p>The text alignment is from RTL.</p>
<p>Menu Place</p>	 <p>The menu is on the right.</p>
<p>Scroll Bar</p>	 <p>The scroll bar is on the left.</p>
<p>Logo</p>	 <p>The logo is on the top right.</p>
<p>Navigation</p>	<p>Didn't support the navigation criteria.</p>

3.3 Culture

This section focuses on culture part by reviewing the working days, currency and Hijri calendar in the selected project management tools as summarized in table 6.

Table 6: Culture Evaluation

	Working Days	Hijri Calendar	Currency
Teamwork PM	No	N/A	Yes
Project Libre	Yes, but manual and only by user.	N/A	Yes
MS Project	Yes	Yes if the Computer's date is set as Hijri.	Yes

As illustrated in table 6, TeamworkPM has the ability to edit the currency; however it does not support changing the working days. Figure 1 shows that there are only two choices either to start the week at Mondays or Sundays and it don't support the Hijri calendar.

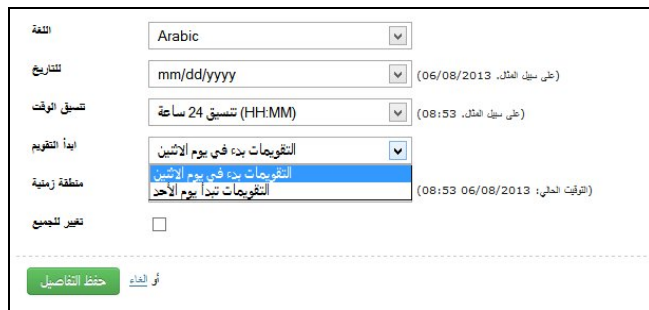


Figure 1: TeamworkPM Editing Calendar Start Date

As for ProjectLibre, it has the ability of changing the currency by default to be compatible with the language of the operating system. For example, S.R. is the currency adopted for Arabic operating system. Also it provides the option to edit the working days after performing some steps which are creating a new calendar and applying the changes on the days; however these changes should be made manually in each day and sort the days as shown in figure 2.



Figure 2: ProjectLibre's Editing the Working Days Feature

Lastly, MS Project has a full feature that supports the Arabian culture, it can be easily configured and the changes will be applied immediately. Moreover, extra features can be obtained by downloading the appropriate plug-ins.

3.4 Word Segmentation

The cursive nature of Arabic writings and the different glyphs of characters are supported in all of the three tools.

4. Recommendations

Arabic presents several important content localization challenges, due to the poor software support and an acute shortage of Arabic translators. The American Translators Association (ATA) lists only 200 English-to-Arabic translators in its online directory, 21 of these 200 are ATA certified translators; this is compared to 2181 English-to-Spanish translators [17]. Moreover, Arabic is one of the more expensive languages to translate as it lacks many of the developments and refinements needed for dealing with modern business and technology. In this sense the Arabic language might be called a technologically under-developed language.

Technology has yet to make significant influence on the Arabic culture as it has in many other parts of the world. Hence, Arabic lacks many linguistic developments needed to deal with more technologically developed languages. Consequently, localizing from a language that is rich with

technical vocabulary for dealing with technical subjects like English, into Arabic entails translation, cultural adaptation of content, and overcoming the linguistic barriers between technologically developed and under-developed languages [18].

While it is not easy to express some computing or technical terms in the Arabic language considering the limited technical vocabulary in Arabic as mentioned in the previous paragraph, a qualified linguist is highly needed in order to create custom Arabic terms that accurately express the exact meanings of the source language terms.

Also, the Arabic language is a meaningful, expressive language, where a text can be expressed in various alternate ways. Therefore, machine translation and untrusted and less reliable linguistic resources does not work well with this language. The Arabic culture is known to have a significant difference from the Western or Asian ones; professional localization services are needed to address the needs of users from this historically rich culture [19]. For example some terms have ambiguous meanings when translated to Arabic using machine translation. Table 7, shows that both the words calculate and compute have the same Arabic translation, although they mean two different things.

Table 7: Ambiguities in Machine Translation

<i>English Word</i>	<i>Arabic (Machine Translation)</i>	<i>English Word</i>	<i>Arabic Meaning</i>
Calculate	حساب	Calculator	آلة حاسبة
Compute	حساب	Computer	حاسوب

While 290 million Arabic-speaking people span over 20 countries, recent studies have indicated that the overall available Arabic content on the Internet is rather small; that's why initiatives like Taghreedat¹ and Arabic Web Days² are focusing on boosting Arabic content on the internet, also to support more Arabic content, AltCity³ and Global Voices⁴, a network of bloggers and translators, launched Translation Fest (T_Fest)⁵.

Collaborating with such initiatives like Taghreedat, which is a regional and international Arabic digital content community building initiative, can help the localization process, as this approach will ensure that the localized content is not machine translated, since the community will have to vote on all the translated terms along with

suggesting other translation to the term if the existing ones are found not suitable. The initiative aims to build an active Arabic digital content community that enrich the quality and quantity of Arabic content on both the web and application translation, through crowd-sourcing to increase Arab users' contribution through original content projects, and projects geared towards localization and Arabization [20].

5. Conclusion

The proliferation of the Arabic language in online content and in software packages supporting Arabic interfaces is increasing, as it is one of the most widely spoken languages in the world [1]. This importance makes the companies and developers focus on this language and provide their products in the mother tongue of more than 290 million people [2]. In addition to the previous factor, many Arabic nations adhere to using the Arabic language in their formal communications and transactions especially in public sectors. In considering project management tools to manage and control enormous deals [5], some of these tools have misunderstanding in the main Arabic language guidelines particularly in translation, grammar and spelling. Moreover, they need to support the Right-to-Left (RTL) layout in the interface, Hijri calendar, working days aligned with the currency as general points related to the culture. Since the Arabic morphology varies (i.e. letters appear differently according to their location within a word), the tools should take care of the words' segmentation.

The aim of this paper is to highlight key issues that are relevant to the design of software project management tools to ensure full support of Arabic interfaces and adaptation to local contexts. To reach this stage, this paper reviews three different types of project management tools, which are TeamworkPM, ProjectLibre and Microsoft Project based on main Arabic language guidelines that were proposed. All these tools have been used and observed to sum up with recommendations that can enhance tools and increase the productivity.

References

- [1] Vivek Kumar Singh. (2012). "Most Spoken Languages in the world." [Online]. Available: <http://www.listsworld.com/top-10-languages-most-spoken-worldwide/>. [Accessed: 01-Jun-2013].
- [2] "World Internet Users Statistics Usage and World Population Stats." [Online]. Available: <http://www.internetworldstats.com/stats.htm>. [Accessed: 01-Jun-2013].
- [3] Vincent, Peter, and Warf, Barney, "Multiple geographies of the Arab Internet," Area (2007), Vol. 39, No.1, pg. 83.[Accessed: 2013].

¹ <http://taghreedat.com/>

² <http://www.arabicwebdays.com/front/index.aspx>

³ <http://www.altcity.me/about/>

⁴ <http://globalvoicesonline.org/lingua/>

⁵ <https://www.facebook.com/events/188199947996388/?ref=2>

- [4] "نبذة عن الملتقى – ملتقى إدارة المشاريع الهندسية المتعثرة" [Online]. Available: <http://www.sspconferences.com/about.html>. [Accessed: 01-Jun-2013].
- [5] M. Baydoun. "Formulating programs of large scale development projects in Saudi Arabia". PM World Journal, vol.1, issue 2, date September 2012. [Accessed: 2013].
- [6] Alawneh, M.F.; Sembok, T.M., "Rule-Based and Example-Based Machine Translation from English to Arabic," 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), pp.343, 347, 27-29 Sept. 2011. [Accessed: 2013].
- [7] Albaloooshi, N.; Mohamed, N.; Al-Jaroodi, J., "The challenges of Arabic language use on the Internet," Internet Technology and Secured Transactions (ICITST), 2011 International Conference for, vol., no., pp.378,382, 11-14 Dec. 2011. [Accessed: 2013].
- [8] Al-Sedrani, A.; Al-Khalifa, H.S., "Design considerations for the localization of Arabic e-commerce websites," 2012 Seventh International Conference on Digital Information Management (ICDIM), pp.331,335, 22-24 Aug. 2012. [Accessed: 2013].
- [9] Mohammed Attia, PavelPecina, YounesSamih, KhaledShaalan, Josef van Genabith, "Improved Spelling Error Detection and Correction for Arabic," COLING 2012, Mumbai, 2012. [Accessed: 2013].
- [10] Franck Portaneri and Fethi Amar, "Arabization of Graphical User Interfaces," in International users interface. New York, John Wiley & Sons, Inc., 1996, pp. 127-150. [Accessed: 2013].
- [11] JOSEPH D. BECKER, "Arabic Word Processing," (1987), pp.600-610.[Accessed: 2013].
- [12] IyadKhaddam and Jean Vanderdonckt. "Flippable User Interfaces for Internationalization," (2011),pp. 223-228.[Accessed: 2013].
- [13] Mohammadi, AkheelaKhanum, Shameem Fatima, MousmiA.Chaurasia, "Arabic Interface Analysis Based on Cultural Markers," (2012),pp. 255-262. [Accessed: 2013].
- [14] Yeo, A., "Software internationalization and localization," Computer Human Interaction, Sixth Australian Conference, page(s): 348- 349, Year: 1996. [Accessed: 2013].
- [15] Zeki, A.M., "The Segmentation Problem in Arabic Character Recognition The State Of The Art," Information and Communication Technologies, 2005. ICICT 2005. First International Conference on, pp.11, 26, 27-28 Aug. 2005. [Accessed: 2013].
- [16] RahimaBentrcia and Ashraf Elnagar. 2008. "Handwriting segmentation of Arabic text," InProceedings of the Fifth IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA '08), Robert Sablatnig (Ed.). ACTA Press, Anaheim, CA, USA, 122-127. [Accessed: 2013].
- [17] "American Translators Association: Online Directories: Directory of Translation and Interpreting Services." [Online]. Available: <http://www.atanet.org/onlinedirectories/individuals.php>. [Accessed: 09-Jun-2013].
- [18] "Why Arabic Is the Most Difficult Language for Localization." [Online]. Available: <http://www.translationdirectory.com/article460.htm>. [Accessed: 09-Jun-2013].
- [19] "Arabic Translation and Localization Challenges." [Online]. Available: <http://blog.globalizationpartners.com/arabic-translation-and-localization-challenges.aspx>. [Accessed: 09-Jun-2013].
- [20] "About Taghreedat | مبادرات تغريدات" [Online]. Available: <http://taghreedat.com/aboutus/>. [Accessed: 09-Jun-2013].

Arwa Al-Amoudi is a teaching assistant at Software Engineering department in the College of Computer and Information Sciences (June 2012 – present). She has earned a Bachelor degree of Computer and Information Sciences in the field of Information Technology in January 2012 at King Saud University. Moreover, Mrs. Alamoudi has published two different papers in local and international conferences proceedings. Also, another publication will be published by Springer CCIS, July 2013. Her current areas of interest are Affective Computing, Human-Computer Interaction, Assistive Technology and Islamic and Arabic related topics.

Hailah AlMazrua is currently a Teaching Assistant in the department of Software Engineering, College of Computer and Information Sciences at King Saud University, Riyadh, Saudi Arabia. She received a bachelor degree of Information Technology, College of Computer and Information Sciences, from King Saud University in 2011. Ms AlMazrua has worked as Researcher/Developer at the Electronics, Communications and Photonics (ECP) Research center in King Abdulaziz City for Science and Technology (KACST) for almost a year (2011-2012). Her current areas of interest and research are Islamic application, Assistive Technology, Web Technologies, and human-computer interaction mainly the field of user experience, usability and Adaptive Technology. She has written and presented a few amount of papers at international conferences on her research areas.

Hebah Al-Moaiqel Received a bachelor degree of Information Technology, College of Computer and Information Sciences, King Saud University (KSU) in 2011. Ms.Al-Moaiqel worked as business/system analyst at SADAD Payment System Company for almost a year and currently as a teaching assistant in the department of Software Engineering, College of Computer and Information Sciences at KSU, Riyadh, Saudi Arabia. Her main interests are improving software's reliability, increasing performance and learning different system methodologies. Ms.Al-Moaiqel also has a growing interest in human-computer interaction especially in the field of user experiences. She has some publications in these domains were published recently in conferences and magazines.

Noura AlOmar has a bachelor degree from the College of Computer and Information Sciences in the field of: Information Technology, King Saud University (KSU), in 2011. Noura's current job is a teaching assistant in the department of Software Engineering, College of Computer and Information Sciences at King Saud University, Riyadh, Saudi Arabia. Ms. AlOmar has few international academic publications which include conference papers and journal papers. In addition, she has published few articles in local newspapers. Her research interests include recommendation systems, community-oriented systems, Web technologies and Bio-Informatics. Furthermore, her research areas in human-computer interaction cover the user-experience testing methods and techniques, affective computing and exploring users' behaviors in online social networks.

Sarah Al-Koblan has a Bachelor degree of Computer and Information Sciences in the field of Information in January 2012 at King Saud University. She is working now as a teaching assistant at the department of Software Engineering in the College of Computer and Information Sciences. Ms. Al-Koblan current research interests are graphic design, information security and artificial intelligence. Additionally, she has published few papers in local and international conferences.

Web-Marketing in Social Networks Using MDA Approach

Lamlili El Mazoui Nadori yasser¹, Mohammed Erramdani², Ibtissam Arrassen³, Esbai Redouane⁴, Mimoun Moussaoui⁵

^{1,5}MATSI Laboratory, Mohammed First University, EST
Oujda 60000, Morocco

²Department of Management, Mohammed First University, EST
Oujda 60000, Morocco

³Department of Mathematics and Computer Sciences, Mohammed First University
Oujda 60000, Morocco

⁴ENCG, Mohammed First University,
Oujda 60000, Morocco

Abstract

the evolution of the Web-marketing meaning the passage of the Web-marketing 2.0 to 3.0, baptized " Semantic Web-marketing ", the main idea being to make easier to the Internet user to use a huge and disrupting stream of information and make it organized and easily accessible.

As so, we have chosen to use the N-tiers applications and the MDA (Model Driving Architecture) transformation. In this paper we are going to present a model-driven approach to the development of N-tiers web applications based on the UML class diagram. The transformation language is the MOF 2.0 QVT (Meta-Object Facility 2.0 Query-View-Transformation) standard which defines the meta-model for the development of model transformation. The transformation rules defined in this paper can generate, from the class diagram, an XML file containing the layers of N-tiers web application respecting a MVC2 (Model-View-Controller), DI (Dependency Injection) and DAO (Data Access Object) patterns. This file can be used to generate the end-to-end necessary code of a web application

Keywords: *Web-Marketing, Transformation by modeling, N-tiers architecture, Transformation rules, MOF 2.0 QVT, Meta-model.*

1. Introduction

In the years following the arrival of internet, many organizations begun to consider MDA (Model-Driven Architecture) as an approach to design and implement enterprise applications. As result, many new trends have appeared under the frameworks, changing the development of classical web applications.

These changes are present in MDA, and help transform a CIM (Computation Independent Model) into a PIM (Platform Independent Model) or to obtain a PSM (Platform Specific Model) from a PIM.

N-tiers (multi-tier) architecture provides a model for developers to create a flexible and reusable application. By breaking up an application into tiers, developers have to modify or add a

specific layer only, rather than rewriting the entire application all over again. There should be a presentation tier, a business or data access tier, and a data tier.

In this work we are going to transform an UML Model concerning an advertisement in a social network to generate a source code.

In a recent work [21], the authors have developed a source and a target meta-models. The first was a PIM meta-model specific to class diagrams. The second was a PSM meta-model for MVC2 (Model-View-Controller) web applications (particularly Struts), then they have elaborated a transformation rules using the approach by programming. The purpose of our contribution is to produce and generate an N-tiers PSM model, implementing MVC2, DI (Dependency Injection) and DAO (Data Access Object) patterns, from the class diagram. In this case, we elaborate a number of transformation rules using the approach by modeling and MOF 2.0 QVT, as transformation language, to permit the generation of an XML file that can be used to produce the required code of the target application. The advantage of this approach is the bidirectional execution of transformation rules.

This paper is organized as follows: We begin in the first section with an introduction. Related works are presented in the second section. The third section permits to develop MDA as architecture. The fourth section presents the N-tiers architecture, the MVC2, DI and DAO patterns and its implementation as frameworks. The approach by modeling and the transformation language MOF 2.0 QVT are the subject of the fifth section. In the sixth section, we have elaborated the UML and N-tiers meta-models. The transformation rules of UML source model to the N-tiers target model, the transformation algorithm and the results of this transformation are presented in the seventh

section. The final section concludes this paper and presents some perspectives.

2. Related work

Many researches on MDA and generation of code have been conducted in recent years. The most relevant are [7][8][4][10][11][13][16][17][18][20][21][22][26][27] [34].

The authors of the work [18], show how to generate JSPs and JavaBeans using the UWE [17], and the ATL transformation language [16]. Among future works cited, the authors considered the integration of AJAX into the engineering process of UWE.

Bezivin et al. [8] propose to use ATL transformation to transform PIMs defined by Enterprise Distributed Object Computing into PSMs for different web services platforms.

Billing et al. [7] define PIM to PSM transformation in the context of EJB by using QVT.

The authors of the work [27] show that the MDA can be considered as a software industrialization pattern (or a software factory). They propose an MDA Engine based on a real case study in an IT services company. It is a proposal for a framework to create custom MDA tools, based on XMI, XSLT and the Visitor Pattern.

The work [10] has proposed a model-driven development approach for E-Learning platform. The authors established the domain model (CIM) through the analysis of business logic, and then they established robustness diagram of the system after the robustness analysis. Then, they stratified on the PIM under the J2EE framework, and proposed the method of transformation from PIM to PSM layer by layer.

The objective of the work [34] is to introduce a new framework for the design of secure Data Warehouses based on MDA and QVT, which covers all the design phases (conceptual, logical and physical) and specifies security measures in all of them.

One approach which has gained much attention in the web-based MDA community is the AndromDA MDA generator [4]. This framework provides a PIM schemes to model and integrate a wide variety of scenarios and comes with a set of plugins, called cartridges.

Two other works follow the same logic and have been the subject of two articles [11] [13]. A metamodel of AJAX has been defined using the AndromDA tool. The generation of AJAX code was made and illustrated by an application that manages CRUD operations of person. The meta-model is very important and we can join it to our meta-models for modeling AJAX user interfaces.

The objective of the work of Nasir et al. [26] is to generate the code of a DotNet application “Student Appointment Management System”. The method used is WebML. The code is generated, applying the MDA approach.

Recently, the work [22] was conducted to model Web MVC2 generation using the ATL transformation language.

This paper aims to rethink and to complete the work presented in the articles [20][21], by applying the standard MOF 2.0 QVT to develop the transformation rules aiming at generating the N-

tiers web according to our target model. Actually, it is the only work for reaching this goal.

3. Model Driven Architecture (MDA)

In November 2000, OMG, a consortium of over 1 000 companies, initiated the MDA approach. The key principle of MDA is the use of models at different phases of application development. Specifically, MDA advocates the development of requirements models (CIM), analysis and design (PIM) and (PSM) code [6].

The MDA architecture is divided into four layers. In the first layer, we find the standard UML (Unified Modelling Language), MOF (Meta-Object Facility) and CWM (Common Warehouse Meta-model). In the second layer, we find a standard XMI (XML Metadata Interchange), which enables the dialogue between middlewares (Java, CORBA, .NET and web services). The third layer contains the services that manage events, security, directories and transactions. The last layer provides frameworks which are adaptable to different types of applications namely Finance, Telecommunications, Transport, medicine, E-commerce and Manufacture, etc.).

The major objective of MDA is to develop sustainable models, those models are independent from the technical details of platforms implementation (J2EE, DotNet, PHP or other), in order to enable the automatic generation of all codes and applications leading to a significant gain in productivity. MDA includes the definition of several standards, including UML [39], MOF [23] and XMI [40].

4. N-tiers architecture

N-tiers application architecture provides a model for developers to create a flexible and reusable application and provides some advantages that are vital to the business continuity of the enterprise. Typical features of a real life N-tiers may include the Security, Availability, Scalability, Manageability, Easy Maintenance and Data Abstraction. To most people, an N-tiers application is anything that is divided into discrete logical parts. The most common choice is a three-part breakdown presentation, business logic, and data access although other possibilities exist.

In this paper, we are using the following layers:

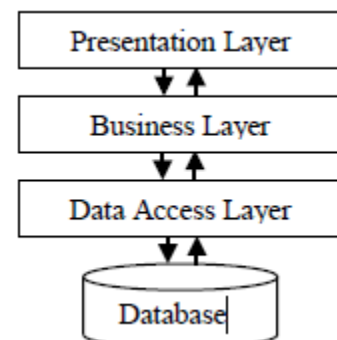


Fig.1 N-tiers Layers

Each Layer can be developed independently of the other provided that it adheres to the standards and communicates with the other layers.

4.1 The presentation Layer with MVC2 pattern

The presentation layer of most applications is often critical to the application's success. After all, the presentation layer represents the interface between the user and the application back-end.

Along time ago, web applications were very simple and the technology that was used to develop them was Common Gateway Interface (CGI). As applications became more complex, the defects and limits of this technology have emerged. Slowness and considerable consumption of memory. Therefore, the J2EE platform applies the architecture MVC2 [3]. In this paradigm, the model represents the information system consisting of javaBeans. The view represents the HTML pages returned to the user, and consists of JavaServerPage (JSP). The Controller is the glue between the two and it is composed of servlets. In short, during the early 80's with smalltalk, MVC was widespread in the field of object development. Many frameworks that implements MVC2 pattern have emerged, among them: Struts [1], PureMVC [29], Gwittir [14], SpringMVC [35], Zend [38], ASP.NET MVC2 [5]. Struts remains the most mature solution that has earned the trust of most developers, that is why we have taken it into account in our source meta-model.

4.2 The Business layer with Data Transfer Object and Dependency Injection patterns

Business logic layer is the Layer of abstraction between the presentation layer and persistence layer to avoid a strong coupling between these two layers and hide the complexity of the implementation of business processing to presentation layer. All business treatments will be implemented by this layer. The implementation of this layer is produced by the DTO pattern to render the result of running the service and the DI pattern to ensure a decoupling between objects.

In an article written in early 2004, Martin Fowler asked what aspect of control is being inverted. He concluded that it is the acquisition of dependent objects that is being inverted. Based on that revelation, he coined a better name for inversion of control: dependency injection [19].

In other words, Dependency Injection is a worthwhile concept used within applications that we develop. Not only can it reduce coupling between components, but it also saves us from writing boilerplate factory creation code over and over again. Many frameworks that implements DI pattern have emerged, among them: Spring [35], Symfony dependency injection [37], Spring.NET [36], EJB [30], PicoContainer [31]. (We have used some Spring classes in our source meta-model).

Recently, with the development of mapping o/r tools, it becomes easier to transfer a model object on the client layer (UI), and the distribution of the service layer, other advantage of the DTOs, is

privileged in N-tiers modern architectures, that is why we have taken it into account in our work.

4.3 The persistence Layer with DAO pattern

This layer is the entry point to the database. All operations required to create, retrieve, update, and delete data in the database are implemented in the components of this layer.

The Data Access Object (DAO) pattern is now a widely accepted mechanism to abstract the details of persistence in an application. In practice, it is not always easy to make our DAO's fully hidden in the underlying persistence layer.

The advantage of this abstraction is that we can change the persistence mechanism without affecting the logic domain. All we need to change is the DAO layer which, if designed properly, is a lot easier to do than changing the entire logic domain. In fact we might be able to cleanly swap in a new data access layer for our new database or alternate persistence mechanism. Many frameworks that implements DAO pattern have emerged, among them: SpringDao [35], JPA [32], Hibernate [15], iBatis [2], NHibernate [28], EJB [30]. We have used Hibernate in our work because it is the most used solution within the java community.

5. The transformation of MDA models

MDA establishes the links of traceability between the CIM, PIM and PSM models due to the execution of the models' transformations.

The models' transformations recommended by MDA are essentially the CIM transformations to PIM and PIM transformations to PSM. In our work, we perform the second transformation PIM to PSM devoted to N-tires web applications.

5.1 Approach by modeling

Currently the transformations of models can be written according to three approaches: Approach by Programming, approach by Template and approach by Modeling. The approach by Modeling is the one used in the present paper. It consists of applying concepts from model engineering to models' transformations themselves. The objective is modeling a transformation, to reach perennial and productive transformation models, and to express their independence towards the platforms of execution. Consequently, OMG elaborated a standard transformation language called MOF 2.0 QVT [24]. The advantage of the approach by modeling is the bidirectional execution of transformation rules. This aspect is useful for the synchronization, the consistency and the models reverse engineering [9].

Figure 2 illustrates the approach by modeling. Models' transformation is defined as a model structured according to MOF2.0 QVT meta-model. The MOF 2.0 QVT meta-model express some structural correspondence rules between the source and target meta-model of a transformation. This model is a perennial and productive model that is necessary to transform in order to execute the transformation on an execution platform.

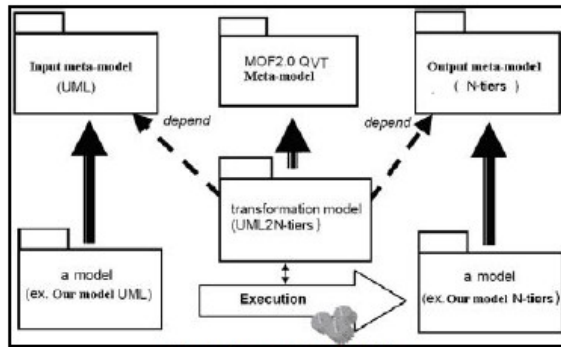


Fig. 2 Approach by Modeling

5.2 MOF 2.0 QVT

Transformations' models are at the heart of MDA, a standard known as MOF 2.0 QVT being established to model these changes. This standard defines the meta-model for the development of transformation model. The QVT standard has a hybrid character (declarative / imperative) in the sense that it is composed of three different transformation languages.

The imperative style languages are better suited for complex transformations including a significant algorithm component. Compared to the declarative style, they have the advantage of optional case management in a transformation. For this reason, we chose to use an imperative style language in this paper.

The imperative QVT component is supported by Operational Mappings language. The vision requires an explicit imperative navigation as well as an explicit creation of target model elements. The Operational Mappings language extends the two declarative languages of QVT, adding imperative constructs (sequence, selection, repetition, etc.) and constructs in OCL edge effect.

This work uses the QVT-Operational mappings language implemented by SmartQVT [33]. SmartQVT is the first open source implementation of the QVT-Operational language. The tool comes as an Eclipse plugin under EPL license running on top of EMF framework. This tool is developed by France Telecom R & D project and partially funded by the European IST Model Ware.

SmartQVT is composed of 3 components:

- **QVT Editor:** helps end users to write QVT specifications.
- **QVT Parser:** converts the QVT concrete textual syntax into its corresponding representation in terms of the QVT meta-model.
- **QVT Compiler:** produces, from a QVT model, a Java program on top of EMF generated APIs for executing the transformation. The format of the input is a QVT specification provided in XMI 2.0 in conformance with the QVT meta-model.

6. UML and N-tiers architecture meta-models

To develop the transformation algorithm between source and target model, we present in this section, the various meta-classes

forming the meta-model UML source and the meta-model N-tiers target.

6.1 Meta-model UML source

The source meta-model structures a simplified UML model based on packages containing data types and classes. Those classes contain typed properties and they are characterized by multiplicities (upper and lower). The classes are composed of operations with typed parameters. Figure 3 illustrates the source meta-model.

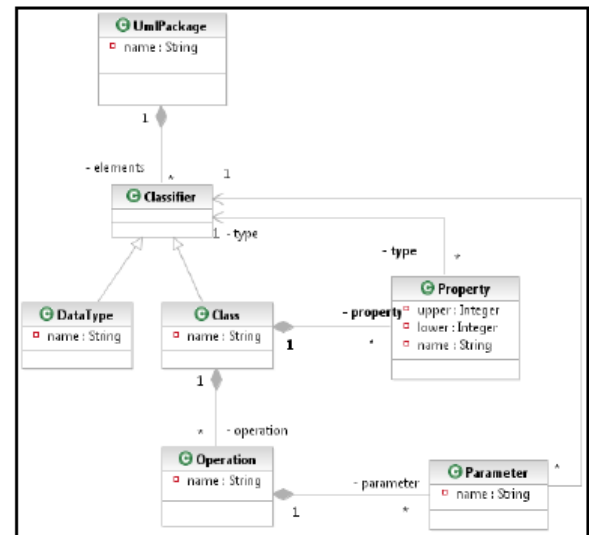


Fig. 3 Simplified UML meta-model

- **UmlPackage:** is the concept of UML package. This meta-class is connected to the meta-class **Classifier**.
- **Classifier:** This is an abstract meta-class representing both the concept of UML class and the concept of data type.
- **Class:** is the concept of UML class.
- **DataType:** represents UML data type.
- **Operation:** is used to express the concept of operations of a UML class.
- **Parameter:** expresses the concept of parameters of an operation. These are of two types, Class or DataType. It explains the link between Parameter meta-class and Classifier meta-class.

- **Property:** expresses the concept of properties of a UML class. These properties are represented by the multiplicity and meta-attributes upper and lower.

6.2 Meta-model N-tiers target

Our target meta-model is composed of three essential part. Figure 4 illustrates the first part of the target meta-model. This meta-model represents a simplified version of the DAO pattern. It presents the different meta-classes to express the concept of DAO contained in the DaoPackage:

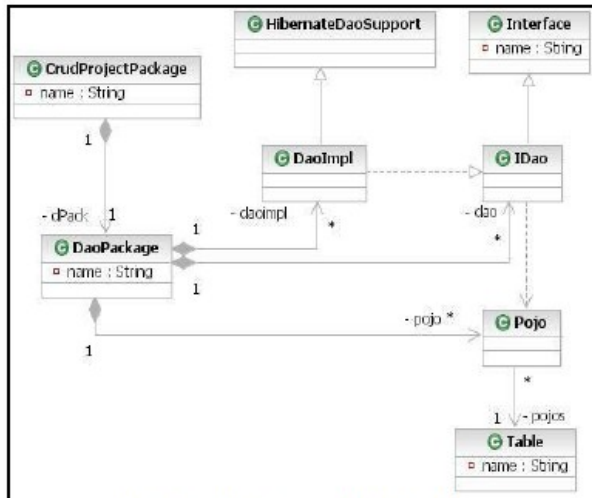


Fig. 4 Simplified meta-model of DaoPackage

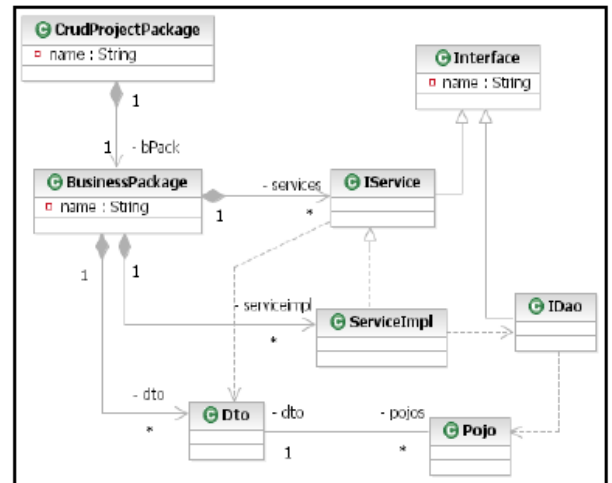


Fig. 5 Simplified meta-model of a BusinessPackage

- **CrudProjectPackage:** represents the project package. This meta-class is connected to the meta-class DaoPackage, BusinessPackage and UIPackage.
- **DaoPackage:** represents package which contains the different meta-classes to express the concept of DAO.
- **HibernateDaoSupport:** expresses the concept of generic class for DAOs, defining template methods for DAO initialization.
- **Interface:** is the concept of UML interface.
- **IDao:** represents the concept of Dao interface containing the methods definition to create, retrieve, update, and delete data in the database.
- **DaoImpl:** expresses the concept of Dao implementation, all methods to create, retrieve, update, and delete data in the database are implemented in this meta-class.
- **Pojo:** represents the concept of pojo. The latter extends the meta-class *Class*. The pojoes represents objects in the area of application. These objects communicate with the tables of relational database, which explains the meta-association with meta-class *Table*.
- **Table:** is the concept of table in the relational databases. It contains a meta-attribute *name* which represents the table name in the database. The meta-class is connected by a meta-association to the meta-class *Column*. Figure 5 illustrates the second part of target meta-model. This meta-model is the business model of the application to be processed. In our case, we opted for components such as DTO and DI pattern. Here, we present the different meta-classes to express the concept of DI contained in the Business Package.

This meta-model structures the models representing the business logic of the target application. This logic is essentially made up of DTO components.

- **BusinessPackage:** represents the package which contains the different meta-classes to express the concept of the business logic of target application.
- **Interface:** (already seen at the DaoPackage meta-model)
- **IService:** represents the concept of service interface containing the methods definition.
- **ServiceImpl:** expresses the concept of service implementation containing the methods representing in IDao meta-class and declared in IService meta-class.
- **IDao:** (already seen at the DaoPackage meta-model)
- **Dto:** represents the concept of business object that needs to be transferred across a process or network boundary. These objects contain all/some attributes of the pojoes, which explains the meta-association with meta-class *Pojo*.
- **Pojo:** (already seen at the DaoPackage meta-model)

Figure 6 illustrates the third part of the target meta-model. This meta-model represents a concept of MVC2 implementation in the user interface.

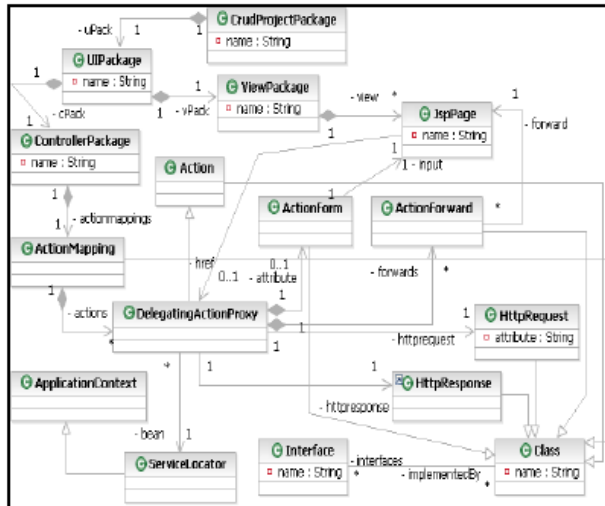


Fig. 6 Simplified meta-model of UIPackage

- **UIPackage:** represents the different meta-classes to express the concept of MVC2. This meta-class is connected to the meta-class *ViewPackage* and *ControllerPackage* which represents respectively View and Controller package.
- **ActionMapping:** Represents the concept of ActionMapping classes. An ActionMapping class contains information to deploy of a class Action. This explains the connection with the meta-class *Action*.
- **Action:** is the concept of action. Class Action contains its own processing of the application; hence it should be linked to the various beans.
- **DelegatingActionProxy:** represents the concept of Proxy for a Spring-managed Struts WebApplicationContext. The proxy is defined in the Struts-config file, specifying this class as the action class. This class will delegate to a Struts Action that is defined in Action bean in the ContextLoaderPlugIn context.
- **ActionForm:** represents the concept of ActionForm classes. An ActionForm represents a form containing the parameters of the request from the view (ViewPackage). This object is used by Action Class (This is particularly one of the four parameters of the operation execute()), which explains the link with the metaclass Action.
- **JspPage:** represents a Jsp page. An action class may be called from a hyperlink in a Jsp. This explains the link between the Jsp page and Action class. The link between ActionForward and Jsp page is trivial. ActionForm is linked to Jsp page because it contains the information that would be transmitted in the request and then filled in the actionForm. The link between Jsp page and HttpRequest expresses the fact that the Jsp page can use the information contained in an HttpRequest object.
- **HttpRequest:** is the concept of *HttpServletRequest* classes.
- **HttpResponse:** represents the concept of *HttpServletResponse* classes.

- **ApplicationContext:** represents the concept of Central interface to provide configuration for an application, An ApplicationContext provides a Bean factory methods for accessing application components and Inheritance from a parent context. Definitions in a descendant context will always take priority. This means, for example, that a single parent context can be used by an entire web application, while each servlet has its own child context that is independent of other servlets.
- **ServiceLocator:** expresses the concept of Service lookup and creation involves complex interfaces and network operations.

This meta-model structures the models representing the view application. In this model, the Servlet invokes the execute() method on the instance of the action class. This method completes its processing and then calls the mapping.findforward() method with a return to a specified Jsp page.

Annexe 1 shows the global view of our meta-model target.

7. Transformation process from UML to N-tiers implementation

CRUD operations (Create, Remove, Update, and Display) are most commonly implemented in all systems. That is why we have taken into account in our transformation rules these types of transactions.

We first developed EMOF models corresponding to our source and target meta-models, and then we implemented the algorithm using the transformation language QVT Operational Mappings. To validate our transformation rules, we conducted several tests. For example, we considered the class diagram (see Figure 7). After applying the transformation on the UML model, composed by the classes User and advertisement, we generated the target model (see Figure 9).



Fig. 7 UML instance model

7.1 Transformation rules

By source model, we mean model containing the various classes of our business model. The elements of this model are primarily classes.

Main algorithm:

```
input umlModel:UmlPackage
output crudModel:CrudProjectPackage
begin
create CrudProjectPackage crud
create DaoPackage daoPackage
```

```
for each e ∈ source model

x = transformationRuleOnePojo(e)
link x to dp
x = transformationRuleOneIDao(e)
link x to dp
x = transformationRuleOneDaoImpl(e)
link x to dp
end for
create BusinessPackage bp;

for each pojo ∈ target model

x = transformationRuleTwoDto(pojo)
link x to bp
end for

for each e ∈ source model

x = transformationRuleTwoIService(e)
link x to bp
x = transformationRuleTwoServiceImpl(e)
link x to bp
end for
create UIPackage uip;
create ViewPackage vp
vp = transformationRuleThreeView(e)
create ControllerPackage cp
cp = transformationRuleThreeController(e)
link vp to uip
link cp to uip
link dp to crud
link bp to crud
link uip to crud
return crud
end
function
transformationRuleOnePojo(e:Class):Pojo
begin
create Pojo pj
pj.name = e.name
pj.attributes = e.properties
return pj
end
function
transformationRuleOneIDao(e:Class):IDao
begin
create IDao idao
idao.name = 'I'+e.name+ 'Dao'
idao.methods = declaration of e.methods
return idao
end
function
transformationRuleOneDaoImpl(e:Class):DaoImpl
begin
create DaoImpl daoImpl
daoImpl.name = e.name+ 'DaoImpl'

for each e1 ∈ DaoPackage

if e1.name = 'I'+e.name+ 'Dao'
put e1 in interfaces
```

```
end if
end for
link interfaces to daoImpl
return daoImpl
end
function
transformationRuleTwoDto(p:pojo):Dto
begin
create Dto dto
dto.name = p.name
dto.attributes = p.attributes
return dto
end
function
transformationRuleTwoIService(e:Class):IService
begin
create IService iservice
iservice.name = 'I'+e.name+ 'Service'
iservice.methods = declaration of e.methods
return iservice
end
function
transformationRuleTwoServiceImpl(e:Class):Service Impl
begin
create ServiceImpl serviceImpl
serviceImpl.name = e.name+ 'ServiceImpl'

for each e1 ∈ BusinessPackage

if e1.name = 'I'+e.name+ 'Service'
put e1 in interfaces
end if
end for
link interfaces to ServiceImpl
return ServiceImpl
end
function
transformationRuleThreeView(e:Class):ViewPackage
begin
create ViewPackage vp

for each e ∈ source model

if e.methods.name ≠ 'remove'
create JspPage page
link page to vp
end if
end for
return vp
end
function
transformationRuleThreeController(e:Class):ControllerPackage
begin
create ControllerPackage cp
create ActionMapping am
for each page viewPackage
link page to actionForward
create actionForm
create Action action
create ActionForward actionForward
actionForm.input=page
actionForm.attribute=action
link page to actionForward
link actionForward to action
put action in am
end for
link am to cp
return cp
```

end

Figure 8 represents the first part of the code of the transformation of UML model source to N-tiers target model.

```
1 transform UML2CRUD(in umlModel:UML, out crudModel:NTIERS);
2
3 main() {
4   umlModel.objects().[UmlPackage]->map UmlPackage2CrudProjectPackage();
5 }
6 mapping UmlPackage:UmlPackage2CrudProjectPackage ( | : CrudProjectPackage |
7   name:= 'crud'+self.name;
8   dPack:= object DaoPackage (
9     name:= 'daoPackage';
10    pojo:= umlModel.objects().[Class]->map class2Pojo();
11    dao:= umlModel.objects().[Class]->map class2IDao();
12    daoimpl:= umlModel.objects().[Class]->map class2DaoImpl();
13  );
14
15  bPack:= object BusinessPackage (
16    name:= 'businessPackage';
17    dao:= crudModel.objects().[Pojo]->map pojo2DTO();
18    services:= umlModel.objects().[Class]->map class2Service();
19    serviceimpl:= umlModel.objects().[Class]->map class2ServiceImpl();
20  );
21
22  uPack:= object UIPackage (
23    name:= 'presentationPackage';
24    vPack := object ViewPackage (
25      name := 'viewPackage';
26      view := umlModel.objects().[Class]->map class2View();
27    );
28    cPack := object ControllerPackage (
29      name := 'controllerPackage';
30      actionmappings:= object ActionMapping (
31        name:= 'actionMappings';
32        actions := umlModel.objects().[Class]->map class2Action();
33      );
34    );
35  );
36 }
```

Fig. 8 A transformation code UML2CRUD

The transformation uses, in entry, a model of the UML type named umlModel, and in output a model of the N-tiers named crudModel.

The entry point of the transformation is the method 'main'. This method makes the correspondence between all the elements of the UmlPackage type of the input model and the element of the CrudProjectPackage type of the output model. The objective of the second part of this code is to transform a UML package into N-tiers package, by creating the elements of type package 'Dao', 'Business' and 'Presentation. It is a question of transforming each class of package UML to Jsp page and Action in the View package, to DTO, IService and ServiceImpl in the Business package, and to Pojo, IDao and DaoImpl in the Dao package, without forgetting to give names to the different packages.

7.2 Result:

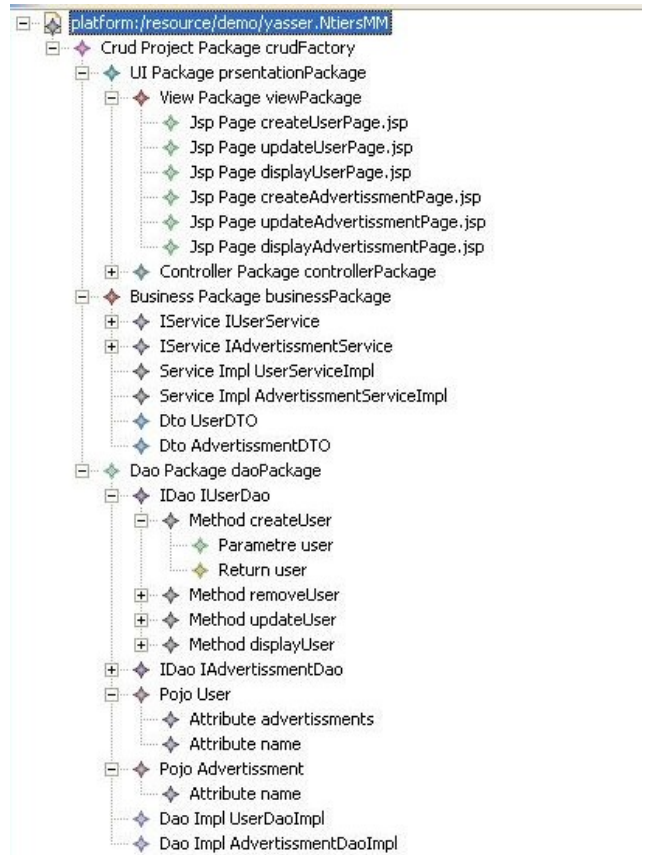


Fig. 9 Generated PSM N-tiers Web model

The first element in the generated PSM model is UIPackage which includes viewPackage that contains the JSPs, namely DisplayUserPage.jsp, AdvertisementPage.jsp, CreateUserPage.jsp, CreateAdvertisementPage.jsp, UpdateUserPage.jsp, and UpdateAdvertisementPage.jsp. Since the operation of the removal requires any form, we'll go to the controllerPackage element, which contains a single element ActionMapping. The latter contains eighteen delegating action proxy whose names are respectively DisplayXAction, CreateXAction, UpdateXAction, RemoveXAction, CreateXEndAction, UpdateXEnd-Action, where X should be replaced by User, and Advertisement. Operations for creation and update, add forms to enter new values. For this reason, we add CreateXEndAction and UpdateXEndAction.

The second element in the generated PSM model is businessPackage which includes three services' interfaces, three services' implementations and three Dtos' objects correspond to the two objects 'User' and 'Advertisement'. The last element in

the generated PSM model is DaoPackage which contains three Pojos' objects that contains their attributes, three Daos' interfaces that contains methods with their parameters and their implementations.

8. Conclusion

In this paper, we applied the MDA approach to generate the N-tiers web application based on UML class diagram to generate a skeleton of a social network and create appropriate advertisements to the users in function of their profiles.

This involves developing all meta-classes needed to be able to generate an N-tiers application respecting a MVC2, DI and DAO patterns, then we applied the approach by modeling and used the MOF 2.0 QVT standard as a transformation language. The transformation rules defined allow browsing the source model instance class diagram, and generating, through these rules, an XML file containing layers of N-tiers architecture according to our target model. This file can be used to produce the necessary code of the target application. The algorithm of transformation manages all CRUD operations. Moreover, it can be re-used with any kind of methods represented in the UML class diagram. In the future, this work should be extended to allow the generation of other components of Web application besides the configuration files. For instance, we will be able to provide part of user interface. Afterward we can consider integrating other execution platforms like PHP and DotNET.

References

- [1] Apache Software Foundation: The Apache Struts Web Application Software Framework (<http://struts.apache.org>).
- [2] Apache Software Foundation: The Apache iBatis Framework (<http://ibatis.apache.org/>).
- [3] Alur, D., Crupi, J., Malks, D., Core J2EE Patterns: Best Practices and Design Strategies (Prentice Hall, 2003).
- [4] AndromDA. <http://www.andromda.org/>.
- [5] ASP.NET MVC site <http://www.asp.net/mvc/>
- [6] Blanc, X., MDA en action : Ingénierie logicielle guidée par les modèles (Eyrolles, 2005).
- [7] Bezivin, J., Busse, S., Leicher, A., Suss, J.G, Platform Independent Model Transformation Based on TRIPLE. In Middleware'04: Proceedings of the 5th ACM/IFIP/USENIX International Conference on Middleware, pages 493- 511,2004.
- [8] Bezivin, J., Hammoudi, S., Lopes, D., Jouault, F., Applying MDA approach for web service platform. In EDOC'04 proceedings of the 8th IEEE International Enterprise Distributed Object Computing Conference, pages 58-70, 2004.
- [9] Czarnecki, K., Helsen, S., Classification of Model Transformation Approaches, in online proceedings of the 2nd OOPSLA'03 Workshop on Generative Techniques in the Context of MDA. Anaheim, October, 2003.
- [10] Cong, X., Zhang, H., Zhou, D., Lu, P., Qin, L., A Model- Driven Architecture Approach for Developing E-Learning Platform , Entertainment for Education. Digital Techniques and Systems Lecture Notes in Computer Science, Volume 6249/2010, 111-122, DOI: 10.1007/978-3-642-14533-9_12, 2010.
- [11] Distant, D., Rossi, G., Canfora, G., Modeling Business Processes in Web Applications: An Analysis Framework. In Proceedings of the 22nd Annual ACM Symposium on Applied Computing (Page: 1677, Year of publication: 2007, ISBN: 1-59593-480-4).
- [12] Eclipse.org. ATLAS Transformation Language (ATL). <http://www.eclipse.org/m2m/atl/>.
- [13] Gharavi, V., Mesbah, A., Deursen, A. V., Modelling and Generating AJAX Applications: A Model-Driven Approach. Proceeding of the 7th International Workshop on Web- Oriented Software Technologies, New York, USA (Page: 38, Year of publication: 2008, ISBN: 978-80-227-2899-7)
- [14] Gwittir Source Web Site <http://code.google.com/p/gwittir/>
- [15] Hibernate Framework (<http://www.hibernate.org/>)
- [16] Jouault, F., Allilaire, F., Bézivin, J., Kurtev, I., ATL: A model transformation tool. Science of Computer Programming-Elsevier Vol. 72, n. 1-2: pp. 31-39, 2008.
- [17] Koch, N., Transformations Techniques in the Model-Driven Development Process of UWE, Proceeding of the 2nd International Workshop Model-Driven Web Engineering, Palo Alto (Page: 3 Year of publication: 2006 ISBN: 1-59593- 435-9).
- [18] Kraus, A., Knapp, A., Koch N., Model-Driven Generation of Web Applications in UWE. Proceeding of the 3rd International Workshop on Model-Driven Web Engineering, CEUR-WS, Vol. 261, 2007
- [19] Fowler, M., Inversion of Control Containers and the Dependency Injection pattern (<http://martinfowler.com/articles/injection.html>)
- [20] Mbarki, S., Erramdani, M., Toward automatic generation of mvc2 web applications, InfoComp - Journal of Computer Science, Vol.7 n.4, pp. 84-91, December 2008, ISSN: 1807- 4545.
- [21] Mbarki, S., Erramdani, M., Model-Driven Transformations: From Analysis to MVC 2 Web Model, International Review on Computers and Software (I.R.E.CO.S.), Vol. 4. n. 5, pp. 612-620, September 2009.
- [22] Mbarki, S., Rahmouni, M., Erramdani, M., Transformation ATL pour la génération de modèles Web MVC 2, 10e Colloque Africain sur la Recherche en Informatique et en Mathématiques Appliquées, Theme5:Information Systems, CARI 2010.
- [23] Meta Object Facility (MOF), version 2.0, January 2006, <http://www.omg.org/spec/MOF/2.0/PDF/>
- [24] Meta Object Facility (MOF) 2.0 Query/View/Transform-ation (QVT), Version 1.1, December 2009. <http://www.omg.org/spec/QVT/1.1/Beta2/PDF/>
- [25] Miller, J., Mukerji, J., al. MDA Guide Version 1.0.1, 2003. <http://www.omg.org/docs/omg/03-06-01.pdf>.
- [26] Nasir, M.H.N.M., Hamid, S.H., Hassan, H., WebML and .NET Architecture for Developing Students Appointment Management System, Journal of applied science, Vol. 9, n. 8, pp. 1432-1440, 2009.
- [27] Ndie, T. D., Tangha1, C., Ekwoje, F. E., MDA (Model- Driven Architecture) as a Software Industrialization Pattern: An Approach for a Pragmatic Software Factories. J. Software Engineering & Applications, pages 561-571, 2010
- [28] NHibernate Framework home site (<http://nhforge.org/>)
- [29] Puremvc framework (<http://puremvc.org/>).
- [30] Panda, D., Rahman, R., Lane, D., EJB3 in action (Manning co., 2007).
- [31] PicoContainer. <http://www.picocontainer.org/>
- [32] Schincariol, M., Keith, M., Pro JPA 2: Mastering the Java Persistence API (Apress, 2009)
- [33] SmartQVT documentation Copyright © 2007, Copyright(c) France Telecom. <http://smartqvt.elibel.tm.fr/doc/index.html>
- [34] Soler, E., Trujillo, J., Blanco, C., Fernandez-Medina, E., Designing Secure Data Warehouses by Using MDA and QVT. Journal of Universal Computer Science, vol. 15, no. 8 pages 1607-1641, 2009.
- [35] Spring Source Web Site (<http://www.springframework.org/>).
- [36] SpringNet Web Site(<http://www.springframework.net/>).
- [37] Symfony open-Source PHP Web Framework Site (<http://www.symfony-project.org/>)
- [38] Zend Framework (<http://framework.zend.com/>).

- [39] UML Infrastructure Final Adopted Specification, version 2.0, September 2003, <http://www.omg.org/cgi-bin/doc?ptc/03-09-15.pdf>
[40] XML Metadata Interchange (XMI), version 2.1.1, December 2007, <http://www.omg.org/spec/XMI/>

Lamili el Mazoui Nadori Yasser is pursuing his Ph.D at Mohammed First University in the Faculty of Sciences. He got a degree of an engineer in Computer Sciences from the National School of Applied Sciences at Oujda. He received his M.Sc. degree in New Information and Communication Technologies from the faculty of sciences and Techniques at Sidi Mohamed Ben Abdellah University. His research activities at the MATSI Laboratory (Applied Mathematics, Signal Processing and Computer Science) have focused on WebMarketing in social networks using MDA (Model Driven Architecture) approach.

Mohammed Erramdani teaches the concept of Information System at Mohammed First University. He got his thesis of national doctorate in 2001. His activities of research in the MATSI Laboratory (Applied Mathematics, Signal Processing and Computer Science) focusing on MDA (Model Driven Architecture) integrating new technologies XML, EJB, MVC, Web Services, etc.

Ibtissam Arrassen Graduate as Computer Science Enginner from the INPT(National Institut of Poste and Telecommunication) and Ph-D-Student at Faculty of Sciences, Laboratory for Computer Science Research, Mohammed First University, Oujda, Morocco.

Redouane Esbai Ph.D at Mohammed First University in the Faculty of Sciences. He got a degree of an engineer in Computer Sciences from the National School of Applied Sciences at Oujda. He received his M.Sc. degree in New Information and Communication Technologies from the faculty of sciences and Techniques at Sidi Mohamed Ben Abdellah University. His research activities at the MATSI Laboratory (Applied Mathematics, Signal Processing and Computer Science) have focused on MDA (Model Driven Architecture).

Mimoun Moussaoui is a Professor, Vice-Director of High School of Technology and Responsible of the MATSI Laboratory (Applied Mathematics, Signal Processing and Computer Science) at Mohammed First University, Oujda, Morocco.

Combining a drug therapy and oncolytic virotherapy to treat cancer: an optimal control approach

Adil El Alami Laaroussi, Mohamed El hia, Mostafa Rachik

Laboratory of Analysis Modeling and Simulation,
Department of Mathematics and Computer Science,
Faculty of Sciences Ben M'Sik, University Hassan II
Mohammedia, BP 7955, Sidi Othman, Casablanca, Morocco.

Abstract

Optimal control theory is applied to a controlled tumor model. Seeking to minimize the infected cells and to maximize uninfected cells, we use two controls representing drug therapy and oncolytic virotherapy. The Pontryagin's maximum principle is used to characterize the optimal controls. The optimal controls are obtained by solving the optimality system. The results are analyzed and interpreted numerically using MATLAB.

Keywords: *Optimal control, mathematical models, oncolytic virus, Pontryagin's maximum principle.*

1. Introduction

Cancer is a disease that begins as a renegade human cell over which the body has lost control. In order for the body and its organs to function properly, cell growth needs to be strictly regulated. Cancer cells, however, continue to divide and multiply at their own speed, forming abnormal lumps, or tumors.

Not all cancers are natural-born killers. Some tumors are referred to as benign because they don't spread elsewhere in the body. But cells of malignant tumors do invade other tissues and will continue to spread if left untreated, often leading to secondary cancers.

Cancers can start in almost any body cell, due to damage or defects in genes involved in cell division. Mutations build up over time, which is why people tend to develop cancer later in life. What actually triggers these cell changes remains unclear, but diet, lifestyle, viral infections, exposure to radiation or harmful chemicals, and inherited genes are among factors thought to affect a person's risk of cancer.

There are over 100 different types of cancer, affecting various parts of the body. Each type of cancer is unique with its own causes, symptoms, and methods of treatment. Like with all groups of disease, some types of cancer are

more common than others. According to the World Health Organization (WHO), cancer is a leading cause of death worldwide and accounted for 7.6 million deaths (13% of all deaths) in 2008. The main types of cancer are: lung (1.37 million deaths), stomach (736 000 deaths), liver (695 000 deaths), colorectal (608 000 deaths), breast (458 000 deaths) and cervical cancer (275 000 deaths). Deaths from cancer worldwide are projected to continue to rise to over 13.1 million in 2030(WHO).

Knowledge about the causes of cancer, and interventions to prevent and manage the disease are extensive. Cancer can be reduced and controlled by implementing evidence-based strategies for cancer prevention, early detection of cancer and management of patients with cancer. Many cancers have a high chance of cure if detected early and treated adequately. The most common types of cancer treatment are surgery chemotherapy, radiation therapy, targeted therapy, and immunotherapy. The efficiency of cancer treatment has improved dramatically in the last decade. However, patients with certain forms of cancer are still left with limited options for therapy. Many tumors also remain completely incurable, creating a need for a broader spectrum of therapeutic strategies. One promising form of treatment is oncolytic virotherapy. This technique employs replication competent viral vectors as agents that preferentially attack and proliferate in cancerous cells, leaving most healthy cells uninjured. The result is the destruction of tumor populations without appreciable damage to normal tissue. Here are many oncolytic viruses which have demonstrated anti-tumor efficacy, including adenoviruses [1] Coxsackieviruses [2], herpes simplex viruses [3], measles viruses [4], Newcastle disease virus [5], reoviruses [6], Seneca Valley virus [7], vaccinia viruses [8], and vesicular stomatitis virus (VSV) [9].

Mathematical modeling of cancer treatment can illuminate the underlying dynamics of therapy systems and can lead to more optimal treatment strategies. A wide variety of models that study the tumors dynamics and treatment have been developed and analyzed (see for example [10, 11, 12, 13] and the reference in it).

In this paper we explore an optimal control strategy of tumor therapy. We use a controlled model of tumor dynamics that includes two controls representing drug therapy and oncolytic virotherapy; our goal is to maximize the number of susceptible cells and to minimize the number of infected cells.

The paper is organized as follows. In section 2, we present a mathematical model with two control terms. The analysis of optimization problem is presented in section 3. In section 4, we give a numerical appropriate method and the simulation corresponding results. Finally, the conclusions are summarized in section 5.

2. Mathematical model of cancers

We consider the model used in [14]. The model considers two types of tumor cells (x) and (y) growing in logistic fashion. (x) is the density of the uninfected tumor cell population and (y) is the density of infected tumor cell population. The model has the following form

$$\begin{cases} \frac{dx}{dt} = r_1 x \left(1 - \frac{x+y}{k}\right) - \frac{bxy}{x+y} \\ \frac{dy}{dt} = r_2 y \left(1 - \frac{x+y}{k}\right) + \frac{bxy}{x+y} - ay \end{cases} \quad (1)$$

With initial conditions: $x(0) = x_0 > 0$ and $y(0) = y_0 > 0$ Here r_1 and r_2 are the maximum per capita growth rates of uninfected and infected cells correspondingly; K is the carrying capacity, b is the transmission rate and a is the rate of infected cell killing by the viruses. All the parameters of the model are supposed to be non-negative.

The model under consideration in this paper comprises of the following

$$\begin{cases} \frac{dx}{dt} = r_1 x \left(1 - \frac{x+y}{k}\right) - (1 - u_1(t)) \frac{bxy}{x+y} \\ \frac{dy}{dt} = r_2 y \left(1 - \frac{x+y}{k}\right) + (1 - u_1(t)) \frac{bxy}{x+y} - u_2(t)y \end{cases} \quad (2.1)$$

The modifications to the original model are the control terms comprising of u_1 and u_2 . Here $u_1 = u_1(t)$ represents the efficiency of drug therapy in blocking new infection. Thus, the infection rate in the presences of drug

is $(1 - u_1(t)) \frac{bxy}{x+y}$. If $u_1 = 1$, the efficiency of drug therapy in blocking new infection is 100%, whereas if $u_1 = 0$ we find the same incidence term in the model (1). On the other hand the original model contained the term a which is assumed to be a constant, but it would be more biologically realistic if this rate depends on the injected dose of virus. Hence, we consider $u_2 = u_2(t)$ as the second control. It represents the efficiency of killing infected cells by the virus (cytotoxicity), which instead of being selected constant values would change over time.

The control function $u_1(t)$ and $u_2(t)$ are bounded, Lebesgue integral function.

3. The optimal problems

In this section we use the optimal control theory to analyze the behavior of the model (2.1), (2.2). Our goal is to maximize the number of the uninfected cells, to minimize the infected cells and the cost of treatment. Mathematically, for a fixed terminal time t_f , the problem is to maximize the objective functional defined by

$$J(u_1, u_2) = \int_0^{t_f} \left\{ x(t) - y(t) - \left[\frac{A_1}{2} u_1^2(t) + \frac{A_2}{2} u_2^2(t) \right] \right\} dt \quad (3)$$

Where t_f is the period of treatment and the parameters $A_1 \geq 0$ and $A_2 \geq 0$ are based on the benefits and costs of the treatment.

In other words, we are seeking optimal control pair (u_1^*, u_2^*) so that

$$J(u_1^*, u_2^*) = \max \{ J(u_1, u_2), (u_1, u_2) \in U \} \quad (4)$$

Where U is the control set defined by

$$U = \left\{ u = (u_1, u_2), u_i \text{ measurable}; 0 \leq u_i(t) \leq 1; t \in [0, t_f]; i = 1, 2 \right\}$$

3.1 Existence of an optimal control

Before to show the existence of the optimal control pair; we prove the existence of the solution for the controlled system (2.1), (2.2).

We rewrite our system (2.1), (2.2) following form

$$\dot{X}_t = AX + F(X) \quad (5)$$

$$\text{Where : } X = (x, y), A = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 - u_2(t) \end{bmatrix},$$

$$F = \begin{bmatrix} -r_1 x \left(\frac{x+y}{k}\right) - (1 - u_1(t)) \frac{bxy}{x+y} \\ -r_2 y \left(\frac{x+y}{k}\right) + (1 - u_1(t)) \frac{bxy}{x+y} \end{bmatrix}$$

And X_t denote derivative of X with respect to time t . Equation (3) is a non-linear system with a bounded coefficient.

We set:

$$D(X) = AX + F(X) \quad (6)$$

And $\|X\| = \max(|x|, |y|)$ we define:

$$\|F(X_1) - F(X_2)\| = \max(|G_1(X_1, X_2)|, |G_2(X_1, X_2)|)$$

Where:

$$G_1(X_1, X_2) = \left\{ r_1 \left(x_2 \left(\frac{x_2+y_2}{k} \right) - x_1 \left(\frac{x_1+y_1}{k} \right) \right) + (1 - u_1(t)) b \left(\frac{x_2 y_2}{x_2+y_2} - \frac{x_1 y_1}{x_1+y_1} \right) \right\}$$

And:

$$G_2(X_1, X_2) = \left\{ r_2 \left(y_2 \left(\frac{x_2+y_2}{k} \right) - y_1 \left(\frac{x_1+y_1}{k} \right) \right) - (1 - u_1(t)) b \left(\frac{x_2 y_2}{x_2+y_2} - \frac{x_1 y_1}{x_1+y_1} \right) \right\}$$

It follows:

$$\begin{aligned} |G_1(X_1, X_2)| &= \left| \frac{r_1}{k} (x_2^2 + x_2 y_2 - x_1^2 - x_1 y_1) + (1 - u_1(t)) b \left(\frac{x_2 y_2 (x_1+y_1) - x_1 y_1 (x_2+y_2)}{(x_2+y_2)(x_1+y_1)} \right) \right| \\ &= \left| \frac{r_1}{k} (x_2^2 + x_2 y_2 - x_1 y_2 + x_1 y_2 - x_1^2 - x_1 y_1) + (1 - u_1(t)) b \left(\frac{x_2 y_2 (x_1+y_1) - x_1 y_1 (x_2+y_2)}{(x_2+y_2)(x_1+y_1)} \right) \right| \\ &\leq \left| \frac{r_1}{k} ((x_2 - x_1)(x_1 + x_2) + y_2(x_2 - x_1) + x_1(y_2 - y_1)) \right| \\ &\quad + (1 - u_1(t)) b \left| \left(\frac{x_2 y_2 x_1 + x_2 y_2 y_1 - x_1 y_1 x_2 - x_1 y_1 y_2}{(x_2+y_2)(x_1+y_1)} \right) \right| \\ &\leq 4r_1 \|X_2 - X_1\| + b \left| \frac{x_2 y_2 x_1 + x_2 y_2 y_1 - x_1 y_1 x_2 - x_1 y_1 y_2}{(x_2+y_2)(x_1+y_1)} \right| \\ &\leq 4r_1 \|X_2 - X_1\| + b \left| \frac{x_2 x_1 (y_2 - y_1) + y_1 y_2 (x_2 - x_1)}{(x_2+y_2)(x_1+y_1)} \right| \\ &\leq 4r_1 \|X_2 - X_1\| + 2k^2 b \|X_2 - X_1\| \\ &\leq (4r_1 + 2k^2 b) \|X_2 - X_1\| \end{aligned}$$

So we will

$$|G_1(X_1, X_2)| \leq (4r_1 + 2k^2 b) \|X_2 - X_1\| .$$

We get

$$|G_2(X_1, X_2)| \leq (4r_2 + 2k^2 b) \|X_2 - X_1\| .$$

Also we get

$$\|F(X_1) - F(X_2)\| \leq M_1 \|X_2 - X_1\|$$

Where

$$M_1 = \max(4r_1 + 2k^2 b, 4r_2 + 2k^2 b) .$$

We get

$$\|D(X_1) - D(X_2)\| \leq M \|X_2 - X_1\|$$

Where

$$M = \max(\|A\|, M_1) .$$

Thus, it follows that the function D is uniformly Lipschitz continuous. From the definition of the control $u_1(t)$, $u_2(t)$, $x(t)$ and $y(t) \geq 0$, we see that a solution of the system (3) exists [15].

Now In order to find an optimal solution pair, we consider the optimal control problem ((2.1), (2.2).)-(4). First we should find the Lagrangian and Hamiltonian for the optimal control problem ((2.1), (2.2).)-(4). Actually, the Lagrangian of the optimal problem is given by

$$L(x, y, u_1, u_2) = x - y - \left(\frac{A_1}{2} u_1^2 + \frac{A_2}{2} u_2^2 \right).$$

We seek the maximal value of the Lagrangian. To accomplish this, we define the Hamiltonian H for the control problem:

$$H(t, X, u, \lambda) = L(x, y, u_1, u_2) + \sum_{i=1}^2 f_i$$

Where f_i is the right side of the differential equation of the i^{th} state variable.

This is written in our case:

$$\begin{aligned} H(x, y, u_1, u_2, \lambda_1, \lambda_2) &= L(x, y, u_1, u_2) \\ &+ \lambda_1 \left(r_1 x \left(1 - \frac{x+y}{k} \right) - (1 - u_1(t)) \frac{bxy}{x+y} \right) \\ &+ \lambda_2 \left(r_2 y \left(1 - \frac{x+y}{k} \right) + (1 - u_1(t)) \frac{bxy}{x+y} - u_2(t) y \right) . \end{aligned} \quad (7)$$

Where λ_1 and λ_2 are the adjoint functions to be determined suitably.

Theorem 1:

There exists an optimal control $(u_1^*(t), u_2^*(t))$ so that

$$J(u_1^*, u_2^*) = \max_{(u_1, u_2) \in U} J(u_1, u_2)$$

subject to the control system (2.1), (2.2).

Proof:

To use an existence result in [16], we must check the following properties.

- 1) The set U of controls and corresponding state variables is nonempty.
- 2) The control set U is convex and closed.
- 3) The right-hand side of the state system is bounded by a linear function in the state and control variables.

- 4) The integrand of the objective functional is concave on U .
- 5) There exist constants $c_1, c_2 > 0$ and $\rho > 1$ so that the integrand $L(x, y, u_1, u_2)$ of the objective functional satisfies:

$$L(x, y, u_1, u_2) \leq c_1 - c_2 (|u_1|^2 + |u_2|^2)^{\frac{\rho}{2}}$$

An existence result by Lukes [17] is used to give the existence of system (2.1), (2.2).with bounded coefficients, which gives condition 1. The control set is convex and closed by definition. Since the state system is bilinear in u_1 and u_2 , the right side of (2.1), (2.2).satisfies condition (3), using the boundedness of the solution. The integrand in the objective functional (4) is concave on U . In addition, we can easily see that there exist a constant $\rho > 1$ and positive numbers c_1 and $c_2 > 0$ satisfying :

$$L(x, y, u_1, u_2) \leq c_1 - c_2 (|u_1|^2 + |u_2|^2)^{\frac{\rho}{2}}$$

3.2 Characterization of the optimal control

In the previous section we showed the existence of the optimal control pair, which maximizes the functional (3) subject to the system (2.1), (2.2).In order to derive the necessary conditions for this optimal control pair, we apply Pontryagin's maximum principle to the Hamiltonian H (7).

If $(x^*(t), u^*(t))$ is an optimal solution of an optimal control problem, then there exists a non-trivial vector function $\lambda(t) = (\lambda_1(t), \lambda_2(t), \dots, \lambda_n(t))$ satisfying the following equalities:

$$\begin{aligned} x'(t) &= \frac{\partial H(t, x^*(t), u^*(t), \lambda(t))}{\partial x}, \\ \lambda'(t) &= -\frac{\partial H(t, x^*(t), u^*(t), \lambda(t))}{\partial \lambda}, \\ 0 &= \frac{\partial H(t, x^*(t), u^*(t), \lambda(t))}{\partial u} \end{aligned}$$

Now, we apply the necessary conditions to the Hamiltonian H . (7)

Theorem 2:

Let $x^*(t)$ and $y^*(t)$ be optimal state solutions with associated optimal control variable $u^*(t) = (u_1^*(t), u_2^*(t))$ for the optimal control problem (2.1), (2.2).and (4). Then, there exist adjoint variables $\lambda_1(t)$ and $\lambda_2(t)$ that satisfy the equations

$$\begin{aligned} \lambda'_1 &= -1 + \lambda_1 \left\{ r_1 \left(-1 + \frac{2x^*+y^*}{k} \right) + (1 - u_1(t))b \left(\frac{y^*}{x^*+y^*} \right)^2 \right\} \\ &+ \lambda_2 \left\{ \frac{r_2 y^*}{k} - (1 - u_1(t))b \left(\frac{y^*}{x^*+y^*} \right)^2 \right\} \end{aligned}$$

$$\begin{aligned} \lambda'_2 &= 1 + \lambda_1 \left\{ \frac{r_1 x^*}{k} + (1 - u_1(t))b \left(\frac{x^*}{x^*+y^*} \right)^2 \right\} \\ &+ \lambda_2 \left\{ r_2 \left(-1 + \frac{x^*+2y^*}{k} \right) - (1 - u_1(t))b \left(\frac{x^*}{x^*+y^*} \right)^2 + u_2(t) \right\} \end{aligned}$$

With transversality conditions:

$$\lambda_i(t_f) = 0, \quad i = 1, 2.$$

Furthermore, the optimal control $u^*(t) = (u_1^*(t), u_2^*(t))$ is given by

$$\begin{aligned} u^*_1 &= \min \left(1, \max \left(0, \frac{(\lambda_1 - \lambda_2) b x^* y^*}{A_1 x^* + y^*} \right) \right) \\ u^*_2 &= \min \left(1, \max \left(0, \frac{-\lambda_2 y^*}{A_2} \right) \right) \end{aligned}$$

Proof:

We use the Hamiltonian (7) in order to determine the adjoint equations and the transversality conditions. By putting $x(t) = x^*(t)$ and $y(t) = y^*(t)$ differentiating the Hamiltonian with respect to x and y we obtain

$$\begin{aligned} \frac{d\lambda_1}{dt} &= -\frac{\partial H}{\partial x} = -1 \\ &+ \lambda_1 \left\{ r_1 \left(-1 + \frac{2x+y}{k} \right) + (1 - u_1(t))b \left(\frac{y}{x+y} \right)^2 \right\} \\ &+ \lambda_2 \left\{ \frac{r_2 y}{k} - (1 - u_1(t))b \left(\frac{y}{x+y} \right)^2 \right\}. \end{aligned}$$

$$\begin{aligned} \frac{d\lambda_2}{dt} &= -\frac{\partial H}{\partial y} = 1 + \lambda_1 \left\{ \frac{r_1 x}{k} + (1 - u_1(t))b \left(\frac{x}{x+y} \right)^2 \right\} \\ &+ \lambda_2 \left\{ r_2 \left(-1 + \frac{x+2y}{k} \right) - (1 - u_1(t))b \left(\frac{x}{x+y} \right)^2 + u_2(t) \right\}. \end{aligned}$$

With transversality conditions:

$$\lambda_i(t_f) = 0, \quad i = 1, 2.$$

And by using the optimality conditions we find

$$\begin{aligned} \frac{\partial H}{\partial u_1} &= -A_1 u_1 + \lambda_1 \frac{bxy}{x+y} - \lambda_2 \frac{bxy}{x+y}. \\ \frac{\partial H}{\partial u_2} &= -A_2 u_2 - \lambda_2 y. \end{aligned}$$

By the bounds in U of the controls, it is easy to obtain u_1^* and u_2^* in the form:

$$\begin{aligned} u^*_1 &= \min \left(1, \max \left(0, \frac{(\lambda_1 - \lambda_2) b x^* y^*}{A_1 x^* + y^*} \right) \right) \\ u^*_2 &= \min \left(1, \max \left(0, \frac{-\lambda_2 y^*}{A_2} \right) \right). \end{aligned}$$

4. Numerical simulations

The optimality system consists of the state system coupled with the adjoint system with the initial and transversality conditions together with the characterization of the optimal control. Utilizing the characterization of the optimal control, we have the following optimality system.

$$\begin{aligned} \frac{dx^*}{dt} &= r_1 x^* \left(1 - \frac{x^*+y^*}{k} \right) - (1 - u_1^*(t)) \frac{b x^* y^*}{x^*+y^*} \\ \frac{dy^*}{dt} &= r_2 y^* \left(1 - \frac{x^*+y^*}{k} \right) + (1 - u_1^*(t)) \frac{b x^* y^*}{x^*+y^*} - u_2^*(t) y^* \end{aligned}$$

$$\lambda'_1 = -1 + \lambda_1 \left\{ r_1 \left(-1 + \frac{2x^* + y^*}{k} \right) + (1 - u_1^*(t)) b \left(\frac{y^*}{x^* + y^*} \right)^2 \right\} + \lambda_2 \left\{ \frac{r_2 y^*}{k} - (1 - u_1(t)) b \left(\frac{y^*}{x^* + y^*} \right)^2 \right\}$$

$$\lambda'_2 = 1 + \lambda_1 \left\{ \frac{r_1 x^*}{k} + (1 - u_1^*(t)) b \left(\frac{x^*}{x^* + y^*} \right)^2 \right\} + \lambda_2 \left\{ r_2 \left(-1 + \frac{x^* + 2y^*}{k} \right) - (1 - u_1^*(t)) b \left(\frac{x^*}{x^* + y^*} \right)^2 + u_2^*(t) \right\}$$

With $u^*_1 = \min \left(1, \max \left(0, \frac{(\lambda_1 - \lambda_2) b x^* y^*}{A_1 x^* + y^*} \right) \right)$
 $u^*_2 = \min \left(1, \max \left(0, \frac{-\lambda_2 y^*}{A_2} \right) \right)$

$\lambda_1(t_f) = 0$, $\lambda_2(t_f) = 0$, $x(0) = x_0$ and $y(0) = y_0$.

In this formulation, there are initial conditions for the state variables and terminal conditions for the adjoints. That is, the optimality system is a two-point boundary value problem, with separated boundary conditions at times $t=0$ and t_f . An efficient method to solve two-point BVPs numerically is collocation. A convenient collocation code is the solver BVP4c Implemented under MATLAB, which can be used to solve nonlinear two-point BVPs. It is a powerful method to solve the two-point BVP resulting from optimality conditions.

The simulations are carried out using the following values taken from [18]:

$r_1=40$, $r_2=2$, $k=100$, and $b=0.02$

The figure 1 shows that, in case without control, the number of uninfected cells decreases sharply. However, it starts to increase since the first days of treatment.

In figure 2, the number of infected cells grows significantly in case without control. While in case with control we observe a steady decrease.

Figures 3-4-5 give the optimal control pair (u_1, u_2) and the optimal value of cost. We observe that the curves drop off steadily which is because of the constant and steady eradication of the infection.

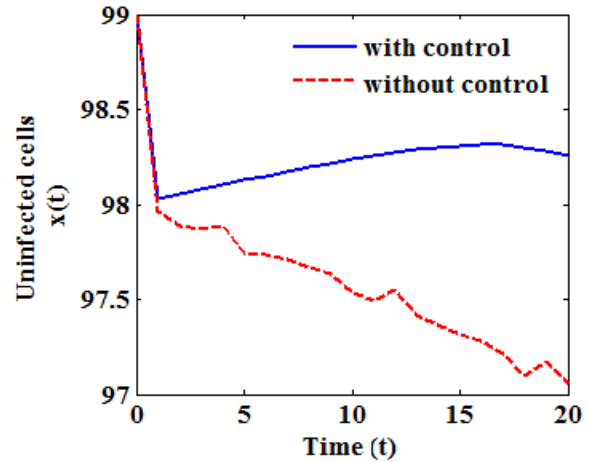


Fig. 1 The function x with and without controls

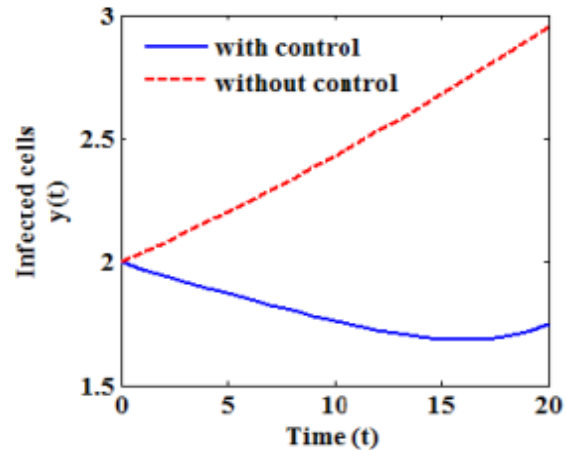


Fig. 2 The function y with and without controls

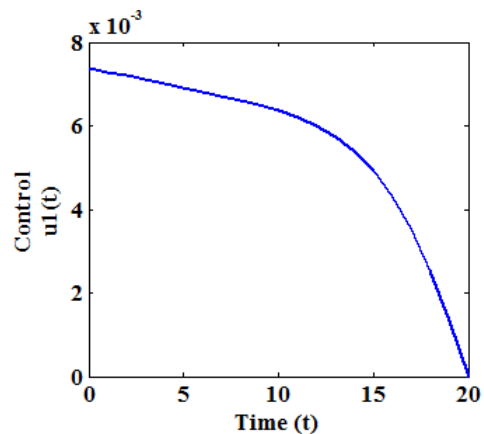


Fig. 3 The optimal control u_1

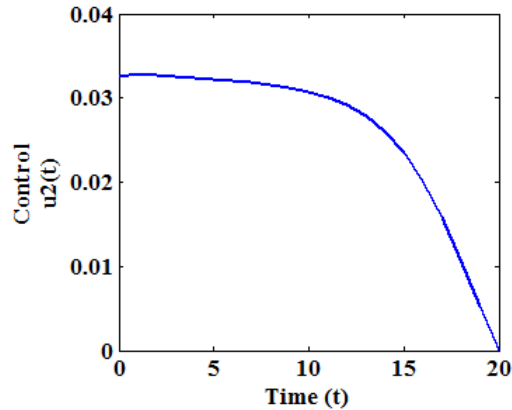


Fig. 4 The optimal control u_2

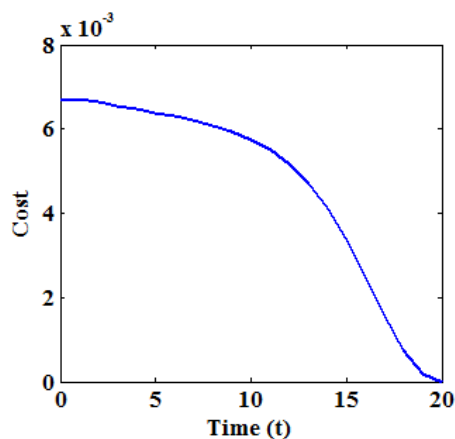


Fig. 5 The optimal value of the cost

4. Conclusion

In this work, we investigate an efficient optimal control strategy of cancer therapy, we use a controlled model of tumor dynamics that includes two controls representing drug therapy and oncolytic virotherapy; our goal is to maximize the number of susceptible cells and to minimize the number of infected cells. The optimal control theory is used to prove the existence and characterize of optimal control pair; the obtained results confirm the performance of our strategy.

Acknowledgments

Research reported in this paper was supported by the Moroccan Systems Theory Network

References

[1] C. Heise, A. Sampson-Johannes, A. Williams, F. McCormick, D.D. Von Hoff, and D.H. Kirn, ONYX-015, an E1B gene-attenuated adenovirus, causes tumor-specific

cytolysis and antitumoral efficacy that can be augmented by standard chemotherapeutic agents, *Nat. Med.* 3 (1997), pp. 639–645.

[2] G.G. Au, A.M. Lindberg, R.D. Barry, and D.R. Shafren, Oncolysis of vascular malignant human melanoma tumors by Cocksackievirus A21, *Int. J. Oncol.* 26 (2005), pp. 1471–1476.

[3] T. Todo, R.L. Martuza, S.D. Rabkin, and P.A. Johnson, Oncolytic herpes simplex virus vector with enhanced MHC class I presentation and tumor cell killing, *Proc. Natl. Acad. Sci. USA* 98 (2001), pp. 6396–6401.

[4] D. Dingli, K.W. Peng, M.E. Harvey, P.R. Greipp, M.K. O'Connor, R. Cattaneo, J.C. Morris, and S.J. Russell, Image-guided radiotherapy for multiple myeloma using a recombinant measles virus expressing the thyroidal sodium iodide symporter. *Blood* 103 (2004), pp. 1641–1646.

[5] A.L. Pecora, N. Rizvi, G.I. Cohen, N.J. Meropol, D. Sterman, J.L. Marshall, S. Goldberg, P. Gross, J.D. O'Neil, W.S. Groene, M.S. Roberts, H. Rabin, M.K. Bamat, and R.M. Lorence, Phase I trial of intravenous administration of PV701, an oncolytic virus, in patients with advanced solid cancers, *J. Clin. Oncol.* 20 (2002), pp. 2251–2266.

[6] C. Comins, J. Spicer, A. Protheroe, V. Roulstone, K. Twigger, C.M. White, R. Vile, A. Melcher, M.C. Coffey, K.L. Mettinger, G. Nuovo, D.E. Cohn, M. Phelps, K.J. Harrington, and H.S. Pandha, REO-10: A phase I study of Intravenous reovirus and docetaxel in patients with advanced cancer, *Clin. Cancer Res.* 16 (2010), pp. 5564–5572.

[7] P.S. Reddy, K.D. Burroughs, L.M. Hales, S. Ganesh, B.H. Jones, N. Idamakanti, C. Hay, S.S. Li, K.L. Skele, A.J. Vasko, J. Yang, D.N. Watkins, C.M. Rudin, and P.L. Hallenbeck, Seneca Valley virus, a systemically deliverable oncolytic picornavirus, and the treatment of neuroendocrine cancers, *J. Natl. Cancer Inst.* 99 (2007), pp. 1623–1633.

[8] B.H. Park, T. Hwang, T.C. Liu, D.Y. Sze, J.S. Kim, H.C. Kwon, S.Y. Oh, S.Y. Han, J.H. Yoon, S.H. Hong, A. Moon, K. Speth, C. Park, Y.J. Ahn, M. Daneshmand, B.G. Rhee, H.M. Pinedo, J.C. Bell, and D.H. Kirn, Use of a targeted oncolytic poxvirus, JX-594, in patients with refractory primary or metastatic liver cancer: A phase I trial, *Lancet Oncol.* 9 (2008), pp. 533–542.

- [9] J. Bell, K. Parato, and H. Atkins, Vesicular stomatitis virus, in *Viral Therapy of Cancer*, K.J. Harrington, R.G. Vile, and H.S. Pandha, eds., Wiley, Hoboken, NJ, 2008.
- [10] Z. Bajzer, T. Carr, K. Josic, S.J. Russell, and D. Dingli, Modeling of cancer virotherapy with recombinant measles viruses, *J. Theor. Biol.* 252 (2008), pp. 109–122.
- [11] N.L. Komarova and D. Wodarz, ODE models for oncolytic virus dynamics, *J. Theor. Biol.* 263 (2010), pp. 530–543.
- [12] A.S. Novozhilov, F.S. Berezovskaya, E.V. Koonin, and G.P. Karev, Mathematical modeling of tumor therapy with oncolytic viruses: Regimes with complete tumor elimination within the framework of deterministic models, *Biol. Direct.* 1 (2006). Available at <http://www.biology-direct.com/content/1/1/6>.
- [13] D. Wodarz, Gene therapy for killing p53-negative cancer cells: Use of replicating versus nonreplicating agents, *Hum. Gene Ther.* 14 (2003), pp. 153–159.
- [14] Artem S Novozhilov, Faina S Berezovskaya, Eugene V Koonin and Georgy P Karev, Mathematical modeling of tumor therapy with oncolytic viruses: Regimes with complete tumor elimination within the framework of deterministic models. *Biology Direct*, 2006.
- [16] G. Birkhoff, G.C. Rota, *Ordinary Differential Equations*, 4th edition, John Wiley & Sons, New York, 1989.
- [15] W. H. Fleming, R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer Verlag, New York, 1975..
- [17] D. L. Lukes, *Differential Equations: Classical To Controlled*, vol. 162 of *Mathematics in Science and Engineering*, Academic Press, New York, NY, USA, 1982.
- [18] Manju Agarwal, Archana S. Bhaduria *Mathematical Modeling and Analysis of Tumor Therapy with Oncolytic Virus*. Lucknow University, Lucknow, India, 2010.

Mixed Immunotherapy and Chemotherapy of Tumors: Optimal Control Approach

SAMIRA ZOUHRI¹, SMAHANE SAADI², ILIAS ELMOUKI³, AMINE HAMDACHE⁴, MOSTAFA RACHIK⁵

¹ Universit Hassan II-Mohammdia, Facult des Sciences Ben M'sik
Dpartement de Mathmatiques, BP.7955, Sidi Othmane, Casablanca, Maroc
samira.zouhrii@gmail.com

² Universit Hassan II-Mohammdia, Facult des Sciences Ben M'sik
Dpartement de Mathmatiques, BP.7955, Sidi Othmane, Casablanca, Maroc
smahanesaadi@gmail.com

³ Universit Hassan II-Mohammdia, Facult des Sciences Ben M'sik
Dpartement de Mathmatiques, BP.7955, Sidi Othmane, Casablanca, Maroc
i.elmouki@gmail.com

⁴ Universit Hassan II-Mohammdia, Facult des Sciences Ben M'sik
Dpartement de Mathmatiques, BP.7955, Sidi Othmane, Casablanca, Maroc
hamdacheamine@gmail.com

⁵ Universit Hassan II-Mohammdia, Facult des Sciences Ben M'sik
Dpartement de Mathmatiques, BP.7955, Sidi Othmane, Casablanca, Maroc
rachik@math.net

Abstract

The aim of this work is to apply optimal control theory to certain cancer treatment strategies which based on combination of multiple cancer therapies, in the form of a system of ordinary differential equations (ODEs), governing cancer growth on cell population level with more than one of therapy, in order to determine the best mix of treatments that minimizes both tumor mass and negative effects upon the health of the patient. Numerical simulations of mixed chemotherapy and immunotherapy shows that neither chemotherapy nor immunotherapy alone are effective in treating the cancer, but in combination the therapies are able to eliminate the entire tumor.

Key words: *chemotherapy, immunotherapy, optimal control theory, Pontryagin's Maximum principle.*

1 Introduction

Cancer is the second cause of death in the world, there were an estimated 12.7 million cancer cases around the world in 2008, of these 6.6 million cases were in men and 6.0 million in women. This number is expected to increase to 21 million by 2030, according to World cancer research fund international[39]. Radiotherapy, chemotherapy, hormone therapy, immunotherapy, gene therapy are the effective treatments for cancer patients. The more recent approach aim is looking into combining immunotherapy and chemotherapy as a way to treat the cancer. The goal of immunotherapy is to strengthen the body's own natural ability to combat cancer by enhancing the effectiveness of the immune system, the use of conventional cancer chemotherapy in combination with immunotherapy was previously not thought to be appropriate because chemotherapy generally reduces immunity and could cancel out the benefits of immunotherapy when given together, although many researchs and theoretical studies and mathematical works showed that combining chemotherapy with immunotherapy can have complementary effects that increases cancer treatment effectiveness, (See for example[1],[40],[41]). The logic behind the development of a combination chemo-immunotherapy strategy is based on using as little chemotherapy drug as possible to effectively kill tumor cells and applying immunotherapy to support the patient's immune system, thus strengthening the body's natural defenses against both the tumor cells and the dangerous side effects of the chemotherapy.

When cancer infects a human cells, the first human body line of defense is the natural killer (NK) cell : white blood cells that actively scan the body for abnormal cells, once found, the fighting strats and the strongest survives. As part of the specific immune response, CD8+ cytotoxic T lymphocytes (CTLs) intervenes to kill the tumor cells. In this work we use an existing model, taken from [1] which describes the interaction of tumor cells, NK cells, CD8+T cells, and Circulating lymphocyte cells, under a combined immunotherapy and chemotherapy, then we apply the method of optimal control theory with quadratic control.

Optimal control has been effectively applied to mathematical models incorporating the interaction between tumors and treatments; for example,[36,37,38] have utilized control theory to maximize the effectiveness of chemotherapy against tumor cell and minimize the toxic effects of treatments, also [42,43] have contributed research on applications to immunotherapy strategies for cancer and HIV. In addition, in the works by de Pillis and radunskaya[44], and de Pillis et al.[1,45] the authors explore various approaches to combining chemo-and immunotherapies through numerical simulation and the implementation of numerical linear controls. The interest in applying quadratic control to this model is determining the best mix of therapy that minimizes both tumor mass and negative effects of such treatments.

There are several different types of immunotherapy which use contrasting methods for fighting off tumor cells, they fall into three main categories: immune response modifiers, monoclonal antibodies, and vaccines (see, for example, [46]). The first category contains substances that affect immune response, such as interleukins (including IL-2), interferons, tumor necrosis factors (TNF), colony-stimulating factors (CSF), and B-cell growth factors. In the next category, monoclonal antibodies are currently being developed to target specific cancer antigens. In the third category are vaccines, which are generally used therapeutically, and are created from tumor cells. In this work, we implement treatment from the first category in the form of mathematical terms that represent IL-2 and tumor infiltrating lymphocyte (TIL) injections.

The outline of this paper is as follows. In section 2, we present the model and discuss the model's assumptions. Section 3 deals with the application of potimal control to the model, beginning with the description of the objective functional. In section 4, we present the Forward Backward Sweep Method to solve the optimality system and we discuss numerical simulations, starting with simulating immunotherapy alone, chemotherapy alone and combinations of these.

2 Model Formulation

In this section, we use an existing immunology model, taken from [1] which describes the dynamic evolution of three populations of immune cells in the presence of tumor and under the combined immunotherapy and chemotherapy treatment. This model involves the following cell populations :

- T, tumor cell population (Units: Number of Cells).
- N, total of natural killer cell (NK) population. These cells are part of the innate immune system and therefore exist even when no tumor cells are present (See, [2]) (Units: Number of Cells per Liter).
- L, total of cytotoxic T lymphocytes (CD8+T) cell population. These cells are active tumor specific cells that are part of the specific immune response. These cells are only present in a large number when tumor cells are present (See, [2],[3]) (Units: Number of Cells per Liter).
- C, number of circulating lymphocytes. This number can be used as a measure of the patient health (See, [12],[11],[10]) (Units: Number of Cells per Liter).
- M, chemotherapy drug concentration in the bloodstream (Units: Milligrams per Liter).
- I, immunotherapy drug: tumor infiltrating lymphocyte (Tils), interleukins (IL-2) concentration in the bloodstream (Units: International Units (IUs) per Liter).

Before going into the model description, it is important to note that, as it is recalled in [1], *there is no universal agreement as to the underlying dynamics or the precise cascades of events that take place in the immune response process. The model recalled and described below is however based on published statements and conjectures as well as reasonable assumptions:*

Tumor Equation(T): The tumor cell population grows logistically, $aT(1-bT)$, in the absence of an immune response, as justified in [13][14]. Death of tumor cells due to natural killer cells takes the form $-cNT$, whereas death due to CD8+T is given by $-DT$ (See, for example, [14],[15],[16]). The presence of tumor cells also stimulates the Natural Killer cells, $g\frac{T^2}{h+T^2}N$, and CD8+T cells, $j\frac{D^2T^2}{K+D^2T^2}L$, (See, for example, [20],[21]).

Natural Killer Cell Equation(N): the source of the NK cell population, eC , is represented as a fraction of the circulating lymphocyte population, a simplification meant to represent the complex cascade of biological events that leads to NK cell stimulation (see, eg, [20]). It's also assumed that a fraction of natural killer cells die when they have interaction with a tumor cell, this gives us the term $-pNT$, (See, [19]).

Tumor Specific T Cell (CD8+T) Equation(L): It's assumed that this population have a linear natural death rate, $-mL$, as well as a quadratic death rate, $-uNL^2$ (see, e.g., [23][24]). The CD8+T cells may also die through interaction with the tumor and this is represented by a mass action term $-qLT$, (See, [19]). Interactions of the tumor with the larger lymphocyte populations, N and C, stimulate CD8+T production, these stimulatory terms are represented by the two positive mass action terms, r_1NT , r_2CT . (See, e.g[22])

Circulating Lymphocyte Equation(C): It's assumed that these cells have a constant source term and a linear death rate.

The Effects of chemotherapy Medicine: Once injected, medicine is assumed to have a linear decay rate. The medicine interacts with each of the four cell populations, T, N, L and C through a term of the form $K_X(1 - e^{-M})X$ (X being T, Nor L), (See,[18]), For each cell population, this term represents cell death due to the medicine. also it's assumed that the fraction of cells killed by chemotherapy depends on the amount of drug present in the system.

The Effects of immunotherapy Medicine (IL-2): Although naturally produced, the cytokine IL-2 is often used to treat cancer(See for example[47]). This model assumes a linear de-

cay rate, additionally, when a CD8+T cells is stimulated by IL-2 it will secrete more IL-2 as represented by $\frac{P_I LI}{g_I + I}$, (See, [3]). The model is governed by the following system of ordinary differential equations :

$$\frac{dT}{dt} = aT(1 - bT) - cNT - DT - K_T(1 - e^{-M})T \quad (1)$$

$$\frac{dN}{dt} = eC - fN + g\frac{T^2}{h + T^2}N - pNT - K_N(1 - e^{-M})N \quad (2)$$

$$\frac{dL}{dt} = -mL + j\frac{D^2 T^2}{K + D^2 T^2}L - qLT + (r_1 N + r_2 C)T - uNL^2 - K_L(1 - e^{-M})L + \frac{P_I LI}{g_I + I} + v_L \quad (3)$$

$$\frac{dC}{dt} = \alpha - \beta C - K_C(1 - e^{-M})C \quad (4)$$

$$\frac{dM}{dt} = -\gamma M + v_M(t) \quad (5)$$

$$\frac{dI}{dt} = -\mu_I I + v_I(t) \quad (6)$$

Where $D = d\frac{(L/T)^i}{s+(L/T)^i}$ and $T(0) = T_0, N(0) = N_0, L(0) = L_0, C(0) = C_0, M(0) = M_0, I(0) = I_0$.

In Table 1, we have provided a summary of equation term descriptions, and in Table 2 we have a list of parameter values taken from experimental results of the patient 9 in [1].

3 The control problem

In this section we consider v_L (TIL treatment), v_M (chemotherapy treatment), v_I (IL-2 treatment) as controls, our aim is to find the best strategy of treatment for fighting cancer. There are various optimal therapy strategies for cancer treatment, it depends on the control objective and the formulation of problem, such as:

- Formulation based on minimizing the drug during the treatment period which also results with the less toxicity effect to the healthy tissues.
- optimal therapy aim at minimizing the drug toxicity and maximizing T-helper

cells(CD4+T), which role is sending signals to other types of immune cells, including CD8+T killer cells, this last destroy and kill the infection or virus.

- Formulation focuses on the health level by allowing a less faster decrease of the tumor size.
- Formulation based on improving the health indicator at price of a gradually longer treatment duration.

In this work we are interested in determining an optimal control, in order to minimize the tumor size and the side effects of the therapy on finite time interval $[0, t_f]$. Limited treatment window $[0, t_f]$ is necessary because the tumor can mutate and develop a resistance to the treatment after some finite time.

The objective functional to be minimized is:

$$J(v_L, v_M, v_I) = \int_0^{t_f} T(t) + A' v_M^2(t) + B' v_I^2(t) + C' v_L^2(t) dt \quad A', B', C' \geq 0 \quad (7)$$

which is quadratic in the three controls and where A' , B' and C' are weight factors.

$$J(V^*) = \min \{J(V) : V \in U\}$$

The main objective is to decrease the number of tumor cells T; More precisely we seek an optimal control $\vec{V}^*(t)$ which minimize the functional subject such that

Where $\vec{V}^*(t) = \begin{pmatrix} v_L^*(t) \\ v_M^*(t) \\ v_I^*(t) \end{pmatrix}$ and U is the control set defined by :

$$U = \left\{ v_L(t), v_M(t), v_I(t) \text{ admissible, } 0 \leq v_L(t) \leq L', 0 \leq v_M(t) \leq M', 0 \leq v_I(t) \leq I', \forall t \in [0, t_f] \right\}$$

There are three constraints associated with this model in which the total drug administered of chemotherapy, IL-2 and Tils are limited by a constant L' , M' , I' , which represents the maximal tolerated dose of immunotherapy and chemotherapy suitable for our patient case(See[1]): $L' = 10^9$ cells is the maximum boost of Tils, $I' = 5 \times 10^6$ cells is the maximum

boost of IL-2, $M' = 5$ mg is the maximum concentration of chemotherapy drug.

Theorem 3.1 Consider the control problem with system equations (1)-(6) and objective functional(7). There exists $V^* \in U$ such that

$$J(V^*) = \min \{J(V) : V \in U\}$$

3.1 Characterization of the Optimal Control

The optimal control is characterized by using Pontryagin's Maximum Principle(See, [25]).

Theorem 3.2 Given an optimal control triple, $V^* = (v_L^*(t), v_M^*(t), v_I^*(t))$, and solutions of the corresponding state system, there exist adjoint variables λ_i for $i=1,2,\dots,6$, satisfying the following equations:

$$\begin{aligned} \frac{d\lambda_1}{dt} &= -1 - \lambda_1 \left(a - 2abT - cN - D + \frac{ds(L/T)^l}{(s+(L/T)^l)^2} - K_T(1 - e^{-M}) \right) - \lambda_2 \left(\frac{2gNT_h}{(h+T^2)^2} - pN \right) - \\ &\lambda_3 \left(\frac{-2DTdlsjLk(L/T)^l}{(k+D^2T^2)^2(s+(L/T)^l)^2} + \frac{2D^2TjLk}{(k+D^2T^2)^2} - qL + r_1N + r_2C \right) \\ \frac{d\lambda_2}{dt} &= \lambda_1 cT - \lambda_2 \left(-f + \frac{gT^2}{h+T^2} - pT - K_N(1 - e^{-M}) \right) - \lambda_3 (r_1T - \mu L^2) \\ \frac{d\lambda_3}{dt} &= \lambda_1 \left(\frac{dsl(L/T)^{l-1}}{(s+(L/T)^l)^2} \right) - \lambda_3 \left(-m - qT - 2\mu NL - K_L(1 - e^{-M}) + \frac{P_I I}{g_I + I} + \frac{2jkDdsl(L/T)^l T^2}{(s+(L/T)^l)^2(k+D^2T^2)^2} + \frac{jD^2T^2}{k+D^2T^2} \right) \\ \frac{d\lambda_4}{dt} &= -\lambda_2 e - \lambda_3 r_2 T + \lambda_4 (\beta + K_C(1 - e^{-M})) \\ \frac{d\lambda_5}{dt} &= \lambda_1 K_T e^{-M} T + \lambda_2 K_N e^{-M} N + \lambda_3 K_L e^{-M} L + \lambda_4 K_C e^{-M} C \\ \frac{d\lambda_6}{dt} &= -\lambda_3 \left(\frac{P_I L g_I}{(g_I + I)^2} + \lambda_6 u_I \right) \end{aligned}$$

Where $\lambda_i(t_f) = 0$ for $i = 1, 2, \dots, 6$. Furthermore we have

$$v_L^*(t) = \min \left(\max(0, \frac{-\lambda_3}{2C'}, L') \right)$$

$$v_M^*(t) = \min \left(\max(0, \frac{-\lambda_5}{2A'}, M') \right)$$

$$v_I^*(t) = \min \left(\max(0, \frac{-\lambda_6}{2B'}), I' \right)$$

Proof

The Lagrangian for our problem is the integrand of the objective functional, coupled with the right hand sides of the state equations (1)-(6) through the adjoint variables, it's given by:

$$\begin{aligned} L = & T(t) + Av_M^2(t) + Bv_I^2(t) + Yv_L^2(t) \\ & +\lambda_1(aT(1 - bT) - cNT - DT - K_T(1 - e^{-M})T) \\ & +\lambda_2(eC - fN + \frac{gT^2N}{h+T^2} - pNT - K_N(1 - e^{-M})N) \\ & +\lambda_3(-mL + j\frac{D^2T^2}{K+D^2T^2}L - qLT + (r_1N + r_2C)T - uNL^2 - K_L(1 - e^{-M})L + \frac{P_I L I}{g_I + I} + v_L(t)) \\ & +\lambda_4(\alpha - \beta C - K_C(1 - e^{-M})C) \\ & +\lambda_5(-\gamma M + v_M(t)) \\ & +\lambda_6(\mu_I I + v_I(t)) \\ & -w_1(t)v_L(t) - w_2(t)(L' - v_L(t)) - w_3(t)v_M(t) - w_4(t)(M' - v_M(t)) - w_5(t)v_I(t) - w_6(t)(I' - v_I(t)) \end{aligned}$$

Where $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6$ are the adjoint variables and $w_1, w_2, w_3, w_4, w_5, w_6$ are penalty multipliers which attach the control constraints and verify the conditions

$$\begin{aligned} w_1(t) \geq 0, w_3(t) \geq 0, w_5(t) \geq 0 \text{ and} \\ w_2(t)(L' - v_L(t)) = 0, \text{ and } w_1(t)v_L = 0, \text{ at } v_L^*(t) \\ w_4(t)(M' - v_M(t)) = 0, \text{ and } w_3(t)v_M = 0, \text{ at } v_M^*(t) \\ w_6(t)(I' - v_I(t)) = 0, \text{ and } w_5(t)v_I = 0, \text{ at } v_I^*(t) \end{aligned}$$

we differentiate the Lagrangian with respect to states, T, N, L, C, M and I, respectively, and the adjoint system can be written as

$$\begin{aligned} \frac{d\lambda_1}{dt} = & -1 - \lambda_1 \left(a - 2abT - cN - D + \frac{ds(L/T)^l}{(s+(L/T)^l)^2} - K_T(1 - e^{-M}) \right) - \lambda_2 \left(\frac{2gNT}{(h+T^2)^2} - pN \right) - \\ & \lambda_3 \left(\frac{-2DTdsjLk(L/T)^l}{(k+D^2T^2)^2(s+(L/T)^l)^2} + \frac{2D^2TjLk}{(k+D^2T^2)^2} - qL + r_1N + r_2C \right) \\ \frac{d\lambda_2}{dt} = & \lambda_1 cT - \lambda_2 \left(-f + \frac{gT^2}{h+T^2} - pT - K_N(1 - e^{-M}) \right) - \lambda_3 (r_1T - \mu L^2) \\ \frac{d\lambda_3}{dt} = & \lambda_1 \left(\frac{dsl(L/T)^{l-1}}{(s+(L/T)^l)^2} \right) - \lambda_3 \left(-m - qT - 2\mu NL - K_L(1 - e^{-M}) + \frac{P_I I}{g_I + I} + \frac{2jkDdsl(L/T)^l T^2}{(s+(L/T)^l)^2(k+D^2T^2)^2} + \frac{jD^2T^2}{k+D^2T^2} \right) \end{aligned}$$

$$\begin{aligned} \frac{d\lambda_4}{dt} &= -\lambda_2 e - \lambda_3 r_2 T + \lambda_4 (\beta + K_C (1 - e^{-M})) \\ \frac{d\lambda_5}{dt} &= \lambda_1 K_T e^{-M} T + \lambda_2 K_N e^{-M} N + \lambda_3 K_L e^{-M} L + \lambda_4 K_C e^{-M} C \\ \frac{d\lambda_6}{dt} &= -\lambda_3 \left(\frac{P_I L g_I}{(g_I + I)^2} + \lambda_6 u_I \right) \end{aligned}$$

Where the transversality conditions $\lambda_i(t_f) = 0$ for $i = 1, 2, \dots, 6$.

To characterize v_L^* , v_M^* , v_I^* , we analyze the necessary optimality condition

$$\frac{\partial L}{\partial v_L} = 0, \text{ at } v_L^*$$

$$\frac{\partial L}{\partial v_M} = 0, \text{ at } v_M^*$$

$$\frac{\partial L}{\partial v_I} = 0, \text{ at } v_I^*$$

we differentiate the Lagrangian with respect to v_L , v_M , v_I , then we obtain

$$\frac{\partial L}{\partial v_L} = 2C' v_L + \lambda_3 - W_1 + W_2$$

$$\frac{\partial L}{\partial v_M} = 2A' v_M + \lambda_5 - W_3 + W_4$$

$$\frac{\partial L}{\partial v_I} = 2B' v_I + \lambda_6 - W_5 + W_6$$

Where

$$\begin{cases} w_2(t)(L' - v_L^*(t)) = 0 & \text{and } w_1(t)v_L^* = 0 \\ w_4(t)(M' - v_M^*(t)) = 0 & \text{and } w_3(t)v_M^* = 0 \\ w_6(t)(I' - v_I^*(t)) = 0 & \text{and } w_5(t)v_I^* = 0 \end{cases}$$

$$v_L = \begin{cases} 0 & \text{if } \frac{-\lambda_3}{2C'} \leq 0 \\ \frac{-\lambda_3}{2C'} & \text{if } 0 < \frac{-\lambda_3}{2C'} < L' \\ L' & \text{if } \frac{-\lambda_3}{2C'} \geq L' \end{cases}$$

$$v_M = \begin{cases} 0 & \text{if } \frac{-\lambda_5}{2A'} \leq 0 \\ \frac{-\lambda_5}{2A'} & \text{if } 0 < \frac{-\lambda_5}{2A'} < M' \\ M' & \text{if } \frac{-\lambda_5}{2A'} \geq M' \end{cases}$$

$$v_I = \begin{cases} 0 & \text{if } \frac{-\lambda_6}{2B'} \leq 0 \\ \frac{-\lambda_6}{2B'} & \text{if } 0 < \frac{-\lambda_6}{2B'} < I' \\ I' & \text{if } \frac{-\lambda_6}{2B'} \geq I' \end{cases}$$

Combining these three cases, the optimal controls are characterized by:

$$v_L^*(t) = \min \left(\max(0, \frac{-\lambda_3}{2C'}), L' \right)$$

$$v_M^*(t) = \min \left(\max(0, \frac{-\lambda_5}{2A'}), M' \right)$$

$$v_I^*(t) = \min \left(\max(0, \frac{-\lambda_6}{2B'}), I' \right)$$

4 Numerical simulations

4.1 Method and Algorithm

For notational simplicity, we express the problem as finding (X, Λ, V) ,

$$\text{where } X = \begin{pmatrix} T \\ N \\ L \\ C \\ M \\ I \end{pmatrix}, \Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{pmatrix}$$

and $V^* = \begin{pmatrix} v_L^* \\ v_M^* \\ v_I^* \end{pmatrix}$

$$\begin{cases} \dot{X} = f(X, V) \in \mathbb{R}^6, & X(0) = X_0, & (4.1) \\ \dot{\Lambda} = g(\Lambda, X) \in \mathbb{R}^6, & \Lambda(t_f) = 0, & (4.2) \\ V = h(X, \Lambda) & & (4.3) \end{cases}$$

our problem can be written as:

$$\begin{cases} \dot{X} = f(X, h(X, \Lambda)) = f_1(X, \Lambda) \\ \dot{\Lambda} = g(\Lambda, X) \end{cases}$$

Letting $\dot{Y} = \begin{pmatrix} \dot{X} \\ \dot{\Lambda} \end{pmatrix} = \begin{pmatrix} f_1(X, \Lambda) \\ g(\Lambda, X) \end{pmatrix} = g_1(X, \Lambda) = g_1(Y) \in \mathbb{R}^{12}$

$$\begin{cases} \dot{Y} = g_1(Y) \\ X(0) = X_0, \\ \Lambda(t_f) = 0, \end{cases} \text{ which is two-point boundary value problem (TBVP)}$$

This process requires The Forward Backward Sweep Method

The Forward Backward Sweep Method first solves the state equation (4.1) with a Runge-Kutta routine, then solves the costate equation (4.2) backwards in time with the Runge-Kutta solver, and then updates the control. This provokes a new approximation of the state, costate, and control. The method continues by using these new updates and calculating new values approximations for (X, λ, V) .

The algorithm describing the approximation method for obtaining the optimal control is the following

Algorithm:

Step 1. Make an initial guess for V over the interval. we choose $V_0 = \begin{pmatrix} L'/2 \\ M'/2 \\ I'/2 \end{pmatrix}$

Step 2. Using the initial condition $X(t_0) = X_0$ and the stored values for V, solve X forward in time according to its differential equation in the optimality system.

Step 3. Using the transversality condition $\Lambda(t_f) = 0$ and the stored values for V and X, solve Λ backward in time according to its differential equation in the optimality system.

Step 4. Update the control by entering the new X and Λ values into the characterization of V.

Step 5. Check convergence. if values of the variables in this interaction and the last iteration are negligibly small, output the current values as solutions. If values are not small, return to Step 2.

The method terminates when there is enough agreement between the states, costates, and controls of two passes through the approximation loop.

For Steps 2 and 3, we use a Runge-Kutta 4 routine, many types of convergence tests exist for Step 5, it is sufficient to require $\|V_{i+1} - V_i\|$ to be small, where V_{i+1} is the vector of estimated values of the control during the current iteration, and V_i is the vector of estimated values from the previous iteration. Both these vectors are of length $n+1$, which is the number of time steps. We require the percentage error to be negligibly small ,i.e.,

$$\|V_{i+1} - V_i\| \leq \epsilon$$

Where ϵ is the accepted tolerance.

4.2 Numerical Experiments

In this section, we test the behavior of the model using parameters taken from experimental results of patient(patient 9 in[1]) from Rosenberg's study on metastatic melanoma. First we present a tumor burden which can be controlled by the immune system and case which this last is not able alone to kill the tumor. We also present a case for which either chemotherapy alone or immunotherapy alone can control the tumor, and case for which a combination therapy is essential

to the survival of the patient.

4.2.1 Immune system response to Tumor:

For this first case we examine an initial tumor burden of 10^5 cells which is controlled easily by the immune system over 130 days(see Figure 1), but a tumor burden of 10^6 cells grows to a dangerous level(see Figure 2).

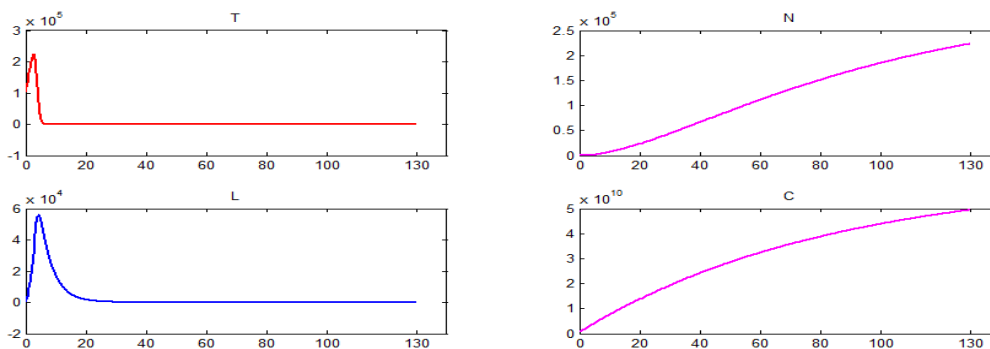


Figure 1: Initial conditions: 10^5 Tumor cells, 10^3 NK cells, 10 CD8+T cells, 6×10^8 circulating lymphocytes.

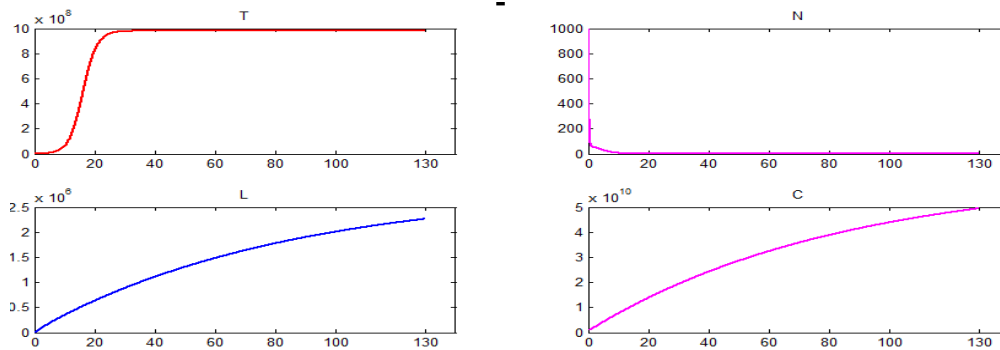


Figure 2: Initial conditions: 10^6 Tumor cells, 10^3 NK cells, 10 CD8+T cells, 6×10^8 circulating lymphocytes.

The innate immune is able to kill a tumor of size 10^5 cells (see figure 1), a size which in many cases is still considered to be below the threshold of clinical detectability in a human, although it's unable to kill a tumor size 10^6 cells (see figure 2) which grows to a dangerous level in the absence of treatment interventions. The body can't fight the virus alone. Parameters for these simulations are provided in Table 2.

4.2.2 Chemotherapy alone:

This experiment represents a situation in which the cancer is large enough considered potentially detectable and the immune system alone is unable to kill the tumor which imposes a treatment interventions. We examine pure chemotherapy for a tumor burden of 2×10^7 cells for 60 days then for 130 days.

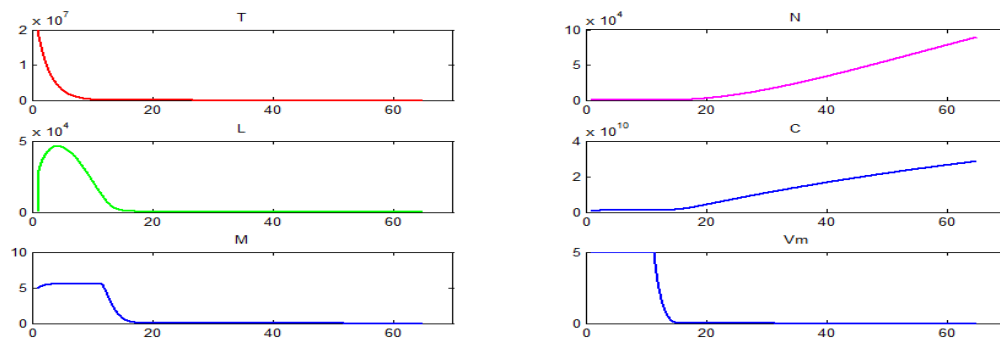


Figure 3: Initial conditions: 2×10^7 Tumor cells, 10^3 NK cells, 10 CD8+T cells, 6×10^8 circulating lymphocytes, V_M is the chemotherapy optimal control.

The chemotherapy treatment interventions given during the initial 20 days from 60 days of treatment is sufficient to eliminate a tumor of size 2×10^7 (Figure 3). Whene we ran the simulation for 130 days, the tumor decreases but survives

despite either chemotherapy treatments (Figure 4). The chemotherapy effectiveness may limited to short time of treatment. Parameters for these simulations are provided in Table 2.

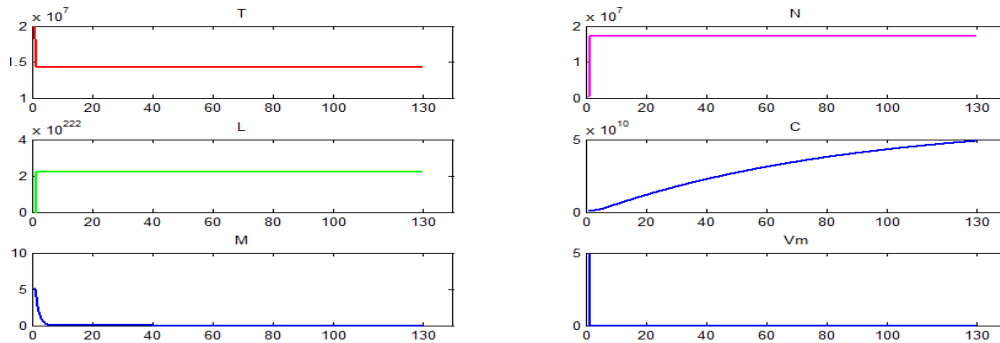


Figure 4: Initial conditions: 2×10^7 Tumor cells, 10^3 NK cells, 10 CD8+T cells, 6×10^8 circulating lymphocytes, V_M is the optimal control.

4.2.3 Immunotherapy alone:

In addition to pure chemotherapy treatments, we examine the effectiveness of pure immunother-

apy treatments with TIL and IL-2 injections for initial tumor size of 10^6 cells, then for tumor burden of 10^7 cells.

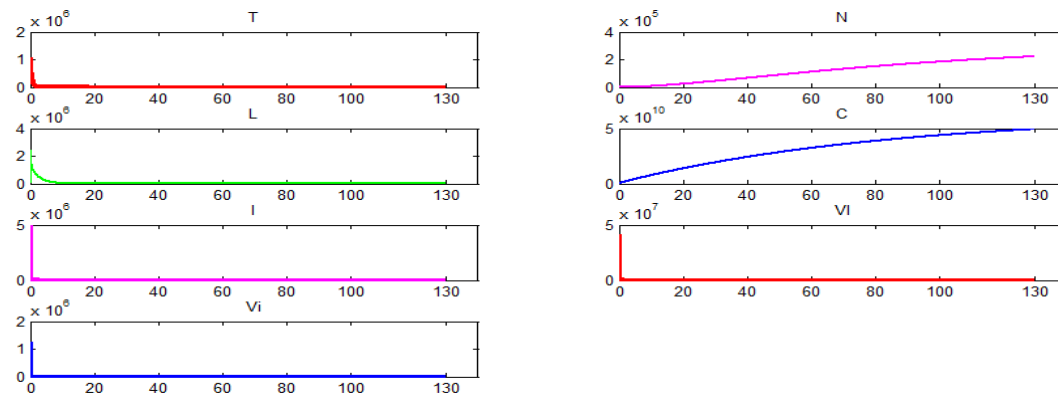


Figure 5: Initial conditions: 10^3 NK cells, 10 CD8+T cells, 6×10^8 circulating lymphocytes, 10^9 Tils, 5×10^6 IL-2. V_I (IL-2 therapy) and V_L (TIL therapy) are the optimal controls and are so small in value that it is barely visible in the graph.

The immunotherapy treatment interventions is sufficient to eliminate a tumor of size 10^6 (Figure 5). For tumor challenge of 10^7 cells, the tumor decreases and survives despite either

immunotherapy intervention (Figure 6). The immunotherapy effectiveness may limited to smaller tumor sizes. Parameters for these simulations are provided in Table 2.

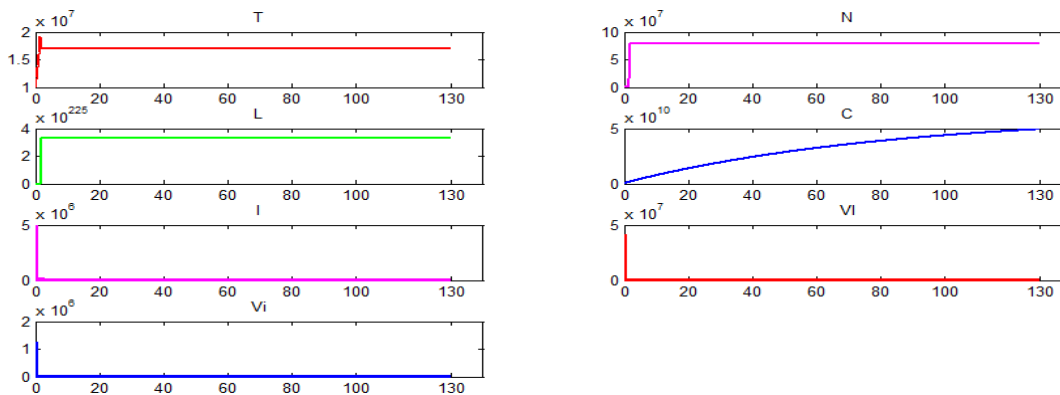


Figure 6: Initial conditions: 10^3 NK cells, 10 CD8+T cells, 6×10^8 circulating lymphocytes, 10^9 Tils, 5×10^6 IL-2, V_I and V_L are the optimal controls and are so small in value that it is barely visible in the graph.

4.2.4 Combination therapy:

We have presented cases for which chemotherapy alone or immunotherapy alone can kill a tumor, also we have presented situations in which these treatments in isolation are not sufficient to eliminate the cancer. Now we combine the separately unsuccessful therapies to examine the ef-

fectiveness of mixed therapy. We measure the patient's immunological health by the number of circulating lymphocytes in the body and do not allow the circulating lymphocytes to drop below a threshold where risk of infection may be too high. In our experiments, we chose that threshold to be the order of 10^8 cells (see,e.g.,[2]), we ran the simulation for 130 days.

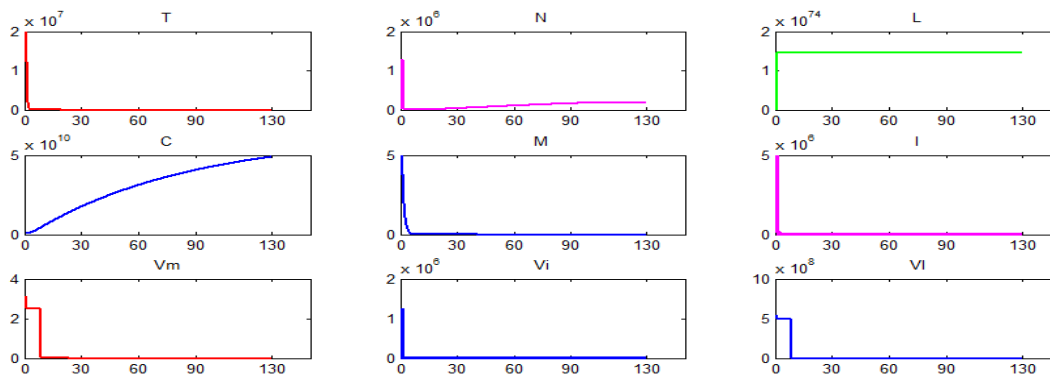


Figure 7: Initial conditions: 2×10^7 tumor cells, 10^3 NK cells, 10 CD8+T cells, 6×10^8 circulating lymphocytes, 10^9 Tils, 5×10^6 IL-2, V_M, V_I and V_L are the optimal controls. Parameters for these simulations are provided in Table 2.

The combination treatment with Til injections, IL-2 injections, and chemotherapy is now able to eliminate a tumor of 2×10^7 cells since the first week of treatment. The chemotherapy V_M and TIL therapy V_L dominate the treatment for the first 10 days of treatment, IL-2 therapy V_I

is so small in value it may be considered practically zero. All initial conditions and parameter values remain the same as in these previous experiments.

5 Conclusion

In this work we have used an existing model of cancer growth, immune response, and treatment that includes activated anticancer-cell transfer (TIL injections), and activation-protein injections (IL-2 injections) in combination with chemotherapy. Our development of combination immunotherapy-chemotherapy protocols demonstrates the following results:

The immune system can control small tumor, al-

though is not sufficient in treating the tumor with high size. Chemotherapy alone administered for short time is effective in killing the tumor. Immunotherapy alone effectiveness may limited to smaller tumor sizes, as is shown in figure 5-6. We use the cases for which chemotherapy alone or immunotherapy alone are not sufficient to kill the tumor, to show that the combination treatment (chemotherapy-immunotherapy) is now capable to eliminate tumor cells.

Eq	Term	Description
$\frac{dT}{dt}$	$aT(1-bT)$	Logistic tumor growth
	$-cNT$	NK-induced tumor death
	$-DT$	CD8+T cell-induced tumor death
	$-K_T(1 - e^{-M})T$	Chemotherapy-induced tumor death
$\frac{dN}{dt}$	eC	Production of NK cells from circulating lymphocytes
	$-fN$	Natural Killer breakdown
	$-pNT$	Natural killer death by exhaustion of tumor-killing
	$g \frac{T^2}{h+T^2} N$	The NK cell recruitment
	$-K_N(1 - e^{-M})N$	Chemotherapy-induced NK death
$\frac{dL}{dt}$	$-mL$	CD8+T cell breakdown
	$-qLT$	CD8+T cell death by exhaustion of tumor-killing resources
	$r_1 NT$	CD8+T cell stimulation by NK-lysed tumor cell debris
	$r_2 CT$	Interactions of the tumor with C cells , stimulate CD8+T production
	$\frac{P_I LI}{g_I + I}$	Stimulatory effect of IL-2((the cytokine interleukin-2)) on CD8+T cells
	$-K_L(1 - e^{-M})L$	Chemotherapy-induced CD8+T death
	$-\mu NL^2$	a quadratic death rate of CD8+T cell
	$j \frac{D^2 T^2}{k+D^2 T^2} L$	The CD8+T cell recruitment
$v_L(t)$	External Tils(tumor infiltrating lymphocyte) controllable	
$\frac{dC}{dt}$	α	a constant source rate
	$-\beta C$	Lymphocyte breakdown
	$-K_C(1 - e^{-M})C$	Chemotherapy-induced Lymphocyte death
$\frac{dM}{dt}$	$-\gamma M$	breakdown of chemotherapy medicine
	$v_M(t)$	External chemotherapy controllable
$\frac{dI}{dt}$	$-\mu_I I$	IL-2 breakdown
	$v_I(t)$	External IL-2 controllable

Table 1: The functional forms for each cell-interaction.

Parameter values

Estimated Value	Description	Source
$a = 4.31 \times 10^{-1}$	Tumor growth rate.	[26]
$b = 1.02 \times 10^{-9}$	1/b is tumor carrying capacity.	[26]
$c = 6.41 \times 10^{-11}$	Fractional (non)-ligand-transduced tumor cell kill by NK cells.	[26][27]
$d = 2.34$	Saturation level of fractional tumor cell kill by CD8+T cells. Primed with ligand-transduced cells, challenged with ligand-transduced cells.	[27]
$e = 2.08 \times 10^{-7}$	Fraction of circulating lymphocytes that become NK cells.	[28]
$l = 2.09$	Exponent of fractional tumor cell kill by CD8+T cells. Primed with ligand-transduced cells, challenged with ligand-transduced cells.	[27]
$f = 4.12 \times 10^{-2}$	Death rate of NK cells.	[28]
$g = 1.25 \times 10^{-2}$	Maximum NK cell recruitment rate by ligand-transduced tumor cells.	[26],[27]
$h = 2.02 \times 10^7$	Steepness coefficient of the NK cell recruitment curve.	[28]
$j = 2.49 \times 10^{-2}$	Maximum CD8+T cell recruitment rate. Primed with ligand-transduced cells, challenged with ligand-transduced cells.	[26],[27]
$k = 3.66 \times 10^7$	Steepness coefficient of the CD8+T cell recruitment curve.	[26],[27]
$m = 2.04 \times 10^{-1}$	Death rate of CD8+T cells.	[29]
$q = 1.42 \times 10^{-6}$	CD8+T cell inactivation rate by Tumor cells.	[28]
$p = 3.42 \times 10^{-6}$	NK cell inactivation rate by Tumor cells.	[26]
$s = 8.39 \times 10^{-2}$	Steepness coefficient of the Tumor-(CD8+T cell) lysis term D. Primed with ligand-transduced cells, challenged with ligand-transduced cells.	[27]
$r_1 = 1.10 \times 10^{-7}$	Rate at which CD8+T cells are stimulated to be produced as a result of tumor cells killed by NK cells.	[29][30]
$r_2 = 6.50 \times 10^{-11}$	Rate at which CD8+T cells are stimulated to be produced as a result of tumor cells interacting with circulating lymphocytes.	[1]
$u = 3.00 \times 10^{-10}$	Regulatory function by NK-cells of CD8+T-cells.	[1]
$K_T = 9.00 \times 10^{-1}$	Fractional tumor cell kill by chemotherapy.	[31]
$K_N = K_L = K_C = 6 \times 10^{-1}$	Fractional immune cell kill by chemotherapy.	[31]
$\alpha = 7.50 \times 10^8$	Constant source of circulating lymphocytes.	[34,35]
$\beta = 1.20 \times 10^{-2}$	Natural death and differentiation of circulating lymphocytes.	[34,35]
$\gamma = 9.00 \times 10^{-1}$	Rate of chemotherapy drug decay.	[33]
$p_I = 1.25 \times 10^{-1}$	Maximum CD8+T-cell recruitment rate by IL-2.	[3]
$g_I = 2.00 \times 10^7$	Steepness of CD8+T-cell recruitment curve by IL-2.	[3]
$\mu_I = 1.00 \times 10^1$	Rate of IL-2 drug decay.	[3]

Table 2: Estimated parameter values.

References

[1] L. G. de Pillis, W. Gu, A. E. Radunskaya, Mixed Immunotherapy and Chemotherapy of tumors: Modeling, Applications and Biological interpretations, 2005.

[2] I.M. Roitt, J. Brostoff, and D.K. Male. Immunology. Mosby, St. Louis, 1993.

[3] D. Kirschner and J.C. Panetta. Modeling immunotherapy of the tumor-immune interaction. J Math Biol., 37(3):23552, September 1998.

[4] L.G. de Pillis, W. Gu, K.R. Fister, T. Head, K. Maples, T. Neal, A. Murugan, K. Yoshida: Chemotherapy for Tumors: An Analysis of the Dynamics and a Study of Quadratic and Linear Optimal Controls,

- 2006.
- [5] L.G. de Pillis, K.R. Fister, W. Gu, Et AL:SEEKING BANG-BANG SOLUTIONS OF MIXED IMMUNOCHEMOTHERAPY OF TUMORS, *Electronic Journal of Differential Equations*, Vol. 2007(2007), No. 171, pp. 124.
- [6] S. Chareyron, M. Alamir: Mixed immunotherapy and chemotherapy of tumors, Feedback design and model updating schemes, *Journal of Theoretical Biology*, 2008.
- [7] M.Rachik, S.Saadi, H.Jourhmane, Y.Rahhou, Z.Tridane: Optimal control of the chemotherapy: HIV model, *Journal of Mathematical control sciences and Applications*, Vol.3, No 1, pp. 79-88, 2010.
- [8] Michael McAseya, Libin Moua, Weimin Hanb: Convergence of the Forward-Backward Sweep Method in Optimal Control.
- [9] Mehmet ITIK, Metin Uymaz SALAMCI, Stephen Paul BANKS: SDRE optimal control of drug administration in cancer treatment.
- [10] R. Glas, L. Franksson, C. Une, M. Eloranta, C. Ohlen, A. Orn, and K. Karre. Recruitment and activation of natural killer (NK) cells in vivo determined by the target cell phenotype: An adaptive component of NK cell-mediated responses.
- [11] B. Melichar, J. Dvorak, P. Jandik, M. Tou-skova, D. Solichova, J. Megancova, and Voboril Z. Intraarterial chemotherapy of malignant melanoma metastatic to the liver. *Hepatogastroenterology*, 48(42):17111715, 2001.
- [12] M.M. Mustafa, G.R. Buchanan, N.J. Winick, G.H. McCracken, I. Tkaczewski, M. Lipscomb, Q. Ansari, and M.S. Agopian. Immune recovery in children with malignancy after cessation of chemotherapy. *J Pediatr Hematol Oncol.*, 20(5):451457, 1998.
- [13] N.F. Britton. *Essential Mathematical Biology*. Springer Verlag, 2003.
- [14] A. Diefenbach, E.R. Jensen, A.M. Jamieson, and D. Raulet. Rae1 and H60 ligands of the NKG2D receptor stimulate tumor immunity. *Nature*, 413:165 171, September 2001.
- [15] Y. Kawarada, R. Ganss, N. Garbi, B. Sacher, T. ad Arnold, and G. Hammerling. NK- and CD8+T cell-mediate eradication of established tumors by peritumoral injection of CpG-containing oligodeoxynucleotides. *J Immunol.*, 167(1):52475253, 2001.
- [16] R.N. Germain. An innately interesting decade of research in immunology. *NatMed.*, 10(12):13071320, December 2004.
- [17] M.A. Cooper, T.A. Fehniger, and M.A. Caligiuri. The biology of human natural killer-cell subsets. *Trends Immunol.*, 22(11):633-640, 2001.
- [18] S. N. Gardner; A mechanistic, predictive model of dose-response curves for cell cycle phasespecific and nonspecific drugs, *Cancer Research*, (60) (2000), 14171425.
- [19] L.G. de Pillis and A.E. Radunskaya. Immune response to tumor invasion. In K.J. Bathe, editor, *Computational Fluid and Solid Mechanics*, volume 2, pages 16611668, 2003. M.I.T.
- [20] M. Alamir. Stabilization of nonlinear systems using receding-horizon control schemes - A parametrized approach for fast systems. Springer, 2006.
- [21] N. Bellomo, A. Bellouquid, and M. Delitala. Mathematical topics on the modelling complex multicellular systems and tumor immune cells competition. *Math. Models Methods Appl. Sci.*, 14:16831733, 2004.
- [22] A.Y.C. Huang, P. Golumbek, M. Ahmadzadeh, E. Jaffee, D. Pardoll, and H. Levitsky. Role of bone marrow-derived cells in presenting MHC class I restricted tumor antigens. *Science*, 264(5161):961-965, May 1994.

- [23] A. Gett, F. Sallusto, A. Lanzavecchia, and J. Geginat. T cell fitness determined by signal strength. *Nat Immunol.*, 4(4):355-360, April 2003.
- [24] S.M. Gilbertson, P.D. Shah, and D.A. Rowley. NK cells suppress the generation of Lyt-2+ cytolytic T cells by suppressing or eliminating dendritic cells. *J Immunology*, 136(10):3567-3571, May 1986.
- [25] L. S. Pontryagin, V.G. Boltyanskii, R. V. Gamkrelidze, E. F. Mishchenko; *The Mathematical Theory of Optimal Processes*, Gordon and Breach, 1962.
- [26] A. Diefenbach, E.R. Jensen, A.M. Jamieson, and D. Raulet. Rae1 and H60 ligands of the NKG2D receptor stimulate tumor immunity. *Nature*, 413:165171, September 2001.
- [27] M.E. Dudley, J.R. Wunderlich, P.F. Robbins, J.C. Yang, P. Hwu, D.J.Schwartzentruber, S.L. Topalian, R. Sherry, N.P. Restifo, A.M. Hubicki, M.R.Robinson, M. Raffeld, P. Duray, C.A. Seipp, L. Rogers-Freezer, K.E. Morton, S.A. Mavroukakis, D.E. White, and S.A. Rosenberg. Cancer regression and autoimmunity in patients after clonal repopulation with antitumor
- [28] V. Kuznetsov, I. Makalkin, M. Taylor, and A. Perelson. Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis. *Bulletin of Mathematical Biology*, 56(2):295321, 1994.
- [29] A. Yates and R. Callard. Cell death and the maintenance of immunological memory. *Discret Contin Dyn S.*, 1(1):4359, 2002.
- [30] A. Lanzavecchia and F. Sallusto. Dynamics of T-lymphocyte responses: Intermediates, effectors, and memory cells. *Science*, 290:9297, 6 Oct 2000.
- [31] M.C. Perry, editor. *The Chemotherapy Source Book*. Lippincott Williams Wilkins, third edition, 2001.
- [32] P. Calabresi and P.S. Schein, editors. *Medical Oncology: Basic Principles and Clinical Management of Cancer*. McGraw-Hill, New York, second edition, 1993.
- [33] L. Bannock. Nutrition. Found at <http://www.doctorbannock.com/nutrition.html>.
- [34] B. Hauser. Blood tests. Technical report, International Waldenstroms
- [35] M. Kim, S. Perry, K. B. Woo; Quantitative approach to the design of antitumor drug dosage schedule via cell cycle kinetics and systems theory, *Ann. Biomed. Engng.*, (5) (1977) 1233.
- [36] G. W. Swan and T. L. Vincent; Optimal control analysis in the chemotherapy of IgG Multiple Myeloma, *Bull. of Math Bio*, (39) (1977) , 317337.
- [37] J. M. Murray, Some optimality control problems in cancer chemotherapy with a toxicity limit, *Math. Biosci.*, (100) (1990), 4967.
- [38] http://www.wcrf.org/cancer_statistics/world_cancer_statistics.php
- [39] H. J. Gogas, J. M. Kirkwood, V. K. Sondak; Chemotherapy for Metastatic Melanoma: Time for a Change, *Cancer*, (109) 3 (2007), 455464.
- [40] A. I. Riker, S. Radfar, S. Liu, Y. Wang, H. T Khong; Immunotherapy of melanoma: a critical review of current concepts and future strategies, *Expert Opinion on Biological Therapy*, (7) 3 (2007), 345-358).
- [41] G. W. Swan, Optimal control applications in biomedical engineering-a survey, *Opt. Control Appl. Meth.*, (2) (1981), 311314.
- [42] V. Kuznetsov and G. D. Knott; Modeling tumor regrowth and immunotherapy *Math and Comp. Modelling*, (33) (2001), 12751287.
- [43] L. G. de Pillis and A. E. Radunskaya; A mathematical tumor model with immune resistance and drug therapy: an optimal control approach, *Journal of Theoretical Medicine*, (3) (2001),79100.

- [44] L. G. de Pillis, A. E. Radunskaya, and C.L. Wiseman; A validated mathematical model of cell-mediated immune response to tumor growth, *Cancer Research*, (61)17 (2005), 79507958.
- [45] E. Rosenbaum and I. Rosenbaum. *Everyones Guide to Cancer Supportive Care: A Comprehensive Handbook for Patients and Their Families*. Andrews McMeel Publishing, 2005.
- [46] *Experience with the Use of High-Dose Interleukin-2 in the Treatment of 652 Cancer Patients*: Steven A. Rosenberg, M.D., PH. D., Michael T. Lotze, M.D., James C. Yang, M.D., Paul M. Aebersold, PH.D., W. Marston Linehan, M.D., Claudia A. Seipp, R.N., and Donald E. White, M.S.

Presentation Methods for Task Migration in Cloud Computing by Combination of Yu Router and Post-Copy

Elham Shamsinezhad¹, Asadollah Shahbahrami², Alireza Hedayati³, Ahmad Khadem Zadeh⁴ and Hamid Baniroostam⁵

¹ Department of Computer Engineering, Guilan Science and Research Branch, Islamic Azad University
Rasht, Guilan, Iran

² Department of Computer Engineering, Guilan University, Rasht, Guilan, Iran

³ Department of Computer Engineering, Central Tehran Branch, Islamic Azad University
Tehran, Tehran, Iran

⁴ Research Institute Information & Communication Technology
Tehran, Tehran, Iran

⁵ Department of Computer Engineering, Guilan Science and Research Branch, Islamic Azad University
Rasht, Guilan, Iran

Abstract

One of the techniques for increasing flexibility and scalability of cloud data centers is Task Migration. The act of migration is performed for different goals like balancing and load sharing, energy management and reducing response time and improvement of Quality of Service. In this paper, different kinds of migrations like Pre-Copy, Post-Copy, TPM triple migration, and related issues are studied. A new approach is being proposed by the combination of post-copy and Yu. First in proposed approach, sub activities have been sent according to Yu routing from source machine to the destination machine and all tasks in source machine are getting stopped and destination machine will continue those tasks. Results of simulation done by MATLAB software shows that proposed approach leads to overload reduction of server system. Finally, to demonstrate the performance of the proposed approach in terms of transforming time and overload will be assessed and compared with methods of Post-Copy and Yu.

Keywords: Migration, Processor, Task, Virtual Machine, Virtualization

1. Introduction

Cloud computing can be considered as the ability of sharing computational resources among a large number of various users. In fact, this is an infrastructure running codes in management model and development capacity. Here management has been defined as allocating reliability due to predefined quality parameters and the elasticity resources in accordance with customers' requirements.

Cloud computing clients do not have real cloud physical infrastructure and they can use cloud resources and their infrastructures only with paying subscription fee to the service providers.

The main body of cloud computing is made up of virtualization technology [1]. Using this technology, physical server machines transform into several virtual machines that each of them meet the needs of multiple users [2]. Using virtual machines inside data centers is not enough to answer the customers' requirements and it is not possible to remain in competition market just with applying this tool. So this is necessary to have some tools in order to displace and determine the best place for every virtual machine. Generally migration from source processors to destination processors increases efficiency in system [3]. Migration algorithms can have a great impact on system performance. Task Migration time from source sub network to destination sub network and task delay time are of the most important criteria of efficiency assessment.

In part 2, the concepts of different migration and migration task are presented. Part 3 is allocated to related works in service migration area and part 4 describes the proposed approach. In fact the proposed approach has been before implemented in mesh networks but in this paper a new approach is proposed through the combination of post-copy and Yu leading to improve the system overhead. Moreover, the performance of the proposed approach will be compared to the other existing methods. Finally in part

5, results of the proposed approach presents and results of the simulation will be expressed.

2. Task Migration

To a better understanding of the proposed approach, the existing literature in task migration area is briefly studied in this part of the paper. Then the way different kinds of task migration methods perform will be described.

Migration of virtual machines: Using virtualization technology today most of data centers increase their system efficiency via increasing the number of customers. This technology provides a virtual environment and factually presents the physical machine with all its modes and data in form of a logical file. Therefore, virtual machines can be transmitted from one place to another. This transmission is called virtual machine migration. One of widely used environments of this technique is data centers which in load and requests amount are dynamically changing.

Different types of task migration are classifies like this:

- Live Migration: In this method, source machine receives users' request without delaying (without blackout Machine) and transmits them to destination machine so that services are provided even during migration process which is one of the advantages of this method.
- Non-live Migration: First virtual machine stops completely in the source and then all processor modes and memory pages are transmitted to destination machine. After receiving all modes and pages from the last mode of machine stored before transmission, destination machine starts up.
- Task Migration in Mesh Networks: In mesh networks, each node is a computer that can include processor, memory, and communicational channels to be in communication with its other neighbor nodes. When a task enters this system, all processors will be allocated to that. Figure 1 shows task migration from one sub network to another sub network. In 3D Mesh network system $M(D,W,H)$ encompasses $N=D*W*H$ nodes distributed on the network. Each processor in this system has some numbers. For instance, a node with the number of $M(D,W,H)$ includes $\{(0,0,0), \dots, (D-1, W-1, H-1)\}$ processors. $SM(d,w,h)$ mesh sub networks consider M as $\{(x1,y1,z1), \dots, (x2,y2,z2)\}$ which in $(x1,y1,z1)$ is its coordinates of the bottom left corner and $(x2,y2,z2)$ represents its coordinates of top right corner. In this case the equations of $h=z2-z1$, $w=y2-y1$, $d=x2-x1$ are dominant there. It is assumed in this paper that Worm Hole Switching method and the next regular routing method such as XY routing have been used. In this method, every node is able to receive and send a message simultaneously [4].

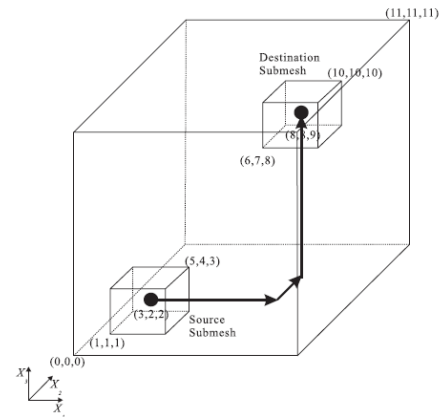


Fig. 1 Task migration from one sub network to another

3. Related Work

Task Migration in multiprocessor networks such as cloud computing are similar to processor migration in distributed systems. Processor migration has been implemented in multiple systems. Also various algorithms have been presented for task migration. These algorithms are totally made up of migration beginning stages, stop running task in source processor, task transmission to destination processor, start running task in destination, task running in destination processor, and removing remained information from source processor. Continuing this part, types of presented migration algorithms will be studied [5].

- Pre Copy: In this algorithm, running processor will not be stopped during its transmission from source machine. pre copy algorithm also has three stages as mentioned bellow [6 and 7].
 - First stage: Repetition of sending data relevant to the virtual machine like memory pages and various kinds of processor modes by source machine to destination one.
 - Second stage: In this stage, the virtual machine goes to suspension mode and then other pages changed during transmission in the first stage and also processor modes necessary to start up destination machine will be sent.
 - Third stage: Destination machine starts up just from the last mode stored in the source before suspension. The point in this algorithm is that the less changes in pages in the first stage, the less time in the second stage the service is inaccessible.
- Post copy algorithm: First in this method all processor modes and the least needed information are sent to destination machine which results in destination machines start up. Then sending memory pages from source to destination will be started. This method

assures that each page will be sent to destination machine only for once so that overhead out of one page's multiple sending found in pre-copy method will be omitted [8 and 9].

- Three Phase Migration (TPM): TPM algorithm has the least suspension time for transmission of the whole system with its all modes. In fact TPM technique is the same with disc pre-copy technique which is also able to displace virtual machine's disc [10].
- CR/TR Motion Algorithm: This algorithm is used for operating synchronization based on processing power in destination machine. It means that log will be sent instead of sending pages. Therefore, sent data amount during two machines' synchronization will be considerably decreased [11].
- Migration algorithm dependence of ware: this algorithm is used when a virtual machine has no interaction with the world outside so that it can transmit a part of data to destination machine just with one time sending. There is a part in this algorithm called administrator which the decision of one or several time displacing processes should be done is made based on its information.

4. The Propose Approach

Presented approach is a combination of Yu routing and post-copy migration. Factually, the results of assessments in the next parts of this paper show that the proposed approach has the advantages of both former methods. Performance of the proposed approach: Using Yu method, the proposed approach first provides the possibility of migration of several sub tasks in parallel through different paths. Consequently there will be maximum parallel and simultaneous transmission of tasks so transmission time of sub tasks and their migration from one virtual machine to another will decrease. Also using method of post-copy reduces the amount of overhead out of repetition of applied changes on memory pages which results in removal of sending data from source machine to destination. In the proposed approach, each computer is considered as a node that can include processor, memory, and communicational channel. With the entrance of any new task, a sub network of nodes is getting allocated to that at the same time routers manages.

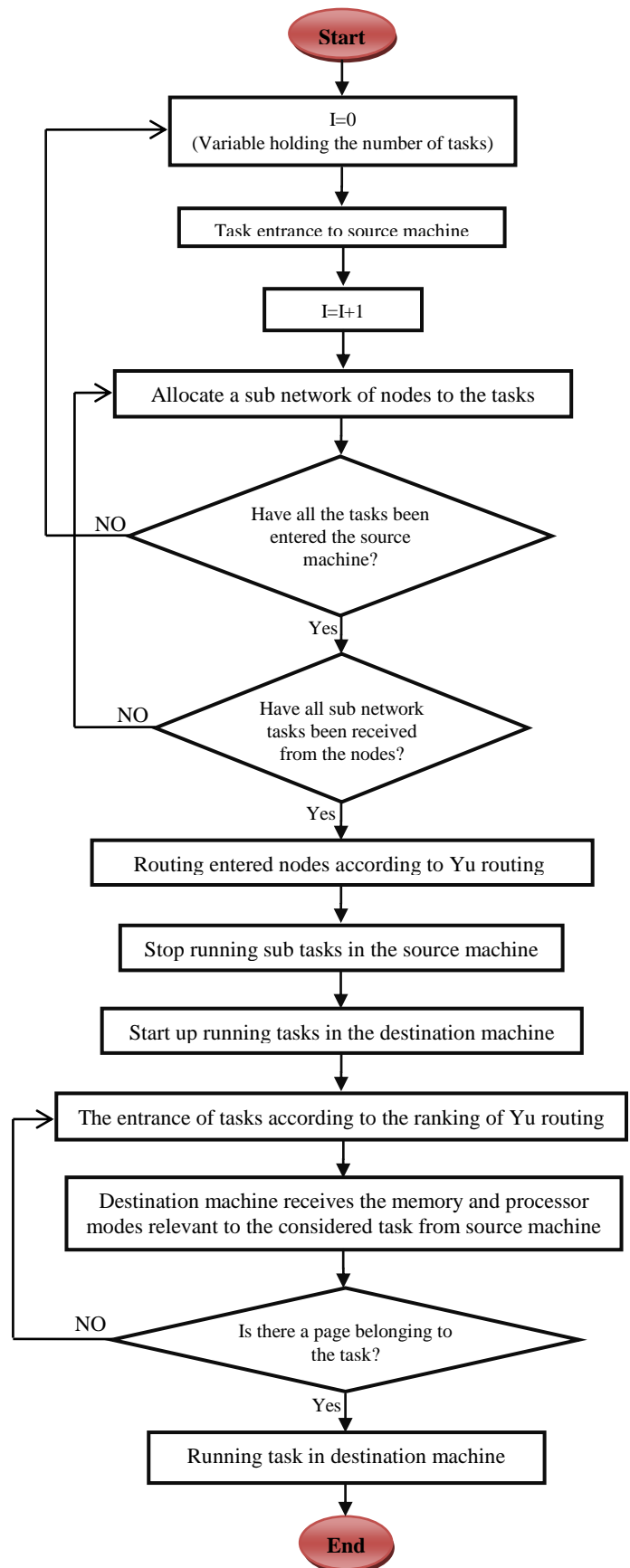


Fig .2 The proposed method flowchart

4-1. The Algorithm of Proposed Approach

Stage1: First sub tasks running in different nodes of source machine are sent to destination sub network based on Yu routing algorithm.

Stage2: According to post-copy algorithm, processor modes will be sent to destination machine. This process causes destination machine start up.

Stage3: Memory pages will be sent from source machine to destination.

Stage4: If one memory page was not found, there will be a delay in destination and that page will be requested from source machine.

Stage5: This stage is the end of running transmitted sub task from source machine to destination machine.

4-2. Comparing Performance of Migration Algorithms

To compare the performance of the proposed approach with the other mentioned Approach, the proposed Approach performance is studied over sent data amount, overhead amount, migration time, and quality of efficiency in this part in table 1.

Table 1: Comparing migration algorithm with the proposed method

Approach	Pre Copy	Post Copy	Three Phase CR/TR Motion Dependence-Aware	Proposed Approach
Sent Data Amount	H	L	H	M
Overhead Amount	L	L	M	L
Migration Time	L	L	L	S
Efficiency	L	L	L	M

Where in table 1, letters H, M, L and S sequentially are stand for High, Medium, Low and Short.

Comparing the proposed approach to the other approaches shows that using the proposed approach reduces sent data amount, overhead amount, and also migration time. Therefore, considering the results of simulation in MATLAB environment, system will have higher efficiency using the proposed approach in accordance to the other ones in table 1.

4-3. Results of Proposed Approach Simulation

Considering weaknesses of past presented methods, to analyze services migration in the proposed approach and also to increase efficiency, three factors bellow have been

considered: Transmission time, Network Traffic Overhead and Services Stop Time.

Results of this experiment done in four real environments of Bit Torrent, SPECWeb 2005, Kernel Compile and NetPerf have been demonstrated in figure 3. Vertical axis shows time measured in second. Actually it shows the whole migration time in the interval of time from the beginning of migration operation till the modes and information in both source and destination machines are fully synchronized. This time decreases because of reduction in transmitted data amount (there is no repetition in post-copy method). In figure 4, vertical axis represents time measured in millisecond shows the very little amount of sent data in the proposed approach. The amount of sent data is low because of no repetition in sending data during migration so the amount of sent data will be low and also data overhead reaches zero. But stop time is still high which can express the weakness of the proposed approach.

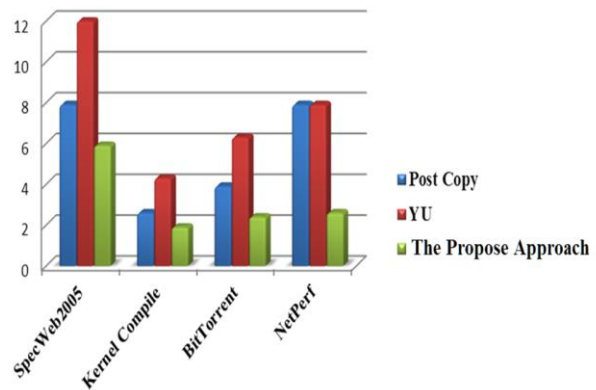


Fig. 3 Tasks migration time from source machine to destination machine

In the proposed approach which in combined algorithm has been also used, sub networks are divided into networks with the size of $p \times q \times r$. In this case, the number of partitions is to be calculated by equation (1).

$$\left\lceil \frac{d}{p} \right\rceil \times \left\lceil \frac{w}{p} \right\rceil \times \left\lceil \frac{h}{r} \right\rceil \quad (1)$$

Consequently in combined method, collected sub tasks migrate to the considered destination machine which in this phase the number of stages follows equation (2).

$$\text{Max}\left(\left\lceil \frac{d}{p} \right\rceil \times \left\lceil \frac{w}{p} \right\rceil \times \left\lceil \frac{h}{r} \right\rceil\right) \quad (2)$$

Considering the results of figures 1 and 2, the proposed approach can largely solves the problem of transmission time and sent data amount.

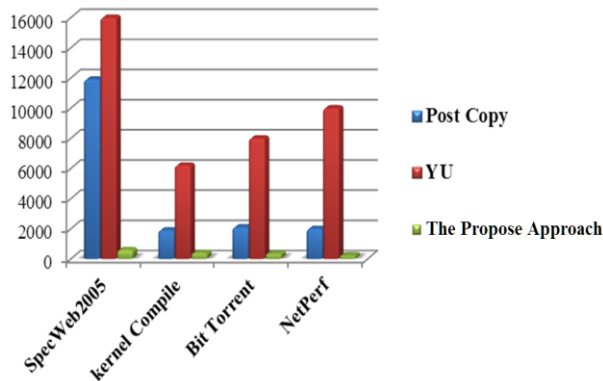


Fig. 4 Sent data amount

5. Conclusion

In this paper migration methods of virtual machines were studied. Also a new approach for task migration in cloud computing was proposed. Via the combination of post-copy algorithm in cloud computing and Yu algorithm in multi processor systems, the proposed approach can have the advantages of both mentioned methods. The proposed approach makes parallel and simultaneous task transmission possible through Yu algorithm which leads to the reduction of task transmission time. Furthermore, considering the results of simulation in MATLAB software, this can be observed that using post-copy method makes data overhead decrease but yet there will be a delay in providing service. Generally this can be concluded that in case of not using migration techniques, there will be unfavorable consequences such as increasing costs, unfavorable quality of service and consequently losing customers of cloud computing providers.

References

- [1] T.Velte and A.Velte and R.Elsenpeter, "Cloud Computing, A Practical Approach, New York", USA, McGraw-Hill, 2010.
- [2] B.P.Rimal and E.Choi and I.Lumb, "A Taxonomy and Survey of Cloud Computing Systems", in Fifth International Joint Conference on INC, IMS and IDC(NCM '09), Aug., 2009, pp. 25-27.
- [3] C.P. Sapuntzakis, R. Chandra and B. Pfaff, J. Chow, M.S. Lam and M. Rosenblum, "Optimization the Migratin of Virtual Computers", In Proceeding of 5th USENIX Symposium on Operating Systems Design and Implementation (OSDI-02), Dec., 2002.
- [4] G. Jong Yu, C. Yung Chan and T. Shi Chen, "Task Migration in Nd Wormhole-Route Mesh Multi computer" ,Journal of Systelns Architecture, Vol. 50, No 4, 2004, pp.177-192.
- [5] K. Sammy, R. Shengbing and Ch. Wilson, "Energy Efficient Security Preserving VM Live Migration In Data Centers For Cloud Computing", International Journal of Computer Science Issues (IJCSI), Vol. 9, Issue 2, No 3, March 2012, pp.33-39.
- [6] C. Clark and K. Fraser and S. Hand and J.G. Hansen and E. Jul, C. Limpach and I. Pratt and A. Warfield, "Live Migration of Virtual Machines", in Proc. of the Second USENIX Symposium on Networked Systems Design and Implementation (NSDI),Boston, MA, USA, May 2005.
- [7] D. Cheriton, "The V distributed system",Communication of the ACM,31 (3), 1988, pp.314-333.
- [8] J. Fejzaj1, I. Tafa and E. Kajo, "The improvement of Live Migration in Data Center in different virtual environment", International Journal of Computer Science Issues (IJCSI), Vol. 9, Issue 6, No 2, November 2012, pp.460- 463.
- [9] M. Richmond and M. Hitchens. "A new process migration algorithm", ACM SIGOPS Operating Systems Review, 31(1), 1997, pp.31 -42.
- [10] Y. Luo, B. Zhang, X. Wang, Z. Wang, and Y. Sun, "Live and incremental whole-system migration of virtual machines using block-bitmap", in Proceedings of Cluster 2008: IEEE International Conference on Cluster Computing. IEEE Computer Society, 2008, pp. 99-106.
- [11] H. Liu, H. Jin, X. Liao, L. Hu and C. Yu, "Live migration of virtual machine based on full system trace and replay", in Proceedings of the 18th International Symposium on High Performance Distributed Computing (HPDC'09), 2009, pp. 101-110.
- [12] A. Nocentino and P. Ruth, "Toward Dependency-Aware Live Virtual Machine Migration", in Proceedings of the 3rd International Workshop on Virtualization Technology in Distributed Computing (VTDC 2009), Barcelona, Spain, Jun. 2009, pp. 59-66.

Assessment of offline Digital Signature Recognition Classification Techniques

Dina Darwish

Assistant Professor, International Academy for Engineering and Media Science – Egypt
6th October city, Egypt

Abstract

The digital signature verification has become an interesting domain, which is widely needed. The usage of online and offline digital signatures has been spreaded worldwide due to the increase of use of bank transactions and user authentication and other similar activities. This requires the creation and the diversification of new online and offline signature verification methods. The signature verification methods contain both online (or dynamic) and offline (or static) signature verification methods. In this paper, an offline digital signature verification technique is proposed, that depends on extracting several features from the signatures to be used during simulation. Some signatures were used for training and others were used for testing only. Different methods such as, vectors manipulation, ensemble classification using boosted trees, and bagged trees, were used in this paper during simulation to obtain results.

Keywords: *Signature Verification, Offline Digital Signature, Features Extraction, Vectors Manipulation, Ensemble Classification, Bagged Trees.*

1. Introduction

The growth in today's online and offline transactions that includes banking transactions has posed the question of how to make secure online and offline signature verification techniques, to eliminate the possibility of personal information theft. There are a different number of personnel characteristics that can be used to identify each person, such as, voice, lip movements, hand geometry, face, iris, retina, fingerprint, and others. These characteristics are called biometrics, and these biometrics can be used to distinguish between one person and another. But the most commonly used biometric nowadays in e-commerce and banking activities is the signature recognition.

The signature can be defined as follows; "the name of a person written with his or her own hand; or the act of signing one's name", according to the American Heritage Dictionary. There is a second definition of signature, which is related to the whole process of signing, it means, that the way the signature is made and the characteristics of the signature, including velocity, pen pressure, stroke, and others are unique to every person.

The first definition, is close to the definition of offline signature, which treats the signature as a two-dimensional image with static characteristics, that does not contain any time-related information. The second definition, is close to the definition of online signature, and is based on dynamic characteristics of the process of signing, such as velocity, pen pressure and others.

The signature verification is a typical pattern recognition task. But both types of signatures; online or offline; use different techniques to verify signatures based on either static or dynamic characteristics.

The task of signature verification includes extracting some characteristics from the recorded information of the signature, and further, comparing them with the characteristics of the reference signature. Let us make a brief survey on different signature recognition and verification techniques used.

Various methods have been implemented for creating features from the signature image, which can be grouped into two main categories: direct methods and transform methods. Direct methods allow generating features directly from image pixels such as grid-based information, pixel density, gray-level intensity, texture... etc. In contrast, transform methods need a transformation of the image into another domain in which features could be created. Fourier, Wavelet, Radon transforms are the most popular methods for creating features [1][5]. Hence, another transform has been proposed namely the contourlet transform (CT) [6].

The main advantage of the CT is the ability to capture significant information about an object.

Furthermore, it offers a flexible multiresolution, local and directional image expansion. These properties are interesting to exploit more specifically for the handwritten signature verification since the signature contains often special characters and flourishes [7].

In [2], this paper describes a method for verification of signatures after extraction of features based on clustering techniques. Clustering involves dividing a set of data points into non-overlapping groups, or clusters, of

points, where points in a cluster are “more similar” to one another than to points in other clusters. In [3], this paper, two methods are proposed to track the variations in signatures. Given the set of training signature samples, the 1st method measures the positional variations of the one-dimensional projection profiles of the signature patterns; and the second method determines the variations in relative stroke positions in the two-dimension signature patterns. In [4], this paper evaluates the performance of an Error Back Propagation (EBP) Artificial Neural Network (ANN) for authenticating the signatures. The work done has provided encouraging results and has re-confirmed the ability of Artificial Neural Networks to recognize patterns and in this case their skill to generalize. An efficient Static Signature Verification (SSV) system that consists of rigorous preprocessing and feature extraction followed by a classifier is used.

In [8], a paper presents a method for verifying handwritten signatures by using NN architecture. Various static (e.g., area covered, number of elements, height, slant, etc.) and dynamic (e.g., velocity, pen tip pressure, etc.) signature features are extracted and used to train the NN. In [9], a paper is primarily focused on skilled forgery detection. It emphasizes on the extraction of the critical regions which are more prone to mistakes and matches them following a modular graph matching approach.

In section 2, we described the features extracted to be used for signature recognition. In section 3, the signature recognition classification techniques were described. In sections 4, the simulation results were discussed and analyzed. In section 5, the conclusion is presented. And finally, the references are cited.

2. Features Extraction for signature recognition

The features needed to be extracted to identify signature are [10]:

1 - The new curve of the signature after rotating the original curve of the signature points around the center x and y coordinates of the original signature curve based on making the original signature curve rotate around its center x and y coordinates, to make the new signature curve that will be used in pattern recognition.

2 - The number of pixels in the signature based on calculating the total number of pixels of the signature.

3 - The occupancy Ratio of the signature to the whole image which is described as :

Occupancy ratio = total number of pixels of the signature/total number of pixels of the signature image * 100

4 - The minimum Eigen value of the signature curve

Where the eigen values of a matrix A are obtained from the solution of the characteristic equation:

$$\det(A - \lambda I) = 0 . \quad (1)$$

where det is the determinant of the matrix (A - λI) and I is the $n \times n$ identity matrix, λ is the eigen value

5 - The maximum height of the signature is based on the following:

Maximum height of the signature = maximum x coordinate of signature – minimum x coordinate of signature

6 - The maximum width of the signature is based on the following

Maximum width of the signature = maximum y coordinate of signature – minimum y coordinate of the signature

7 - The Euclidean distance between every two consecutive points in the signature curve

8 - The angle between every two consecutive points in the signature curve

9 - The height to width ratio of the signature

3. Assessment of Signature Recognition classification techniques

Three different signature recognition classification techniques were used to recognize signatures. These techniques were:

- 1) Vectors manipulation
- 2) Ensemble classification using boosted trees
- 3) Tree Classification using bagged trees.

Assessment of these techniques is based on simulation in which we used 500 signatures for 100 persons, each person has 5 signatures. We used 60% of the signatures for training, and the other 40% were used for testing. We used MATLAB 2011 during simulation.

3.1 Vectors Manipulation Technique

Vectors manipulation is based on finding the differences between each vector to be tested and each pattern vector used to identify one signature. Each tested vector is compared with the 300 reference patterns representing the 100 persons, the smallest difference between any tested vector and reference pattern vector, means that the tested vector belongs to the N person having this reference pattern vector.

For each vector the following rules applies,

```

For i=1:vector length
  For j=1:300
     $D_{ij} = abs ( VP_{ij} - VT_i )$ 
  End
End
    
```

Where, *VP* is the vector reference pattern
VT is the tested pattern

We calculate the sum of each vector in the difference matrix

$$S_j = sum(D_{ij})$$

Where, *i* represent rows from 1 to vector length and *j* represent columns from 1 to 300

S_j is a vector containing the sum of each column

Then,

$$Find \min(S_j)$$

Where, *j* represents the column of the person *y* for example

By this way, we classify all the 500 patterns, either being only test patterns or patterns used as references by finding the least difference between any test pattern and any reference pattern.

Figure (1) shows the proposed simulation for assessment of vector manipulation technique.

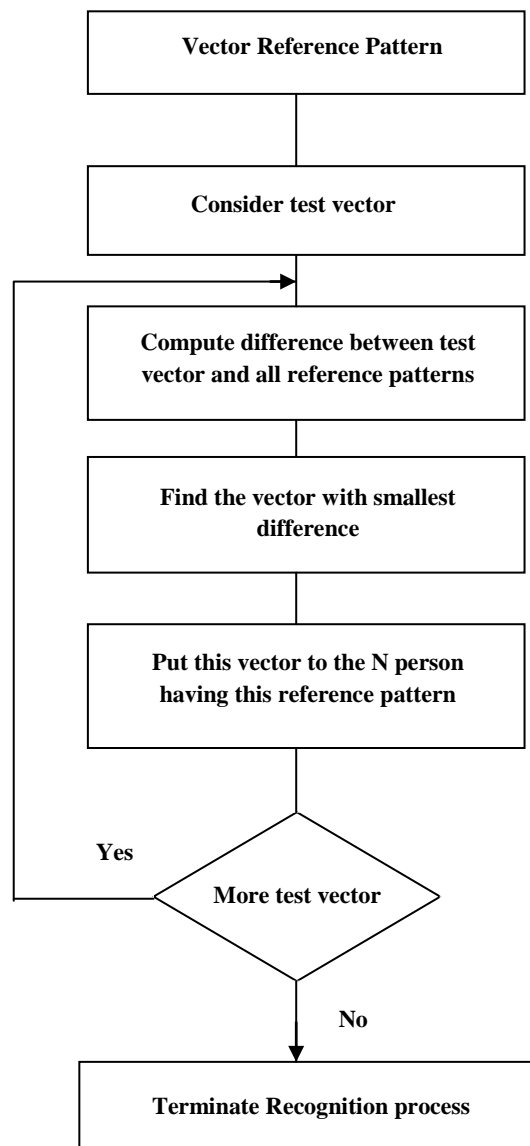


Fig.1 Simulation chart of vector manipulation technique

3.2 Ensemble classification using boosted trees

(A) Common types of ensembles

1. Bayes optimal classifier

The Bayes Optimal Classifier is an optimal classification technique. It is an ensemble of all the hypotheses in the hypothesis space. On average, no other ensemble can outperform it, so it is the ideal ensemble. Each hypothesis is given a vote proportional to the likelihood that the training dataset would be sampled from a system

if that hypothesis were true. To facilitate training data of finite size, the vote of each hypothesis is also multiplied by the prior probability of that hypothesis. The Bayes Optimal Classifier can be expressed with following equation [11]:

$$y = \operatorname{argmax}_{c_j \in C} \sum_{h_i \in H} P(c_j|h_i)P(T|h_i)P(h_i) \quad (2)$$

where y is the predicted class, C is the set of all possible classes, H is the hypothesis space, P refers to a probability, and T is the training data. As an ensemble, the Bayes Optimal Classifier represents a hypothesis that is not necessarily in H . The hypothesis represented by the Bayes Optimal Classifier, however, is the optimal hypothesis in *ensemble space* (the space of all possible ensembles consisting only of hypotheses in H).

2. Bootstrap aggregating (bagging)

Bootstrap aggregating, often abbreviated as *bagging*, involves having each model in the ensemble vote with equal weight. In order to promote model variance, bagging trains each model in the ensemble using a randomly drawn subset of the training set. As an example, the random forest algorithm combines random decision trees with bagging to achieve very high classification accuracy.

3. Boosting

Boosting involves incrementally building an ensemble by training each new model instance to emphasize the training instances that previous models mis-classified. In some cases, boosting has been shown to yield better accuracy than bagging, but it also tends to be more likely to over-fit the training data. By far, the most common implementation of Boosting is Adaboost, although some newer algorithms are reported to achieve better results.

4. Bucket of models

A "bucket of models" is an ensemble in which a model selection algorithm is used to choose the best model for each problem. When tested with only one problem, a bucket of models can produce no better results than the best model in the set, but when evaluated across many problems, it will typically produce much better results, on average, than any model in the set.

The most common approach used for model-selection is cross-validation selection. It is described with the following pseudo-code [11]:

For each model m in the bucket:

Do c times: (where 'c' is some constant)

Randomly divide the training dataset into two datasets: A, and B.

Train m with A

Test m with B

Select the model that obtains the highest average score

Cross-Validation Selection can be summed up as: "try them all with the training set, and pick the one that works best".

5. Stacking

The crucial prior belief underlying the scientific method is that one can judge among a set of models by comparing them on data that was not used to create any of them. This same prior belief underlies the use in machine learning of *bake-off contests* to judge which of a set of competitor learning algorithms is actually the best fit in selected domains.

This prior belief can also be used by a single practitioner, to choose among a set of models based on a single data set. This is done by partitioning the data set into a *held-in* data set and a *held-out* data set; training the models on the held-in data; and then choosing whichever of those trained models performs best on the held-out data. This is the cross-validation technique, mentioned above.

Stacking (sometimes called *stacked generalization*) exploits this prior belief further. It does this by using performance on the held-out data to combine the models rather than choose among them, thereby typically getting performance better than any single one of the trained models. It has been successfully used on both supervised learning tasks (regression) and unsupervised learning (density estimation). It has also been used to estimate Bagging's error rate.

Because the prior belief concerning held-out data is so powerful, stacking often out-performs Bayesian model-averaging. Indeed, renamed *blending*, stacking was extensively used in the two top performers in the recent Netflix competition.

(B) Simulation of Ensemble Classification Technique

MATLAB 2011 is used to simulate the ensemble technique for which figure (2) shows the information necessary to create an ensemble [12].

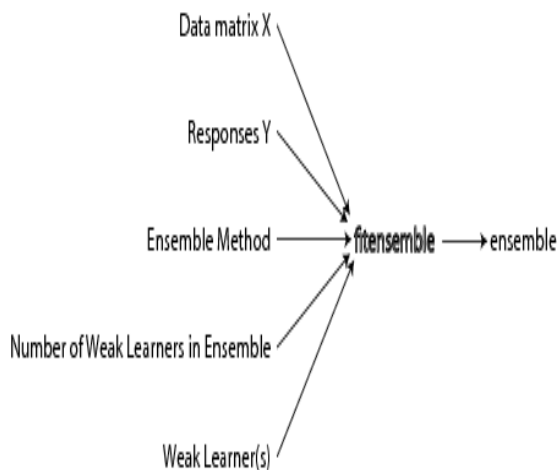


Fig. 2 Elements of the simulation technique

For all classification or nonlinear regression problems, follow these steps to create an ensemble [12]:

1. Put Predictor Data in a Matrix
2. Prepare the Response Data
3. Choose an Applicable Ensemble Method
4. Set the Number of Ensemble Members
5. Prepare the Weak Learners
6. Call fitensemble

Ensemble Algorithms

- AdaBoostM1
- AdaBoostM2
- Bag
- GentleBoost
- LogitBoost
- LPBoost
- LSBoost
- RobustBoost
- RUSBoost
- Subspace
- TotalBoost
- AdaBoostM1

Where, LSBoost is the used algorithm, and is described as follows:

LSBoost (least squares boosting) fits regression ensembles. At every step, the ensemble fits a new learner to the difference between the observed response and the aggregated prediction of all learners grown previously. The ensemble fits to minimize mean-squared error.

You can use LSBoost with shrinkage by passing in the LearnRate parameter. By default this parameter is set to 1, and the ensemble learns at the maximal speed. If you set LearnRate to a value from 0 to 1, the ensemble fits every new learner to $y_n - \eta f(x_n)$, where

y_n is the observed response.

$f(x_n)$ is the aggregated prediction from all weak learners grown so far for observation x_n .

η is the learning rate. Figure (3) shows the proposed simulation for Ensemble classification technique.

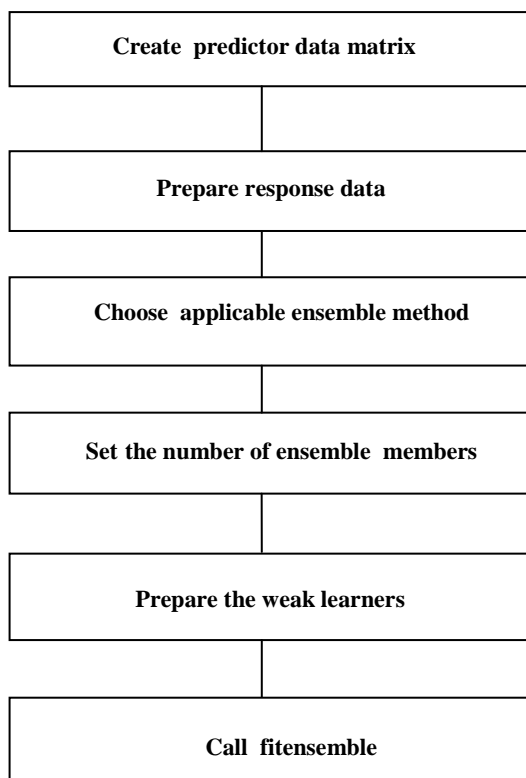


Fig. 3 Simulation chart for ensemble classification technique

3.3 Tree classification using bagged trees

Classification trees and regression trees [14] predict responses to data. To predict a response, follow the decisions in the tree from the root (beginning) node down to a leaf node. The leaf node contains the response. Classification trees give responses that are nominal, such as 'true' or 'false'. Regression trees give numeric responses.

Each step in a prediction involves checking the value of one predictor (variable). Figure (4) is a simple classification tree:

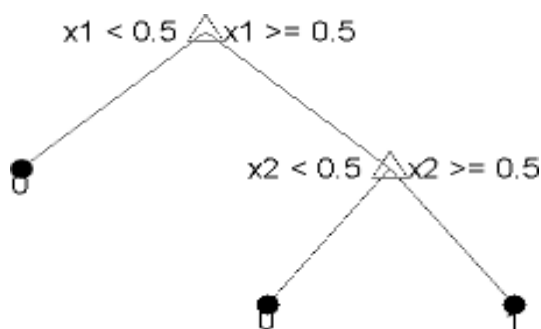


Fig. 4 Classification tree

This tree predicts classifications based on two predictors, x_1 and x_2 . To predict, start at the top node, represented by a triangle (Δ). The first decision is whether x_1 is smaller than 0.5. If so, follow the left branch, and see that the tree classifies the data as type 0.

If, however, x_1 exceeds 0.5, then follow the right branch to the lower-right triangle node. Here the tree asks if x_2 is smaller than 0.5. If so, then follow the left branch to see that the tree classifies the data as type 0. If not, then follow the right branch to see that the tree classifies the data as type 1.

The classification tree and the regression tree methods perform the following steps to create decision trees:

1. Start with all input data, and examine all possible binary splits on every predictor.
2. Select a split with best optimization criterion.
3. If the split leads to a child node having too few observations (less than the minimum leaf parameter), select a split with the best optimization criterion subject to the minimum leaf constraint.
 - ✓ Impose the split.
 - ✓ Repeat recursively for the two child nodes.

The explanation requires two more items: description of the optimization criterion, and stopping rule.

Stopping rule: Stop splitting when any of the following hold:

- The node is *pure*.
 - For classification, a node is pure if it contains only observations of one class.
 - For regression, a node is pure if the mean squared error (MSE) for the observed response in this node drops below the MSE for the observed response in the entire data multiplied by the tolerance on quadratic error per node (qetoler parameter).
- There are fewer than minimum parent observations in this node.
- Any split imposed on this node would produce children with fewer than minimum leaf observations.

Optimization criterion:

- Regression: mean-squared error (MSE). Choose a split to minimize the MSE of predictions compared to the training data.
- Classification: One of three measures, depending on the setting of the split criterion name-value pair provided in MATLAB 2011:
 - 'gdi' (Gini's diversity index, the default)
 - 'twoing'
 - 'deviance'

For a continuous predictor, a tree can split halfway between any two adjacent unique values found for this predictor. For a categorical predictor with L levels, a classification tree needs to consider $2L-1$ splits. To obtain this formula, observe that you can assign L distinct values to the left and right nodes in $2L$ ways. Two out of these $2L$ configurations would leave either left or right node empty, and therefore should be discarded. Now divide by 2 because left and right can be swapped. A classification tree can thus process only categorical predictors with a moderate number of levels. A regression tree employs a computational shortcut: it sorts the levels by the observed mean response, and considers only the $L-1$ splits between the sorted levels.

The classification tree splits nodes based on either *impurity* or *node error*. Impurity means one of several things, depending on your choice of the split criterion name-value pair in MATLAB 2011:

- Gini's Diversity Index (gdi) —The Gini index of a node is [14]

$$1 - \sum_i p^2(i), \tag{3}$$

where the sum is over the classes i at the node, and $p(i)$ is the observed fraction of classes with class i that reach the node. A node with just one class (a *pure* node) has Gini index 0; otherwise the Gini index is positive. So the Gini index is a measure of node impurity.

- Deviance ('deviance') —With $p(i)$ defined as for the Gini index, the deviance of a node is [14]

$$-\sum_i p(i) \log p(i). \tag{4}$$

A pure node has deviance 0; otherwise, the deviance is positive.

- Twoing rule ('twoing') —Twoing is not a purity measure of a node, but is a different measure for deciding how to split a node. Let $L(i)$ denote the fraction of members of class i in the left child node after a split, and $R(i)$ denote the fraction of members of class i in the right child node after a split. Choose the split criterion to maximize

$$P(L)P(R) \left(\sum_i |L(i) - R(i)| \right)^2, \tag{5}$$

where $P(L)$ and $P(R)$ are the fractions of observations that split to the left and right respectively. If the expression is large, the split made each child node purer. Similarly, if the expression is small, the split made each child node similar to each other, and hence similar to the parent node, and so the split did not increase node purity.

- Node error — The node error is the fraction of misclassified classes at a node. If j is the class with largest number of training samples at a node, the node error is

$$1 - p(j). \tag{6}$$

Figure (5) shows the proposed simulation for tree classification technique.

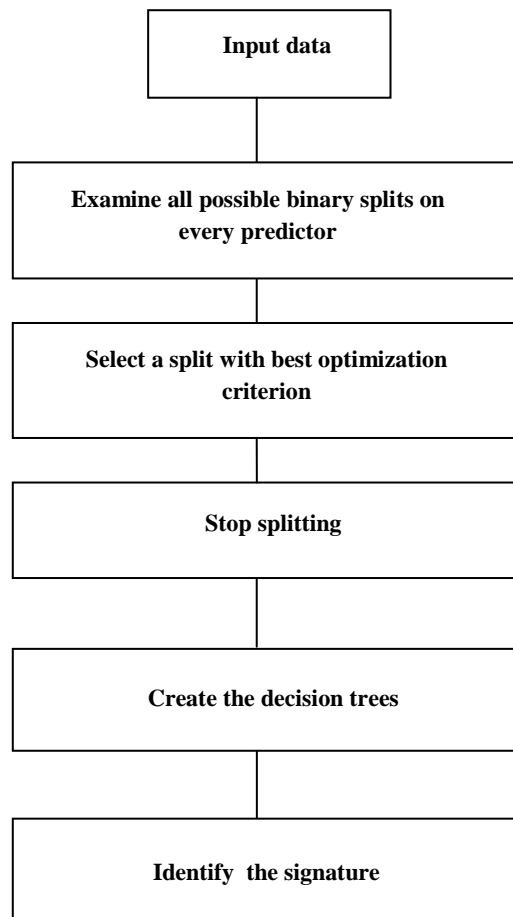


Fig. 5 Simulation chart for tree classification technique

Bagged Decision Trees

Bagging [13], which stands for "bootstrap aggregation," is a type of ensemble learning. To bag a weak learner such as a decision tree on a dataset, generate many bootstrap replicas of this dataset and grow decision trees on these replicas. Obtain each bootstrap replica by randomly selecting N observations

out of N with replacement, where N is the dataset size. To find the predicted response of a trained ensemble, take an average over predictions from individual trees.

Bagging works by training learners on resampled versions of the data. This resampling is usually done by bootstrapping observations, that is, selecting N out of N observations with replacement for every new learner. In addition, every tree in the ensemble can randomly select predictors for decision splits—a technique known to improve the accuracy of bagged trees.

By default, the minimal leaf sizes for bagged trees are set to 1 for classification and 5 for regression. Trees grown with the default leaf size are usually very deep. These settings are close to optimal for the predictive power of an ensemble. Often you can grow trees with larger leaves without losing predictive power. Doing so reduces training and prediction time, as well as memory usage for the trained ensemble.

Another important parameter is the number of predictors selected at random for every decision split. This random selection is made for every split, and every deep tree involves many splits. By default, this parameter is set to a square root of the number of predictors for classification, and one third of predictors for regression.

Several features of bagged decision trees make them a unique algorithm. Drawing N out of N observations with replacement omits on average 37% of observations for each decision tree. These are "out-of-bag" observations. You can use them to estimate the predictive power and feature importance. For each observation, you can estimate the out-of-bag prediction by averaging over predictions from all trees in the ensemble for which this observation is out of bag. You can then compare the computed prediction against the observed response for this observation. By comparing the out-of-bag predicted responses against the observed responses for all observations used for training, you can estimate the average out-of-bag error. This out-of-bag average is an unbiased estimator of the true ensemble error. You can also obtain out-of-bag estimates of feature importance by randomly permuting out-of-bag data across one variable or column at a time and estimating the increase in the out-of-bag error due to this permutation. The larger the increase, the more important the feature. Thus, you need not supply test data for bagged ensembles because you obtain reliable estimates of the predictive power and feature importance in the process of training, which is an attractive feature of bagging.

Another attractive feature of bagged decision trees is the proximity matrix. Every time two observations land on the same leaf of a tree, their proximity increases by 1. For normalization, sum these proximities over all trees in the ensemble and divide by the number of trees. The resulting matrix is symmetric with diagonal elements equal to 1 and off-diagonal elements ranging from 0 to 1. You can use this matrix for finding outlier observations and discovering clusters in the data through multidimensional scaling.

4. Simulation Results

500 patterns were used for simulation, representing 500 signatures for 100 persons, each person having 5 signatures. We used 60% of the patterns for training and the rest 40% for testing. We tested all the 500 patterns, which are all vectors of the same size. Each vector representing a signature. The results obtained are summarized in table (1).

Table 1: Simulation results

Simulation environment	Percentage of correctly classified signatures
Vectors Manipulation	77.4%
Ensemble classification using boosted trees	61.2%
Tree classification using Bagged trees	79.8%

Figure (6) shows the percentage of correctly classified signature for the three techniques. From table (1) and figure (6), we noticed that the percentage of correctly classified signatures using the ensemble classification using boosted trees is 61.2%, which represents the lowest percentage of correctly classified patterns among the three simulation environments. Using the bagged trees classification, 79.8% of the signatures were correctly classified, which represents the highest percentage of correctly classified patterns among the three simulation environments. Using the vectors manipulation, the percentage of the correctly classified patterns was 77.4%, which is closest to the percentage of the bagged trees classification.

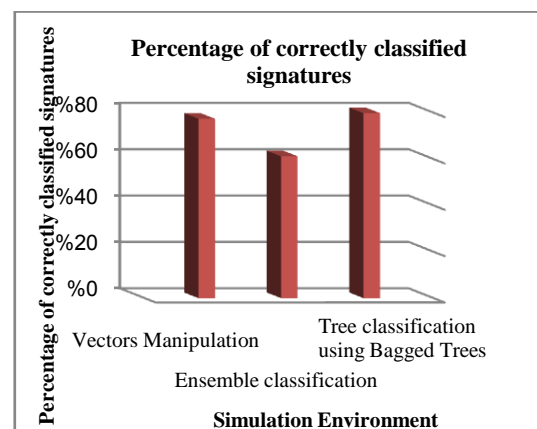


Fig. 6 Percentage of correctly classified signature

5. Conclusions

From the implemented simulation, the tree classification using bagged trees showed the best performance with signature recognition ratio of 79.8%, then, the vectors manipulation technique follows it with a signature recognition ratio of 77.4%, then, comes the ensemble classification using boosted trees with a signature recognition ratio of 61.2%, which is the least recognition ratio among the three simulation environments. ^b

References

- [1] D. Impedovo and G. Pirlo, "Automatic Signature Verification: The State of the Art", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 38, 2008, pp. 609–635.
- [2] S. Biswas, T.Kim and D. Bhattacharyya, "Features Extraction and Verification of Signature Image using Clustering Technique", International Journal of Smart Home, Vol. 4, 2010.
- [3] B. Fangaet al., "Off-line signature verification by the tracking of feature and stroke positions", Pattern Recognition, Vol. 36, 2003, pp. 91 – 101.
- [4] K.Dial and S. Mahesh, "Off-line Handwritten Signature Verification using Artificial Neural Network Classifier", International Journal of Recent Trends in Engineering, Vol. 2, 2009.
- [5] H. Nemmour and Y. Chibani, "Handwritten Arabic word recognition based on Ridgelet transform and support vector machines", in International Conference on High Performance Computing and Simulation (HPCS), Istanbul, July 2011, pp. 357–361.
- [6] M. N. Do and M. Vetterli, "The Contourlet Transform: An Efficient Directional Multi resolution Image Representation", Image Processing, Vol. 14, 2005, pp. 2091-2106.
- [7] M. R. Pourshahabi, M. H. Sigari and H. R. Pourreza, "Offline Handwritten Signature Identification and Verification Using Contourlet Transform", in International Conference of Soft Computing and Pattern Recognition, Malacca, December 2009, pp. 670–673.
- [8] S. Dewangan, (2011) "Neural Network-based Offline Handwritten Signature Verification System using Hu's Moment Invariant Analysis", International Journal of Engineering and Advanced Technology (IJEAT), Vol. 1, 2011.
- [9] A.Gupta, G.Khandelwal and S.Chakraverty, "Offline Signature Verification Using Critical Region Matching", Journal of Signal Processing, Image Processing and Pattern, Vol. 2, 2009.
- [10] Dina Darwish, "Simulation and evaluation of Signature Recognition Techniques", Egyptian Computer Journal, Vol.37, 2013.
- [11] http://en.wikipedia.org/wiki/Ensemble_learning
- [12] <http://www.mathworks.com/help/stats/ensemble-methods.html>
- [13] C.Sutton, Classification and Regression Trees, Bagging, and Boosting, Handbook of Statistics, Elsevier B.V., 2005.
- [14] <http://www.mathworks.com/help/stats/classification-trees-and-regression-trees.html>

Dina Darwish received the B.Sc. in 2004 and the M.Sc. in 2006 with honors degree from Arab Academy for Science and Technology, Egypt. She received the Ph.D. degree from Cairo University, Egypt, 2009. Her main interests include communications systems, computer networks, internet technology, and multimedia systems. She is assistant professor of communications and computer networks, International Academy for Engineering and Media Science (IAEMS), Egypt, since September 2009.

A Model for Optimizing Data Caching of Dual Mode Handheld Devices

Mohamed Grida¹, Hasnaa Soliman² and Mohamed Hassan²

¹Industrial Engineering Department, Faculty of Engineering, Zagazig University
Zagazig, Egypt

²Information System Departments, Faculty of Computers and Informatics, Zagazig University
Zagazig, Egypt

Abstract

Handheld apparatus are widely used especially when equipped with an automatic data capturing solutions. They are usually connected to the main database server using either an economic fixed location mode or a more expensive mobile mode. In order to quantify the gap of the two modes, a metric called mobility coefficient is introduced. Then, a mathematical model is developed to optimize the handheld data caching. This model considers several factors affect the cache performance such as the data item size, its access frequency, its update rate, the cache size, and the caching cost. Finally, a series of simulation experiments are conducted in order to validate of the proposed model under various system configurations.

Keywords: *Handheld performance, Caching optimization, dual mode gadget.*

1. Introduction

With the rapid evolution in mobile computing and wireless technologies, more smart handheld devices (such as laptops, palmtops, and personal digital assistants (PDAs)) are supported and used all over the world, and they become a significant part of everyday life and work. These devices were defined as small, mobile, and battery-powered computing devices, which have no mass storage system and all persistent data have to be kept in a limited battery-buffer RAM [1] [2][3][4]. The main motivation for using these devices is their ability to reduce data entry errors while increasing the data entry rate [1] [4]. At the same time, current handheld devices support various ways for communication with the mass data storage devices such as serial cables or Universal Serial Bus (USB) for direct connection, short range wireless connections such as Bluetooth, IrDA and Wi-Fi (IEEE 802.11) and cellular networks for long distances through GPRS or 3G [3] [5]. Handheld devices can be connected with the mass data storage devices (such as the database server) through two main modes, which are economic base mode (for connectivity inside the organization) and costly mobile mode (for connectivity outside the organization). Within the base mode, handheld can be connected to the database

server through serial cables, USB, Bluetooth, or Wi-Fi by which they can obtain bulky amount of data with low cost, high bandwidth and little power consumption. On the other hand, within the mobile mode, a handheld can be connected to the database server through cellular network via GPRS or 3G technologies, which are costly and provide limited bandwidth [6] [7][8][9][10][11][12] [13]. In order to overcome the bandwidth and the cost limitation of the mobile mode, data caching at the handheld local memory is considered as an effective solution to overcome these limitations and to reduce the network traffic [14] [6][15][8][7][16][17] [13]. Due to the limited storage capacity and processing power of the handheld devices, it will be impossible to hold all the accessed data items at the handheld local memory. Consequently, a model for optimizing the handheld data caching is needed by which a user can determine which database items can be cached at handheld local memory in the base mode. This model should aim to reduce the total cost and traffic between the handheld device and the database server.

In the next sections, an overview of related and previous researches on caching area will be presented. After that, a model for optimizing the handheld data caching will be developed based on a brief description of the handheld system architecture. Finally, the performance of the proposed scheme will be examined before concluding and highlighting future work possibilities.

2. Literature survey

Through the past few years, cache optimality has been addressed by number of researchers to decide which data items should be kept in or evicted from the cache. In this section, a brief overview of the caching replacement schemas will be introduced along with the factors affecting the cache performance. Most of previously described cache replacement policies were represented as a function of the different factors such as access frequency, update rate and so on.

In 2000, Xu et al. [18] proposed a cache replacement policy called Stretch Access Rate Inverse Update – frequency (SAIU), which was represented by a gain function as shown in Eq.(1).

$$gain(i) = \frac{A_{delay(i)} \cdot P_{A_i}}{S_i \cdot U_i} \quad (1)$$

This policy considered only four factors affect cache replacement decision of a data item i , which were access probability (P_{A_i}), update frequency (U_i), retrieval delay ($A_{delay(i)}$), and data size (S_i). It aimed to remove the data item with the least gain function value from the cache [18]. In 2004, Xu et al. added cache consistency to the former policy due to its importance for some applications such as financial transactions. An updated SAIU replacement policy was introduced by including (cache validation delay caused by the underlying cache invalidation scheme) as cache consistency metric. The updated policy was called Minimum Stretch integrated with Access rates, Update frequencies, and cache validation Delay (Min-SAUD) [15]. They introduced a new gain function considering two new factors, which are the cache validation delay ($V_{delay(i)}$) and the ratio of update rate (\bar{u}_i) to access rate (\bar{a}_i) for a data item i (X_i) as shown in Eq. (2).

$$gain(i) = \frac{P_{A_i}}{S_i} \left(\frac{A_{delay(i)}}{1 + X_i} - V_{delay(i)} \right) \quad (2)$$

In 2005, Yin et al. introduced a cache replacement policy based on the generalized cost function expressed in Eq. (3).

$$Cost(i) = P_{A_i} (A_{cost(i)} - V_{cost(i)} - P_{U_i} U_{cost(i)}) \quad (3)$$

This policy brought in three additional factors affecting the decision of caching a data item or not, which were the cost of fetching a data item to the cache ($A_{cost(i)}$), the cost of validating the cache consistency ($V_{cost(i)}$) and the cost of getting the updated data from the server ($U_{cost(i)}$). It aimed to evict the data items with the least value of $[cost(i) / S_i]$ from the cache [8].

In 2007, Chand et al. developed a cache replacement policy called Least Utility Value with Migration (LUV-Mi), which aimed to maximize the benefit of caching the data items by measuring a utility value for each data item kept in the cache as shown in Eq. (4). The data item with a high utility value will be kept in the cache. This policy introduced an additional factor affecting the decision of

replacing a data item, which is time to life value (TTL_i) that indicates the freshness or validity of a data item. Moreover, it introduced the distance (δ_i) that calculates the number of hops between the requesting client and the responding client (data source) as a form of the data transmission cost [19].

$$utility_i = \frac{P_{A_i} \cdot \delta_i \cdot TTL_i}{S_i} \quad (4)$$

Based on the generalized cost function proposed by Yin et al. (2005), Chand et al. (2009) proposed a novel cache replacement policy called Least Profitable Value (R-LPV), which based on a profit function expressed by Eq. (5).

$$PF_i = A_i \times (A_{delay(i)} - V_{delay(i)} - P_{U_i} \times U_{delay(i)}) \quad (5)$$

This replacement policy divided the delay factor into three sub factors, which are the data access delay ($A_{delay(i)}$), the cache validation delay ($V_{delay(i)}$) and data update delay ($U_{delay(i)}$) affecting the decision of caching the data items. Moreover, it used the access frequency (A_i) as a metric of the data access probability (P_{A_i}) representing the expected number of accesses to item (i) per a certain period of time. This aimed to maximize the total profit value for cached data items and it evicted an item (i) with minimum $[PF_i / S_i]$ value from the cache [20].

In 2010, Pant et al. used only two of previously factors in order to remove the item with least important value from the cache and to introduced a cache replacement policy for wireless sensors networks called item replacement policy (IRP) represented by Eq. (6)[21].

$$Imp_i = P_{A_i} \times TTL_i \quad (6)$$

In 2011, Dimokas et al. used some of previously described parameters and developed a cost based cache replacement policy for wireless sensor network. The data items with the greatest costs were removed from the cache [17]. This policy was represented as cost function as shown in Eq. (7).

$$cost(i) = \frac{A_{delay(i)} \times S_i}{TTL_i \times A_i} \quad (7)$$

Finally, all of the previously described cache replacement policies were summarized in Table 1.

Table 1: The most commonly factors affecting the cache performance along with the author who used each factor in a study

Decision Factors	Authors						
	Xu et al. (2000)	Xu et al. (2004)	Yin et al. (2005)	Chand et al. (2007)	Chand et al. (2009)	Pant et al. (2010)	Dimokas et al. (2011)
Item Size(s_i)	Yes	Yes	Yes	Yes	Yes		Yes
Access probability (P_{A_i}) or Access frequency (A_i)	Used P_{A_i}	Used P_{A_i}	Used P_{A_i}	Used P_{A_i}	Used A_i	Used P_{A_i}	Used A_i
Update probability (P_{U_i}) or Update frequency (U_i)	Used U_i	Used P_{U_i}	Used P_{U_i}		Used P_{U_i}		
Distance (δ_i) between the requesting client and the data source				Yes			
Data retrieval delay	Yes	Yes			Yes		Yes
Data retrieval cost			Yes				
Time-To-Live value associated with each item(TTL_i)				Yes		Yes	Yes

3. The proposed model

The mobile computing system for the proposed model mainly consists of a central server/s, which stores a large database and a number of mobile clients (handheld devices), which host limited size mobile databases as shown in Fig. 1. These devices may not be always connected to the server due to their operational constraints. Therefore, these devices can connect to the database server periodically in order to obtain their required data using one of two modes the base economic mode and the costly mobile mode.

According to the previously described survey, there are seven factors affect the decision of caching data or request them from the server when required through the mobile networks. These factors can be categorized into two main categories. First, factors related to system, which are the data retrieval delay and data retrieval cost in addition to the distance between the mobile client and the server. Second, factors related to data, which are the data item size, its access frequency, its update rate and its time to live. The proposed model addresses the first category through introducing a metric called mobility coefficient in order to quantify the system connectivity effectiveness. Then, it utilizes the second category in order to optimize the handheld data caching.

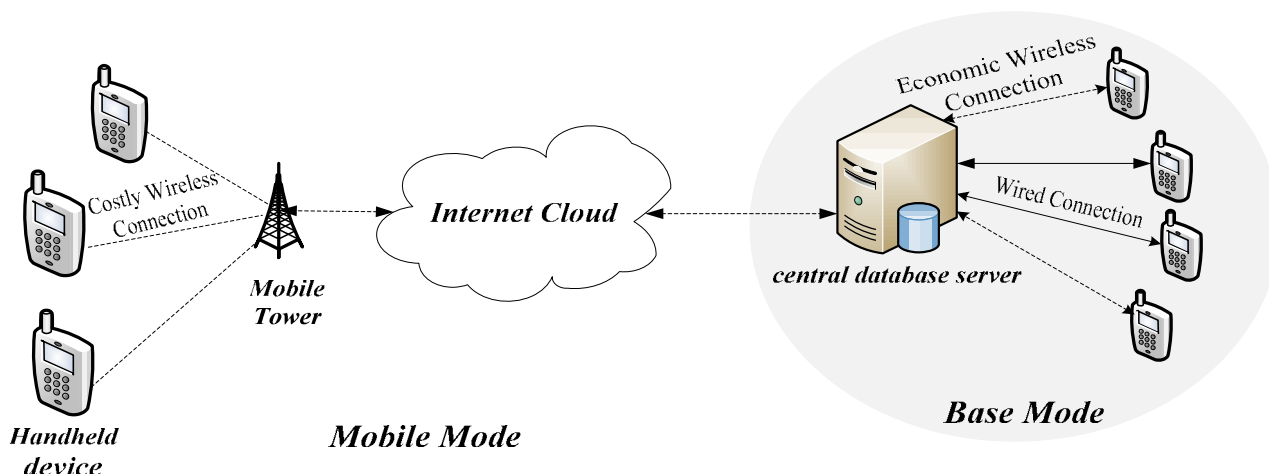


Fig.1 The proposed system architecture

3.1 Mobility coefficient

According to system architecture, a mobile client with a handheld device can connect to the central database server either through the base mode or through mobile mode. The operational conditions of the mobile computing environment may have some obstacles that may not exist in non-mobile environments. In order to quantify such obstacles, a metric called Mobility Coefficient will be introduced. Based on this measure, an organization can decide the return on its investment in caching resources. Thus, mobility coefficient can be represented by an equation annotated by M_C as shown in Eq. (8).

$$M_C = \frac{\text{Attributes of mobile mode}}{\text{Attributes of base mode}} \quad (8)$$

According to previously described survey, there are three important factors affecting the connectivity between the handheld device and the database server either in base mode or mobile mode, which are data retrieval delay, data retrieval cost and the distance between requesting node and the server. As this distance was introduced as a form of data transmission cost, it will not be used in the proposed model. Moreover, [6] [8] [9] [10][11] [7] [16][12] illustrated that there are two reasons for the data retrieval delay, which are the bandwidth (transmission speed) and the availability of the service. Consequently, it can be concluded that there are three factors affecting the connectivity between the handheld device and the server, which are the data transmission cost, the transmission bandwidth and the service availability. Therefore, the mobility coefficient can be represented as a function of the three previously mentioned factors as shown in Eq. (9).

$$M_C = \left(\frac{BW_b}{BW_m}\right) \times \left(\frac{COST_m}{COST_b}\right) \times \left(\frac{AV_b}{AV_m}\right) \quad (9)$$

Where:

- BW_m : Bandwidth of mobile mode.
- BW_b : Bandwidth of base mode.
- $COST_m$: Transmission cost for the mobile mode.
- $COST_b$: Transmission cost for the base mode.
- AV_m : Availability for mobile mode.
- AV_b : Availability for base mode.

According to Eq. (9), there is one positive relationship between the connectivity effectiveness and the ratio of transmission cost for mobile mode to the transmission cost for the base mode $\left(\frac{COST_m}{COST_b}\right)$. While, there are two negative relationships between the connectivity effectiveness and the ratio of the bandwidth of two modes $\left(\frac{BW_m}{BW_b}\right)$, and the ratio of availability of two modes $\left(\frac{AV_m}{AV_b}\right)$.

Based on the value of M_C , the organization can decide to size of the handheld cache. For systems with high mobility coefficient ($M_C \gg 1$), the organization should invest the hardware of handheld devices in order to increase the amount of data being cached and reduce the connection to the server using mobile mode. While for systems with low mobility coefficient ($M_C \cong 1$), smaller cache size should be considered because requesting data from the server through mobile mode instead of caching them will not affordable.

3.2 The proposed Caching model

In this section, a simple client side data-caching model will be introduced. Fig.3 shows the data flow diagram (DFD) of model. In the base mode, the handheld connect to the server to acquire the data to be cached. In mobile mode, it connects to the server either to request non-cached data or to update the cached data.

The previously described survey illustrated that there are four important factors related to data items and can affect the decision of caching a data item, which are the data item size (S_i), its access frequency (A_i), its update rate (U_i), and its time-to-live (TTL_i). However, the value of TTL_i can be expressed in term of update rate for each data item (U_i). There is a negative relationship between the gain and the data item size and its update rate resulting in reducing the opportunity of caching data [18]. However, there is a positive relationship between the gain and the access frequency resulting in increasing the opportunity of caching. Moreover, increasing the data item size will result in increasing the traffic between the client and the server. Consequently, the proposed model aims to reduce the network data traffic denoted by " $M_{traffic}$ ", which can be calculated as following:

$$\begin{aligned} M_{traffic} &= [(1 - X_i) * S_i * A_i] + [X_i * S_i * U_i] \\ &= S_i * A_i - X_i * S_i * A_i + X_i * S_i * U_i \\ &= S_i * A_i + X_i(S_i * U_i - S_i * A_i) \\ &= S_i * A_i + (S_i * X_i(U_i - A_i)) \end{aligned} \quad (10)$$

- X_i : A binary caching decision variable.

If a data item (i) is cached in base mode ($X_i = 1$), it will need to be updated periodically by the server. If it isn't cached ($X_i = 0$), it will be requested from the server when required through mobile networks.

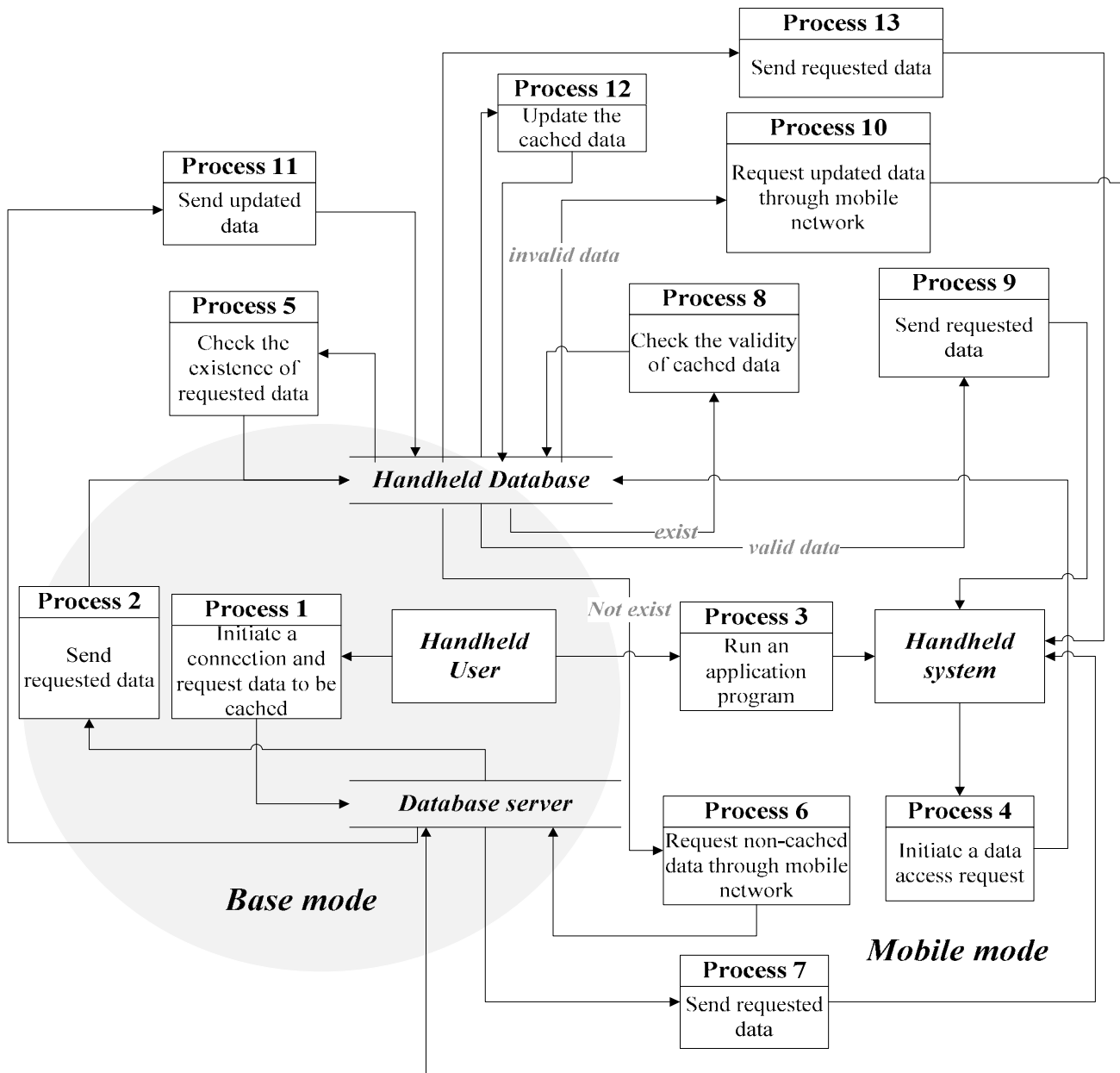


Fig. 2 The DFD Diagram for the system model

As the Eq. (9) for the mobility coefficient and Eq. (10) for computing mobile traffic affect the decision of caching a data item at the handheld local memory or not, both equations should be integrated into one formula in order to introduce the final representation for the proposed model. This formula is denoted by weighted mobile traffic as follows:

Weighted Mobile Traffic = Mobility Coefficient × Mobile Traffic

$$= M_c * \left[\sum_{i=1}^N (S_i * A_i + (S_i * X_i (U_i - A_i))) \right] \quad (11)$$

The literature introduced an additional factor affecting the decision of caching data locally at the handheld or not which is the size of the handheld cache [15] [8] [19][10] [11] [20] [16] [17] [22] [13]. Therefore, Eq. (11) representing the proposed model should be updated to include the new factors as follows:

Total operating cost

$$= [Tran_{cost} * \text{Weighted Mobile Traffic}] + [Cach_{cost} * Cach_{size}]$$

$$= Tran_{cost} * M_C * \left[\sum_{i=1}^N (S_i * A_i + (S_i * X_i (U_i - A_i))) \right] + [Cach_{cost} * (\sum_{i=1}^V X_i * S_i)] \quad (12)$$

In order to minimize the total operating cost, Eq. (12) is differentiated with respect to X_i as following:

$$\frac{d(\text{Total operating cost})}{d(X_i)} = 0$$

$$(A_i - U_i) = \frac{Cach_{cost}}{Tran_{cost} * M_C} \quad (13)$$

The right side term of the above equation tend to be very small due to the great value of M_C . Therefore, it is clear that the value of $(A_i - U_i)$ is critical to decide whether to

cache a data item into the handheld or not. Items with update rate higher than its access frequency should not be cached and vice versa.

4. Case Study

In order to test the proposed model validity, a set of Matlab functions was developed to simulate the whole system along with the proposed cache optimization model. A set of data representing a two mode handled operation of a food distribution company in Cairo, Egypt was obtained in order to test the system, as shown in Table 2. The company maintains its main Enterprise Resource Planning (ERP) database within its headquarter, where all the sales and distribution representatives connects their handheld units through a Wi-Fi-network. After that, the team used to connect to the ERP server in order to update the products, roots, orders, and customers info through 3G mobile network.

Table 2: Default system parameter settings

<i>General parameters notations</i>	<i>Description</i>	<i>Default value</i>
N	Number of data items in the database	28 data items
BW_m	The bandwidth for mobile mode	14.4 Kbps
BW_b	The bandwidth for base mode	115 Kbps
$COST_m \text{ or } Tran_{cost}$	The cost for transmitting 1GB of Data in mobile mode	10 dollars
$COST_b$	The cost for transmitting 1GB of Data in base mode	1 dollar
AV_m	Service availability for the mobile mode	30 %
AV_b	Service availability for the base mode	70%
Max_{cache}	Maximum cache size for handheld	32kB
S_i	The data item size	Ranging from .78 to 10000 KB
U_i	The update rate for each data item per day	Ranging from 1 to 1000 times daily
A_i	The access frequency for each data item per day	Ranging from 1 to 300 times daily
$Cach_{cost}$	The cost of acquiring and running 1GB of handheld cache	1 dollar

4.1 Using default system parameters

In this section, the performance of the proposed model is investigated by using default system parameters shown in Table 2 and the simulation results are displayed in Fig. 3. According to Fig. 3, it is obvious that the proposed model

doesn't cache all data items but only a small subset of the data items whose value of cache coefficient $(U_i - A_i)$ is less than zero. Moreover, some of the negative cache coefficient data items were not cached due to the limited size of the cache.

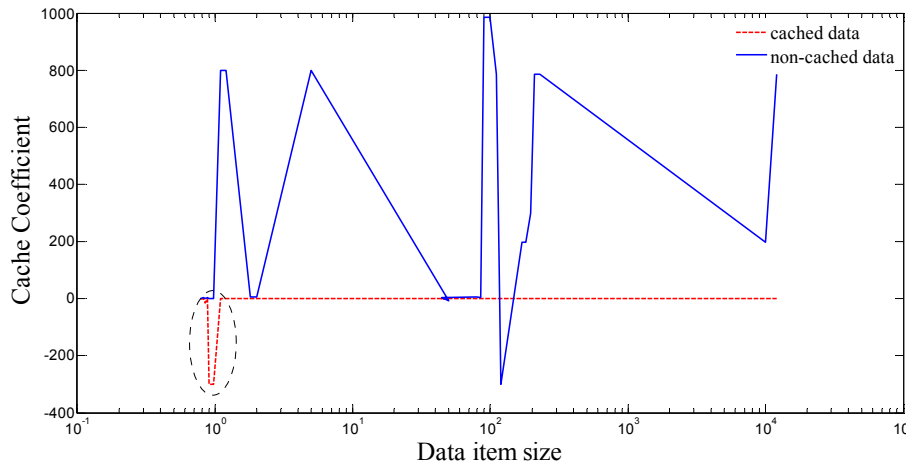


Fig. 3 The Proposed model performance using default data items' parameter settings.

4.2 The impact of increasing the cache size

According to the proposed model, not only the data item's access frequency and its update rate but also the cache size can affect the cache performance. In order to investigate the impact of the cache size on the proposed model performance, in this section, the proposed model is examined using a different cache size. The simulation results are shown in Fig. 4. By comparing the simulation results in Fig. 3 and Fig. 4, it can be concluded that increasing the cache size to 32GB instead of 32 KB results

in caching all of more frequently accessed data items whose Cache coefficient value less than zero. Therefore, it can be concluded that the proposed model balances between caching more frequently accessed items and the cache size. When the cache size is limited, only the more frequently accessed small sized data items are cached. On the other hand, when a plenty of cache is available, all of the frequently accessed data items rare cached regardless their size as shown in Fig. 4.

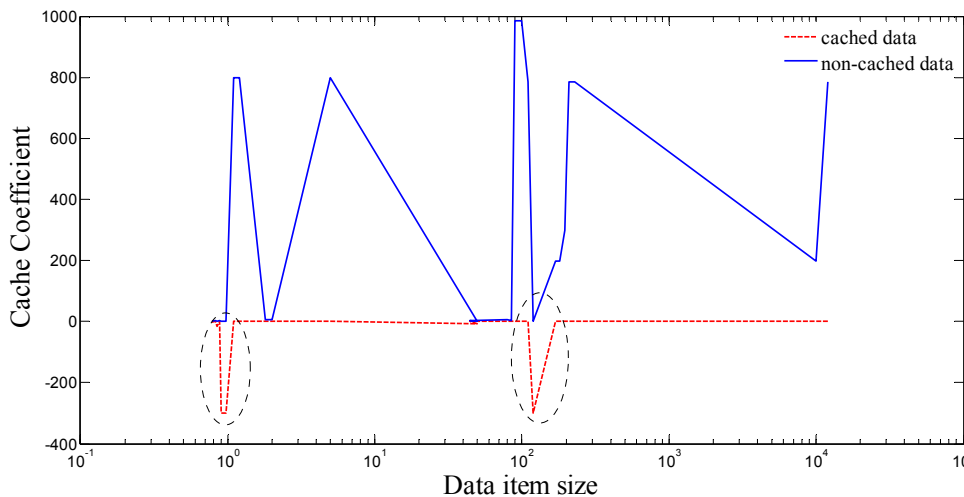


Fig.4 The impact of the cache size on the proposed model performance

4.3 Impact of the data transmission cost

According to Eq. (12) and Eq. (13), it can be concluded that not only the cache size, the data access frequency and update rate affect the caching decision, but also other

factors such as data transmission cost can have an impact on the caching decisions. Therefore, the proposed model is examined with a theoretically very low transmission cost in order to force the right term side of Eq. (13) to have a positive value away from zero. The simulation results are

displayed in Fig. 5. The proposed model cached only the data items with negative caching coefficient that has an absolute value greater than the caching-transmission weighted value shown in the right side of Eq. (13).

5. Conclusions

A model for optimizing the handheld data caching in a dual mode environment is proposed. That model objective is to reduce traffic between the handheld used and the server. This model is based on a literature survey that determined the factors affecting the caching decisions, which are the data item size, its access frequency, its update rate and the cache size and the transmission cost.

In order to evaluate the performance of the proposed model, a set of MATLAB functions is created to simulate the system and proposed optimization model. These functions were used along with a data set obtained from an Egyptian food distribution company in order to test the validity the system outcomes. The tests showed that the system was able to determine the appropriate caching decisions under different caching sizes and transmission costs. There is a need to test the model with data obtained from different domains and system configurations in order to exploring its applicability with these configurations.

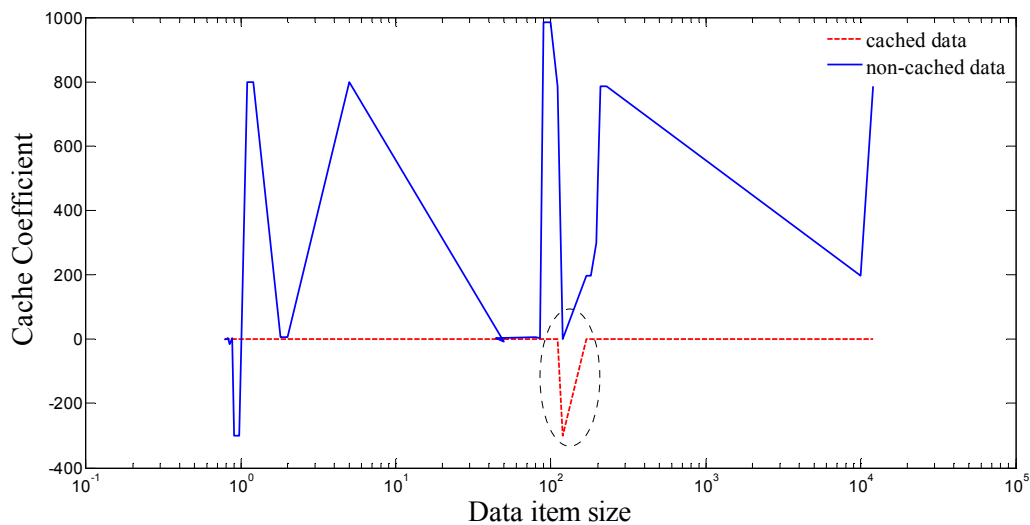


Fig.6 The impact of the data transmission cost on the proposed model performance under large cache size (32GB).

References

- [1] J. A. Blackman, "The Usefulness of Handheld Computers in a Surgical Group Practice", M.S. thesis, School of Medicine, Oregon Health Sciences University, May 1999.
- [2] F. A. Adatia, and P. L. Bedard, "Palm reading": 1. Handheld hardware and operating systems, Canadian Medical Association Journal, Vol. 167, Issue 7, 2002 October 1, PP. 775-780.
- [3] W. C. Hu, J. h. Yeh, H. J. Chu, and C. w. Lee, "Internet-Enabled Mobile Handheld Devices for Mobile Commerce", Contemporary Management Research, Vol. 01, No. 01, September 2005, PP.13-34.
- [4] Y. C. Lu, Y. Xiao, A. Sears, and J.A. Jacko, "A review and a framework of handheld computer adoption in healthcare", International Journal of Medical Informatics, Vol.74, Issue 5, June 2005, pp. 409-422.
- [5] A. A. Economides and N. Nikolaou, "Evaluation of Handheld Devices for Mobile Learning", International Journal of Engineering Education, 2008.
- [6] S. K. Madria, M. Mohania, S. S. Bhowmick, and B. Bhargava, "Mobile Data and Transaction Management", Information Sciences Journal, Vol.141, Apr 2002, PP. 279-309.
- [7] R. B. Patel, and Anu, "A Mobile Transaction System for Open Networks," BVICAM's International Journal of Information Technology, Vol.1, Issue 1, January-July 2008, PP. 67-78.
- [8] L. Yin, G. Cao and Y. Cai, "A Generalized Target-Driven Cache Replacement Policy for Mobile Environments", Journal of Parallel and Distributed Computing, Vol. 65, No. 5, 2005, pp. 583-594.
- [9] E. Pitoura, and P. K. Chrysanthis, "Caching and Replication in Mobile Data Management", IEEE Data Eng, 2007, PP. 13-20.

- [10] S. Yi, S. Jung, and J. Suh, "Better mobile client's cache reusability and data access time in a wireless broadcast environment", *Data & Knowledge Engineering*, Vol. 63, Issue 2, November 2007, PP. 293–314.
- [11] H. K. Chavan, and S. D. Patil, "A Survey of Mobile Database Cache Replacement Policies".
- [12] S. Fageeril, R. Ahmad, and A. A. Mu'azu, " Mobile Data availability Similarity between Cooperative Caching and Replication", *IOSR Journal of Computer Engineering*, Vol. 1, Issue 4, May-June 2012, PP. 09-13.
- [13] M. Akon, M.T. Islam, X. Shen, and A. Singh, "A Bandwidth and Effective Hit Optimal Cache Scheme for Wireless Data Access Networks with Client Injected Updates", *Computer Networks*, Vol.56, Issue 7, 2012, PP. 2080–2095.
- [14] I. CHUNG, L. GRUENWALD, C.HWANG, and S.Y. JUNG, "Adaptive Transaction Management Protocols for Mobile Client Caching DBMSs", *Journal of Information Science and Engineering*, Vol. 18, No.3, May 2002, PP.445-457.
- [15] J. Xu, Q. Hu, W. Lee, and D. Lee, "Performance Evaluation of an Optimal Cache Replacement Policy for Wireless data Dissemination", *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, No. 1, Jan. 2004, pp. 125-139.
- [16] Y. Du, S. K. S. Gupta, G. Varsamopoulos, "Improving On-Demand Data Access Efficiency in MANETs with Cooperative Caching", *Elsevier Ad Hoc Journal*, No.7, May 2009, PP.579–598.
- [17] N. Dimokas, D. Katsaros, L. Tassiulas, and Y. Manolopoulos, "High performance, low complexity cooperative caching for wireless sensor networks", *Wireless Networks*, Vol.17, Issue 3, April 2011, PP.717-737.
- [18] J. Xu, Q. L. Hu, D.L. Lee, W. C. Lee, "SAIU: An Efficient Cache Replacement Policy for Wireless on-Demand Broadcasts," *Proc. Ninth ACM Int'l Conf. Information and Knowledge Management*, Nov. 2000, PP. 46-53.
- [19] N. Chand, R.C. Joshi and M. Misra, "Cooperative Caching Strategy in Mobile Ad Hoc Networks Based on Clusters", *International Journal of Wireless Personal Communications special issue on Cooperation in Wireless Networks*, Vol.43, Oct 2007, PP.41–63.
- [20] N. Chand, R.C. Joshi, and M. Misra, "Energy Efficient Cache Replacement in Mobile Computing Environment", *Ubiquitous Computing and Communication Journal*, 2009, pp. 1-12
- [21] S. Pant, N. Chauhan, and P. Kumar, "Effective Cache based Policies in Wireless Sensor Networks: A Survey", *International Journal of Computer Applications*, Vol.11, No.10, December 2010.
- [22] L. Jia, "A Mobile Ad-hoc Network Data Cache Invalidation Method", *Procedia Engineering*, Vol.23, 2011, PP.150–154.

Mohamed Osama Khalil Grida is an assistant professor in Industrial Engineering Department in Faculty of Engineering in Zagazig University.

Hasnaa Raafat Hamed Soliman is a teaching assistant in Information System Department in Faculty of Computers and Informatics in Zagazig University.

Mohamed Monier Hassan is a professor in Information System Department in faculty of Computers and Informatics in Zagazig University.

On the Limits of Perfect Security for Steganographic System

Khan Farhan Rafat, M. Sher

Department of Computer Science, International Islamic University
Islamabad, 44000, Pakistan

Abstract

Until now the discussion on perfect security for steganographic systems has remained confined within the realm of mathematicians and information theory experts whose concise and symbolic representation of their philosophies, postulates, and inference thereafter has made it hard for the naïve academics to have an insight of the concepts. This paper is an endeavor not only to appraise on the limitations of one of such pioneer comprehensions but also to illustrate a pitfall in another scheme that asserts on having perfect security without the use of public or secret key. Goals set are accomplished through contrasting test results of a steganographic scheme that exploits English words with corresponding acronyms for hiding bits of secret information in chat - a preferred way to exchange messages these days. The misapprehension about perfect security and reign in characteristic of stego key in bit embedding process are unfolded respectively by launching elementary chosen-message and chosen-cover attack, and through proposed enhancement of target scheme.

Keywords: *Perfect Security, Conceal, Covert Channel, Deception, Oblivious Communication, Unobtrusive.*

1. Introduction

Organized data serves as a single point of reference that makes prediction about an event a much simpler task than without it i.e., provisioning of some sort of meaningful *information* for decision making, the essence of which, however, cannot be measured but the severity of which can be realized from the fact that in today's world information is being regarded as a *double-edged sword* [1] capable of imparting devastating impact on the privacy of an individual and as well as nation (at global level) more withering than what was faced by Hiroshima and Nagasaki during World War - II.

Some of the most recent incidents including *Blaster worm* attack (2003), electronic cutoff of Estonia from rest of the world (2007), attack by *Stuxnet* virus on Iranian nuclear installations (2010) and Wiki leaks (2011) calls for a daring need, more than ever before, to guard information security frontiers by evolving new and analyzing and modifying/updating existing security schemes from falling into the hands of hostile.

Steganography is referred to as art and science [2] for covert communication. The name was first cited in a work presented by *Trithemus* (1462-1516) entitled *Steganographia* and is of Greek origin where the words $\sigma\tau\epsilon\gamma\alpha\nu\acute{o}-\varsigma$ (Steganos) and $\gamma\rho\alpha\phi\text{-}\epsilon\iota\nu$ (Graphos) are put together as single English word which means covered/Hidden Writing [3]. The essence of steganography is to hide the very existence of information [4] in contrast to cryptography whose rationale is to make information incomprehensible [5]. By virtue of being seamless steganography has emerged as a preferred choice for information hiding these days.

1.1 Paper Plot

This paper is organized as follows: Section 2 expound on the limitations of perfect security and consequence of its realization without using stego key. Section 3 elaborates on bit embedding scheme with the help of which short comings of prevalent apprehensions will be emphasized. Section 4 highlights on tools used to contrast resemblance/dissimilarity between cover text and stego object. Analysis of the target scheme using test cases along with test results are shown in Section 5. Reign in attribute of stego key towards system's security is discussed in Section 6 where an enhancement in context of information theoretic security is proposed. Section 7 concludes on our argument.

2. Perfect Security

Subsequent discussion expands on the notion of perfect security and allied misapprehensions. However, as already stated, deliberate effort has been made to forgo complex mathematical illustrations, elaborate and exemplify the concept in simple terms for easy understanding.

2.1 Cachin's Conception

Cachin [6] was first to come up with the idea of perfect security by suggesting for an information theoretically secure scheme where seamless bit embedding in randomly

selected cover text renders a stego object that has the same probability distribution as that of the former and equated it as follows:

$$D(P_c || P_s) \leftarrow \varepsilon \leftarrow \sum_{q \in Q} P_c(q) \log_2 \frac{P_c(q)}{P_s(q)} \quad (1)$$

, where P_c & P_s are probability distributions for cover text and stego object respectively.

It is obvious that $\frac{P_c(q)}{P_s(q)} = 0$ for $\frac{0}{P_s(q)}$, and equals ∞ for $\frac{P_c(q)}{0}$.

2.1.1 Precincts

- i. Evidently for $\varepsilon = 0$, the system is perfectly secure. But the question is whether such system exist? According to Cachin such a system does exist -at least theoretically- called *One Time Pad* (OTP) but realized that how *Wendy* (an observer) would let go such an output (stego object).
- ii. It is apparent from equation (1) that bit embedding has an implicit binding to render an output that must be from within the random oracle constituting the cover texts or else $\varepsilon \rightarrow \infty$ i.e. alphabet's bound constraint.
- iii. Notwithstanding above the trait to undermine *Wendy's* ability to detect such communication seems fictitious in its eternity.

2.2 Perfect Security without Public or Secret Key

Boris Ryabko and Daniil Ryabko [7] opted for generating all possible fixed length sequences of a given cover text and then transmitting only the sequence number that corresponds to the secret bits of the message where the block length of cover text must also be \leq message bit stream.

2.2.1 Limitations

- i. Besides having obvious memory constraint the provision of '*given cover text*' at both ends covertly accentuate on a shared secret (i.e. *key*) without which it cannot be regarded as an oblivious communication (*A contradiction of the postulate in itself*).
- ii. Knowledge of the algorithm alone is sufficient to retrieve hidden sequence number to exact on the message bits.
- iii. Contradicts Kerchoff's Principle [8].

3. Target Steganographic Scheme

Table 1 – List of Words along with their Acronyms

<u>Acronyms</u> Column Label (1)	<u>Words/Phrases</u> Column Label (0)
&	and
2	To
2NIGHT	Tonight
4	For
...	...
4U	For you
...	...
A3	Anytime, anywhere, anyplace
AAMOF	As a matter of fact
ABH	Anyone but him
ACTO	According to
...	...
ADD	Address
ADR	Ain't doin' right
AE	Almost every
AFAIC	As far as I'm concerned
...	...

To precisely apprehend on our revelation we selected the chat-lingo of the so called 'thumb generation' which is easy to grasp and does not arouse suspicion. Acronyms are contractions for comparatively long or frequently used words/phrases like *As soon as possible* which is abbreviated as *ASAP*. [9] Suggested using acronyms together with corresponding words / phrases of English language to hide bits of secret information in chat. The scheme works by arranging words/phrases in one of the two column table, the other column of which is populated with corresponding acronyms. The column containing words/phrases are labeled as "0" while acronyms are headed by label "1". Table 1 indicates one such arrangement.

Next, text cover composed of words/phrases and acronym from the predefined table is prepared. Secret information to be hidden inside the body of text cover is translated into bits. Text cover is then iterated till its end to search for words/phrase or acronym matching those in the table till end. Each time when a word/phrase or acronym is encountered corresponding secret message bit (in sequence) is examined with reference to column heads (label) of the table. Suppose if secret message bit is 0 and the corresponding matching text in cover is word/phrase i.e., having column labeled as 0, the cover text remains unchanged. However, if the secret message bit is 1 and the corresponding matching text in cover is word/phrase, the word/phrase in cover text is replaced by its corresponding acronym. In short, binary message bit 0 corresponds to having word/phrase in the stego object while binary message bit 1 corresponds to having acronym in place of words/phrases.

4. Tools to Realize Quality of Test Results

To maintain transparency and for better understanding and visual substantiation of test results we favored for probability distribution plots between cover text and stego object using Minitab 16 [10] and presented our reader with quantified output through Hamming [11], Levenshtein (Edit Distance) [12], and Jaro-Winkler [13] distance.

5. Analysis

To analyze our target scheme we arranged English Language words and their corresponding acronyms derived from [14] in two columns and labeled those as (0) and (1) respectively as in Table 1.

Next was the choice of secret message to unveil limitations of the idea of perfect security. Since, the target steganographic scheme allows for embedding of secret message bits (0/1) inside text-cover, hence, we favored for testing the scheme using message bits which were comprised of all 0's and developed three test cases as under:

Case – I: Cover text comprising of regular text but without acronyms (i.e. cover text devoid of acronyms):

It is evident from Table 1 that for text cover comprising of all words (*exclusive of words having acronyms*), a chosen-message of all 0-bits resulted in stego object which is an exact replica of the cover text.

a. Quantified Test Results

- A. Hamming, Levenshtein, and Jaro-Winkler distance for cover text and stego Object are computed as shown in Table 2.

Table 2 – Quantifying closeness of Cover Text & Stego Object

Distance	Difference
Hamming	0
Levenshtein	0
Jaro-Winkler	0

- B. To contrast probability distribution plots (Fig. 1 refers) of cover text and stego object, their mean, variance and standard deviation are computed as shown in Table 3.

Table 3 – Quantified Similarity between Cover Text & Stego Object

	Cover Text	Stego Object	Difference
Mean	0.089	0.089	0
Variance	0.976	0.976	0
STD	0.031	0.031	0

b. Envisaged Test Results

Figure 1 illustrates an exact match between cover text and stego object that going with the notion of perfect security renders $\epsilon = 0$ i.e. appears as **PERFECTLY SECURE**. We, however, envisage the test result as frightening as $\epsilon = 0$ **does not account for 'perfect security'**.

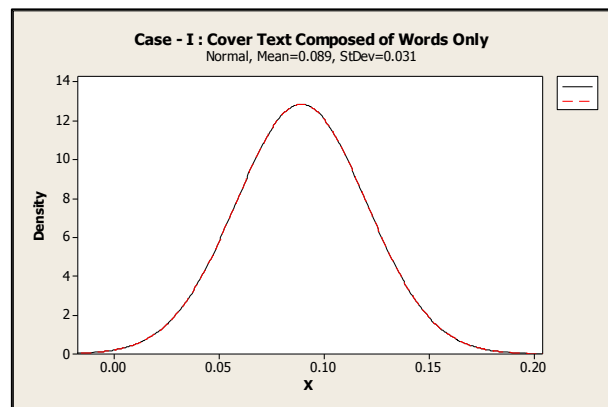


Fig. 1 Probability Distribution Graphs Contrasting Cover Text & Stego Object

Case – II: Cover text comprising of regular text and all acronyms in place of their corresponding words:

We repeated the preceding course and test results obtained thereafter are elaborated as under:

a. Quantified Test Results.

- A. Hamming, Levenshtein, and Jaro-Winkler distance for cover text and stego (Text) Object are computed as shown in Table 4.

Table 4 – Quantifying closeness of Cover Text & Stego Object

Distance	Difference
Hamming	Different file lengths
Levenshtein	228
Jaro-Winkler	0.443

- B. To contrast probability distribution plots (Fig. 2 refers) of cover text and stego object, their mean, variance and standard deviation are computed as shown in Table 5.

Table 5 – Quantified Similarity between Cover Text & Stego Object

	Cover Text	Stego Object	Difference
Mean	0.058	0.089	0.031
Variance	0.315	0.976	0.661
STD	0.017	0.031	0.014

b. Envisaged Test Results

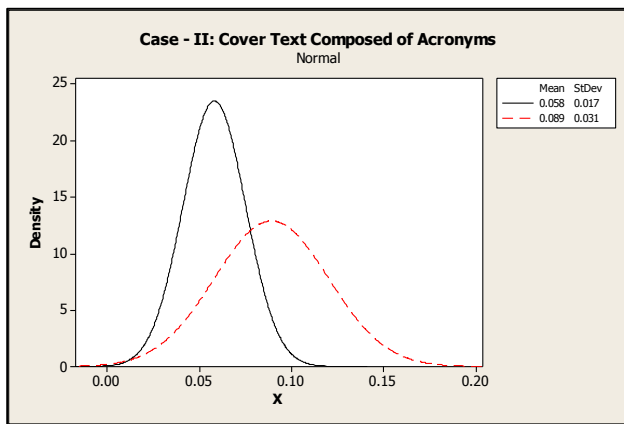


Fig. 2 Probability Distribution Graphs Contrasting Cover Text & Stego Object

The similarity between cover and stego object can be realized from Fig. 2 where the two probability distribution graphs vary for their mean and standard deviation by values 0.031 and 0.014 respectively. However, *interestingly all acronyms in cover text were replaced by their corresponding word/phrase.*

But is this scheme secure? We, however, doubt it since **a chosen message alone is sufficient to expose bit embedding methodology i.e. words having acronyms conceal secret binary message bit ‘0’.**

Case – III: *Cover text comprising of regular English text and mix of words with their corresponding acronyms*

Likewise, we experimented by composing cover text with a mix of words along with their corresponding acronyms whose test results follow as under:

a. Quantified Test Results

A. Hamming, Levenshtein, and Jaro-Winkler distance for cover text and stego (Text) Object are computed as shown in Table 6.

Table 6 – Quantifying closeness of Cover Text & Stego Object

Distance	Difference
Hamming	Different file lengths
Levenshtein	326
Jaro-Winkler	0.902

B. To contrast probability distribution plots (Fig. 3 refers) of cover text and stego object, their mean, variance and standard deviation are computed as shown in Table 7.

Table 7 – Quantified Similarity between Cover Text & Stego Object

	Cover Text	Stego Object	Difference
Mean	0.082	0.080	0.031
Variance	0.988	0.955	0.661
STD	0.031	0.030	0.014

b. Envisaged Test Results

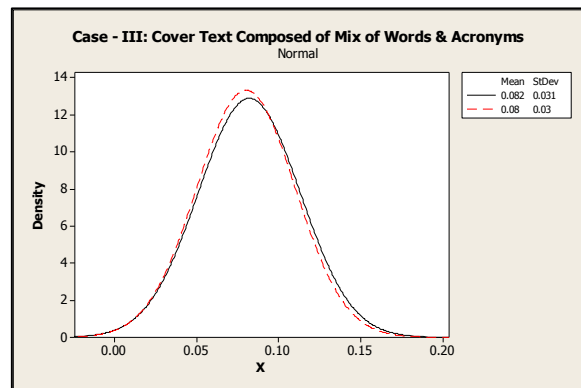


Fig. 3 Probability Distribution Graphs Contrasting Cover Text & Stego Object

The similarity between cover and stego object can be realized from Fig. 3 where the probability distribution graphs vary for their mean and standard deviation by values 0.031 and 0.014 respectively. Attention, however, is invited to the fact that *here words and acronyms in cover text were replaced by their corresponding acronyms and words in stego object respectively.*

Above, however, once again reflects on the limitation of the existing scheme and reconfirms on the deficiency of some shared secret (i.e. stego key etc.) between communicating parties.

Similarly, repeating the aforesaid cases with a chosen-message comprising of all binary bits ‘1’, and subsequently by contrasting differences in known-cover and resultant stego object, reconfirmed our findings.

6. Proposed Enhancement

The trait of aforementioned scheme also fell well short of Kerckhoff’s principle as knowledge of algorithm or cover or experimenting with choice of message alone suffices to unveil bit embedding methodology.

Based on its logical formation we opted for information theoretic model [15] where pre-processing of covert text just before bit embedding is proposed and ensued as follows:

6.1 Evolved Algorithm

Prepared cover text comprising of English words including acronyms. Generated 256-bit stego key (FIPS publication 198) [16]. Obtained 256-bit HASH of the stego key using SHAH-256 [17]. We then iterated through the entire cover text once. Searching for words or acronyms in Table 1 and swapping those with their corresponding acronyms or words respectively in the cover text if corresponding binary bit of the HASH is ‘1’ i.e. *random pre-processing of the cover text*.

Next we iterated and swapped column entries of Table 1 for corresponding binary bit ‘1’ of the stego key and obtained a new Table 8 as in [18].

Table 8 – Mixed List of Words and their Corresponding Acronyms

<u>Acronyms</u> Column Label (1)	<u>Stego</u> Key	<u>Words/Phrases</u> Column Label (0)
and	1	&
2	0	To
Tonight	1	2NIGHT
For	1	4
...
4U	0	For you
...
Anytime, anywhere, anyplace	1	A3
As a matter of fact	1	AAMOF
ABH	0	Anyone but him
ACTO	0	According to
...
ADD	0	Address
ADR	0	Ain't doin' right
Almost every	1	AE
As far as I'm concerned	1	AFAIC
...

Since entries in Table 1 are far more than 256, hence the pointer to stego key bits is re initialized to 1 each time it

exceeds the value 256 i.e. *random shuffling of the contents of pre-agreed list*.

On the analogy of Section 5, we composed (chose) a secret message encompassing all ‘0’-bits and proceeded with the three test cases respectively by iterating through the preprocessed cover text and substituting words or acronyms with words/acronyms from column labeled as ‘0’ of Table 8 – corresponding to chosen-message bit ‘0’.

6.2 Test Results

Interestingly the outcomes of the three test cases have almost proportionate distribution of words and acronyms for the same message which were spread over the entire stego object, similar to those envisage in Table 7, thereby affixing on the suitability of our proposed enhancement even for worst case scenarios and assuring its decoding only by those in possession of stego key (adherence to Kerchoff’s principle) or via brute force attack. . *The preprocessing stage introduces an uncertainty that is difficult to comprehend*.

6.3 Future Work

Following in sequence may further add operational ease along with entailing Wendy’s efforts to extract the hidden secret out of the cover text:

- i. Type of message i.e. text, image etc. and message length be appended before secret message as header.
- ii. Stego key dependent bisection of the message along with header.
- iii. Compression of bisected secret message.
- iv. Encryption of compressed data before bit embedding.
- v. Adding digital signatures will ensure message integrity and non-repudiation.
- vi. Use of PKI etc.to exchange stego keys.

7. Conclusion

Security can neither be measured in terms of length and breadth nor can it be quantified. *The notion on perfect security for finite sequence and without a shared secret between communicating parties seems fictitious* as no matter what, a finite sequence will always get decoded constraint only by time, resources and determination of the adversary e.g. by launching brute force attack. Hence the safest and most pragmatic approach is not to underestimate Wendy’s ability to precise on cover text carrying hidden

message but rather randomness be made a part of stego key dependent bit embedding process (*Stego key* be preferably unique in its perpetuity) to have varied outcomes even for the same message and using same cover text. It is only through this postulation that we are ought to devise information theoretically secure schemes that can prolong Wendy's efforts for a time equal to or longer than that where the hidden information losses its vitality.

Foresaid above, ours is also the first known attempt to scrutinize the cogency of 'perfect security' by launching basic chosen-message and chosen-cover attacks on an unpretentious steganographic scheme.

References

- [1] Kendall Hoyt, Stephen g. Brooks, A Double-Edged Sword: Globalization and Biosecurity. International Security, Vol. 28, No. 3, pp. 123-148, Winter 2003/04.
- [2] Chincholkar A.A. and Urkude D.A. (2012) Design and Implementation of Image Steganography. Journal of Signal and Image Processing, ISSN: 0976-8882 & E-ISSN: 0976-8890, Volume 3, Issue 3, pp.-111-113.
- [3] Alexander A. Sherstov. The Communication Complexity of Gap Hamming Distance. Theory of Computing, Vol. 8, 2012, pp. 197-208. www.theoryofcomputing.org
- [4] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn, Information Hiding? A Survey, Proc. IEEE, 1999
- [5] Al.Jeeva,V.Palanisamy and K. Kanagaram, Comparative Analysis Of Performance Efficiency And Security Measures Of Some Encryption Algorithms, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com, Vol. 2, Issue 3, May-Jun 2012, pp.3033-3037.
- [6] Cachin, C. An Information-Theoretic Model for Steganography, in Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecture Notes in Computer Science, Springer, 1998, pp. 306-318.
- [7] Boris Ryabko and Daniil Ryabko, Constructing perfect steganographic systems, Information and Computation 209 (2011) 1223–1230, © 2011 Elsevier Inc.
- [8] Cloudflare. A note about Kerckhoff's Principle. June 19, 2012. Internet:<http://blog.cloudflare.com/a-note-about-kerckhoffs-principle>
- [9] Shirali-Shahreza, M.H. Text Steganography in chat, 3rd IEEE/IFIP International Conference in Central Asia on Internet, (ICI 2007), 2007, pp. 1-5.
- [10] Minitab 16. <http://www.facebook.com/Minitab> [June 6, 2012]
- [11] Hamming, Richard W. (1950), Error detecting and error correcting codes, Bell System Technical Journal 29 (2): 147–160, MR 0035935.
- [12] Levenshtein VI (1966). Binary codes capable of correcting deletions, insertions, and reversals, Soviet Physics Doklady 10: 707–10.
- [13] M.A. Jaro. Probabilistic linkage of large public health data files (disc: P687-689). Statistics in Medicine, 1995, pp. 14:491-498.
- [14] The List of Chat Acronyms & Text Message Shorthand, <http://www.netlingo.com/acronyms.php>
- [15] J.Zöllner, H.Federrath, H.Klimant, A.Pfutzmann, R.Piotraschke, A.Westfeld, G.Wicke, and G.Wolf, Modeling the security of steganographic systems, In the Proc. 2nd Workshop on Information Hiding, April 1998, Portland, LNCS 1525, Springer-Verlag, 1998, 345-355.
- [16] FIPS PUB 198-1, http://csrc.nist.gov/publications/fips/fips198-1/FIPS-198-1_final.pdf
- [17] T Hansen - 2006, US Secure Hash Algorithms (SHA and HMAC-SHA), <http://tools.ietf.org/html/rfc4634>
- [18] Rafat, K. F, Enhanced Text Steganography in SMS, Computer, Control and Communication, 2009, IC4 2009, 2nd International Conference, pp. 1-6



KHAN FARHAN RAFAT is a Ph.D. Scholar at International Islamic University, Islamabad – Pakistan. He did MCS from Gomal University, D.I.K. followed by MS in Telecommunication Engineering from UMT, Lahore – Pakistan. As a veteran of information security with almost 24 years of hands-on experience he has worked in varied roles in areas not limited only to programming, evaluation & analysis of Software/Hardware based security modules, and formulating security policies.



Professor Dr. Muhammad Sher is Dean Faculty of Basic and Applied Sciences at International Islamic University, Islamabad – Pakistan. He received B.Sc. degree from Islamia University Bahawalpur and M.Sc. degree from Quaid-e-Azam University, Islamabad, Pakistan. His Ph.D. is from TU Berlin, Germany in Computer Science and Electrical Engineering. His area of research is Next Generation Networks Security. An eminent Scholar who has a number of research publications to his credit.

A Review of Fault Detection Techniques for Wireless Sensor Networks

Er. Saurabh, Dr. Rinkle Rani Aggarwal

Department of Computer Science & Engineering, Thapar University, Patiala.

ABSTRACT

Today wireless sensor networks (WSNs) emerge as a revolution in all aspects of our life. WSNs have unique specifications of themselves that describe them different from other networks. Fault tolerance is one of the most significant of many challenges in these networks. Five key features need to be considered when developing WSN solutions: scalability, security, reliability, self-healing and robustness. In this paper the main objective is to provide a comparative study of fault detection techniques using different approaches. Sensor nodes have various energy and computational constraints. To provide quality service by coverage protocols, there arises a need for developing protocols to provide fault tolerance, event reporting, and maintain energy efficiency.

Key words, of the Abstract- wireless sensor network (WSN); fault tolerance; cluster head; fault tolerant systems; fault diagnosis;

1. INTRODUCTION

1.1. Motivation

The reliability of computer, communication, and storage devices was recognized in the initial times as one of the key issues in computer systems. Since the 1950's, techniques that

enhance the reliability of computer and communication systems were developed both in academia and industry. It has been also recognized that as computers complexity and number of communication devices increases, fault-tolerance will be in great demand. Surprisingly, fault tolerance has never been the major design objective. While there are a number of reasons for this situation, the most important is that the reliability of individual components has been increasing at a much more rapid pace than it was expected. The rapid growth of the Internet in the last 10 years was the first major facilitator of the renewed interest in fault tolerance and related techniques such as self-repair. Internet requires the constant mode of operation and therefore special effort has been placed to develop fault tolerant data centers. Emergence of wireless sensor networks will further increase the importance of fault tolerance. At the same time, wireless sensor networks will impose a number of unique new conceptual and technical challenges to fault-tolerance researchers. There are at least three major groups of reasons why research in fault tolerant sensor networks should receive a significant attention. The first one is related to the technology and implementation aspects. Two components of a sensor node, sensors and actuators, directly interact with the environment and will be subject to a variety of physical, chemical, and biological

forces. Therefore, lower intrinsic reliability is expected than integrated circuits in fully enclosed packaging. Wireless sensor networks will be often deployed as consumer electronic devices that will put significant constraints on the cost and therefore, quality of used components. More importantly, nodes operate under strict energy constraints that will make energy budget dedicated to testing and fault tolerance very limited. The second reason is that applications will be equally as complex as the involved technology and architectures. More importantly, sensor networks will often operate in an autonomous mode without a human in the loop. In addition, security and privacy concerns will often prevent extensive testing procedures. Lastly, and maybe most importantly, many applications of sensor networks will be safety critical and can have very adverse impact on humans and the environment, in particular when the actuators are used. The final reason is that wireless sensor networks themselves are a new scientific and engineering field and it is not still quite clear as to what is the best way to address a particular problem. At this level, it is also difficult to accurately predict the best way to treat fault tolerance within a particular wireless sensor network approach. Additionally, both technology and applications for wireless sensor networks are changing at a rapid pace. Therefore, with respect to fault tolerance, it is important to consider schemes that conduct error detection using only local information at their own level or, to design fault tolerant techniques that do not significantly increase the communication overhead. On the other hand if the computation energy is significantly higher than the

communication requirements, it is a good idea to support communication resources at one node with the computation resources at other nodes. It is preferable to develop fault tolerant sensor fusion approaches that require little additional computation regardless of any additional communication requirements.

1.2 Sensor Network

A wireless sensor network is a collection of nodes organized into a cooperative network [1, 2]. A wireless sensor network (WSN) consists of tiny, low-powered sensors communicating with each other possibly through multihop wireless links and collaborating to accomplish a common task. A wireless sensor network is a system of small, wirelessly communicating nodes where each node is equipped with multiple components [5]. The nodes communicate wirelessly and often self-organize after being deployed in an ad hoc fashion. Such a network is envisioned to integrate the physical world with the Internet and computations. The power supply on each node is relatively limited, and replacement of the batteries is frequently often not practical due to the large number of the nodes in the network. Each node consists of may contain multiple types of memory (program, data and flash memories), processing capability (one or more microcontrollers, CPUs or DSP chips), have a RF transceiver (usually with a single omnidirectional antenna), have a power source (e.g., batteries and solar cells), and accommodate various sensors and actuators. Sensor nodes collaborate with each other to perform tasks of data sensing, data communication, and data processing [2]. Systems of 1000s or even 10,000

nodes are anticipated. Such systems can revolutionize the way we live and work. Advances in sensor technology and wireless communications have enabled the design and development of inexpensive, large-scale sensor networks that are suitable for different applications, such as health monitoring, environmental monitoring, and battlefields surveillance. A fundamental aspect in the design of WSNs is to keep them functional as long as possible. Because of scarce battery power (or energy), sensors may entirely deplete energy or have remaining energy below some threshold that is required for the sensors to function properly. Those sensors are called faulty as they cannot perform any monitoring task properly. A WSN is said to be functional if at any time there is at least one communication path between every pair of non faulty sensors in the network. The existence of communication paths between pairs of sensors, however, is related to another fundamental property of WSNs, called vertex-connectivity (or simply connectivity). In general, sensing applications are required to be fault-tolerant, where any pair of sensors is usually connected by multiple communication paths. Therefore, network functionality and hence network fault tolerance strongly depends on connectivity. Figure 1 below represents the common architecture of Wireless Sensor Networks and their nodes.

The Wireless Sensor Networks are capable of sensing and forwarding the sensed data, and performing reactions based on received data appropriately. The WSN's consists of sensor nodes and sink nodes. The sensor nodes usually have low costs, limited energy supply and

limited transmission range; they are responsible

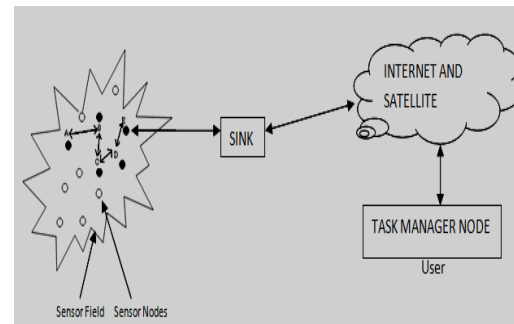


Figure 1: WSN communication Architecture for detecting events or sensing environmental data. The sink nodes are resource-richer nodes with abundant energy sources, higher communication and computation capability, and the ability to perform powerful reactions. When the sink node performs some action then these nodes are called actor nodes. When a sensor node detects some data to be delivered in its monitoring area, it will transmit the event to neighboring nodes, which in turn will forward the event one hop further. The hardware components of a sensor node have been shown in figure 2. In this way, the event reaches the sink. Once the sink node receives the data, it will perform corresponding reactions appropriately. WSNs enable some realistic applications, such as military, phenomenon monitoring, and attack detection [1].

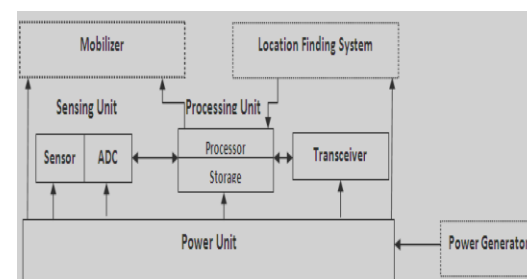


Figure 2: Components of a sensor node

Currently, wireless sensor networks are beginning to be deployed at an accelerated pace. It is not unfair to expect that in coming 10-15 years world will be covered with wireless sensor networks having access to them via the Internet. This can be equivalent being Internet becoming a physical network. This new technology is exciting with unlimited potential for numerous application areas including environmental, medical, military, transportation, entertainment, crisis management, homeland defense, and smart spaces. Since a wireless sensor network is a distributed real-time system a natural question is how many solutions from distributed and real-time systems can be used in these new systems? Unfortunately, very little prior work can be applied and new solutions are necessary in all areas of the system. The main reason is that the set of assumptions underlying previous work has changed dramatically. Most past distributed systems research has assumed that the systems are wired, have unlimited power, are not real-time, have user interfaces such as screens and mice, have a fixed set of resources, treat each node in the system as very important and are location independent. In contrast, for wireless sensor networks, the systems are wireless, have scarce power, are real-time, utilize sensors and actuators as interfaces, have dynamically changing sets of resources, aggregate behavior is important and location is critical. Many wireless sensor networks also utilize minimal capacity devices which places a further strain on the ability to use past solutions. Even though sensor networks are a special type of ad hoc networks, the protocols designed for ad hoc networks

cannot be used as it is for sensor networks due to the following reasons:

- a) The number of nodes in sensor networks is very large and has to scale to several orders of magnitude more than the ad hoc networks and thus require different and more scalable solutions.
- b) The data rate is expected to be very low in WSN and is of statistical in nature. But mobile ad hoc network (MANET) is designed to carry rich multimedia data and is mainly deployed for distributed computing.
- c) A sensor network is usually deployed by a single owner but MANET is usually run by several unrelated entities. [4]
- d) Sensor networks are data centric i.e. the queries in sensor network are addressed to nodes which have data satisfying some conditions and unique addressing is not possible as they do not have global identifiers. But MANET is node centric, with queries addressed to particular nodes specified by their unique addresses.
- e) Sensor nodes are usually deployed once in their life time and those nodes are generally stationary except a few mobile nodes, while nodes in MANET move in an ad hoc manner.
- f) Like MANET sensor nodes are also designed for self configuration, but the difference in traffic and energy consumption require separate solutions. In comparison to ad hoc networks, sensor nodes have limited power supply and recharge of power is impractical considering the large number of nodes and the environment in which they are deployed. Therefore energy consumption in WSN is an important metric to be considered.

g) Sensor networks are application specific. One can't have a solution that fits for all the problems.

1.2.1 WSN Design Factors

There are number of design factors for designing an effective and efficient wireless sensor networks. Some of them have been discussed here: [1]

- Fault Tolerance
- Scalability
- Production Costs
- Hardware Constraints
- Sensor Network Topology
- Environment
- Transmission Media
- Power Consumption

1.3 Fault Tolerance

Fault-tolerance or **graceful degradation** is the property that enables a system (often computer-based) to continue operating properly in the event of the failure of (or one or more faults within) some of its components. Fault tolerance is the ability of a system to deliver a desired level of functionality in the presence of faults [8]. Nodes in WSNs are prone to failure due to energy depletion, hardware failure, communication link errors, malicious attack, and so on. If its operating quality decreases at all, the decrease is proportional to the severity of the failure, as compared to a naïvely-designed system in which even a small failure can cause total breakdown. Fault-tolerance is particularly sought-after in high-availability or life-critical systems. A WSN is said to be fault tolerant if it

remains functional in spite of $\kappa - 1$ sensor failures, where κ is network connectivity. Another important issue in the design of WSNs is what is called sensing coverage, a good indicator of the quality of surveillance of a field of interest [6]. Some sensing applications demand full coverage here every location in the field is covered by at least one sensor. Moreover, to cope with the problem of faulty sensors, duplicate coverage of the same region is desirable. Sensor redundancy is strongly related to the degree of sensing coverage requested by sensing applications, that is, the maximum number of sensors simultaneously covering any location in the field. Notice, however, that sensing coverage and network connectivity are not totally orthogonal concepts. While sensing coverage depends on the sensing range, connectivity relates to the communication range of the sensors. Sensing coverage becomes meaningless if the sensed data cannot be exchanged by the sensors so they reach a central gathering point, called the sink, for further analysis. Thus, for a network to function properly, both sensing coverage and network connectivity should be maintained.

Fault-tolerance is not just a property of individual machines; it may also characterize the rules by which they interact. For example, the Transmission Control Protocol (TCP) is designed to allow reliable two-way communication in a packet-switched network, even in the presence of communications links which are imperfect or overloaded. It does this by requiring the endpoints of the communication to expect packet loss, duplication, reordering and corruption, so that these conditions do not

damage data integrity, and only reduce throughput by a proportional amount.

1.3.1 Fault Tolerance at Different Levels

Five levels of fault tolerance were discussed in [14]. They are physical layer, hardware layer, system software layer, middleware layer, and application layer. On the basis of study, we classify fault tolerance in WSNs into four levels from the system point of view. More specifically, fault tolerance in a WSN system may exist at hardware layer, software layer, network communication layer, and application layer.

Hardware Layer

Faults at hardware layer can be caused by malfunction of any hardware component of a sensor node, such as memory, battery, microprocessor, sensing unit, and network interface (wireless radio).

Software Layer

Software of a sensor node consists of two components: system software, such as operating system, and middleware, such as communication, routing, and aggregation. Software bugs are a common source of errors in WSNs.

Network Communication Layer

Faults at network communication layer are the faults on wireless communication links. Link

faults can be caused by surrounding environments or by radio interference of sensor nodes.

Application Layer

Fault tolerance can be addressed also at the application layer. For example, finding multiple node-disjoint paths provides fault tolerance in routing. The system can switch from an unavailable path with broken links to an available candidate path.

1.3.2 The Need for Fault Tolerant Protocols and Design Issues

Sensor networks share common failure issues (such as link failures and congestion) with traditional distributed wired and wireless networks, as well as introduce new fault sources (such as node failures). Fault tolerant techniques for distributed systems include tools that have become industry standard such as SNMP and TCP/IP, as well as more specialized and/or more efficient methods that have been extensively researched [14]. The faults in sensor networks cannot be approached in the same way as in traditional wired or wireless networks due to the following reasons:

- a) traditional network protocols are generally not concerned with energy consumption, since wired networks are constantly powered and wireless ad hoc devices can get recharged regularly;
- b) traditional network protocols aim to achieve point-to-point reliability, whereas wireless sensor networks are concerned with reliable event detection;

- c) in sensor networks, node failures occur much more frequently than in wired, where servers, routers and client machines are assumed to operate normally most of the time; this implies that closer monitoring of node health without incurring significant overhead is needed;
- d) traditional wireless network protocols rely on functional MAC layer protocols that avoid packet collisions, hidden terminal problem and channel errors by using physical carrier sense (RTS/CTS) and virtual carrier sense (monitoring the channel).
- b) Fault detection: this is to use different metrics to collect symptoms of possible faults;
- c) Fault isolation: this is to correlate different types of fault indications (alarms) received from the network, and propose various fault hypotheses;
- d) Fault identification: this is to test each of the proposed hypotheses in order to precisely localize and identify faults;
- e) Fault recovery: this is to treat faults, i.e., reverse their adverse effects.

Many of the recent fault detection algorithms have either vaguely defined fault models or an overly general fault definition. [6], briefly listed selected faults, and develop a cross validation method for online fault detection based on very broad fault definitions. Looking beyond fault detection and correction techniques, there has been relevant work that frames our thrust to provide fault taxonomy.

1.3.3 Taxonomy of Fault Tolerant Techniques

Recent research has developed several techniques that deal with different types of faults at different layers of the network stack. To assist in understanding the assumptions, focus, and intuitions behind the design and development of these techniques, the taxonomy of different fault tolerant techniques used in traditional distributed systems [15] was given as:

- a) Fault prevention: this is to avoid or prevent faults;

Fault identification and isolation, sometimes are collectively referred to as fault diagnosis. Note that there do exist some techniques that address a combination of all these aspects. In fact, these techniques operate at different layers of the network protocol stack. Most fault avoidance techniques operate in the network layer, adding redundancy in routing paths; a majority of fault detection and recovery techniques operate at the transport layer; and a few fault recovery techniques perform at the application layer, concealing faults during online data processing.

2. RELATED WORK

2.1 Fault Detection: An Overview

Fault detection is the first phase of fault management, where an unexpected failure should be properly identified by the network system. The existing failure detection approaches in WSNs can be classified into two types: centralized and distributed approach.

2.1.1 Centralized Approach

Centralized approach is a common solution to identify and localize the cause of failures or suspicious nodes in WSNs. Usually; a geographically or logically centralized sensor node (in terms of base station [5, 17, and 18], central controller or manager [4], sink) takes responsibility for monitoring and tracing failed or misbehavior nodes in the network. Most these approaches consider the central node has unlimited resources (e.g. energy) and is able to execute a wide range of fault management maintenance. They also believe the network lifetime can be extended if complex management work and message transmission can be shifted onto the central node. The central node normally adopts an active detection model to retrieve states of the network performance and individual sensor nodes by periodically injecting requests (or queries) into the network. It analyzes this information to identify and localize the failed or suspicious nodes. In [17], the base station uses marked packets (containing geographical information of source and destination locations etc) to probe sensors. It relies on nodes response to identify and isolate the suspicious nodes on the routing paths when an excessive packet drops or compromised data has been detected. In addition, the central manager provides a centralized approach to prevent the potential failure by comparing the current or historical states of sensor nodes against the overall network information models (i.e. topology map, and energy map). As a summary, the centralized approach is efficient and accurate to identify the network faults in certain ways.

2.1.2 Distributed Approach

Distributed approach encourages the concept of local decision-making, which evenly distributes fault management into the network. The goal of it is to allow a node to make certain levels of decision before communicating with the central node. It believes the more decision a sensor can make, the less information needs to be delivered to the central node. In the other word, the control centre should not be informed unless there is really a fault occurred in the network. Others address the use of decision fusion centre (i.e. several fusion nodes across the network) to make the final decisions on suspicious nodes in the network [11, 12, 14, 16].

* Node Self-Detection

A self detection model to monitor the malfunction of the physical components of a sensor node via both hardware and software interface has been proposed by number of researchers. Self-detection of node failure is somehow straightforward as the node just observes the binary outputs of its sensors by comparing with the pre-defined fault models. In data dissemination protocols which deliver large segments of data to the entire (or part of the) network, the destination nodes are responsible for detecting the missing packet or the window of missing packets, and communicating the feedback to the source using NACK messaging.

* Neighbor Coordination

Failure detection via neighbor coordination is another example of fault management distribution. Nodes coordinate with their neighbors to detect and identify the network

faults (i.e. suspicious node or abnormal sensor readings) before consulting with the central node. For example, in a decentralized fault diagnosis system [12], a sensor node can execute a localized diagnosis algorithm in steps to identify the causes of a fault. In addition, a node can also query diagnostic information from its neighbors (in one-hop communication range). This allows the decentralized diagnostic framework to scale easily to much larger and denser sensor networks if required. Alternatively, suspicious (or failed) nodes can be identified via comparing its sensor readings with neighbor's median readings. With this motivation [9], developed a localized algorithm to identify suspicious node whose sensor readings have large difference against the neighbors. Although this algorithm works for large size of sensor networks, the probability of sensor faults needs to be small. If half of the sensor neighbors are faulty and the number of neighbors is even, the algorithm cannot detect the faults as efficient as expected. In addition, this approach also requires each sensor node to be aware of its physical location by equipped with expensive GPS or other GPS-less technology. [7, 8] address the accuracy of failure detection via a two-phase neighbor coordination scheme. Similar approach in [6], where a node can listen on its neighbor using WATCHDOG. If data packets have not been transmitted properly by the neighbors of a node it is currently routing to, fail or misbehaving neighbors can be easily detected.

* Clustering Approach

Clustering [14] has become an emerging technology for building scalable and energy

balanced applications for WSNs. [18], derived an efficient failure detection solution using a cluster-based communication hierarchy to achieve scalability, completeness, and accuracy simultaneously. They split the entire network into different clusters and subsequently distribute fault management into each individual region. Intracluster heartbeat diffusion is adopted to identify failed nodes in each cluster. While, [13] adopt an event-driven detection via a manager-agent model supported by management architecture MANNA [3]. In this approach, agents are executed in the cluster-heads with more resources than common nodes. A manager is located externally to the WSN where it has a global vision of the network and can perform complex management tasks and analysis that would not be possible inside the network. Every node checks its energy level and sends a message to the manager or agent whenever there is a state change. The manager then uses this information to build topology map and network energy model for monitoring and detecting the potential failure of the network in future. Furthermore, random distribution and limited transmission range capability of common-node and cluster-heads provides no guarantee that every common-node can be connected to a cluster head. In addition, the transmission costs for network state polling has not been considered in this approach.

Distributed Detection

The basic idea of Distributed Detection is to have each node make a decision on faults (typically binary data of abnormal sensor reading). This approach is especially energy-efficient and ideal for data centric sensor

applications. However, there remain various research challenges in order to achieve a better balance between fault detection accuracy and the energy usage of the network. Usually, the efficiency of such failure detection schemes is counted in terms of node communication costs, precision, detection accuracy and the number of faulty sensor nodes tolerable in the network. In

Clouqueurs work [15], fusion sensors (in terms of manager nodes) coordinate with each other to guarantee that they obtain the same global information about the network before making a decision, as faulty nodes may send them inconsistent information.

3. CONCLUSIONS

Mobile computing is an emerging trend in distributed computing for several applications. The mobility of mobile hosts (MHs), limited

battery power on MH, limited wireless bandwidth, noisy wireless environment, handoff, and limited (or lack stable storage on MH present challenging problems in providing fault-tolerance to such mobile computing systems. Due to the potential deployment in uncontrolled and harsh environments and due to the complex arch, wireless sensor networks are and will be prone to a variety of malfunctioning. The goal of this paper is to identify the most important types of faults, techniques for their detection and diagnosis, and to summarize the first techniques for ensuring efficiency of fault resiliency mechanisms. In addition to a comprehensive overview of fault tolerance techniques in general, and in particular in sensor networks, techniques that ensure fault resiliency during sensor fusion as well as the approach for heterogeneous built-in-self-repair fault tolerance were also discussed.

Figure 3: Comparative Chart for Existing Fault Detection Techniques in Wireless Sensor Networks

Name of Technique	Working Principle	Advantages	Disadvantages
On-line Fault Detection	Approach applied on arbitrary type of fault model, with probability based identification of faulty nodes.	Accuracy in presence of Gaussian noise even for relatively sparse networks.	Effort restricted only to faults in sensors rather than taking other communication and computation units of a node into consideration.
Centralized Fault Detection	Centralized sensor node takes responsibility of identifying and locating the failed or misbehaved node.	Accurate and Fast for identifying faulty node.	Central node becomes single point of data traffic concentration and also causes high volume of message and quick energy depletion
Sympathy [5]	Message flooding approach to pool event data and current states from sensor nodes to a Sympathy node which further transmits to sink node	Fetches data to a sympathy node rather than each node sending directly to sink node.	Message broadcasting creates redundancy of data at sympathy node.
WATCHDOG [6]	A node can listen on its neighbor if data packets have not been transmitted properly by its neighbors it is currently routing to.	Encourages concept of local decision making. More decision a node makes the less will be required to deliver to sink node.	Slow and error prone as it is always difficult to keep an eye on all its neighbors.
FT-DSC Protocol	Clustered based approach in which CH receives info from members only when event of interest occurs	Energy saving by not delivering messages to CHs in every time slot of a frame	Selection of cluster head is always done on basis of level of energy remaining.
FREM [17]	Only requires the touch set on the destination node for quick restart. the remainder of image is transferred after process is restarted on destination.	Allows fast restart of a failed process without requiring the availability of entire checkpoint image.	Issues with this are how to accurately identify the touch set, how to set the tracking window, how to load partial image or destination node.

REFERENCES

- 1] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci, "A Survey on Sensor Networks", proceedings of IEEE Communications Magazine, August 2002.
- 2] Ian F. Akyildiz, Ismail H. Kasimoglu, "Wireless sensor and actor networks research challenges", Elsevier Ad Hoc Networks2, pp. 351–367, 2004.
- 3] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, "Wireless sensor network survey", Elsevier Computer Networks 52, pp. 2292–2330, 2008.
- 4] L. Paradis and Q. Han, "A survey of fault management in wireless sensor networks", Journal of Network System Management, pp. 171-190, 2007.
- 5] N. Ramanathan, K. Chang, R. Kapur, L. Girod, E. Kohler, and D. Estrin, "Sympathy for the sensor network debugger", in SenSys '05: Proceedings of the 3rd international conference on Embedded networked sensor systems, pp. 255-267, 2005.
- 6] A. Mahmood, E. J. McCLUSKEY, "Concurrent Error Detection Using Watchdog Processors", IEEE TRANSACTIONS ON COMPUTERS, pp. 160-174, 1988.
- 7] F. Koushanfar, M. Potkonjak, and A. Angiovanni-Vincentell, "Fault tolerance techniques for wireless ad hoc sensor networks", Sensors 2002, Proceedings of IEEE, pp. 1491-1496, 2002.
- 8] S. Harte, A. Rahman, and K. Razeeb, "Fault tolerance in sensor networks using self- diagnosing sensor nodes", Intelligent Environments, 2005, The IEEE International Workshop, pp. 7-12, June 2005.
- 9] W. L. Lee, A. Datta, and R. Cardell-oliver, "Winms: Wireless sensor network-management system, an adaptive policy-based management for wireless sensor networks", School of Computer Science and Software engineering, University of Western Australia, 2006.
- 10] A. Perrig, R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler, "Spins: security protocols for sensor networks", Wireless Networks, pp. 521-534, 2002.
- 11] Y. Sankarasubramaniam, O. B. Akan, and I. F. Akyildiz, "Esrt: event-to-sink reliable transport in wireless sensor networks", in MobiHoc '03: Proceedings of the 4th ACM International symposium on Mobile ad hoc networking & computing, pp. 177-188, ACM, 2003.
- [12] Q. Han, I. Lazaridis, S. Mehrotra, and N. Venkatasubramanian, "Sensor data collection with expected reliability guarantees", Pervasive Computing and Communications Workshops, pp. 374-378, March 2005.
- [13] L. B. Ruiz, I. G. Siqueira, L. B. e. Oliveira, H. C. Wong, J. M. S. Nogueira, and A. A. F. Loureiro, "Fault Management in Event-Driven Wireless Sensor Networks", in MSWiM '04:

- Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, pp. 149-156, ACM, 2004.
- [14] Ramakrishna Gummadi, Todd Millstein, and Ramesh Govindan, "Declarative Failure Recovery for Sensor Networks," AOSD '07, March 12-16 2007.
- [15] Youngbae Kim, James S. Plank, Jack J. Dongarra, "Fault Tolerant Matrix Operations for Networking of Workstations Using Multiple Checkpointings", High Performance Computing on the Information Superhighway, HPC Asia '97 IEEE, pp. 460-450, 1997.
- [16] Rana Ejaz Ahmed, and Abdul Khaliq, "On the Role of Base Station in Fault-Tolerant Mobile Networks", Electrical and Computer Engineering, Canadian Conference 2004, pp. 473-476, 2004.
- [17] Rana Ejaz Ahmed, and Abdul Khaliq, "A Low-Overhead Checkpointing Protocol for Mobile Networks", Electrical and Computer Engineering, IEEE CCECE 2003, pp. 1779-1782, 2003.
- [18] Yawei Li, Zhilling Lan, "A Fast Restart Mechanism for Checkpoint/Recovery Protocols in Networked Environments", Dependable Systems and Networks with FTCS and DCC, 2008, pp. 217-226, 2008.
- [19] Anas Abu Taleb, Dhiraj K. Pradhan, Taskin Kocak, "A Technique to Identify and Substitute Faulty Nodes in Wireless Sensor Networks", Sensor Technologies and Applications, SENSORCOMM'09, pp. 346-351, 2009.

New Efficient Technique for Compression of ECG Signal

Nidhal K. El Abbadi¹ Abbas M. Al-Bakry²

1 University of kufa
Najaf, Iraq

2 University of Babylon
Babylon, Iraq

Abstract

Data compression is a common requirement for most of the computerized applications. There are number of data compression algorithms, which are dedicated to compress different data formats. This paper examines lossless data compression algorithm for ECG data by using new method to process the ECG image strip, and compares their performance.

We confirming that the proposed strategy exhibits competitive performances compared with the most popular compressors used for ECG compression.

Key words: Data compression, ECG, compression ratio, image compression, lossy compression , lossless compression.

1. Introduction

An ECG is simply a representation of the electrical activity of the heart muscle as it changes with time, usually printed on paper for easier analysis. Like other muscles, cardiac muscle contracts in response to electrical *depolarization* of the muscle cells. It is the sum of this electrical activity, when amplified and recorded for just a few seconds that we know as an ECG.

The amplitude, or voltage of the recorded electrical signal is expressed on an ECG in the vertical dimension and is measured in millivolts (mV). On standard ECG paper 1mV is represented by a deflection of 10 mm. An increase in the amount of muscle mass, such as with left ventricular hypertrophy (LVH), usually results in a larger electrical depolarization signal, and so a larger amplitude of vertical deflection on the ECG.

An essential feature of the ECG is that the electrical activity of the heart is shown as it varies with time. In other words we can think of the ECG as a graph, plotting electrical activity

on the vertical axis against time on the horizontal axis. Standard ECG paper moves at 25 mm per second during real-time recording. This means that

when looking at the printed ECG a distance of 25 mm along the horizontal axis represents 1 second in time.

ECG paper is marked with a grid of small and large squares. Each small square represents 40 milliseconds (ms) in time along the horizontal axis and each larger square contains 5 small squares, thus representing 200 ms. Standard paper speeds and square markings allow easy measurement of cardiac timing intervals. This enables calculation of heart rates and identification of abnormal electrical conduction within the heart (Figure 1).

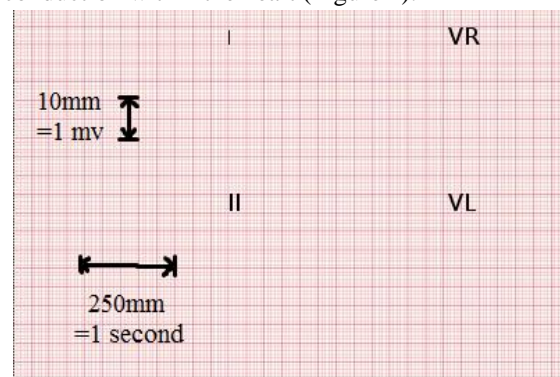


Fig 1: sample of ECG strip

Electrocardiogram (ECG) compression has been the object of numerous research works. Their main objective is to reduce the amount of digitized ECG data as much as possible with a reasonable implementation complexity while maintaining a clinically acceptable signal. Consequently, reduction of digitized ECG data allows improvement of storage capacity in the memory and/or reduces the cost of transmission.

The central goal of electrocardiogram (ECG) data compression techniques is to preserve the most useful diagnostic information while compressing a signal to an acceptable size (Al-Shrouf et al, 2003). Lossless compression is the best choice as long as the compression ratio is acceptable, but it cannot usually offer a satisfactory compression ratio (CR). To obtain significant signal compression, lossy compression is preferable to a lossless compression (AHMED et al, 2007). In this case, compression is accomplished by applying an invertible orthogonal transform to the signal, and one tries to reduce the redundancy present in the new representation. Due to its decorrelation and energy compaction properties and to the existence of efficient algorithms to compute it, discrete cosine transforms and modified discrete cosine transform have been widely investigated for ECG signal compression. Over the years, a variety of other linear transforms have been developed which include discrete Fourier transform (DFT), discrete wavelet transform (DWT) and many more, each with its own advantages and disadvantages (Daubechies, 1998).

One of the most difficult problems in ECG compression and reconstruction is defining the error criterion that measures the ability of the reconstructed signal to preserve the relevant information. As yet, there is no mathematical structure to this criterion, and all accepted error measures are still variations of the mean square error or absolute error, which are easy to compute mathematically, but are not always diagnostically relevant.

ECG signals contain a large amount of information that requires large storage space, large transmission bandwidth, and long transmission time. Therefore, it is advantageous to compress the signal by storing only the essential information needed to reconstruct the signal as in fig 2.

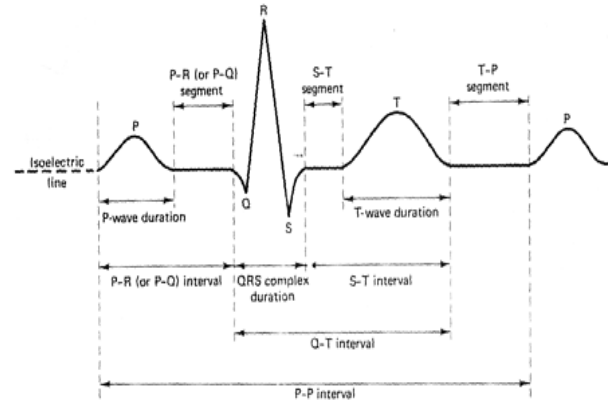


Fig 2: the essential information in ECG strips

Thus, in ECG signal compression, the objective is to represent the signal using fewer bits per sample, without losing the ability to reconstruct the signal. ECG data compression techniques are typically classified into three classes (Cardenas and Lorenzo, 1999). These classes are: direct compression, transform coding, and parameter extraction methods. In the direct compression techniques, redundancy in a data sequence is reduced by examining a successive number of neighboring samples. An example of this approach is the coordinate reduction time encoding system (CORTES). In the transform coding techniques, redundancy is reduced by applying linear transformation to the signal and then compression is applied in the transform domain rather than in the time domain. Examples of this type are Fourier transforms and wavelet transforms. In the parameter extraction techniques, the signal can be reconstructed by extracting a set of parameters from the original signal, which are used in the reconstruction process (Nave and Cohen, 1993).

This paper is organized as follow. Section 2 shows the related work. Section 3 presents an idea about the compression measures. Section 4 displays the research methodology. Finally, the paper is concluded in section 5.

2. ECG Compression Algorithms

Many existing compression algorithms have shown some success in electrocardiogram compression; however, algorithms that produce better compression ratios and less loss of data in the reconstructed data are needed. This project will provide an overview of

several compression techniques and will formulate new emerging algorithms that should improve compression ratios and lessen error in the reconstructed data. Following some of these algorithms:

(Ahmed et al, 2007), present compression technique for ECG signals using the singular value decomposition (SVD) combined with discrete wavelet transform (DWT). The central idea is to transform the ECG signal to a rectangular matrix, compute the SVD, and then discard small singular values of the matrix. The resulting compressed matrix is wavelet transformed, threshold and coded to increase the compression ratio. The results showed that data reduction with high signal fidelity can thus be achieved with average data compression ratio of 25.2:1.

(Chawla, 2009), in this paper Principal Component Analysis (PCA) is used for ECG data compression, denoising and decorrelation of noisy and useful ECG components or signals signal-to-noise ratio is improved

(ALSHAMALI, 2010), this paper proposes a new wavelet-based ECG compression technique. It is based on optimized thresholds to determine significant wavelet coefficients and an efficient coding for their positions. Huffman encoding is used to enhance the compression ratio.

(Bendifallah et al, 2011), An improvement of a discrete cosine transform (DCT)-based method for electrocardiogram (ECG) compression is presented. The appropriate use of a block based DCT associated to a uniform scalar dead zone quantizes and arithmetic coding show very good results.

(Anubhuti et al, 2011), A wide range of compression techniques based on different transformation techniques like DCT, FFT; DST & DCT2 were evaluated to find an optimal compression strategy for ECG data compression. Wavelet compression techniques were found to be optimal in terms of compression.

(ALSHAMALI, 2011), adaptive threshold mechanism to determine the significant wavelet coefficients of an electrocardiogram (ECG) signal is proposed. It is based on estimating thresholds for different sub-bands using the concept of energy packing efficiency (EPE). Then thresholds are optimized using the particle swarm optimization (PSO) algorithm to achieve a target compression ratio with minimum distortion.

3. Compression measures

The size of compression is often measured by CR, which is defined as the ratio between the bit rate of the original signal (boriginal) and the bit rate of the compressed one (bcompressed) (Jalaleddine et al, 1990).

$$\text{Compression Ratio} = \frac{\text{Size after compression}}{\text{Size before compression}}$$

$$\text{Compression Factor} = \frac{\text{size before compression}}{\text{size after compression}}$$

$$\text{saving percentage} =$$

$$\frac{\text{size before compression} - \text{size after compression}}{\text{size before compression}}$$

The problem of using the above definition of CR is that every algorithm is fed with an ECG signal that has a different sampling frequency and a different number of quantization levels; thus, the bit rate of the original signal is not standard. Some attempts were made in the past to define standards for sampling frequency and quantization, but these standards were not implemented, and developers of the algorithms still use rates and quantizes that are convenient to them. The number of bits transmitted per sample of the compressed signal has been used as a measure of information rate. This measure removes the dependency on the quantize resolution, but the dependence on the sampling frequency remains. Another way is to use the number of bits transmitted per second as a compression measure. This measure removes the dependence on the quantizes resolution as well as the dependence on the sampling frequency.

4. Proposed Algorithm

There are many different devices used for ECG, all shared to provide data to physicians to help them analyze ECG data to detect abnormalities, all of these devices draw ECG waveform on the specific paper, the physician can read the ECG strip to decide whether the heart normal or not, and determined if there is an imbalance or diseases facing the patent heart.

In the proposed method, the ECG data used in the image form, and treated as image data, then compressed by three stages. It is very easy to get digital image from ECG devices.

In general, the ECG image (strip) consists from the image background (baselines) and the ECG waveform draws on this strip according to heart activity (essential information).

The baselines are standard, and the distance between the lines are fixed as explained in previous section.

This work aims to isolate the ECG waveform data from background (baselines).

It is clear that the ECG waveform represent the useful data needed for physicians, as opposite of baselines which represent an assistance shape help to interpret the ECG waveform which change according to patient status. Usually the ECG waveform generally draws on baselines with dark color.

The research focused on possibility to isolate ECG waveform data from baselines, and retrieve it's later without loss data (lossless or almost lossless compression).

Isolation of ECG waveform data from baselines data not easy work, due to interference between ECG waveform data and baselines data, this causes either to lost some of ECG waveform data or save some of baselines data with ECG waveform data and make it noisy data, both cases confuse the physician and not help him in diagnosis.

ECG image data (strip) represent as a matrix of pixels, each pixel consist of three bytes (one for red color, other for green color and the last one for blue color), it is possible to imagine these matrix as three channels (channel for each base color).

The experiment proves that the isolation of ECG waveform from the origin image not useful due to lose a lot of data in addition to noise, same thing happened when the origin image converted to gray scale image.

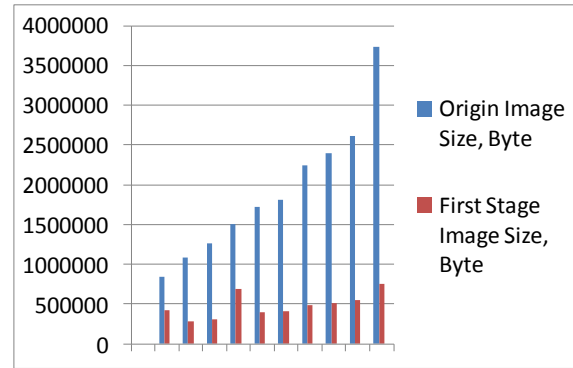


Fig 3: comparing compression ratio in first and second stage

The best way to isolate the ECG waveform data done by divide the origin image data to three sub-images data, each for one color channel, then process the red sub-image to isolate the ECG waveform data, and neglect the other two sub-images, this step reduce size to one third of origin size.

The red sub-image processed to isolate ECG waveform data by filtering followed by applying Sobel edge detection algorithm. This step will reduce the image size to more than 80% of origin size. The Sobel edge detection algorithm is the best algorithm among the other edge detection algorithms for this work to process red channel image.

The result of this stage is binary image, this produced by converting the back ground color (baselines to black color) and the ECG waveform color converted to white color (black and white image need only one bit to represent it's).

The last step is to compress the binary image by using DEFLATE algorithm which is lossless compression algorithm.

The result of average compression percent from applying the proposed algorithm on 9 different ECG images was 99.125% as shown in fig 4.

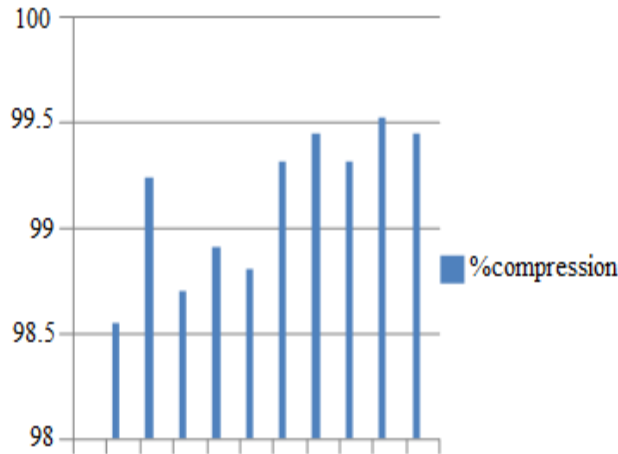


Fig 4 compression rate for 9 different ECG slips

Table 1: Performance of Compression Techniques (Om Prakash, 2012)

Method	CR	CF	SP
RLE	0.384	2.60	61.60
HUFFMAN	0.313	3.19	68.70
LZW	0.224	4.64	77.64
DCT	0.096	10.42	91.68
FFT	0.104	9.62	89.572
DST	0.148	6.76	70.407
DCT-II	0.042	23.81	94.28
FANO	0.684	1.46	31.625
PROPOSED ALGORITHM	0.018	52.8	98.1

and draws it (project) as black dot in standard paper at the same coordinate.

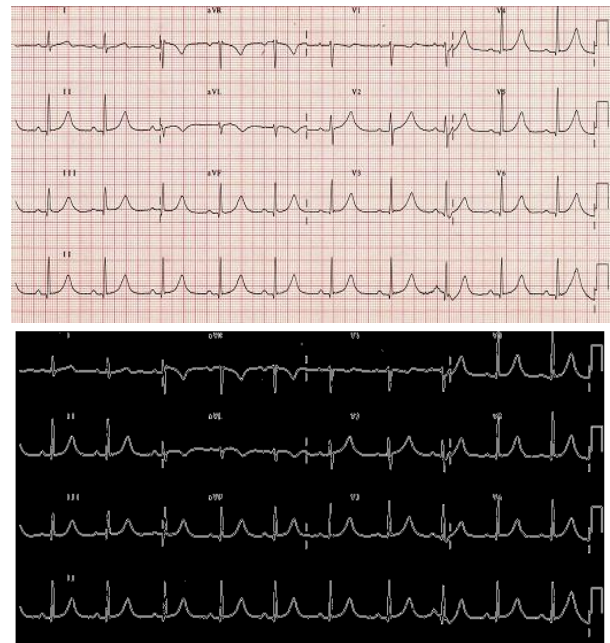
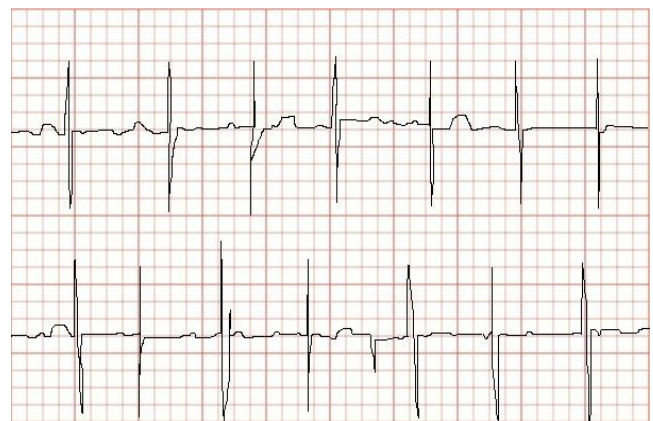


Fig 5: ECG strip and the corresponding binary image



A- Origin image

The decompression process achieved by two steps first one is to reconstruct the binary image. While the second step focus on projection the data of white color in binary image on the standard ECG paper (paper with baselines). This accomplished by reading the coordinates for each white dot in binary image



B- Binary image for image 1



C- Image after decompressed (reconstruct image)

Fig 6: three images before and after compression, and the intermediate step (binary image).

Note: in the Fig 6, Image (A) is drawing image to simulate the ECG strip, and not origin ECG image, just used to measure the performance of algorithm and the quality of the resulted ECG image after decompression..

The quality metrics of proposed compression-decompression algorithm was as in table 2. Where:

PSNR: is the peak signal-to-noise ratio in decibels (dB). The PSNR is only meaningful for data encoded in terms of bits per sample, or bits per pixel.

MSE: The mean square error (MSE) is the squared norm of the difference between the data and the approximation divided by the number of elements.

MAXERR: is the maximum absolute squared deviation of the data (real value signal), from the approximation (reconstructed image).

L2RAT: is the ratio of the squared norm of the signal or image approximation (reconstructed image), to the input signal or image (original image).

Table 2: quality metrics for ECG strip after decompression

Argument	First stage decompression reconstructed binary image	Final stage decompression ECG strip reconstructed
PSNR	∞	63.0108
MSE	0	0.0325
Maxerror	0	1
L2RAT	1	0.6286

There was essentially no false positive diagnosis made on either the compressed or the uncompressed strips, so it can be concluded that the compressor which we evaluated has no measureable influence on diagnostic specificity.

ECG signals that are clean and have a high signal-to-noise ratio (SNR) are relatively easy to interpret, both by a computer and a human healthcare provider.

The new algorithm introduces promise result as highly compression ratio and almost without loss of information visually as fig 6 confirmed. Also, table 2 confirms the similarity of images before and after compression.

With some improvement to this method we can introduce new lossless compression method with high compression ratio.

Digital ECG recording and ECG strip offers potentially higher quality than can be obtained from Holter tape recording, since this method is not subject to wow, flutter, and poor signal-to-noise ratio and low frequency response.

References

- ALSHAMALI, 2010, "Wavelet based ECG compression with adaptive thresholding and efficient coding", *Journal of Medical Engineering & Technology*, Vol. 34, Nos. 5–6, 335–339.
- ALSHAMALI, M. AL-AQIL, 2011, "ECG compression using wavelet transform and particle swarm optimization", *Journal of Medical Engineering & Technology*, Vol. 35, No. 3–4, 149–153.
- Al-Shrouf, A., Abo-Zahhad, M. and Ahmed, S.M., 2003, "A novel compression algorithm for electrocardiogram signals based on the linear prediction of the wavelet coefficients". *Digital Signal Processing*, 13, 604–622.
- Anubhuti Khare, Manish Saxena, Vijay B. Nerkar, 2011, "ECG Data Compression Using DWT", *International Journal of Engineering and Advanced Technology (IJEAT)*, Volume-1, Issue-1.
- Bendifallah, R. Benzid and M. Boulemden, 2011, "Improved ECG compression method using discrete cosine transform", *3191ELECTRONICS LETTERS* Vol. 47 No. 2. DOI 10.1049/el.2010.
- Cardenas-Barrera, J.L. and Lorenzo-Ginori, J.V., 1999, "Mean-shape vector quantizer for ECG signal compression", *IEEE Transactions on Biomedical Engineering*, 46, 62 – 70.
- Daubechies, I., 1988, "Orthonormal bases of compactly supported wavelets", *Communications on Pure & Applied Mathematics*, 41, 909–996.
- M. P. S. Chawla, 2009, "A comparative analysis of principal component and independent component techniques for electrocardiograms", *Neural Comput & Applic*, 18:539–556. DOI 10.1007/s00521-008-0195-1.
- Morteza Moazami-Goudarzi, Mohammad H. Moradi, Ali Taheri, 2005, "Efficient Method for ECG Compression Using Two Dimensional Multiwavelet Transform", *World Academy of Science, Engineering and Technology* 2.
- Nave G., and Cohen, A., 1993, "ECG compression using long-term Prediction", *IEEE Transactions on Biomedical Engineering*, 40, 877 – 885.
- Om Prakash, Vivek Chandra, Pushpendra Singh, 2012, "Design and Analysis of an efficient Technique for Compression of ECG Signal", *International Journal of Soft Computing and Engineering (IJSCE)*, ISSN: 2231-2307, Volume-1, Issue-5.
- S. M. AHMED, A. F. AL-AJLOUNI, M. ABO-ZAHHAD, and B. HARB, 2009, "ECG signal compression using combined modified discrete cosine and discrete wavelet transforms", *Journal of Medical Engineering & Technology*, Vol. 33, No. 1, 1–8.
- S. M. AHMED, Q. AL-ZOUBI, and M. ABO-ZAHHAD, 2007, "A hybrid ECG compression algorithm based on singular value decomposition and discrete wavelet transform", *Journal of Medical Engineering & Technology*, Vol. 31, No. 1, 54 – 61.
- S. M. S. Jalaleddine, C. G. Hutchens, R. D. Strattan and W. A. Coberly. 1990, "ECG Data Compression Techniques – A Unified Approach", *IEEE Trans. on Biomedical Eng.*, vol. 37, 4, 329-341.



Nidhal El Abbadi, received BSc in Chemical Engineering, MSc, and PhD in computer science, worked in industry and many universities, he is general secretary of colleges of computing and informatics society

in Iraq, Member of Editorial board of Journal of Computing and Applications, reviewer for a number of international journals, has many published papers and three published books (Programming with Pascal, C++ from beginning to OOP, Data structures in simple language), his research interests are in image processing, biomedical, and steganography, He's Associate Professor in Computer Science in the University of Kufa – Najaf, IRAQ.



Abbas M. Al-Bakry, Graduate from Computer Science Dept., University of Technology-Baghdad in 1989. Get PhD. In computer science at 2003, from august 2005 to february 2010 become head of computer science department ,and in 2010 till now work as associate

dean of the Information Technology College. Editor in chief topic for the International Journal of Network Computing and Advanced Information Management.

Performance Evaluation of Naive Bayes and Decision Stump Algorithms in Mining Students' Educational Data

Ayinde A.Q¹, Dr Adetunji A.B², Bello M³ and Odeniyi O.A⁴

1. Computer Science Department, Osun State College of Technology,
Esa-Oke, 234035/South West, Nigeria

2 Computer Science and Engineering Department, LAUTECH,
Ogbomoso, 23402/South West, Nigeria

3. Computer Science and Engineering Department, LAUTECH
Ogbomoso, 23402/South West, Nigeria

4. Computer Science Department, Osun State College of Technology,
Esa-Oke, 234035/South West, Nigeria.

Abstract

Educational data mining is an emerging trend, concerned with developing methods for exploring the huge data that come from the educational system. This data is used to derive the knowledge which is useful in decision making which is known as Knowledge Discovery in Databases (KDD). EDM methods are useful to measure the performance of students, assessment of students and study students' behavior etc. In recent years, Educational data mining has proven to be more successful at many of the educational statistics problems due to enormous computing power and data mining algorithms. The main objective of this research is to find out interesting patterns in the educational data that could contribute to predicting student performance. This paper describes how to apply the main data mining methods such as prediction and classification to educational data.

Keywords: Educational Data Mining, Data Mining, Knowledge Discovery in Databases, Algorithms

1. Introduction

The data mining has attracted a great deal of attention in the information technology industry, due to availability of large volume of data which is stored in various formats like files, texts, records, images, sounds, videos, scientific data and many new data formats. There is imminent need for turning such huge data into meaningful information and knowledge. The data collected from various applications require a proper data mining technique to extract the knowledge from large repositories for decision making. Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large volume of data [1].

Data mining and knowledge discovery in databases are treated as synonyms, but data mining is actually a step in the process of knowledge discovery. The sequences of steps indentified in extracting knowledge from data are shown in Figure 1.

The main functionality of data mining techniques is applying various methods and algorithms in order to discover and extract patterns of stored data. These interesting patterns are presented to the user and may be stored as new knowledge in knowledge base. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making. Data mining has been used in areas such as database systems, data warehousing, statistics, machine learning, data visualization, and information retrieval. Data mining techniques have been introduced to new areas including neural networks, patterns recognition, spatial data analysis, image databases and many application fields such as business, economics and bioinformatics. The main objective of this paper is to predict the students' grade by application of classifiers to educational data in Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria. The first section is used to describe the history and current trends in the field of Educational Data Mining (EDM). The second section is the review of past works on educational data mining. The third section covers the research methodology. The fourth section covers the results and discussions. The fifth section covers the conclusion.

2. Reviews on Educational Data Mining

The educational data mining community [2] defines educational data mining as, "Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the setting which they learn in". There are increasing research interests in using data mining techniques in educational filed. This new emerging field, EDM, concerns with developing methods that discover knowledge from data originating from educational environments.

Educational data mining techniques often differ from traditional data mining techniques, in explicitly exploiting the multiple levels of meaningful hierarchy in educational data.

EDM focuses on collection, archiving, and analysis of data related to students' learning and assessment. The analysis performed in EDM research is often related to techniques drawn from variety of literatures [3], including psychometrics, machine learning, data mining, educational statistics, information visualization and computational modeling.

Reviews pertaining to not only the diverse factors like personal, socio-economic, psychological and other environmental variables that influence the performance of students but also the models that have been used for the performance prediction are available in the literature and a few specific studies are listed below for reference.

Walters and Soyibo [4] conducted a study to determine Jamaican high school students' (population n=305) level of performance on five integrated science process skills with performance linked to gender, grade level, school location, school type, student type, and socio-economic background (SEB). The results revealed that there was a positive significant relationship between academic performance of the student and the nature of the school.

Khan [5] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a main objective to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream. The selection was based on cluster sampling technique in which the entire population of interest was divided into groups, or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socio-economic status had relatively higher academic achievement in general.

Hijazi and Naqvi [6] conducted a study on the student performance by selecting a sample of 300 students (225

males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance.

Kristjansson, Sigfusdottir and Allegrante [10] made a study to estimate the relationship between health behaviors, body mass index (BMI), self-esteem and the academic achievement of adolescents. The authors analyzed survey data related to 6,346 adolescents in Iceland and it was found that the factors like lower BMI, physical activity, and good dietary habits were well associated with higher academic achievement.

Moriana et al. [11] studied the possible influence of extra-curricular activities like study-related (tutoring or private classes, computers) and/or sports-related (indoor and outdoor games) on the academic performance of the secondary school students in Spain. A total number of 222 students from 12 different schools were the samples and they were categorized into two groups as a function of student activities (both sports and academic) outside the school day. Analysis of variance (ANOVA) was used to verify the effect of extracurricular activities on the academic performance and it was observed that group involved in activities outside the school yielded better academic performance.

Bray [12], in his study on private tutoring and its implications, observed that the percentage of students receiving private tutoring in India was relatively higher than in Malaysia, Singapore, Japan, China and Srilanka. It was also observed that there was an enhancement of academic performance with the intensity of private tutoring and this variation of intensity of private tutoring depends on the collective factor namely socio-economic conditions. Modeling of student performance at various levels is discussed in [10], [11], and [12]. Ma, Liu, Wong, Yu, and Lee [4] applied a data mining technique based on association rules to find weak tertiary school students (n=264) of Singapore for remedial classes. Three scoring measures namely Scoring Based on Associations (SBA-score), C4.5-score and NB-score for evaluating the prediction in connection with the selection of the students for remedial classes were used with the input variables like sex, region and school performance over the past years. It was found that the predictive accuracy of SBA-score methodology was 20% higher than that of C4.5 score, NB-score methods and traditional method.

Kotsiantis, et al. [8] applied five classification algorithms namely Decision Trees, Perception-based Learning,

Bayesian Nets, Instance-Based Learning and Rule-learning to predict the performance of computer science students from distance learning stream of Hellenic Open University, Greece. A total of 365 student records comprising several demographic variables like sex, age and marital status were used. In addition, the performance attribute namely mark in a given assignment was used as input to a binary (pass/fail) classifier. Filter based variable selection technique was used to select highly influencing variables and all the above five classification models were constructed. It was noticed that the Naïve-Bayes algorithm yielded high predictive accuracy (74%) for two-class (*pass/fail*) dataset.

Al-Radaideh, et al. [13] applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. They used 12 predictive variables and a 4-class response variable for the model construction. Three different classification methods namely ID3, C4.5, and the Naïve Bayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models with the predictive accuracy of 38.33% for four-class response variable.

Cortez and Silva [9] attempted to predict failure in the two core classes (Mathematics and Portuguese) of two secondary school students from the Alentejo region of Portugal by utilizing 29 predictive variables. Four data mining algorithms such as Decision Tree (DT), Random Forest (RF), Neural Network (NN) and Support Vector Machine (SVM) were applied on a data set of 788 students, who appeared in 2006 examination. It was reported that DT and NN algorithms had the predictive accuracy of 93% and 91% for two-class dataset (*pass/fail*) respectively. It was also reported that both DT and NN algorithms had the predictive accuracy of 72% for a four-class dataset.

From these specific studies, we observed that the student performance could depend on diversified factors such as demographic, academic, psychological, socio-economic and other environmental factors.

3. Research Methodology

The data mining research was based on the CRISP-DM (Cross-Industry Standard Process for Data Mining) research approach. The open source software tool WEKA (Knowledge Flow Interface) was used for the research implementation. During the **Business Understanding Phase** the specific University management needs are identified. In the **Data Understanding Phase** the students' educational results at 200level and 500level with their corresponding biodata were collected for observation. During the **Data Preprocessing Phase**, student data collected from the Department of Computer Science and

Engineering at Ladoko Akintola University of Technology Ogbomosho, Oyo State, Nigeria, were organized in a new data mart.

The research sample includes data about 473 students, described by 10 parameters (gender, age, mode of admission (MOA), religion, pre-degree score, student matriculation number, state of origin, 200 level CGPA, 500level GPA and student grade. The provided data is subjected to many transformations –removing parameters that are considered useless (e.g. fields with one value only), replacing fields containing free text with nominal variable (with a number of distinct values), transforming numeric to nominal variables, etc. The data is also being studied for missing values (very few and not important), and obvious mistakes (corrected).

The data mining task is to develop and validate a predictive model that predicts the students' university performance based on the student examination results. The target variable was the "student class", it was constructed as a categorical variable, based on the numeric values of the "student total university score" attribute which has five distinct values - "First Class" (4.5-5.00), "Second Class Upper" (4.49-3.50), "Second Class Lower" (3.49-2.50), "Third Class" (2.49-1.50) and "Pass"(1.49- 1.00). The dataset contains 473 instances (10 classified as First Class, 188 classified as Second Class Upper, 182 classified as Second Class Lower, 64 classified as Third Class, and 30 classified as pass), each described with 10 attributes (1 output and 9 input variables), nominal, numeric and categorical.

During the **Modeling Phase**, two different classification algorithms are selected and applied. Popular WEKA classifiers (with their default settings unless specified otherwise) are used, including a common decision tree algorithm (Decision Stump) and Bayesian classifiers (Naïve Bayes).

4. Tables, Figures and Equations

4.1 Table and Figure

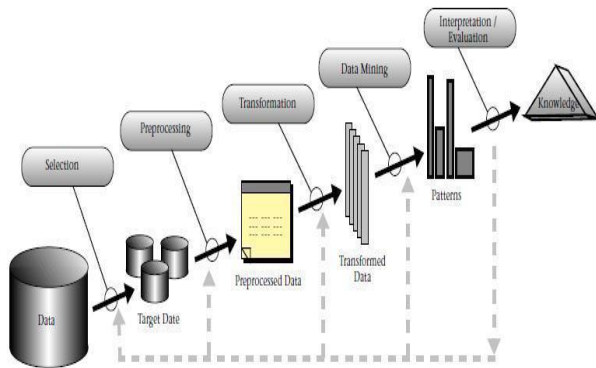


Fig. 1 The steps for extracting knowledge from data.

Table1: Performance evaluation of the classifiers

	1 ST CLAS S		2 ND CLASS UPPER		2 ND CLASS LOWER		THIRD CLASS		PASS	
	NB	DS	NB	DS	NB	DS	NB	DS	NB	DS
TPrate	1	0	0.6	0.9	0.8	0.8	0.8	0.9	0.6	0
FPrate	1	0	0.0	0.5	0.1	0.7	0.1	0.1	0.0	0
PRN	1	0	0.7	0.4	0.7	0.6	0.6	0.6	0.6	0

N.B: Naïve Bayes D.S: Decision Stump TP Rate: True Positive Rate FP Rate: False Positive Rate PRN: Precision

4.2 Equations

$$\text{Bayes theorem: } P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (1)$$

$$\text{Naive Bayes Classifier: } VNB = \arg_{\max} P(V_j) \prod P(a_i | V_j) \quad (2)$$

$$\text{CART: } \phi \left(\frac{s}{t} \right) = 2P_i P_R \sum_{j=1}^{\# \text{ classes}} P \left(\frac{j}{t_L} \right) - P \left(\frac{j}{t_R} \right) \quad (3)$$

4. Results and Discussions

The WEKA Knowledge flow application was used at this stage. Each classifier was applied for two testing options - cross validation (using 10 folds) and percentage split (2/3

of the dataset used for training and 1/3 – for testing). The results for the overall accuracy of the applied classifiers, including True Positive Rate and Precision (the average values for the 10-fold cross validation and split options) are presented in Table I. The results for the classifiers' performance on the five classes are presented on Table.1

The achieved results revealed that the Bayesian classifier (Naïve Bayes) performs best because it was able to predict for all the grades while the decision tree classifier (Decision Stump) cannot predict for First Class Grade and Pass Grade. On the average, the Naïve Bayes precision was above 77percent while Decision Stump precision was 57percent on the average.

Decision Stump inability to predict for First Class Grade and Pass Grade were due to insufficient dataset (training data and test data) and improper calibration of the cross validation fold maker.

Further research efforts will be directed at achieving higher accuracy of the classifiers' prediction by additional transformations of the student educational record, reconstruction of the target variable, tuning of the classification algorithms' parameters, working with larger dataset for training and testing etc.

5. Conclusion

Frequently used Decision tree classifiers and Bayesian classifiers are studied and the experiments are conducted to choose two classifiers for retention data to predict the student's educational performance. On working on student's performance, many attributes have been tested, and some of them are found effective on the prediction. The 200level CGPA was the strongest attribute and the 500level GPA had little effect on the student's final grade. Machine learning algorithms such as the Naïve Bayes and Decision Stump can learn effective predictive models from the student educational data accumulated from the preceding years, because the precision value for student performance was 77%. The practical results show that we can produce an accurate prediction list for the student performance, purposely by applying the predictive models to the records of the new set of 200level students. This study will also work to identify those students which need extraordinary attention to perform well in their discipline.

References

- [1] Usama Fayyad, Gregory Piatesky-Shapiro, and Padhraic Smyth, "From Data Mining to Knowledge Discovery in Databases", American Association for Artificial intelligence, pp. 37-54, 1997.
- [2] Baker R.S.J.D., "Data Mining For Education. In International

- Encyclopedia of Education (3rd edition)", B. MCGAW, PETERSON, P., BAKER Ed. Elsevier, Oxford, UK, 2009. Forman, G. 2003. An extensive Empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3, pp. 1289-1305, 2003.
- [3] www.educationaldatamining.org.
- [4] Y. B. Walters, and K. Soyibo, "An Analysis of High School Students' Performance on Five Integrated Science Process Skills", *Research in Science & Technical Education*, Vol. 19, No. 2, 2001, pp.133 – 145.
- [5] Z. N. Khan, "Scholastic Achievement of Higher Secondary Students in Science Stream", *Journal of Social Sciences*, Vol. 1, No. 2, 2005, pp. 84-87.
- [6] S. T. Hijazi, and R. S. M. Naqvi, "Factors Affecting Student's Performance: A Case of Private Colleges", *Bangladesh e-Journal of Sociology*, Vol. 3, No. 1, 2006.
- [7] Y. Ma, B. Liu, C.K. Wong, P.S. Yu, and S.M. Lee, "Targeting the Right Students Using Data Mining", *Proceedings of KDD, International Conference on Knowledge discovery and Data Mining*, Boston, USA, 2000, pp. 457-464.
- [8] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques", *Applied Artificial Intelligence*, Vol. 18, No. 5, 2004, pp. 411-426.
- [9] P. Cortez, and A. Silva, "Using Data Mining To Predict Secondary School Student Performance", In *EUROSIS*, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.
- [10] A. L. Kristjansson, I. G. Sigfusdottir, and J. P. Allegrante, "Health Behavior and Academic Achievement Among Adolescents: The Relative Contribution of Dietary Habits, Physical Activity, Body Mass Index, and Self-Esteem", *Health Education & Behavior*, (In Press).
- [11] J. A. Moriana, F. Alos, R. Alcalá, M. J. Pino, J. Herruzo, and R. Ruiz, "Extra Curricular Activities and Academic Performance in Secondary Students", *Electronic Journal of Research in Educational Psychology*, Vol. 4, No. 1, 2006, pp. 35-46.
- [12] M. Bray, *The Shadow Education System: Private Tutoring And Its Implications For Planners*, (2nd ed.), UNESCO, PARIS, France, 2007.
- [13] Q. A. Al-Radaideh, E. M. Al-Shawakfa, and M. I. Al-Najjar, "Mining Student Data using Decision Trees", *International Arab Conference on Information Technology (ACIT'2006)*, Yarmouk University, Jordan, 2006.
- A.Q Ayinde** obtained his B.Tech (Computer Science) from LAUTECH (2008). He is a research student in the Department of Computer Science and Engineering, Ladoke Akintola University of Technology (LAUTECH) and a Lecturer in the Department of Computer Science at Osun State College of Technology, Esa-Oke. His research areas include Data Mining, ICT, Soft Computing and Database. He is a member of the following professional bodies Redhat Linux (RHCE Certified) Microsoft (MCITP Certified) and Information and Technology in Infrastructure Library (ITIL Certified).
- A.B Adetunji** obtained her B.Sc (computer Science) from University of Ibadan (1988), M.Sc (Computer Science) from Obafemi Awolowo University Ile-Ife (2002) and obtained her PhD in Computer Science from LAUTECH (2010). She is a Senior Lecturer, presently working in Computer Science and Engineering Department, LAUTECH. She is a member of Computer Professionals Registration Council of Nigeria (CPN). Her research areas include Database, Artificial Intelligent and Data Mining.
- M. Bello** obtained his B.Tech (Computer Science) from LAUTECH (2008). He is a research student in the Department of Computer Science and Engineering, Ladoke Akintola University of Technology (LAUTECH) Presently working as a Programmer in Mathematical Department at Kogi State University, Ayingba. His research areas include Data Mining, ICT, Artificial Intelligent and Database. He is a member of the following professional bodies Redhat Linux (RHCE Certified) and Microsoft (MCITP Certified)
- O.A Odeniyi** obtained his B.Tech (Computer Science) from LAUTECH (1996). He obtained his PGD in Education from National Teachers' Institute Kaduna (2006). He is a research student at the Department of Computer Science and Engineering in LAUTECH. He is a Lecturer 1 at the Department of Computer Science, Osun State College of Technology, Esa Oke. He is a member of the Computer Professionals Council of Nigeria (CPN) and Nigeria Computer Society (NCS). His research areas include Soft Computing, ICT, Artificial Intelligence and Database.

Effectiveness of Online Job Recruitment System: Evidence from the University of the East

Mary Grace G. Ventura¹ and Rex P. Bringula²

^{1,2} College of Computer Studies and Systems, University of the East
Manila, 1008, Philippines

Abstract

The purpose of the study was to develop an online recruitment software that would facilitate the fast and accurate selection of qualified applicants. The Modified Waterfall Model was utilized in the development of the software. The developed software was then evaluated by six groups of respondents to determine the effectiveness of the system in terms of performance, reliability, security, and cost-effectiveness. Descriptive statistics revealed that the software was effective. Analysis of variance showed that there was no significant difference in the evaluation of the six groups on the effectiveness of the developed online recruitment system. This indicated that the six groups of respondents had the same rating on the effectiveness of the software. It implies that the software would perform its function effectively by selecting qualified applicants within a shorter period of time. Thus, it was recommended that the system be adopted by the University. Recommendations to improve the software were also offered.

Keywords: *Cost-effectiveness, Online Recruitment, Performance, Recruitment, Reliability, Security.*

1. Background of the Study

Normally, the quality of people hired is the key metric for measuring the effectiveness of the employment function, but in certain circumstances, the speed of hiring may actually be a more significant contributor to quality hiring [1]. A good productive hiring is not a matter of putting an advertisement in the newspaper, setting up some chairs and tables on the appointed day, and then taking in some resumes to be followed up later. The purpose of a hiring event is to reach out to prospective employees and bring the specific kind of skills and experiences in the organization especially those which cannot be built from within [2]. The question is how this could be made possible.

Basic traditional recruiting begins with processing an application form. In addition, detailed employment standards are set and a job description for each position is offered. Application forms and resumes should be verified with the scope of verification increasing according to the

importance of the position to be filled. Finally, performance tests and other evaluation aids can help in finding the best person for the job [3]. However, this method may take much time and effort.

The traditional hiring process (i.e., from advertisement of job positions to hiring of successful applicants) has its own inherent weaknesses. The advertisement of job positions alone may pose a problem. It is costly to advertise through print media (e.g., newspapers or magazines). Thus, the publication of job positions can only be advertised for a short period of time. Also, the system of submission requires the physical presence of the applicant to hand over the resume. This hampers the application of competent yet geographically secluded job seekers. It is worth noting that the Philippines is a country composed of 7,100 islands. Thus, old recruitment practices might not be suitable in today's competitive environment [2].

This problem can be addressed by an online recruitment system. Online recruitment is set to change the way in which companies recruit their workers. Online recruitment, as a fundamental business process, is the removal of complex and unnecessary paper works, and the introduction of streamlined workflow systems, reliable database applications, and efficient communication channels between job seekers and managers. "At a relatively low cost, the Internet offers employers and job searchers access to detailed and up-to-date information about job searchers and job vacancies in different locations around the world" [4, p. 94]. In this manner, companies can commit themselves to equal opportunities as job providers and can attract new and qualified candidates [2].

The University of the East (UE) is committed to this advocacy. UE over the last half-century evolved into one of the largest private institutions for higher learning in the country. It has three campuses which are situated in Manila, Caloocan, and Quezon City. One of the institutional objectives of the University is to program its

educational offerings to meet the needs of modernizing society as well as the demands of the emerging world of the 21st century (Faculty Manual of UE, 2002). To achieve this goal, much will depend on the discovery, creation, integration, transmission, and application of knowledge, as faculty members teach, conduct research, and perform academic, clinical and public services, as well as get involved in outreach and extension activities. The management and its staff also play a vital role in maintaining, managing, and organizing the list of responsibilities and tasks that this organization should accomplish. Hand-in-hand, they work hard to put into action their shared vision and mission.

To realize the desired outcomes, the University hires only competent and dedicated faculty members and non-teaching staff. Challenged by the present problems in recruiting qualified applicants, this study was conceived. It aimed to develop an online job recruitment system that would be utilized in the University of the East. The effectiveness of the software was evaluated in terms of performance, reliability, security, and cost-effectiveness. Furthermore, it determined whether the evaluation of the effectiveness of the software significantly differed as perceived by the six groups of respondents.

It was hypothesized that there is no significant difference in the respondents' evaluation on the effectiveness of the developed Online Job Recruitment System for the University of the East in terms of performance, reliability, security, and cost-effectiveness.

2. Literature Review

Parry and Tyson [5] conducted a study on the recruitment activities of corporations for a period of six years with the use of survey and interview methods, questions were asked as to why the respondents utilized or did not employ online recruitment, whether they predicted their use of the Internet for recruitment to change, and what impact they expected Internet recruitment to have on the use of other recruitment methods. Human Resource directors and managers, finance directors, managing directors and recruitment specialists from a sample of UK organizations with over 25 employees were the respondents of the study. There were 25,524 responses in the survey and twenty (20) HR or resourcing managers were interviewed.

The results of the survey showed that the most common reasons of using corporate or commercial websites in their recruitment were cost-effectiveness (75%), ease of use for

candidates (64%), a larger candidate pool (53%), ease of use for the organization (52%), speed to hire (52%), and company policy (50%). On the other hand, the less common reasons were success in finding candidates (44%) and keeping ahead of competitors (32%).

The interview results supported these findings. The interviewees disclosed that the company used online recruitment in order to reduce recruitment costs and to improve the efficiency of the recruitment process. These were the significant drivers for the adoption of both corporate and commercial websites. In the same study, about half of the interviewees suggested that the need to "move with the times" or to "keep up with other organizations" be the primary motivation to adopt online recruitment.

It was also found out that interviewees who had positive experiences of online recruitment believed that this form of recruitment could minimize the time taken to hire employees because posting advertisements on the Internet was faster. Interviewees also described the online recruitment as cost-effective since it reduced the use of paper. As regards the success of the online recruitment, mixed results were reported. Interviewees described the success of online recruitment in terms of sufficient generation of shortlist candidates or the ability to attract good-quality applicants. However, this was not true for all who were interviewed.

Kar and Bhattacharya [6] conducted a similar study. They determined the factors that could contribute to the effectiveness of the job portals and the elements of the job portal that could help increase the users' satisfaction on the use of the portal. Survey method and personal interviews were conducted to meet these objectives. Two hundred fifty (250) purposively selected respondents participated in the study.

The study established that the age group 18-22 years old and 33 years old and above had more likely to search jobs through job portals. The curriculum vitae distribution and the face-to-face interaction were the key factors contributing to the popularity of job portals. Respondents believed that they would be more satisfied with the job portals if chat facility, online test, and help desk/call center facilities were available at the job portals.

Sylvia and Mol [7] examined the perceptions of applicants towards web-based procedures. There were 1,360 respondents who were applicants for jobs in multinational financial services organizations in the United Kingdom, the Netherlands, and Belgium. With respect to the

demographics, it was disclosed that external applicants (as opposed to the internal applicants), Belgian (as opposed to Dutch), and Internet savvy (as opposed to less savvy) candidates were more satisfied with the online application procedure. It was also revealed that the features of the website, perceived efficiency, and user-friendliness were the most important determinants of applicant satisfaction.

Haroon and Zia-ur-Rehman [8] also investigated online recruitment in Pakistan. A total of sixty-five (65) respondents from small and large firms of the different sectors of the industries in Pakistan participated in the study. Data were collected through telephone interviews. Haroon and Zia-ur-Rehman [8] showed that preference was given to small firms as compared to large firms in terms of using internet recruitment. They also showed that large firms had their own websites and use them for recruitment as compared to small firms. They also revealed that online recruitment became a new medium that was going to replace the other traditional sources of recruitment because online recruitment offered reduced recruitment costs, time-saving capability, quick response features in checking application status, and online resume development.

Lastly, the effect of e-recruitment on the design of the recruitment process was also examined. Holm [9] made three explorative case studies in three large organizations in Denmark from 2008 to 2010. The companies selected were all multinational corporations originating from, and with headquarters in, Denmark. The study investigated the possible changes in the tasks, subtasks, and activities of the business process of recruiting which was attributed to the use of e-recruitment. Using in-depth, face-to-face, semi-structured interviews with a number key informants (e.g., Human Resource partners and employer-brand managers, recruitment planners, and other people who were involved in the recruitment process), it was shown that e-recruitment transformed the traditional recruitment process into a time- and space-independent, collaborative hiring process. These findings were consistent with those in the studies previously presented. It was also shown that the most significant changes in the process were in the sequence and increased divisibility of the main recruitment tasks and subtasks.

3. Methodology

3.1 Research Design, Subjects, Sample, and Sampling

Design

Descriptive approach was used to collect data to determine if the objectives of the study were met. The respondents of the study came from the following groups.

- Human Resource Department (HRD) staff were responsible in operating the system. In the said department, only four (4) among the staff performed the screening of job application forms.
- The head of the different departments with teaching and non-teaching staff knew the manpower needs of the department and they issued requests for vacant positions to be filled up. Of the different departments in the university, fifty-two (52) were appointed as department heads. Only twenty percent (20%) or ten (10) were chosen as respondents of the study.
- Newly hired teaching and non-teaching staff used the developed system in their job application in the University of the East. A total of one hundred fifty-five (155) was newly hired in 2010. Only twenty percent (20%) or thirty-one (31) were chosen as respondents of the study.
- Regular/full-time teaching and non-teaching staff evaluated the online recruitment system in terms of speed, ease, and cost-effectiveness. There was a total of six hundred fifty-four (654) regular teachers and staff. Ten percent (10%) or sixty-five (65) regular/full-time teaching and non-teaching staff were chosen as respondents of this study.

Purposive sampling method was utilized in choosing the respondents. The following criteria were set in choosing the respondents.

- For the Human Resource Department (HRD) staff, their job description includes processing of job application form.
- For the Department Heads, they were authorized to issue requests for manpower needs.
- For the newly hired teaching and non-teaching staff, whose bachelor's degrees were somewhat related to I.T. and who had a background on how online recruitment worked, they would attest to the significance of the developed online recruitment system.
- For the regular/full-time teaching and non-teaching staff who applied and who were hired through the use of the previous system, they could compare the previous

method of job application system with the present developed online job recruitment system.

3.2 Research Instrument

A survey questionnaire was utilized as the research instrument. It consisted of two parts. The first part gathered the demographics of the respondents, such as sex, age, gender, length of service in the University, educational attainment, designation, level of computer literacy, and frequency of Internet usage. The second part gathered the perceptions of the respondents on the effectiveness of the online job recruitment system. A 5-point Likert-scale type was used to measure the effectiveness of the online job recruitment system. The scale, mean range, and the verbal interpretation are shown in Table 1.

Table 1. Rating Scale, Mean Range, and Verbal Interpretation

Scale	Mean Range	Verbal Interpretation
5	4.51-5.00	Very Effective
4	3.51-4.50	Effective
3	2.51-3.50	Moderately Effective
2	1.51-2.50	Slightly Effective
1	1.00-1.50	Not Effective

The content of the questionnaire was validated by Computer Science faculty members, Human Resource Department personnel, an Advertising expert, and a Psychology professor, all of whom were not involved in the study. Comments and suggestions were taken into consideration to improve the questionnaire.

3.3 Data-Gathering Procedure

An observation was conducted on how the University was practicing its traditional way of hiring. This observation prompted the researcher to strengthen the need to develop a new and an effective way of attracting the best applicants for the job positions.

An interview was done with the HRD staff who have worked with the University for at least five (5) years. First, a letter was sent to the Human Resource Department Head for approval to do the interview. Then, a face-to-face interview was conducted on the scheduled date. A questionnaire was prepared to make sure that essential data needed for the study were gathered. The interview was held at the Office of the Head of the Department.

3.4 Software Process Model, the Developed Software, and Software Evaluation

The development of the software followed the Modified Waterfall Model [10]. This model was adapted since it ensured that all flaws were addressed before proceeding to the next step of the software development life cycle. This was highly applicable to the context of the study since it would minimize the time needed to interview the respondents of the study. In this manner, their job would not be interrupted.

With the use of the developed software, the Human Resource Department could publish their job advertisements. All advertisements were supported by agreed job descriptions, person specification, and further information about the department or section in which the job was located. Department Heads of different colleges looking for candidates in the field of teaching were also capable of posting their job requirements and person specifications for the job vacancy. Both concerned parties could create job-ads using templates.

The system automatically shortlisted the application forms according to years of experience, skills, and educational attainment relevant to the job vacancy. The system was able to provide sending of automated responses to qualified applicants. On the other hand, applicants could view further particulars of the post, which included background information concerning the University, faculty, department, the job description, and person specifications. They could fill out the application form online, and easily update their profiles. Sample screen shot is shown in Figure 1.

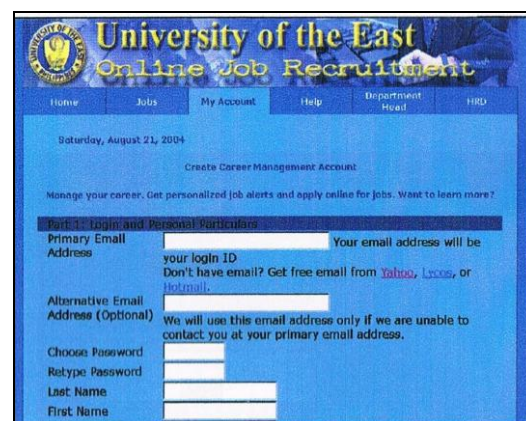


Fig. 1 A sample screen shot

However, the system was not capable of doing such task as detecting errors on entered data of applicants. The system could not determine whether the data provided by the applicants were correct. Typographical errors were not detected by the system. The system could not also verify if all skills, knowledge, experience, and education specified by the users were true.

Software evaluation was conducted to find out the acceptability of the software. Software evaluation was based on FURPS (Functionality, Usability, Reliability, Performance, and Security) Quality Factors [11]. Only three quality indicators (Performance, Reliability, and Security) were selected from FURPS because only these three criteria were deemed applicable in this study. Moreover, Performance and Functionality were treated as one criterion since both referred to the effective processes of the software. Lastly, Cost-effectiveness was included since it was asserted that the software could save costs to the University.

Performance refers to the capability of online recruitment system to generate a list of applicants suitable for the job vacancy in a quick manner. The software could facilitate efficient filtering of qualified candidates and could provide a well-matched shortlist of qualified applicants. Moreover, the system came up with recruitment policies which were free from errors and inconsistencies.

Reliability is the extent to which a program can be expected to perform its intended function [12]. The system could produce the required results with precision in terms of the quality of the applicant being hired for the job vacancy. It could filter out applicants' forms without anomalies and any form of discrimination. Moreover, it provided correct results with minimum expenditure of time and effort.

Security is the availability of mechanisms that control or protect programs and data. It had the mechanism to protect and control programs and data, and it provided its users passwords as security against unauthorized users [12]. It was capable of securing the confidentiality of data received from different types of applicants.

Cost-effectiveness refers to the justification on the amount spent for the investment which commensurate with effectiveness. Online recruitment was more inexpensive compared to the other means of recruitment such as print media advertising. It could provide sufficient space to outline enough information for the candidate to understand the position and the institution. Moreover, it

was capable of reaching potential employees at a longer period of time.

3.4 Statistical Treatment of Data

Frequency counts, percentage, and arithmetic mean were used to describe the data. Analysis of Variance (ANOVA) was utilized to find the significant differences on the evaluation of the six groups of respondents on the effectiveness of the system. A 5% level of probability and 95% reliability were employed to determine the significance of the findings.

4. Findings and Discussion

4.1 Demographic Profile of the Respondents

Table 2 presents the percentage distribution of the respondents' demographic profile. As can be gleaned from the table, there were four (4) or 4% HRD personnel, ten (10) or 9% department heads, thirty-five (35) or 32% regular faculty members, twenty-one (21) or 19% newly hired faculty members, thirty (30) or 27% regular non-teaching staff and ten (10) or 9% newly hired non-teaching staff chosen as respondents of the study. Majority of the respondents were regular faculty members. The number of respondents for each designation was not randomly selected. It was based on purposive sampling through inquiry before the respondents were allowed to evaluate the system. Those who fell under the criteria specified by the study were chosen.

As to gender, fifty-nine (59) or 54% of the respondents were female, while fifty-one (51) or 46% were male. In terms of age of the respondents, majority of the respondents were below thirty (30) years old ($f = 54$, 49%) and about 47% ($f = 52$) belonged to the age bracket of 31-40 years old. With regard to the length of service to the University, most of the respondents ($f = 58$, 53%) have just been newly hired or have just started working at the University. As per educational attainment, majority of the respondents, which was equal to forty-six (46) or 42%, were bachelor's degree holders and were enrolled in the graduate school. About thirty-three (33) or 30% were master's degree holders. Only one (1) or 1% of the respondent was a doctoral degree holder.

Table 2. Demographic Profile of the Respondents

DEMOGRAPHIC PROFILE		
	<i>f</i>	%
Designation		
HRD	4	4
DH	10	9
RFM	35	32
NHFM	21	19
RNTS	30	27
NHNTS	10	9
Gender		
Male	51	46
Female	59	54
Age		
Below 30 years old	54	49
31-40 years old	52	47
41-50 years old	3	3
51-60 years old	1	1
Length of Service		
0-5 years	58	53
6-10 years	40	36
11-15 years	12	11
16 years and above	0	0
Educational Attainment		
Bachelor's Degree	17	15
Bachelor's Degree with Masters	46	42
Master's Degree	33	30
Master's Degree with Doctoral Units	13	12
Doctoral Degree	1	1
Level of Computer Literacy		
Beginner	29	26
Average	65	59
Expert	16	15
Frequency of Internet Usage		
Always	29	26
Often	46	42
Seldom	35	32
Never	0	0
TOTAL		110
		100
Legend:		
HRD	= Human Resource Department Personnel	
DH	= Department Head	
RFM	= Regular Faculty Members	
NHFM	= Newly Hired Faculty Members	
RNTS	= Regular Non-Teaching Staff	
NHNTS	= Newly Hired Non-Teaching Staff	
<i>f</i>	= frequency	
%	= percentage	

Regarding computer literacy, majority of the respondents ($f = 65, 59\%$) were of average level. There was only a small percentage (15%) of respondents who were experts. They were professional end-users who had formal knowledge and education about computer. Nonetheless, most of the respondents ($f = 81, 74\%$) were computer literate and were able to use the proposed software. Lastly,

Table 2 also shows that all respondents accessed the Internet. Thus, the respondents of the study could easily adapt to the proposed software.

4.2 Effectiveness of the Developed Online Job Recruitment System

Table 3 presents the status of the online recruitment system in terms of Performance, Reliability, Security, and Cost-effectiveness of the software as perceived by the respondents. In terms of Performance, NHFM and NHNTS gave a “very effective” rating on this criterion. NHNTS gave the highest mean rating (mean = 4.77, Verbal Interpretation (V.I.) = “Very Effective”) while HRDP gave the lowest rating (mean = 3.93, V.I. = “Effective”). Nonetheless, the rating of HRDP was still “effective”. Thus, all respondents agreed that the software was acceptable in terms of the performance of the software. This implies that the software could generate a list of applicants for a job vacancy in a quick manner and could also facilitate efficient filtering of qualified candidates.

Table 3. Effectiveness of the Developed Online Recruitment System in terms of Performance, Reliability, Security, and Cost-Effectiveness

Software Criteria	Respondents	Mean	V.I.
Performance	HRDP	3.93	Effective
	DH	4.20	Effective
	RFM	4.50	Effective
	NHFM	4.65	Very Effective
	RNTS	4.49	Effective
	NHNTS	4.77	Very Effective
Reliability	HRDP	4.63	Very Effective
	DH	4.48	Effective
	RFM	4.30	Effective
	NHFM	4.46	Effective
	RNTS	4.23	Effective
	NHNTS	4.22	Effective
Security	HRDP	3.75	Effective
	DH	3.97	Effective
	RFM	4.11	Effective
	NHFM	4.19	Effective
	RNTS	4.04	Effective
	NHNTS	4.07	Effective
Cost-Effectiveness	HRDP	4.00	Effective
	DH	4.07	Effective
	RFM	4.08	Effective
	NHFM	3.78	Effective
	RNTS	4.16	Effective
	NHNTS	4.30	Effective

The reliability of the software was also “effective” based on the evaluation of DH (mean = 4.48), RFM (mean = 4.30), NHFM (mean = 4.46), RNTS (mean = 4.23), and NHNTS (mean = 4.22). It is important to note that HRDP rated the software as “very effective” (mean = 4.63). This is a good indication that the software met its intended function with precision without using too much time and

effort. The security features of the software were also evaluated as “effective” by all the respondents (HRDP, mean = 3.75; DH, mean = 3.97; RFM, mean = 4.11; NHFM, mean = 4.19; RNTS, mean = 4.04; and NHNTS, mean = 4.07). In other words, the mechanisms of the software to control or protect data from unauthorized access were effective. Thus, it can be assured that the applicants’ information could not be exploited and their privacy was maintained.

As regards the Cost-effectiveness of the developed online recruitment system, NHNTS gave the highest mean rating of 4.30 while NHFM gave the lowest mean rating of 3.78. Nonetheless, all of the evaluation had verbal rating of “effective”. This revealed that they perceived the software to be “effective” in reducing the cost of the recruitment process. This finding supported the studies of Brencic and Norris [4], Parry and Tyson [5], and Haroon and Zia-ur-Rehman [8].

Table 4 shows the summary of results on the effectiveness of the developed online recruitment system. The four (4) criteria were rated by the respondents as “effective” with the overall mean rating of 4.22. Though Security got the lowest mean rating, its verbal interpretation was still “effective”. On the other hand, Performance of the software got the highest mean rating of 4.42. This indicates that the developed Online Recruitment System was working effectively based on all criteria used in the study.

Table 4. Summary of Results on the Effectiveness of the Developed Online Recruitment System

Software Criteria	Composite Mean	Verbal Interpretation
Performance	4.42	Effective
Reliability	4.39	Effective
Security	4.02	Effective
Cost-Effectiveness	4.08	Effective
Overall Mean	4.22	Effective

4.3 Differences in the Evaluation of the Respondents as regards the Effectiveness of the Developed Online Job Recruitment System

Analysis of Variance (ANOVA) was used on the perception of the respondents on the different software criteria. As shown in Table 5, there was no significant difference in the perceptions of the respondents in the software criteria in terms of Performance (F -value = 2.28; $p < 0.05$), Reliability (F -value = 1.76; $p < 0.05$), Security (F -value = 0.37; $p < 0.05$), and Cost-effectiveness (F -value = 0.77; $p < 0.05$).

The findings of ANOVA disclosed that the perceptions of the six groups of respondents in terms of the different software criteria did not differ from one another. In other words, they all had the same evaluation of the software in terms of the four criteria. This implies that they all perceived that the system was effective in terms of performance, reliability, security, and cost-effectiveness.

Table 5. Differences in the Evaluation of the Respondents as regards the Effectiveness of the Developed Online Job Recruitment System

Software Criteria	Respondents	Mean	F -value	p -value	Significance
Performance	HRDP	3.93	2.28	0.062	Not Significant
	DH	4.20			
	RFM	4.50			
	NHFM	4.65			
	RNTS	4.49			
	NHNTS	4.77			
Reliability	HRDP	4.63	1.76	0.128	Not Significant
	DH	4.48			
	RFM	4.30			
	NHFM	4.46			
	RNTS	4.23			
	NHNTS	4.22			
Security	HRDP	3.75	0.37	0.866	Not Significant
	DH	3.97			
	RFM	4.11			
	NHFM	4.19			
	RNTS	4.04			
	NHNTS	4.07			
Cost-Effectiveness	HRDP	4.00	0.77	0.574	Not Significant
	DH	4.07			
	RFM	4.08			
	NHFM	3.78			
	RNTS	4.16			
	NHNTS	4.30			

It is also suggested that upon the implementation of the software, the users of the software (i.e., the respondents of the study) adopt the system. Since the users have the same level of user-experience in the system, fewer errors may be committed while using the software. It can also be argued that because of the respondents’ positive evaluation on the system, there will be less resistance in changing the recruitment process. Nonetheless, it is recommended that follow-up studies be conducted to shed more light on this matter.

5. Conclusions and Recommendations

In view of the foregoing findings, the null hypothesis stating that there is no significant difference in the respondents’ evaluation on the effectiveness of the developed Online Job Recruitment System for the University of the East in terms of performance, reliability, security, and cost-effectiveness was accepted. It is concluded that the developed software was effective in selecting qualified applicants within a shorter period of

time. Hence, it would become a significant contributor to quality hire. It is also concluded that Performance, Reliability, Security, and Cost-effectiveness could be utilized as criteria in evaluating online recruitment software.

Based on the conclusions drawn, it is recommended that the software be implemented in the company. After implementation, it is proposed that the impact of the system on the recruitment processes in the University be determined and that the changes in the recruitment processes with introduction of the online recruitment software be investigated. It is suggested that studies on the user-experience, adoption or non-adoption of the software, and errors committed in using the software be conducted and that Usability criterion of FURPS be assessed.

In terms of enhancement of the software, it is strongly recommended that an online exam be incorporated in the recruitment and that extra security features such as the level of access classified according to the position in the company also be incorporated in the software.

Acknowledgments

The authors are greatly indebted to Dr. Ester A. Garcia, Dr. Linda P. Santiago, Dr. Olivia C. Caoili, Dr. Leila Gano, Dr. Socorro R. Villamejor, and Dean Rodany A. Merida. This paper is partially funded by the University of the East.

References

- [1] J. Sullivan, "Gaining a competitive advantage through increasing the speed of hire", <http://www.gatelyconsulting.com/PP15JS28.HTM>
- [2] O. Shafique, "Recruitment in the 21st Century", *International Journal of Contemporary Research in Business*, Vol. 4, No. 2, 2012, pp. 887-901.
- [3] R. L. Bintliff, *Crime proofing your business*, New York: McGraw Hill, Inc., 1994.
- [4] V. Brencic, and J. B. Norris, "Employers' online recruitment and screening practices", *Economic Inquiry*, Vol. 50, No. 1, 2012, pp. 94-111.
- [5] E. Parry, and S. Tyson, "An analysis of the use and success of online recruitment methods in the UK", *Human Resource Management Journal*, Vol. 18, No. 3, 2008, pp. 257-274.
- [6] A. Kar, and S. Bhattacharya, "E-recruitment and customer satisfaction: An empirical study in and around Kolkata", *The Icfai Journal of Management Research*, Vol. 8, No. 2, 2009, pp. 34-54.
- [7] H. Sylva, and S. T. Mol, "E-recruitment: A study into applicant perceptions of an online application system", *International Journal of Selection and Assessment*, Vol. 17, No. 3., 2009, pp. 311-323.
- [8] M. Haroon, and M. Zia-ur-Rehman, "E-recruitment: Across the small and large firms in Pakistan", *Interdisciplinary Journal of Contemporary Research in Business*, Vol. 2, No. 1, 2010, pp. 179-191.
- [9] A. B. Holm, "E-recruitment: Towards an ubiquitous recruitment process and candidate relationship management", *Zeitschrift für Personalforschung*, Vol. 26, No. 3, 2012, pp. 241-25.
- [10] N. Mo. A. Munassar, and A. Govardhan, "A comparison between five models of software engineering", *International Journal of Computer Science Issues*, Vol. 7, Issue 5, 2010, pp. 94-101.
- [11] R. B. Grady, and D. L. Caswell, *Software metrics: Establishing a company-wide program*, New Jersey, USA: Prentice Hall, 1987.
- [12] R. S. Pressman, *Software engineering: A practitioner's approach*, 5th edition. New York: McGraw-Hill, 2001.

Mary Grace G. Ventura received her bachelor's degree in Computer Science from City College of Manila and her Master's degree in Information Technology at Adamson University. Currently, she is pursuing her doctorate in Technology Management at the Technological University of the Philippines. She is presently the Chairperson of Entertainment and Multimedia Computing of the College of Computer Studies and Systems (CCSS), University of the East. She has been teaching Web Page design, Database Management, Systems Quality Assurance and Computer Ethics for almost 10 years at CCSS. Her research interests include blended learning, e-learning, and web-based systems such as e-recruitment.

Rex P. Bringula, Ph.D., is a professor at the University of the East (UE), College of Computer Studies and Systems. He received his BS Computer Science degree from UE as a Department of Science and Technology scholar. He obtained his Master in Information Technology and Ph.D. in Technology Management degrees from the Technological University of the Philippines. He has been involved in conducting school- and government-funded research projects, and he has been presenting papers in local and international conferences. He is an active editorial board member of local and international journals. His research interests are in computer science/IT education, affective computing, Internet studies, cyber-behavior, web usability, and environmental issues. Recently, his paper received a MERIT Paper Award given by the International Conference on Software Engineering 2013 at Hong Kong, China.

Evaluation of Different Query Expansion Techniques by using Different Similarity Measures in Arabic Documents

Hayel Khafajeh¹ and Nidal Yousef²

¹ CIS Department, Zarqa University
Zarqa, 009625, JORDAN

² CIS Department, AL- ISRA University
AMMAN, 00926, JORDAN

Abstract

Millions of users search daily for their needs using internet and other information stores, they search by writing their queries. Unfortunately, these queries may fail to reach to their needs, this fail known as word mismatch. One way of handling this Word mismatch is by using a thesaurus, that shows (usually semantic) the relationships between terms. The main goal of this study is to design and build an automatic Arabic thesaurus using Local Context Analysis technique that can be used in any special field or domain to improve the expansion process and to get more relevance documents for the user's query. This technique can be used in any special field or domain to improve the expansion process and to get more relevant documents for the user's query. Results of this study were compared with the classical information retrieval system.

Two hundred and forty two Arabic documents and 59 Arabic queries were used for building the requirements of the thesaurus, such as inverted File, indexing, term-term co-occurrence matrix, etc. All of these documents involve computer science and information system vocabulary.

The system was implemented in ORACLE 11 g environment and run on Pentium-4 laptop with 2.13GHz speed, 2.86MB RAM memory, and hard disk capacity of 500GB.

The study has shown that the Local Context Analysis technique improved the retrieval in a remarkable way better than the classical retrieval method.

Keywords: *Query Expansion, Local Context Analysis technique, Co-Occurrence, Similarity, Thesaurus, Indexing, Natural language (NL), Synonyms.*

1. Introduction

Information retrieval (IR) deals with the representation, storage, organization and access of information items. The representation and organization of the information items

should provide the user with easy access to the information that he is interested in.

Unfortunately, characterization of the user information-need is not a simple task because of the language of the user. [13]

The word thesaurus has Greek and Latin origins and is used as a reference to a treasury of words. [7]

The Thesaurus involves some normalization of the vocabulary and includes a structure much more complex than a simple list of words and their synonyms, the popular thesaurus published by Peter Roget [26].

A thesaurus can be automatically established by analyzing the relationships among documents and statistics of term co-occurrences in the documents. From the thesaurus constructed in this way, one will be able to obtain synonyms or related terms given a user query. Thus, these related terms can be used for supplementing users' original queries [14].

Is a statistical approach where the occurrences of terms in documents, chapters or some other unit are computed? The closer the words occur, the more significant is the co-occurrence. Many automatic indexing methods do not consider how closely words occur, just if they occur in the same document [6].

2. RELATED WORK

Many researchers discussed the co-occurrence analysis of the documents text such as Chen and Lynch [8], Crouch [7], and Salton [27].

The limitation of the popular symmetric similarity functions (such as cosine, Dice, and Jaccard's) have been reported by Peat and Willett [24]. Their research showed that similar terms identified by symmetric co-occurrence function tended to occur very frequently in the database that is being searched and thus did little or nothing to improve the discriminatory power of the original query. They concluded that this can help explaining Sparck Jones finding that the best retrieval results were obtained if only the less

frequently occurring terms were clustered and if the more frequently occurring terms were left UN clustered.

The co-occurrence analysis used by Schutze and Pedersen in their research was based on number of times a word co-occurs with other words in a document. Schutze and Pedersen described this matrix as a "term-by-term matrix" (Schutze and Pedersen, 1997) [27].

Topical or semantic similarity between two words can then be defined as the cosine between the corresponding columns of the matrix.

The assumption is that words with similar meanings will occur with similar neighbors if enough text material is available. (Schutze & Pedersen 1997, p.311) [29] there are efficiency problems with this approach: the matrix that is used to Compare each word in the vocabulary to all other words in the vocabulary tend to be quite large, and it takes quite a long time to process the word comparisons, depending on the size of the vocabulary.

Hidetsugu anba in [15] introduced a new method to build a thesaurus in English language and Japanese language. They also define four kinds of relations between language words to build these thesauri. The first of these relations is: Hypernym/ Hyponym: used to extract the terms that are related by the Hypernym/Hyponym relation. The second relation is a defining what they call abbreviation extraction. It depends on the relation between the terms and their abbreviations. The third relation is what they call synonym extraction. To find these kinds of relations, the researchers concentrate on the citation relation between terms. The fourth kind of relations is what they call related terms extraction. In the results they presented that the suggested system has improved the information retrieval process through query expansion.

Although Crouch and Yang (1992) [6] automatically generated thesaurus classes from text keywords, which can subsequently be used to index documents and queries. Crouch's approach is based on Salton's vector space model and the term discrimination theory. Documents are clustered using the complete link clustering algorithm (agglomerative, hierarchical method). Ekmekcioglu et al. [12] tested retrieval performances for 110 queries on a database of 26,280 bibliographic records using four approaches: original queries and query expansion using co-occurrence data, Soundex code (a phonetic code that assigns the same code to words that sound the same), and string similarity measure (based on similar character microstructure), respectively. The four approaches produced 509 (original queries), 526 (term co-occurrence), 518 (Soundex), and 534 (string) documents, respectively. They concluded that there were no significant differences in retrieval effectiveness among these expansion methods and initial queries. However, a close examination of their results revealed that there was a very small degree of overlap between the retrieved relevant documents

generated by the initial queries and those produced by the co-occurrence approach (19% overlap using the Dice coefficient). This suggests that search performance may be greatly improved if a searcher can select and use the terms suggested by a co-occurrence thesaurus in addition to the terms he/she has generated.

Several research groups have experimented with an algorithmic approach to cross-domain term switching recently. Chen et al. experimented extensively in generating, integrating, and activating multiple thesauri (some were existing thesauri, others automatically generated, all in computing-related areas) [9] [11]. Both Kim and Kim [20] and Chen et al. [9] proposed treating (automatic and manually-created) thesauri as a neural network or semantic network and applying spreading activation algorithms for term-switching. Despite questions about the usefulness of automatic thesaurus browsing heuristics [17], our recent experiment revealed that activation-based term suggestion was comparable to the manual thesaurus browsing process in document recall and precision, but that the manual browsing process was much more laborious and cognitively demanding [11].

3. COLLECTING TERMS

Thesaurus construction requires collecting a set of terms. Some of these will end up becoming preferred terms and others may not appear in the thesaurus at all in their original form, but they may suggest concepts that need to be covered in some way.

In the local query expansion, local context concepts are selected based on co-occurrence with query terms, concepts are chosen from the top ranked documents [18], and the best passages are used instead of whole documents. Local context analysis involves only by the top ranked documents that have been retrieved by the query, i.e. the top ranked documents for a query were proposed as a source of information, so the most frequent 3 terms (none stop words) from the top ranked are added to the query.

In a global strategy as in [30] the query expansion technique presented explored the lexical-semantic links in Wordnet in order to expand hierarchically related terms to the original query. In a local strategy, the top-ranked documents retrieved for a given query are examined to determine terms for query expansion.

Apart from this expansion has been carried out by replacing or adding thesaurus words or synonyms to the existing query. Research pioneer Voorhees [Voorhees, 1994] has shown that this mechanism decreases the IR performance. However, her research points out that a manually built corpus specific thesaurus can give better results.

The automatic methods to build the thesaurus are characterized by the high precision in the determination of

relations between words, and the possibility of using the same method for more than one language and a great number of corpuses can be used to build the automatic thesaurus [23]. However, the automatic thesauri encounter a problem which is the difficulty of the results' evaluation. The thesauri are widely used in query expansion through adding new words to the query before beginning with the retrieval process

4. ARABIC LANGUAGE PROBLEMS

The problems of Arabic language that are related to our project are:

- 1) A word may take several meanings, depending on its position on the text and if the text is pointed or not, so that it makes an ambiguous view.
- 2) several words "حاسوب" (computer), "حاسبات" (Computers), "حوسبة" (Computing), "حساب" (Computations) and "محاسبة" (accounting), have the same root "حسب" (Compute), in spite of that their meaning is different, and our calculation is based on root only.
- 3) Some words may have more than one root. "مخافة" (Fear) it has two roots "خفى", (Hid), and "أخفى" (Cached).
- 4) When we deal with pointed text is a big problem?

5. Local context Analysis

Local context analysis is a statistical approach, where the occurrences of terms in documents Term co-occurrence analysis is one of the approaches used in IR research for forming multi-phrase terms. Local Context Analysis, implemented as term-suggestion devices. The closer the words occur, the more significant is the co-occurrence.

Any IR system performs the following tasks [5]:

- 1- Deleting the stop word from the documents.
- 2- Extracting Stems for each term in the documents.
- 3- Creating the inverted file based on the root of each document. (The root technique used is suffix prefix removal).

Two hundred and forty two Arabic documents were used to build the database of the thesaurus. These documents contain 2499 distinct terms. An inverted file of nearly size 22478 records was built. The problems faced in building the thesaurus were:

- 1) Compute the weight of each term in each document.
- 2) Compute the weight of each two terms in the same document.
- 3) Compute the similarity between each two terms (Compute the cluster weights).

After terms were identified in each document, we first computed the term frequency and the document frequency for each term in a document. Term frequency, tf_{ij} , represents the Number of occurrences of term j in document i . Document frequency, df_j , represents the Number of documents in a collection of n documents in which term j occurs. A few changes were made to the standard term frequency and inverse document frequency measures.

Usually terms identified from the title of a document are more descriptive than terms identified from the abstract of the document. In addition, terms identified by the user Filters are usually more accurate than terms generated by automatic indexing. This is due to the fact that terms generated by automatic indexing are relatively noisy [10].

We then computed the combined weight of term j in document i , d_{ij} , based on the product of "term frequency" and "inverse document frequency" as follows:

$$d_{ij} = tf_{ij} * \log \frac{N}{df_j}$$

Where N : represents the total number of documents in the collection.

We then performed term co-occurrence analysis based on the asymmetric "Cluster Function" developed by Chen and Lynch [8]. We have shown that this asymmetric Similarity function represents term association better than the popular cosine function.

The weighting-factor appearing in the equations below is a further improvement of our Cluster algorithm.

$$ClusterWeight(T_j, T_k) = \frac{\sum_{i=1}^n d_{ijk}}{\sum_{i=1}^n d_{ij}} \times WeightingFactor(T_k)$$

These equations indicate the similarity weights from term T_j to term T_k , d_{ij} and d_{ik} were calculated based on the equation in the previous step. d_{ijk} represents the combined weight of both Terms T_j and T_k in document i . d_{ijk} is defined similarly as follows:

$$d_{ijk} = tf_{ijk} * \log \frac{N}{df_{jk}}$$

Where tf_{ijk} represents the number of occurrences of both term j and term k in document i (The smaller number of occurrences between the terms was chosen).

df_{jk} represents the Number of documents (in a collection of N documents) in which terms j and k occur together.

In order to penalize general terms (terms which appeared in many places) in the local context analysis, we developed the following weighting schemes which are similar to the inverse document frequency function:

$$WeightingFactor(T_k) = \frac{\log \frac{N}{df_k}}{\log N}$$

Terms with a higher df_k value (more general terms) had a smaller weighting factor value; this caused the occurrence probability to become smaller. [10]

So here weight cluster is like the similarity in similarity thesaurus, applying the local context analysis and finding the weight factor between each two terms.

6. EXPANSION PROCESS

The local context analysis started with computations of each term's document frequency (the number of documents in a collection in which a word occurs) and term frequency (the frequency of occurrence of a word in a document). Terms appearing in the title of a document were assigned higher weights than terms in the abstract or other parts of the document. Terms that had been identified by the object filters in the first step were also assigned higher weights than those identified in the automatic indexing process. The inverse document frequency was then computed with some extra features. Multiple-word terms were assigned higher weights than single word terms since the former usually convey more precise semantic meaning than the latter.

Our local context analysis thesaurus was based on co-occurrence with query terms, terms are chosen from the top ranked documents [18], and the best passages are used instead of whole documents. Local context analysis involves only by the top ranked documents that have been retrieved by the query, i.e. the top ranked documents for a

query were proposed as a source of information. In our research we expanded the greater 3 terms associated with greater weight cluster and we consider its terms as an expanded term and we expand the original Query, and after we retrieve the documents we rank it, (Lu et al., 2008) suggest that ranking by relevance can result in better retrieval performance. Thus, we computed TF-IDF scores for retrieved documents (Kim and Wilbur, 2005; Lu et al., 2008) and then ranked them based on these scores. A document with a higher TF-IDF score is returned earlier in a list.

7. DISCUSSION

One of the major problems of the modern IR systems is the word mismatch that concerns the discrepancies between terms used for describing documents and the terms used by the searchers to describe an information need. A way of handling the word mismatch is by using a thesaurus, which shows (usually semantic) relationships between terms. Thesauri can aid the indexer or the indexing system in choosing the correct terms to describe the contents of documents, and in normalizing the terms so that all terms are e.g. presented in singular form. In the searching process, thesauri can help the searcher to find terms to refine a query, by expansion of the original query.

Some of the relationships between terms that are handled by thesauri are narrower term (NT), broader term (BT), and related term (RT). There are some obvious problems with manually constructing thesauri. It is an expensive and time-consuming process that requires a domain-expert or an expert at document description. In domains where new research fields develop frequently, thesauri become out of date, and need to be updated, which again is time-consuming and expensive. By using documents published in the domain in question as a corpus, a thesaurus can be created and updated automatically. The terminology of the researchers of the field will be the basis of the indexing process and the assignment of index terms. There are a number of different approaches available for automatically creating thesauri, among others different kinds of statistical local context analysis. A way of following up this paper would be to go in deeper on the different approaches, and/or select the one most interesting for my future thesis project.

This study is implemented on Oracle 11G, and the project was tested on the 242 Arabic documents that were used by Hmeidi and Kanaan (1997) [16]. The user query was 59 Arabic queries in many general and scientific fields (mostly were related to computer science field) [7].

The following results were found from the study:-

- 1- The recall is better when using the local context analysis thesaurus than using the classical IR system. This result is also reported by Qiu and Frei [25].
- 2- The precision is almost better when using classical IR system than using local context analysis thesaurus in a very small range.
- 3- On average recall/precision levels, the local context analysis thesaurus makes a good effect on the last 8 levels (0.3 to 1). While it has limitation on the first 2 level (0 to 0.2) this mean that local context analysis is better in the last 8 level (Figure 3)
- 4- Many researchers concluded that the effective of the retrieval process when we using a thesaurus will increase, when we increase the number of documents in the collection.

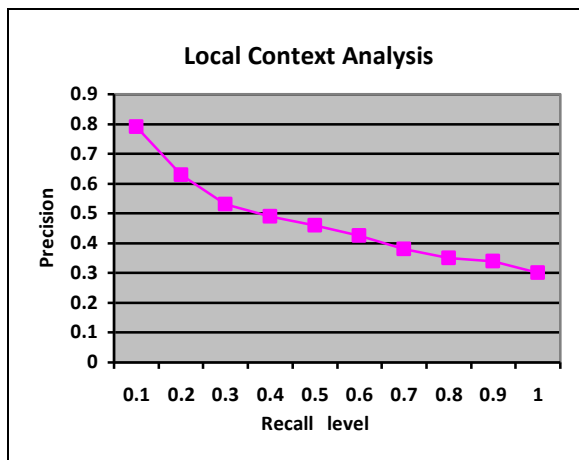


Fig. 1: Using local context analysis

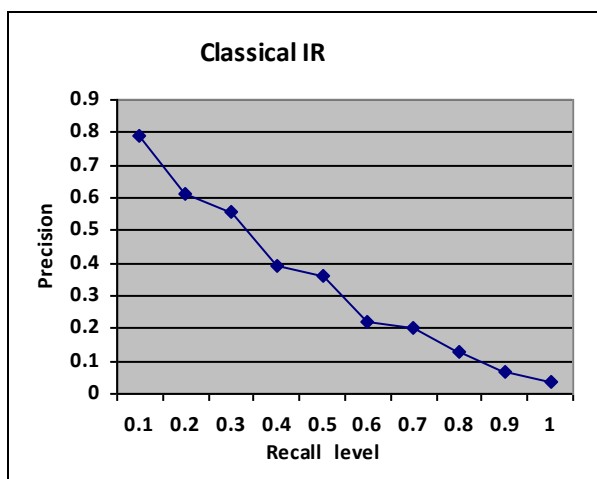


Fig 2: Classical information retrieval

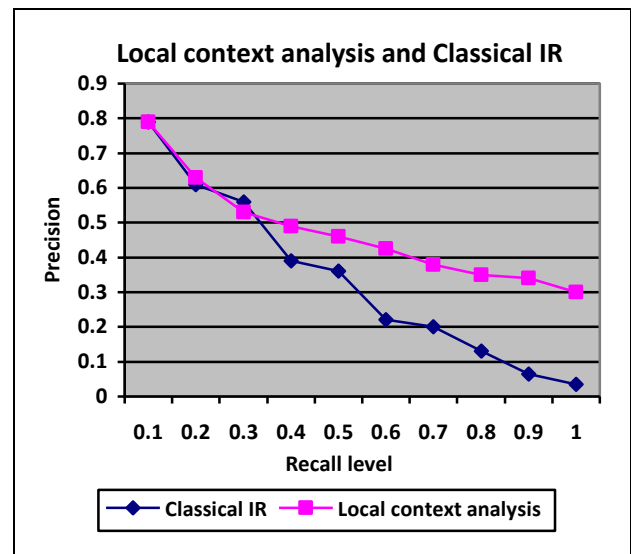


Fig. 3: Comparison between local context analysis and classical IR model

8. CONCLUSIONS

In a world of increasing facing information overload, where the issue is not how many documents can be found in a particular research subject, but rather how to weed through thousands of documents on a topic to find the most relevant ones. Based on the results of this study, the following conclusions may be drawn:

- 1) The local context analysis improves the recall in a good manner
- 2) The precision is almost better when using local context analysis than using classical IR system in a very small range.
- 3) The local context analysis based on recall/precision level improves the effective of the retrieval task of the system especially on the last 8 levels. Qui and Frei [25] deferent with this conclusion. They reported that most of query expansion methods (including co-occurrence) failed to improve the retrieval process. On another hand Khafajeh [18] showed that using Association thesaurus in Arabic language retrieving system has been improved the effective of the retrieval task of the system.
- 4) The experiments results showed that using the stemmed words improved the retrieval process when they were used by local context analysis. While when the full words were used in the traditional system, the system's performance was the worst in the continuous retrieval process, because the precision values

decreased in a remarkable way. When the recall values increased, mostly the precision values reached to zero. But, in the same system with using the stemmed words, its performance degraded less sharply.

- 5) Finally, we present some of the future works that can be achieved. These works are related to another techniques for using query expansion. Especially, there are many query expansion methods that are not applied on the Arabic corpora. Continuing our program of studying different methods of query expansion in Arabic information retrieval (AIR), we may examine the effects of varying methods of term suggestion for user-controlled query expansion such as local context analysis and Relevance Feedback, and improving automatic method to build Arabic corpus.

Acknowledgments

This research is funded by the Deanship of Research and Graduate Studies in Zarqa University /Jordan.

References

- [1] Abdelali, A. Localization in Modern Standard Arabic. Journal of the American Society for Information Science and technology Volume 55, Number 1, 2004.
- [2] Al-Shalabi, R. Kannan, G., Al-Jaam, J., Hasnah A., and Helat, E., Stop-word Removal Algorithm for Arabic Language, processing of the 1st International Conference on Information & Communication Technologies: from theory to Applications-ICTTA, Damascus, 2004.
- [3] Amir Hossein Jadidinejad, Hadi Amiri, "Local Cluster Analysis as a Basis for High-Precision Information Retrieval", In Proceeding of INFOS2008 International Conference on Informatics and Systems, Egypt, 2008
- [4] Baeza-Yates, Ricardo And Berthier Ribeiro-Neto. "Modern Information Retrieval". Addison-Wesley, New York City, NY, ACM Press, 1999.
- [5] Baeza-Yates, R. and Ribeiro-Neto, B. (1999) "Modern Information Retrieval". Addison Wesley.
- [6] Crouch, C. and Yang, B. (1992) "Experiments in Automatic Statistical Thesaurus Construction" In Proceedings of the Fifteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, pp 77-88, Copenhagen, Denmark, June 21-24.
- [7] Crouch, C. J. (1990) "An approach to the automatic construction of global thesauri, Information Processing and Management," Vol. 26, No. 5, pp. 629-640
- [8] Chen, H. and Lynch, K. J. (1992) "Automatic construction of networks of concepts characterizing document databases" IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, No. 5, pp. 885-902.
- [9] Chen, H., and Lynch, K., et al.(1993) "Generating, integrating, and activating thesauri for concept-based document retrieval," IEEE EXPERT, Special Series on Artificial Intelligence in Text-based Information Systems, Vol. 8, No. 2, PP. 25-34.
- [10] Chen, H., Ng, T. D., et al. (1997). "A Concept Space Approach to addressing the word mismatch in scientific Information Retrieval: An experiment on the Worm Community System." Journal of the American Society for Information Science, Vol. 48, No. 1, pp. 17-31.
- [11] Chen, H., and Ng, D., (1995) "An algorithmic approach to concept exploration in a large knowledge network (automatic thesaurus consultation): symbolic branch-and-bound vs. connectionist Hopfield net activation," Journal of the American Society for Information Science, Vol. 46, No. 5, PP. 348-369.
- [12] Ekmekcioglu, F., Robertson, A., et al. (1992) "Effectiveness of query expansion in ranked-output document retrieval systems," Journal of Information Science, Vol. 18, PP. 139-147.
- [13] <http://www.ai.bpa.arizona.edu/papers>.
- [77] Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma, Query Expansion by Mining User Logs, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 15, NO. 4, JULY/AUGUST 2003
- [15] Hidetsugu anba. "Query Expansion using an Automatically Constructed Thesaurus" Proceedings of NTCIR-6 Workshop Meeting, May 15-18, Tokyo, Japan 2007.
- [16] Hmeidi, I., Kanaan, G., Evans, M. (1997) "Design and Implementation of Automatic Indexing for information Retrieval with Arabic Documents. Journal of the American Society of information science. Vol. 48, No. 10, pp. 867-881.
- [17] Jones, S., et al. (1995) "Interactive thesaurus navigation: intelligent rules OK?" Journal of the American Society for Information Science, Vol. 46, No. 1, PP. 52-59.
- [18] Khafajeh, H. et al. Automatic Query Expansion for Arabic text retrieval based on Association and Similarity thesaurus, European Mediterranean & Middle Eastern Conference on Information Systems (EMCIS 2010), April 12-13, 2010,
- [19] Kim W, Wilbur WJ. A strategy for assigning new concepts in the MEDLINE database. AMIA Annu, Symp Proc 2005;395-399. [PubMed: 16779069], Lindberg D, Humphreys B, Mccray A. The unified medical language system. Methods Inf Med 1993;32, (4):281-291. [PubMed: 8412823]
- [20] Kim, Y., and Kim, J. (1990) "A model of knowledge based information retrieval with hierarchical concept graph," Journal of Documentation, Vol. 46, PP. 113-116.
- [21] Lu Z, Kim W, Wilbur WJ. Evaluating relevance ranking strategies for medline retrieval. in press, 2008
- [22] Miller, U. (1997) "Thesaurus construction: problems and their roots." Information Processing and Management Vol. 33, No. 4, pp. 481-493.
- [23] Nakayama, K. and Hara, T. and Nishio, S. "A thesaurus construction method from large scale web dictionaries", Advanced Information Networking and Applications, AINA'07. 21st International Conference on IEEE, pages 932-939, 2007.
- [24] Peat, H. J. and Willett, P. (1991) "The limitations of term co-occurrence data for query expansion in document

- retrieval systems," Journal of the American Society for Information Science, Vol. 42, No. 5, pp. 378-383.
- [25] Qiu, Y. and Frei, H. P. (1993) "Concept based query expansion". In Proc. 16th Int'l Conference on R&D in IR (SIGIR), pp. 160-169.
- [26] Roget, P. (1988) Roget's II the new Thesaurus, Houghton Mifflin Company, Boston, USA.
- [27] Salton, G. (1989) Automatic Text Processing. Addison-Wesley Publishing Company, Inc., Reading, MA, USA.
- [28] S. Battiato, G. M. Farinella, G. Impoco, O. Garretto, and C. Privitera "Cortical Bone Classification by Local Context Analysis", Conference: MIRAGE - MIRAGE , pp. 567-578, 2007.
- [29] Schutze, H., and Pedersen, J. (1994) "A Co occurrence-based Thesaurus and Two Applications to Information Retrieval," In Proceedings of RIAO, pp. 266-274.
- [30] Voorhees, E. (1994). Query Expansion using lexical-semantic relations. In Proceedings of ACM SIGIR International Conference on research and development in Information Retrieval, pp. 61-69.



Assistant Professor HAYEL KHAFAJEH received the degree of B.Sc. in 1985, he earned his Master degree of M.Sc. in IT, A Ph.D. in CIS was received in 2008, he joined Zarqa University in Jordan in 2009, in 2010 he was the head of the CIS department and for two years, and he still Assistant

Professor in the same University.

Assistant Professor Hayel Khafajeh has been worked for 23 years in the Educational field such as programmer, teacher for the pioneer students, teachers supervisor, head of IT division, and Manager of ICDL Center all of that were for the ministry of Education, Jordan. He has published 6 research papers in international journal and conferences. And there are another 3 papers under processing. He has published many educational computer books for the Ministry of education, where the last book has been published is JAVA PROGRAMMING in 2013.



Assistant Professor, Nidal YOUSEF received the degree of B.Sc. in 1997, he earned his Master degree in IT, A Ph.D. was received in 2008 in CIS, He joined King Abdulaziz University, in KSA in 2009, as an Assistant Professor and in 2010 he moved to AL-Esra University in Jordan as Assistant professor in the

college of computing and information technology. He has published 4 research papers in international journal and conferences.

Design and Implementation of Feynman Gate in Quantum-dot Cellular Automata (QCA)

Md. Anisur Rahman¹, Fatema Khatun², Angona Sarkar³ and Md. Fazlul Huq^{4*}

¹Department of Information and Communication Technology
 Mawlana Bhashani Science and Technology University
 Santosh, Tangail – 1902, Bangladesh

²Department of Information and Communication Technology
 Mawlana Bhashani Science and Technology University
 Santosh, Tangail – 1902, Bangladesh

³Department of Information and Communication Technology
 Mawlana Bhashani Science and Technology University
 Santosh, Tangail – 1902, Bangladesh

⁴Department of Information and Communication Technology
 Mawlana Bhashani Science and Technology University
 Santosh, Tangail – 1902, Bangladesh

*Corresponding author

Abstract

Quantum cellular automata (QCA) have been used widely to digital circuits and systems. QCA technology is a promising alternative to CMOS technology. It is attractive due to its fast speed, small area and low power consumption, higher scale integration, higher switching frequency than transistor based technology. Various QCA circuits, Multivalve Reversible Logic (MVL) Circuit as well as Feynman gate have been proposed in this paper. The QCA offers a novel electronics paradigm for information processing and communication. In this paper, a Feynman gate circuit is proposed based on QCA logic gates: the Maj3, Maj AND gate, Maj OR based on QCA logic gates. The proposed circuit is a promising future in constructing of nano-scale low power consumption information processing system and can stimulate higher digital applications in QCA.

Keywords: Quantum Cellular Automata, QCA Logic Gates, Feynman gate in QCA.

1. Introduction

QCA is a novel emerging technology in which logic states are not stored as voltage levels, but rather the position of individual electrons. Conceptually, QCA represents binary information by utilizing a bitable charge configuration rather than a current switch. A QCA cell can be viewed as a set of four “dots” that are positioned at the corners of a square. A quantum dot is a site in a cell in which a charge can be localized. The cell contains two extra mobile electrons that can

quantum mechanically tunnel between dots, but not cells. In the ground state and in the absence of external electrostatic perturbation [1], the electrons are forced to the corner positions to maximize their separation due to Coulomb repulsion. As shown in Figure 1, the two possible charge configurations are used to represent binary “0” and “1”. Note that in the case of an isolated cell, the two polarization states are energetically degenerate. However the presence of other charges (neighbor cells) breaks the degeneracy and one polarization state becomes the cell ground state [1]. Polarization P measures the extent to which the charge distribution is aligned along one of the diagonal axes. If the charge density on dot i is ρ_i , then the polarization is defined as [2, 3],

$$P = \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_1 + \rho_2 + \rho_3 + \rho_4} \quad (1)$$

The tunneling between dots implies that ρ_i may not be integers as polarization values.

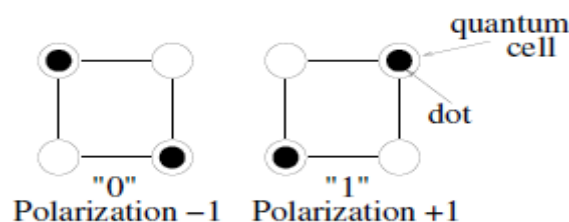


Fig. 1 QCA Cell

The QCA cells themselves comprise the interconnecting wires as described in [4]. An example of a QCA wire is shown in Figure 2. In this example, a value of 1 is transmitted along the wire. Only a slight polarization in a cell is required to fully polarize its neighbor. The direction for the flow of information through a gate or a wire is controlled by a four stage clocking system described in [5] which raises and lowers barriers between the cells.

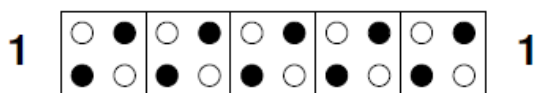


Fig. 2 QCA Wire.

Described in [3] were other logic gates formed by restricting the polarity of one input to the 3-input majority gate to be a constant value. Figure 3 illustrates a 2-input AND gate and a 2-input OR gate formed in this manner. By replacing input C with a cell having a fixed polarity of 0, the 3-input majority gate functions as an AND gate. By replacing input C with a cell having a fixed polarity of 1, the 3-input majority gate functions as an OR gate. In this case C is a control input. In the example, OR gate on the left side of Figure 3, $Out = A+B$. Similarly, replacing input C with a cell having a fixed polarity of 1 creates a 2-input OR gate. In the example AND gate on the right side of Figure 3, $Out = AB$.

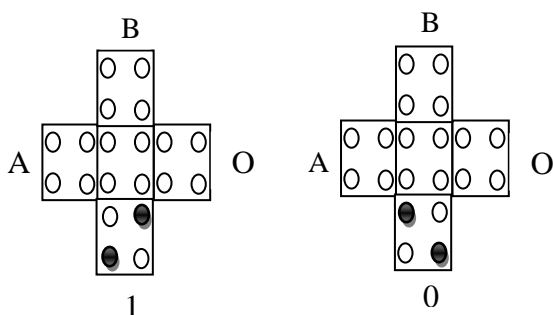


Fig. 3 2-input OR && 2-input AND gates.

Quantum arithmetic components need reversible logic circuits for their construction. Reversible logic circuits find wide application in low power digital design, quantum computing and nanotechnology. An optimized and low quantum cost one digit carry skip BCD adder, has been designed using new FHNG reversible logic gates [6]. It is an important logic gate in reversible logic. In this paper we have designed and simulated a Feynman gate in QCA technology.

2. QCA Implementation

There have been several proposals for physically implementing QCA: Micro-sized QCA devices have been fabricated with metal which operate at 50mK [7] [8] and an extensive literature has been reported on developing molecular implementations of QCA [9] [10].Magnetic QCA (MQCA) has been investigated and fabricated [11] [12] for room temperature operation. In this section, a brief background on Metal, Molecular, and Magnetic QCA is provided.

3. Proposed Circuit and Presentation

3.1 Feynman gate

Figure 4 shows a 2 x 2 Feynman gate [13]. The input vector is I (A, B) and the output vector is O (P, Q) and the relation between input and output is given by $P=A$, $Q = A \oplus B$. Since it is a 2 x 2 gate, it has a quantum cost of 1 [14]. It is used to copy the input without producing garbage bits.

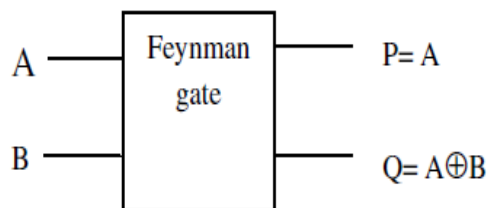


Fig. 4 Feynman gate.

3.2 Feynman gate in QCA

The block diagram of QCA is the Feynman gate shown in Figure 5 .The fundamental logic gate for QCA is the Feynman gate shown in Figure 6 that is composed of Sixty (60) cells. Two of these, representing the inputs to the cell, are labeled A and B. using the terminology of [3], the center cell is the “device cell” that performs the calculation. The remaining cell, labeled out, provides the output. The circuit shown in Figure 6 performs the Boolean function $Out2 = A \oplus B$ and $out1 = A$;

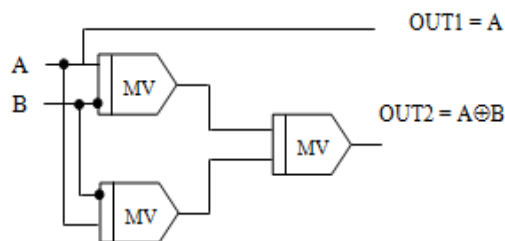
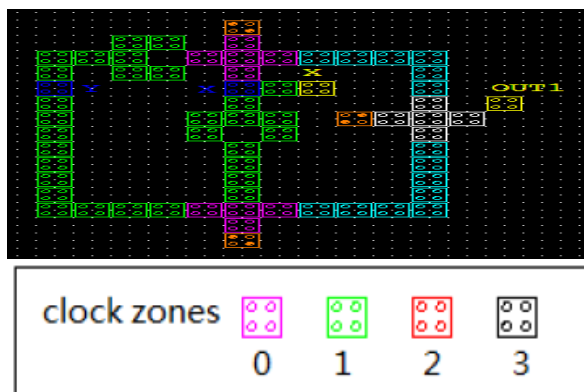


Fig. 5 Block Diagram of QCA is the Feynman gate.



The Feynman gate layout has been simulated using QCA Designer version 2.0.3; a layout and simulation tool for QCA. The simulation results for a Feynman gate are shown in Figure 7.

Fig. 6 Design QCA is the Feynman gate.

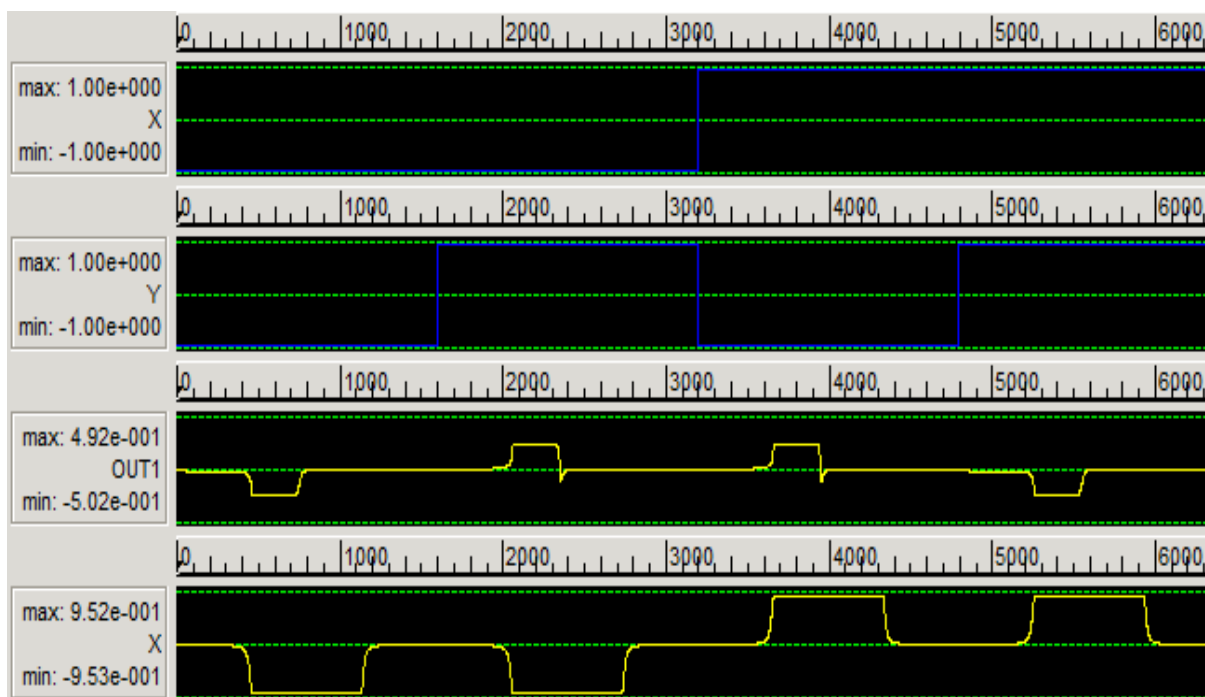


Fig.7 Simulated waveforms for Feynman gate circuit.

In this Simulation we have used the coherence vector computational engine and the following parameters: (10 nm × 10 nm) cell size, 2.5 nm cell-to-cell distances, and 2.5 nm dot size and 40 nm radius of influence.

4. Conclusion

This paper present a Feynman gate based on QCA does logic gates. The design is very useful for future computing techniques like ultra low power digital circuits and quantum computers. MVs provide a functionally complete logic set for QCA. This QCA circuit design provide a new functional paradigm for information encoding. In addition, QCA binary logic functions and the associated new nano-technology will provide high-speed computing, high-density applications. It is believed that QCA will become a more practical ways to create a faster and denser circuit.

References

- [1] Lent, C. S., P. D. Tougaw and W. Porod, "Quantum Cellular Automata: The Physics of Computing with Arrays of Quantum Dot Molecules," PhysComp '94: Proceedings of the Workshop on Physics and Computing, IEEE Computer Society Press, 1994, pp. 5-13.
- [2] Lent, C. S., et al., "Quantum cellular automata," Nanotechnology, Vol. 4, No. 1, 1993, pp. 49-57.
- [3] Tougaw, P. D. and C. S. Lent, "Logical Devices Implemented Using Quantum Cellular Automata," Journal of Applied Physics, Vol. 75, No. 3, 1994, pp. 1818-1825.
- [4] C. S. Lent and P. D. Tougaw, "Lines of interacting quantum-dot cells: A binary wire,"

- Journal of Applied Physics, vol. 74, no. 10, pp. 6227-6233, November 15, 1993.
- [5] C. S. Lent and P. D. Tougaw, "A device architecture for computing with quantum dots," Proceedings of the IEEE, vol. 85, no. 4, pp. 541-557, April 1997.
- [6] Md.Belayet Ali, Md.Samiur Rahman and Tahmina parvin "Optimized digit carry skip BCD adder using new reversible logic gates " IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012. 1694-0814
- [7] Amlani, I., et al., "Demonstration of a Six-dot Quantum Cellular Automata System," Applied Physics Letters, Vol. 72, No.17, 1998, pp. 2179-2181.
- [8] Orlov, A. O., et al., "Realization of a Functional Cell for Quantum-Dot Cellular Automata," Science, Vol. 277, No. 5328, 1997, pp. 928-930.
- [9] Lieberman, M., et al., "Quantum-dot Cellular Automata at aMolecular Scale," Analysis of the New York Academy of Science, Vol. 960, 2002, pp. 225-239.
- [10] Lieberman, M., et al., "Quantum-Dot Cellular Automata at a Molecular Scale," Annals of the New York Academy of Sciences, Vol. 960, 2002, pp. 225-239.
- [11] Cowburn, R. P. and M. E. Welland, "Room Temperature Magnetic Quantum Cellular Automata", Science, Vol. 287, 2000, pp. 1466-1468.
- [12] Parish, M. C. B., "Modeling of Physical Constraints on Bistable Magnetic Quantum Cellular Automata", Ph.D. Thesis, University of London, UK, 2003.
- [13] R. Feynman, "Quantum Mechanical Computers," Optics News, Vol.11, pp. 11-20, 1985.
- [14] Majid Mohammadi, Mohammad Eshghi, Majid Haghparast and Abbas Bahrololoom, (2008) "Design and Optimization of Reversible BCD Adder/Subtractor Circuit for Quantum and Nanotechnology Based Systems", World Applied Sciences Journal, vol. 4, no. 6, pp. 787-792.

Biographical notes:

Md. Anisur Rahman is currently completing his M. Sc in the Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail – 1902, Bangladesh. He has completed his B. Sc in the Department of Computer Science and Engineering at the same university. His area of research interest is cloud computing, reversible logic synthesis etc.

Fatema Khatun is currently completing her B. Sc in the department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail – 1902, Bangladesh. Her area of research interest is reversible logic synthesis.

Angona Sarkar is currently completing her B. Sc in the department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail – 1902, Bangladesh. Her area of research interest is photonic crystal fiber.

Md. Fazlul Huq is currently working as a lecturer in the Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail – 1902, Bangladesh. He has completed his B. Sc and M. Sc in the Department of Applied Physics Electronics and Communication Engineering, University of Dhaka, Dhaka – 1000, Bangladesh. He has 09 published national and international research papers. His area of research interest is artificial neural network, cloud computing, embedded system design, nanotechnology etc.

Face Recognition Using SVM Based on LDA

Anissa Bouzalmat¹, Jamal Kharroubi² and Aarsalane Zarghili³

¹ Department of Computer
Science faculty of Science and
Technology, Sidi Mohamed Ben Abdellah
University, Route d'Imouzzer Fez, 2202/30000 Morocco

² Department of Computer
Science faculty of Science and
Technology, Sidi Mohamed Ben Abdellah
University, Route d'Imouzzer Fez, 2202/30000 Morocco

³ Department of Computer
Science faculty of Science and
Technology, Sidi Mohamed Ben Abdellah
University, Route d'Imouzzer Fez, 2202/30000 Morocco

Abstract

We present a method for face recognition that investigate the overall performance of linear ,polynomial and RBF kernel of SVM for classification based on global approach and used images having different expression variations, pose and complex backgrounds. In the first we reduce dimensional feature vector by LDA method , the result of vectors feature propagates to a set of SVM classifier, we trained SVM classifier with linear and non linear kernel for each dataset (face94, face96, grimaces)[1,2,3] in the database . Experiments demonstrate that use the LDA method combined with SVM classifier and the choice of a suitable kernel function with optimal parameters can produce high classification accuracy compared to KNN classifier on a variety of images on different Database.

Keywords: *Face Recognition, SVM, LDA, PCA, KNN.*

1. Introduction

A face recognition system recognizes a face by matching the input image against images of all faces in a database and finding the best match.

It can be roughly divided into two main categories: local and global approaches. In local feature approaches a number of fiducially or control points are extracted and used for classification, while in global approaches the whole image serves as a feature vector, the techniques was developed this approach are Eigen faces ,Linear Discriminate Analysis (LDA)[4] that method outperforms PCA in terms of class discrimination, neural networks [5] and Support Vector Machines (SVM) [6], is considered easier to use and performs particularly well with high dimensional feature vectors and in case of lack of training

data ,these factors which may significantly limit the performance of most neural networks [7].

The features constitute of the global image in case of the approach global are a high dimension, so it is difficult to use it without applying reduction method such high dimension of vectors have proven to be one of the biggest problems of face recognition systems then it is necessary to apply a method to reduce dimension of vectors that will perform the process for face recognition algorithm. As one of feature extraction methods for face recognition problem, linear discriminate analysis (LDA) were applied for the images this drastically reduced the number of attributes of feature vectors. By choosing the two of the most popular machine learning algorithms SVM method and K-nearest neighbors that is one of the simplest but effective in many cases in machine learning algorithms [8] then comparing the accuracy for face classification of these two classifiers, We implement a recognition system using SVM and KNN classifiers based on LDA, The procedure is described as follows: The outline of the paper is as follows: Section 2 Description of the proposed method. In section 3 contains experimental results. Section 4 concludes the paper.

2. The Proposed Method

The present study proposed in this paper (Fig 1) is designed for face recognition. The system consists of: a) The whole image serves as a feature vector and reduces it by linear discriminate analysis b) Transform theses feature vectors to the format of an SVM and scale them. Finally, the obtained feature vectors are used as input of classifier support vector machine with different kernel and K-nearest neighbor classifier.

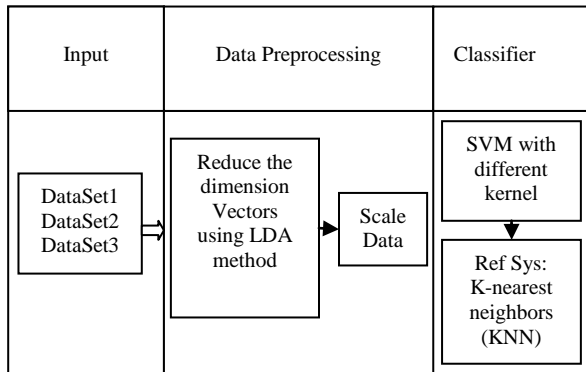


Fig. 1: Architecture description of the proposed approach.

2.1 Construction the features vectors

When the whole images of database were treated as features, we have a large features vector in this situation, the computational expense will increase, especially for non-linear classifiers so

It is important to reduce the dimensionality of features before classification by applying LDA method which is significantly better to perform dimensionality reduction while preserving as much of the class discriminatory information as possible [9].

2.1.1 Linear discriminate analysis (LDA)

Linear discriminate analysis (LDA) is powerful tools used to reduce dimension of feature vectors without loss of information and used as a feature extraction step before classification [4], it is generally believed that algorithms based on LDA are superior to those based on PCA in the lower dimensional subspace [10].

LDA try to maximize class seperability witch determines a subspace in which the between-class scatter (extra personal variability) is as large as possible, while the within-class scatter (intrapersonal variability) is kept constant. In this sense, the subspace obtained by LDA optimally discriminates the classes-faces. The Objective of LDA seeks to reduce dimensionality while preserving as much of the class discriminatory information as possible.

We have a set of C-class and D-dimensional samples $\{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$, N_1 of which belong to class w_1 , N_2 to class w_2 and N_c to class w_c , in order to find a good discrimination of these classes we need to define a measure of separation, we define a measure of the within-class scatter by Eq. (1):

$$S_w = \sum_{i=1}^c S_i$$

$$S_i = \sum_{x \in w_i} (x - \mu_i)(x - \mu_i)^T \quad (1)$$

$$\mu_i = \frac{1}{N_i} \sum_{x \in w_i} x_i$$

And the between-class scatter Eq. (2) becomes:

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (2)$$

$$\mu = \frac{1}{N} \sum_{\forall x} x = \frac{1}{N} \sum_{i=1}^c N_i \mu_i$$

Matrix $S_T = S_B + S_W$ is called the total scatter similarly; we define the mean vector and scatter matrices for the projected samples as:

$$\tilde{S}_w = \sum_{i=1}^c \sum_{y \in w_i} (y - \tilde{\mu}_i)(y - \tilde{\mu}_i)^T$$

$$\tilde{S}_B = \sum_{i=1}^c N_i (\tilde{\mu}_i - \tilde{\mu})(\tilde{\mu}_i - \tilde{\mu})^T, \quad \tilde{\mu}_i = \frac{1}{N_i} \sum_{y \in w_i} y, \quad \tilde{\mu} = \frac{1}{N} \sum_{\forall y} y$$

From our derivation for the two-class problem, we can

write:
$$\begin{cases} \tilde{S}_B = W^T S_B W \\ \tilde{S}_w = W^T S_w W \end{cases}$$

Recall that we are looking for a projection that maximizes the ratio of between-class to within-class scatter. Since the projection is no longer a scalar (it has C-1 dimensions), we use the determinant of the scatter matrices to obtain a scalar objective function Eq. (3):

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_w|} = \frac{W^T S_B W}{W^T S_w W} \quad (3)$$

And we will seek the projection matrix w^* that maximizes this ratio It can be shown that the optimal projection matrix w^* is the one whose columns are the eigenvectors corresponding to the largest eigen values of the following generalized eigen value problem Eq. (4):

$$w^* = \left[w_1^* | w_2^* | \dots | w_{c-1}^* \right] = \arg \max \frac{|W^T S_B W|}{|W^T S_w W|}$$

$$\Rightarrow (S_B - \lambda_i S_w) W_i^* \quad (4)$$

S_B is the sum of C matrices of rank ≤ 1 and the mean

vectors are constrained by:
$$\frac{1}{c} \sum_{i=1}^c \mu_i = \mu$$

Therefore, S_B will be of rank $(C-1)$ or less and this means that only $(C-1)$ of the eigen values λ_i will be non-zero. The projections with maximum class separability information are the eigenvectors corresponding to the largest eigen values of $S_W^{-1}S_B$ we seek $(C-1)$ projections $[y_1, y_2, \dots, y_{c-1}]$ by means of $(c-1)$ projection vectors w_i arranged by columns into a projection matrix $W=[w_1|w_2|\dots|w_{c-1}]$:

$$y_i = w_i^T x \Rightarrow y = W^T x$$

The LDA be useful in order to reduce dimensionality and speed up the classifier for training.

Footnotes should be typed in singled-line spacing at the bottom of the page and column where it is cited. Footnotes should be rare.

2.2 Support Vector Machines (SVM)

SVMs (Support Vector Machines) are a useful technique for data classification and are still under intensive research [11],[12]. Although SVM is considered easier to use than Neural Networks, In addition, its true potential is highlighted when the classification of non-linearly separable data becomes possible with the use of a kernel function, which maps the input space into a possibly higher dimensional feature space in order to transform the non-linear decision boundary into a linear one. There exists a range of kernel functions, where a particular function may perform better for certain applications. For more details we describe the theory of support vector machines as a combination of two main concepts: kernel functions and maximal margin hyperplanes.

2.2.1 Scaling Data

After the low-dimensional feature vector were obtained we scale attributes of training data to the range $[-1, 1]$ linearly, then scale the attributes of the test data using the same scaling function of the training data that is very important which is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation.

2.2.2 Kernels functions

The simplest form of SVM is introduced as a hyper plane maximizing the margin of separation on linear separable data sets, in many real-world problems, noisy data will render linear separation impossible so no feasible solution to the margin maximization problem can be found, in this case we consider the non linear classifiers which can be overcome by an approach called kernel technique, it was introduced as the method of potential functions [13]. The general idea of the kernel is to map the input data to a high dimensional space and separate it there by a linear classifier. This will result in a classifier nonlinear in input space.

The mapping $\Phi: R^p \rightarrow F$ is applied to each example before training and the optimal separating hyper plane is constructed in the feature space F (Fig 2).

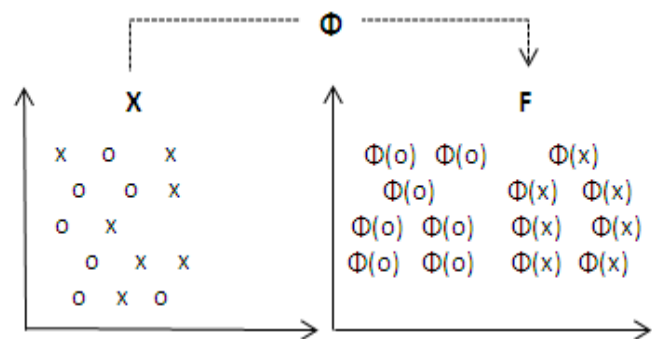


Fig. 2 : The mapping of features to feature space F.

There are several kernels are being proposed by researchers, we lists the kernel expressions and corresponding parameters. The three basic kernels as follow: linear, polynomial, radial basis function (RBF).

The Linear kernel (5) is the simplest kernel function. It is given by the inner product. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterparts.

$$k(x_i, x_j) = \langle x_i, x_j \rangle \quad (5)$$

Note that $\langle x_i, x_j \rangle$ represents dot product, where x_i and x_j denote two arbitrary feature vectors.

For the Polynomial kernels (6) are a non-stationary kernels. They are well suited for problems where all the training data is normalized.

$$k(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + \text{coef0})^{\text{degree}} \quad (6)$$

degree, γ , coef0 respectively denote degree of the polynomial, coefficient of the polynomial function, and the coadditive constant.

The Gaussian kernel (7) is an example of radial basis function kernel.

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (7)$$

The parameter γ denotes the width of the Gaussian radial basis function where: $\gamma = \frac{1}{2\sigma^2}$

The parameter σ plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand. If over estimated, the exponential will behave almost linearly and the higher-dimensional projection will start to lose its non-linear power. In the other hand, if underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data.

2.2.3 Maximal Margin Hyperplanes

After we change the representation of the training examples by mapping the data to a features space F where the optimal separating hyper plane (OSH) is constructed Fig 3, we limited our study to the case of two-class discrimination [14] and we consider the training data S a set of l vectors features each vector has n dimension, where each point x_i belongs to one of two classes identified by the label -1 or 1 Eq 8.

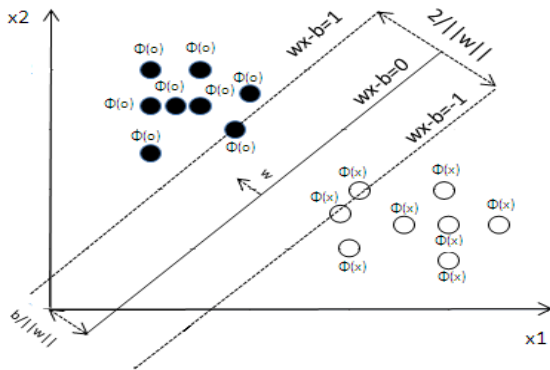


Fig. 3: Maximum-margin hyper plane for SVM trained with samples from two classes.

$$S = \left\{ (x_i, y_i) \mid x_i \in R^n, y_i \in \{-1, 1\} \right\}_{i=1}^l \quad (8)$$

We have solving a quadratic optimization problem with linear constraints that can be interpreted in terms of the Lagrange multipliers calculated by quadratic programming Eq: 9

$$\left\{ \begin{array}{l} \max(\alpha_i) : \tilde{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \text{(for any } i = 1, \dots, n) \quad 0 \leq \alpha_i \leq c \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{array} \right\} \quad (9)$$

α_i are the Lagrange multipliers parameters to be adjusted , c is the penalty parameter of the classification error term it must be adjusted because the data are rarely completely separable, the x_i are the training examples.

The solution of the optimization problem will be a vector $w \in F$, that can be written as a linear combination of the training inputs Eq: 10

$$w = \sum \alpha_i y_i x_i \quad (10)$$

(w, b) define the hyperplane and b is the bias Eq:11

$$OSH = \{x : w \cdot x + b = 0\} \quad (11)$$

we use the separating (OSH), once we have trained it on the training set, The (OSH) divides the R^n into two regions: one where $w \cdot x_i + b \geq 0$ and one where $w \cdot x_i + b \leq 0$. To use the maximal margin classifier, we determine on which side the test vector lies and assign the corresponding class label. Hence, the predicted class of a test point x is the output of the decision function Eq 12.

$$d(x) = \text{sgn} \left(\sum_{i=1}^l \alpha_i y_i k(x_i, x) + b \right) \quad (12)$$

2.2.4 Multiclass Classification

In real world problems we often have to deal with $d \geq 2$ classes. Our training set will consist of pairs $(x_i; y_i)$, where $x_i \in R^n$ and $y_i \in \{1, \dots, d\}$; $i = 1.. l$. for solving this problem we construct decision functions of the form Eq:13

$$x \rightarrow \arg \max_{c \in \{1, \dots, d\}} [\langle w_c, \phi(x) \rangle + b_c] \quad (13)$$

Here, $\Phi : x \rightarrow H, \Phi = k(x, \cdot)$, is a feature map into a reproducing kernel Hilbert space H with corresponding kernel k , and $w_1, \dots, w_d \in H$ are class-wise weight vectors.

Two basic strategies to extend SVMs to multi-category classification can be distinguished. One approach is to combine separately trained binary SVM classifiers after training as done in the prominent one-versus-all method [15,16] .In the second method, a single optimization problem considering all classes is derived and solved at once, in [17] all-in-one approach perform significantly better than the one-vs-all method.

Crammer & Singer [18] proposed an alternative multi-class SVM. They also take all class relations into account at once and solve a single optimization problem, however, with fewer slack variables. The main reason for this modification of the Weston & Watkins approach [19] primal problem was to speed-up the training, because the Weston & Watkins approach turned out to be too slow for many applications. The Crammer & Singer classifier is trained by solving the primal problem Eq:14

$$\min_{w_c} \frac{1}{2} \sum_{c=1}^d \langle w_c, w_c \rangle + C \sum_{n=1}^l \xi_n \quad (14)$$

Subject to Eq 15:

$$\begin{cases} \forall n \in \{1, \dots, l\}, \forall c \in \{1, \dots, d\} \setminus \{y_n\} \\ \langle w_{y_n} - w_c, \phi(x_n) \rangle \geq 1 - \xi_n \\ \text{and } \forall n \in \{1, \dots, l\} : \xi_n \geq 0 \end{cases} \quad (15)$$

For learning structured data, Crammer & Singer's method is usually the SVM algorithm of choice because is the one sometimes denoted as simply multi-class SVM.

2.2.5 K-Fold Cross-validation

In order to find the optimal models it has been performed a k-fold cross validation on the data set. In particular, we performed the parameters optimization on the Experimental training set using Cross-validation method that consists to divide the training set into k subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining k-1 subsets. Thus; each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. The advantage of k-fold Cross validation is that all the examples in the Dataset are eventually used for both training and testing.

2.2.6 Selection parameters of SVM model

The process of determining the SVM model is greatly influenced by the selection of kernel. First of all, we need to choose our kernel. This is a parameter in itself. Each kernel has a different set of parameters, and will perform differently, so in order to compare kernels you will have to optimize each kernel's parameters.

The search for selected these parameters was conducted using a grid-search method [20] for all three kernel functions, the grid search technique is not always an exhaustive method (depending on the grid resolution) such that is the stable technique than adaptive search which can save not so much time and lead to some strange results, other automatic methods, such as [21,22,23], exist but are iterative and can be computationally expensive.

the selection of optimal parameters for different kernel type (linear, RBF, polynomial) and find the best model selection is an important research issue in the data mining area .For each kernel chosen we search the optimal parameters and we vary these parameters around values intervals and calculate the classification accuracy.

The parameters that could be varied are: C is the parameter for SVM, regardless of kernel chosen; it has nothing to do with the kernel. It determines the tradeoff between a wide margin and classifier error it was varied over Eq: 16

$$C \in \{2^{-5}, 2^{-3}, \dots, 2^{15}\} \quad (16)$$

For the polynomial kernel, gamma serves as inner product coefficient in the polynomial Eq:17.

$$k(x_i, x_j) = (\text{gamma} \langle x_i, x_j \rangle + \text{coef0})^{\text{degree}}$$

$$\text{gamma} \in \{1, 2, \dots\} \quad (17)$$

In the case of the RBF kernel, γ determines the RBF width. In both cases , γ and gamma was varied over Eq:18:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

$$\gamma \in \{2^{-15}, 2^{-13}, \dots, 2^{-5}, 2^3\} \quad (18)$$

The degree is the main parameter of polynomial kernel, but we can also vary gamma and coef0 to make the kernel non-symmetric Eq :19.

$$\text{degree} \in \{1, 2, 3, \dots\} \quad \text{coef} \in \{1, 2, 3, \dots\} \quad (19)$$

We take the parameters of classifier to minimize the generalization error of that classification problem, we try small and large values for these parameters then we decide which are better for the data by cross validation, the parameters are tuned using 5-fold cross-validation estimation performance. We select the optimal parameters by a two-step grid search.

In linear kernel we vary the parameter C, in the case of polynomial kernel we change the two parameters degree and gamma while keeping C and coef constant, and C and coef while keeping degree and gamma constant and for RBF kernel we vary the parameters C and γ .

First we do a coarse grid search for each of the kernel using the following sets of these values,

Thus combinations of parameters are tried in this step, An optimal pair $(C_0, \gamma_0, \text{gamma}_0, \text{degree}_0, \text{coef}_0)$ is selected from this coarse grid search. In the second step, a fine grid search is conducted around $(C_0, \gamma_0, \text{gamma}_0, \text{degree}_0, \text{coef}_0)$ with Eq:20

$$\begin{aligned}
 C &\in \{0.1c_0, 0.2c_0, \dots, 6c_0, 8c_0\} \\
 \gamma &\in \{0.1\gamma_0, 0.2\gamma_0, \dots, 6\gamma_0, 8\gamma_0\} \\
 coef &\in \{0.1coef_0, 0.2coef_0, \dots, 6coef_0, 8coef_0\} \quad (20) \\
 degree &\in \{0.1degree_0, 0.2degree_0, \dots, 0.6degree_0, 0.8degree_0\} \\
 gamma &\in \{0.1gamma_0, 0.2gamma_0, \dots, 6gamma_0, 8gamma_0\}
 \end{aligned}$$

All together combinations of C , γ , $gamma$, $degree$, $coef$ are tried in this step and we selected the ones providing the best accuracy. The final optimal hyperparameter is selected from this fine search. After those parameters are found, the whole training set is trained again to generate the final classifier.

2.2.7 K-nearest neighbors (KNN)

K-Nearest Neighbor algorithm (KNN) is part of supervised learning and it is one of the simplest using machine learning algorithms. Besides its simplicity, KNN is a widely used technique, being successfully applied in many applications in the field of data mining, statistical pattern recognition and many others [8, 24].

An object is classified by the “distance” from its neighbors, with the object being assigned to the class most common among its k distance-nearest neighbors. It is usual to use the Euclidean distance, though other distance measures such as the Manhattan distance can be used. To calculate distance between two vector positions in the multidimensional space.

The training process for KNN consists only of storing the feature vectors and class labels of the training samples and requires a parameter K , which is the number of near neighbors to consider. K -selection is the important problem we should take into account is how to choose a suitable K for making the classification more successful. Generally, according to Shakhnarovich et.al [25], larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. For the purpose of this assignment the parameter is searched by sequentially trying all of the possible values $(1, \dots, N)$, N = size of the dataset). This process is computationally expensive since N classification of the full dataset has to be made in order to find the value that gives less error as possible.

The only heuristic used to reduce the computation time is the assumption that the best value of K will be at most half of the size of the dataset $(1 < K < N/2)$.

Another important assumption was made about the classifier’s parameter: the K that performs the best on the training set is likely to perform well on the test set.

3. Experimentation and Results

Our experiments were performed on three face databases: face94, face96 and grimaces database. The datasets have both male and female subjects, and have representatives from 4 different races.

We split each DataSet on the training and testing sets both contains 10 classes, each class have 14 images for training and 6 images for testing Fig 4. The number of features is the dimension of image width*height=180*200.

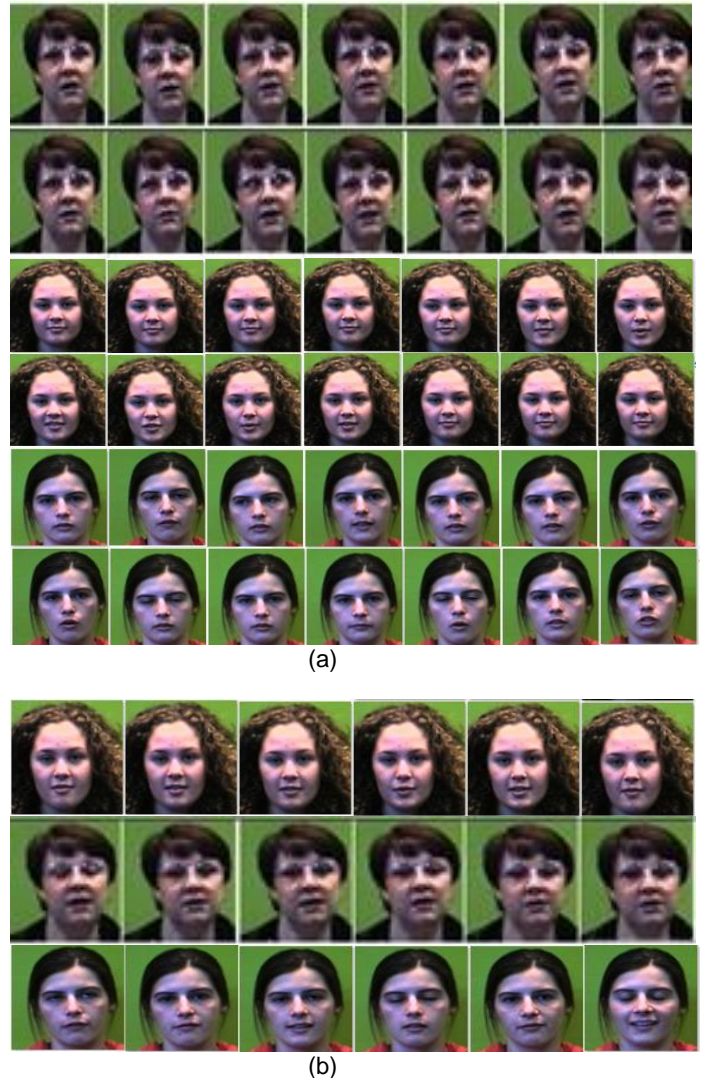


Fig. 4: Examples of (a) training and (b) test images (of database Face94) which used for classification by SVM kernels and KNN



Fig. 5 : Examples of (a) training and (b) test images (of database Grimace) which used for classification by SVM kernels and KNN

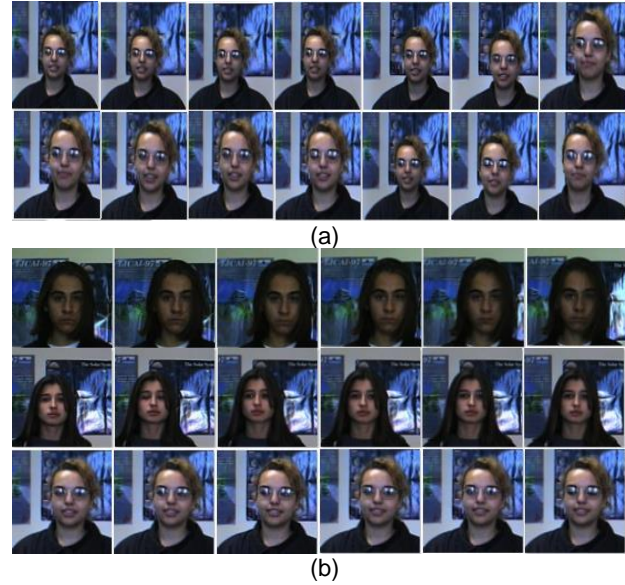
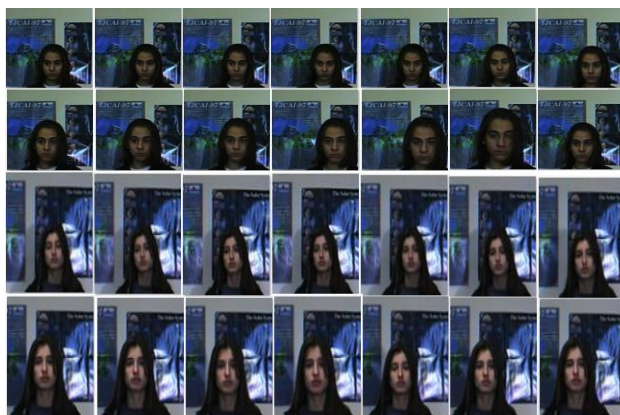


Fig. 6 : Examples of (a) training and (b) test images (of database Face96) which used for classification by SVM kernels and KNN

In the first the whole feature data set was reduced by LDA then we train the classifiers SVM and KNN .for SVM adopting different kernel we train the data and a grid search method with 5-fold cross-validation exercise was applied for selecting the best parameters in the final the models SVM was created. In the KNN the training process through 5 -fold cross-validations consists of searching the parameter k that gives the best performance on the training set.

In table 1,2,3,4, for each Datasets(Face94,Face96 and Grimace) it can be seen the optimal parameters of models SVM constructed in the training stage, each model requires the selection of optimal parameters for different kernel type and the best neighbors K for KNN.

Table 1: The optimal parameters for linear kernel SVM in different Database.

<i>DataSet</i>	<i>Linear</i>	
	Parameter : c	Cross validation accuracy
Face94	c = 0.011	90%
Face96	c =0.0312	90%
Grimace	c =0.217	91.42%

Table 2: The optimal parameters for polynomial kernel SVM in different Database.

<i>DataSet</i>	<i>Polynomial</i>	
Face94	Parameters: c, gamma, degree, coef	Cross validation accuracy
	(c =2) , (gamma = 1) , (degree= 2),(coef = 1)	73.4%
Face96	(c = 1) , (gamma = 3) , (degree= 3),(coef = 9)	60.4%
Grimace	(c = 2.25) , (gamma = 11) (degree= 3.3),(coef = 13)	80.4%

Table 3: The optimal parameters for RBF kernel SVM in different Database.

<i>DataSet</i>	<i>RBF</i>	
Face94	Parameters: c, gamma	Cross validation accuracy
	(c=0,26), (gamma=0.25)	97.14%
Face96	(c=2), (gamma=0.4525)	95.1%
Grimace	(c =2), (gamma =0. 16)	100%

Table 4: the best K neighbors of KNN in different Database.

<i>DataSet</i>	<i>Ref System : KNN</i>	
Face94	Best neighbors K	Cross validation accuracy
	(k=41)	82.08%
Face96	(k =19)	80.3%
Grimace	(k =33)	91.24%

In order to test the performance of classifiers SVM and KNN on the different Datasets, we compare the prediction accuracies of SVM adopting different type of kernel function with these selected parameters with the prediction accuracies of KNN.

Table 5 shows the classification accuracy on Testing Set using the classifiers SVM with different kernel and KNN on different dataset face94, face96, Grimace.

Table 5: The accuracy values obtained on three test Datasets while changing the SVM kernel (linear, polynomial, RBF) Vs Ref Sys (KNN).

<i>DataSet</i>	<i>Linear</i>	<i>Polynomial</i>	<i>RBF</i>	<i>Ref System : KNN</i>
Face94	78.34	73.2	90.8	57.29 (with k=41)
Face96	74.4	71.6	88.24	50.6 (with k=19)
Grimace	85.25	82.25	93.7	64.8 (with k=33)

For a given dataset, the classification accuracy varies significantly among different kernels function. This indicates that the classification accuracy vary with the SVM kernel .The testing accuracy of the RBF kernel is more efficient compared to the other two kernels, that may be due to a number of reasons: it can determine a non-linear decision boundary (not possible for a linear kernel), it has fewer parameters than the polynomial kernel and is consequently simpler to tune, and it faces less numerical difficulties (polynomial kernel value may go to infinity). and the other side the SVM classifier performs overall better than the KNN classifier ,The best overall performance was 93.7% obtained by SVM with RBF kernel and 64.8% by KNN so the SVM classifier outperforms the KNN when using the suitable kernel with the best parameters.

4. Conclusion

The proposed method is performed with less amount of memory which includes the efficient techniques as the use of LDA method gives a real performance on reducing data and the computational time involved in the training stage of cross-validation and grid search process which can improve SVM accuracy a little. And on the other hand the SVM classification is powerful for classification comparing with KNN, It was shown that considerably high classification accuracy can be achieved by selecting optimal set of parameters for the RBF kernel .In the future work will look into efficient technique using the Fisher kernel of SVM and automatic selection of optimal kernel parameters using a large dataset.

References

- [1]Computer Vision Science Research ml, 2007. Projects,<http://cswww.essex.ac.uk/mv/allfaces/faces94.ht>
- [2] Computer Vision Science Research ml, 2007. Projects, <http://cswww.essex.ac.uk/mv/allfaces/faces96.html>.
- [3] Computer Vision Science Research ml, 2007. Projects, <http://cswww.essex.ac.uk/mv/allfaces/grimace.html>
- [4] K. Etemad, R. Chellappa, Discriminant Analysis for Recognition of Human Face Images, Journal of the Optical Society of America A, Vol. 14, No. 8, August 1997, pp. 1724-1733.
- [5] Lawrence, S., Giles, C.L., Tsoi, A.C., and Back, A.D. (1998).Face Recognition: A Convolutional Neural Network Approach IEEE Transactions on Neural Networks, 8(1):98-113.
- [6] G. Guo, S.Z. Li, K. Chan, Face Recognition by Support Vector Machines, Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition, 26-30 March 2000, Grenoble, France, pp. 196-201.
- [7] D. Michie, D. J. Spiegelhalter, C. C. Taylor, and J. Campbell, editors. Machine learning, neural and statistical classification. Ellis Horwood, Upper Saddle River,NJ, USA, 1994. ISBN 0-13-106360-X.

- [8] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001).
- [9] A. Martinez, A. Kak, "PCA versus LDA", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, 2001.
- [10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.
- [11] Vapnik, V., 1995: The Nature of Statistical Learning Theory. Springer, N.Y.
- [12] Vapnik, V., 1998: Statistical Learning Theory. Wiley, N.Y.
- [13] Aizerman, M. A., Braverman, E. M., and Rozonoer, L. I., 1964: Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. Autom. Remote Control 25.
- [14] V. Vapnik, Statistical Learning Theory, John Wiley and Sons, New York, 1998.
- [15] V. Vapnik. Statistical Learning Theory. Wiley, 1998.
- [16] R. Rifkin and A. Klautau. In defense of one-vs-all classification. Journal of Machine Learning Research, 5:101-141, 2004.
- [17] C. Demirkesen and H. Cherifi. A comparison of multiclass SVM methods for realworld natural scenes. In J. Blanc-Talon and other, editors, Advanced Concepts for Intelligent Vision Systems (ACIVS 2008), volume 5259 of LNCS, pages 763-763. Springer, 2008.
- [18] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel based vector machines. Journal of Machine Learning Research, 2:265-292, 2002.
- [19] J. Weston and C. Watkins. Support vector machines for multi-class pattern recognition. In M. Verleysen, editor, Seventh European Symposium on Artificial Neural Networks (ESANN), pages 219-224, 1999.
- [20] C.W. Hsu, C.C. Chang, and C.J. Lin. A practical guide to support vector machines. Technical report, Department of Computer Science & Information Engineering, National Taiwan University, July 2003.
- [21] C. Staelin. Parameter selection for support vector machines. Technical report, HP Labs, Israel, 2002.
- [22] O. Chapelle and V. Vapnik. Model selection for support vector machines. In Proc. Advances in Neural Information Processing Systems (NIPS), 1999.
- [23] J.H. Lee and C.J. Lin. Automatic model selection for SVM. Technical report, Department of Computer Science & Information Engineering, National Taiwan University, November 2000.
- [24] LI LiLi, ZHANG YanXia and ZHAO YongHeng, "K-Nearest Neighbors for automated classification of celestial objects," Science in China Series G-Phys Mech Astron, Vol.51, no.7, July 2008, pp. 916-922.
- [25] Shakhnarovich, Darrell and Indyk, Nearest Neighbor Methods in learning and vision, MIT-press, 2005.

Anissa Bouzalmat is a PhD student at Sidi Mohammed Ben Abdellah University, Laboratory Intelligent Systems and Applications ISA in Morocco. He received his Master in Computer Science from the University of Sidi Mohammed Ben Abdellah in 2008. His current research interests are face recognition and detection.

Jamal Kharroubi is a professor at Sidi Mohammed Ben Abdellah University, Laboratory Intelligent Systems and Applications ISA (Communication Systems and Knowledge Processing group) in Morocco. His current research interests are Biometric, face recognition and detection..etc.

Arsalane Zarghili is a professor at Sidi Mohammed Ben Abdellah University, Laboratory Intelligent Systems and Applications ISA (Artificial Vision & Embedded Systems group) in Morocco. His current research interests are Biometric, face recognition and detection..etc.

Blind Fake Image detection

Nidhal El Abbadi¹, Adil Mohamad Hassan², Mudher Mohammed AL-Nwany³

¹Computer Science Dep., University of Kufa
Najaf, Iraq

²Mathematical Dep., University of Kufa
Najaf. Iraq

³Mathematical Dep., University of Kufa
Najaf. Iraq

Abstract

With the great convenience of computer graphics and digital imaging, it becomes much easier to alter the content of images than before without any visually traces to catch these manipulations. Many fake images are produced whose content is feigned. Thus, the images cannot be judged whether they are real or not visually. In order to detect fake images, this paper proposes a blind detection uses singular value decomposition (SVD) as a classifier to make a binary decision on whether an image is fake or real. This work is an improvements process to an existence method to detect fake image using SVD. The experimental results prove the effectiveness of this algorithm to detect any small changes in image even with one dot of real image.

Key words: SVD, fake image, singular value decomposition, image processing.

1. Introduction

Cameras are regarded as trustworthy devices and photos traditionally imply truth. Nowadays, digital photos have been widely used as historical records and as evidences of real happenings in applications from journalist reporting, police investigation, law enforcement, insurance, medical and dental examination, military, and museum to consumer photography.

While digital photos are conveniently used, their credibility has been severely challenged due to numerous fraudulent cases involving image forgeries, e.g. the fake results on human stem-cell research [5].

With the availability of powerful image editing tools, numerous image retouching techniques have become practical, which can be used to create great artistic works.

However, malicious modification of image content forms a serious threat to the secure and legal usage of digital images.

By skillful manipulation, forgery may be very difficult to recognize by the naked eye. Therefore, automatic detection of image forgery has attracted much research interest. In recent years, many image forgery detection techniques have been proposed, especially passive approaches which do not require any additional information besides the image itself [5] [6].

Some published methods make use of lighting abnormality [1], blur moment invariants, and similarity/dissimilarity of color and structural characteristics [3].

Digital tools have enabled easy image creation, modification and distribution, which make fraudulent image forgeries easier than ever.

Fakes are created either by merging two or more photos or altering an existing image. Because image manipulation happens at the pixel level, detection is not as easy as it was before the digital era. Tricky fakes can be exposed by algorithms that detect discrepancies or statistical irregularities at the bit level.

An image is authentic if it represents a witness to an actual event, place, or time.

A definition of image authenticity should enable us to distinguish an authentic image from the fake images, such as the 2D composite images and the 3D computer graphics images.

It is still a problem how to detect whether digital images are fake or real. Generally, there is an obvious boundary between the fake area and the real area, with the improvement of desktop photograph manipulation software, which cannot be used to distinguish fake images and real images. There are a many studies related to detect the fake images. In [5], a blind detection of photomontage is introduced using higher order statistics, where photomontage is a

similar concept with image fakery. In [6], a model based on bipolar signal perturbation is introduced to detect spliced images. These two papers used a statistics model and bi-coherence features to detect image forgery, and they are often used to detect human speech signal. In [1], Popescu and Farid introduced some techniques of exposing digital forgeries by detecting traces of resampling, which also tried to resolve the similar problem. Mahdian and Saic [2] used periodicity due to interpolation to perform blind image authentication. They introduced Radon transform on the basis of second derivative to detect rotation without estimation of the rotation angle.

In this paper, a SVD based fake image detection scheme is developed, which uses the change of the direction of the eigenvector in orthogonal subspace to detect the evidence of image fakery.

2. Singular Value Decomposition

In linear algebra, the **singular value decomposition (SVD)** is a factorization of a real or complex matrix, with many useful applications in signal processing and statistics.

SVD is based on a theorem from linear algebra which says that a rectangular matrix A can be broken down into the product of three matrices - an orthogonal matrix U , a diagonal matrix S , and the transpose of an orthogonal matrix V . The theorem is usually presented something like this:

$$A_{m \times n} = U_{m \times m} \cdot S_{m \times n} \cdot V_{n \times n}^T \quad \dots \dots \dots 1$$

2.1 Proposition:

If $A = USV^T$ and

$$v' = v + \alpha v_1 \frac{\|v\|}{\|v_1\|}, \quad A' = US'V'^T,$$

then $A = A'$ when $\alpha \rightarrow 0$.

Proof:

$$\lim_{\alpha \rightarrow 0} v' = \lim_{\alpha \rightarrow 0} \left(v + \alpha v_1 \frac{\|v\|}{\|v_1\|} \right) = v$$

$$\text{Then } v' = v \text{ when } \alpha \rightarrow 0 \quad S = S'$$

And this implies to $A = A'$

$$\text{since } [A = USV^T \text{ and } A' = US'V'^T]$$

2.2 Theorem:

$$\frac{\|v-v'\|}{\|v'\|} \frac{\|v_1+v_2\|}{\|v_1\|} = \sqrt{2} * \frac{\|v-v'\|}{\|v'\|}, \text{ and it's unique.}$$

Proof:

$$\frac{\|v_1+v_2\|}{\|v_1\|} = \frac{\sqrt{2} * \|v_1\|}{\|v_1\|} = \sqrt{2} \quad (\text{since } v_1 \text{ and } v_2 \text{ are orthonormal}),$$

$$\text{Then } \frac{\|v-v'\|}{\|v'\|} \frac{\|v_1+v_2\|}{\|v_1\|} = \sqrt{2} * \frac{\|v-v'\|}{\|v'\|}$$

To prove it's unique

Since v is unique (as **properties of SVD**),

And v' is unique (since $v' = v + \alpha v_1 \frac{\|v\|}{\|v_1\|}$)

So $\sqrt{2} * \frac{\|v-v'\|}{\|v'\|}$ is unique.

2.3 Corollary:

$$\frac{\|v-v'\|}{\|v'\|} \frac{\|v_1+v_2\|}{\|v_1\|} \rightarrow 0 \text{ when } \alpha \rightarrow 0.$$

Proof:

By theorem 2.2 we have

$$\frac{\|v-v'\|}{\|v'\|} \frac{\|v_1+v_2\|}{\|v_1\|} = \sqrt{2} * \frac{\|v-v'\|}{\|v'\|}$$

Then when $\alpha \rightarrow 0$ we have

$$\|v-v'\| \rightarrow 0$$

$$\therefore \sqrt{2} * \frac{\|v-v'\|}{\|v'\|} \rightarrow 0$$

$$\therefore \frac{\|v-v'\|}{\|v'\|} \frac{\|v_1+v_2\|}{\|v_1\|} \rightarrow 0 \quad \text{when } \alpha \rightarrow 0.$$

2.4 Corollary:

$A = B$ If and only if

$$\sqrt{2} * \frac{\|v_{A-v'_A}\|}{\|v'_A\|} = \sqrt{2} * \frac{\|v_{B-v'_B}\|}{\|v'_B\|}$$

Proof:

By theorem 2.2

$(\sqrt{2} * \frac{\|v_A - v'_A\|}{\|v'_A\|})$ is unique and,

$(\sqrt{2} * \frac{\|v_B - v'_B\|}{\|v'_B\|})$ is unique,

$\therefore \sqrt{2} * \frac{\|v_A - v'_A\|}{\|v'_A\|} = \sqrt{2} * \frac{\|v_B - v'_B\|}{\|v'_B\|}$ If and only if

$A = B$

3. Methodology

The proposed method to detect the fake image can achieve by processing image in many steps as follow:

1. First the original image A is transformed using SVD:

$$A = USV^T \dots\dots\dots 2$$

Where U and V are the orthogonal matrices, V^T denotes the transpose of V , and S is a diagonal matrix whose diagonal elements can form a column vector v .

2. Two secret column vectors v_1 , and v_2 are constructed, which satisfy

$$\|v_1 \cdot v\| = 0, \quad \|v_2 \cdot v\| = 0, \quad \text{and} \quad \|v_1 \cdot v_2\| = 0$$

Where ' \cdot ' denotes the inner product.

3. The main goal is to protect the image before publishing it to the public; this will be achieved by changing the diagonal of S matrix result from relation (2) with new elements counted by the following equation:

$$v' = (v + \alpha v_1) \times \frac{\|v\|}{\|v_1\|} \dots\dots\dots 3$$

Where α is a scalar factor, which is set to 0.0001 for the purpose of this research.

The vector v' from relation (3) is restored as new diagonal elements into zero matrix S' correspondingly, for that new image will be constructed (A') as a protected image from the following relation:

$$A' = US'V^T \dots\dots\dots 4$$

A' is the preprocessed image and publish to public. SVD is robust to slight alteration of images, i.e., the vector v is stable under slight alteration of the image. In proposed image preprocessing procedure, the alteration of the vector v in relation (2) is very small, keep $v' \approx v$ by **Proposition 2.2**, so the image preprocessing does not change the quality of origin image significantly (which will be demonstrated in the later examples).

4. Another suggestion in current research is using auto threshold (T_{th}) by **Theorem 2.3**. instead of constant threshold (0.01) for all images as in previous researches, which mean for each image

there is specific threshold counted by the following relation:

$$\text{Threshold} = \sqrt{2} \times \frac{\|v - v'\|}{\|v\|} \dots\dots\dots 5$$

3.1 Fake Image Detection

When we have an image A^\wedge and need to check whether it's fake or not, it will be decompose using SVD as follow:

$$A^\wedge = U^\wedge S^\wedge (V^\wedge)^T \dots\dots\dots 6$$

Then the vector v^\wedge is extract from the diagonal elements of S^\wedge . The proposed fake image detection can be given as follow:

$$P = \left\| \frac{v^\wedge}{\|v^\wedge\|} \cdot \frac{v_1}{\|v_1\|} \right\| + \left\| \frac{v^\wedge}{\|v^\wedge\|} \cdot \frac{v_2}{\|v_2\|} \right\| \dots\dots\dots 7$$

Where P is the detection value, which denotes the fake factor of the test image.

if $P > T_{th}$, then the tested image is fake, otherwise it isn't fake by **Corollary 2.5**. In the current proposed scheme, the two secret vectors, v_1 and v_2 , are the key construction, on which the detection result depends.

We supposed

$$\|v_1\| = \|v_2\| \quad \text{and} \quad v^\wedge = v' + v_f \dots\dots\dots 8$$

However, the vector v^\wedge is composed of two vectors, one is the original vector v' , and the other is the fake vector v_f .

Relation (7) can rewrite by using the equivalents' in relations (8) to get:

$$P = \left\| \frac{v' + v_f}{\|v^\wedge\|} \cdot \frac{v_1}{\|v_1\|} \right\| + \left\| \frac{v' + v_f}{\|v^\wedge\|} \cdot \frac{v_2}{\|v_2\|} \right\| \dots\dots\dots 9$$

Relation (9) can be computed approximately as follow:

$$P \approx \left\| \frac{v_f \cdot (v_1 + v_2)}{\|v^\wedge\| \|v_2\|} \right\| = \frac{\|v_f\| \|v_1 + v_2\|}{\|v^\wedge\| \|v_2\|} = \sqrt{2} \frac{\|v_f\|}{\|v^\wedge\|} \dots\dots\dots 10$$

by Corollary 2.4

It is obvious that the detection factor P depends on the fake vector v_f where $v_f = v' - v^\wedge$ and since v', v^\wedge are unique (**properties of SVD**) so that v_f is unique, while v_f denotes the fake vector of the tested image, so the detection value P can reflect the status of fake image. Also, there are two secret vectors in our proposed process, v_1 and v_2 , which do not have influence on the absolute value of detection result.

4. The result

To prove the proposed method we will take some of image and make intended change on it to see how this algorithm works.

4.1 The first image is the famous image (Lenna) as shown in fig 1, the origin image will protected by applying relations (3, and 4), it is clear the origin image have no perceptual difference from the protected image. Intended changing made on the protected image by changing the face of Lanna. The auto threshold for this image which is (1.3555e-004) counted by relation 5, also the fake factor counting using relation (10) which is equal (0.0185), it is clear ($P > T_{th}$), then the tested image (C) is fake.



Fig 1: (A) origin image. (B) Protected image (image 1). (C) Fake image

4.2 The second example is to test the effect of rotation on the result of this method. Protected image of Lenna (image B in fig 1) is rotated with 90 degree.

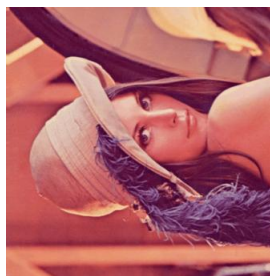
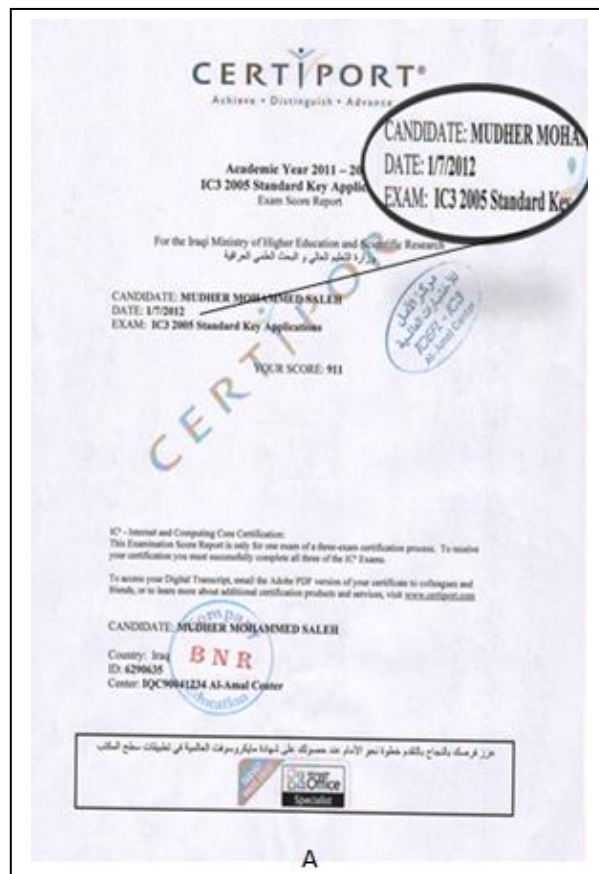
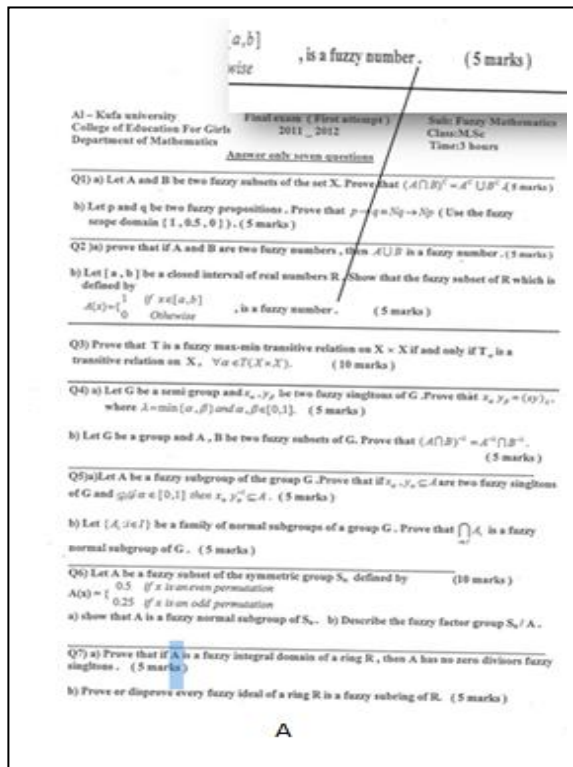


Figure 2: rotated image for test (image2)

Then the counted fake factor P of image in fig 2 was (1.3462e-004), so we decided by **Corollary 2.5**. this image is original ($P < T_{th}$).

4.3 Third example to test small change in copy of IC3 certificate, the change is made by changing one number only in the date (1/7/2012 change to 1/7/2013). Fig 3 shows the two IC3 images.





The P of this image (2.4055e-004) greater than threshold T_{th} of this image (1.0191e-004). So it's fake by Corollary 2.5.

4.4 The fourth example done on the copy of questions (Fuzzy mathematics exam) which we change just one dot on it as shows in fig 4. The threshold of the protected image was (1.4142e-004) while the fake factor $P = 1.9060e-004$, which is greater than threshold of this image. So the image is fake by Corollary 2.5.

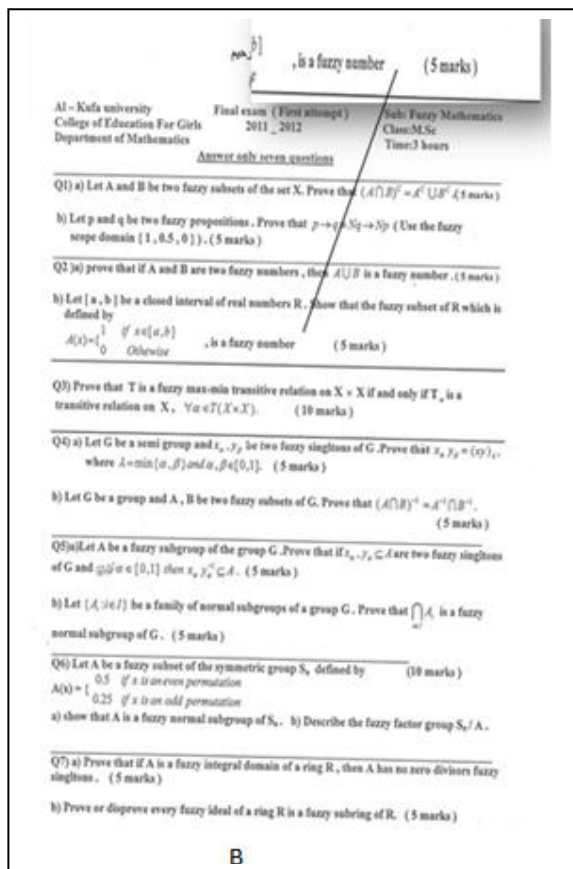


Fig 4: (A) the protected image (image 4). (B) Fake image.

4.5 To summarize the results from the above examples, and to prove the results, table 1 compares the results from the proposed method with the result by the paper of [7].

Table 1: Results of experiments compared with Lai Chung method [7].

Method	No. of image	1	2	3	4
Previous method	T_{th}	0.01			
	PSNR	Average = 38			
Proposed method	Fake factor (P)	0.0185	1.3462e-004	2.4055e-004	1.9060e-004
	T_{th}	1.3555e-004	1.3555e-004	1.0191e-004	1.4142e-004
	PSNR	107.0874	107.0874	91.3923	87.6442
	Decision	Fake	Origin	Fake	Fake
Reality of Images	Fake	Origin	Fake	Fake	

5. Conclusions

In this paper, we used SVD technique for fake image detection Scheme. Before the images are published to public, some assistant information is inserted into them. Recurring to the secret information, the work in [7] introduced the mathematical SVD operation in fake image.

In this paper a modification of pervious work especially in [7] is achieved.

We calculated the two secret vectors in new method. The important improvement in this work is the scalar factor expanding from 0.01 to 0.0001.

Also there is another improvement related to threshold, where the threshold in [7] was constant for all images, while in this work the threshold will be different for each image, auto threshold determine for each image instead of constant threshold (0.01).

These improvements enhance the detection efficiency and eliminate false detection, in [7] the false positive rate was 0.8% when checking 1000 image while in this paper the rate decreases to 0.0%.

We must note that all method of detect fake image need the original image and the fake image to make recognition between them but in current method the origin image no longer needed.

A comparison between previous work and this work is studied where the sample test taken as color images.

We can make a decision that our SVD scheme is very excellent in detecting fake image and it's sensitive for any small area modified in any image.

References

- [1] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling", IEEE Trans. Signal Process., vol.53, no.2, pp.758–767, Feb. 2005.
- [2] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation", IEEE Trans. Inf. Forensics Security, vol. 3, no. 3, pp. 529–538, Sep. 2008.
- [3] Steve J. Leon; "Linear Algebra with Applications", Macmillan Publishing Company, New York; 1996.
- [4] T. Konda, Y. Nakamura, "A new algorithm for singular allude composition and its parallelization", Parallel Comput. (2009), doi:10.1016/j.parco.2009.02.001
- [5] T.T. Ng, S.F. Chang, and Q. Sun, "Blind detection of photomontage using higher order statistics", IEEE Int. Symp. Circuits Syst. (ISCAS), pp.688–691, May 2004.
- [6] T.T. Ng and S.F. Chang, "A model for image splicing", IEEE Int. Conf. Image Process. (ICIP), pp.1169–1172, Oct. 2004.
- [7] Wei LU, Fu-Lai Chung, and Hongtao LU, "Blind Fake Image Detection Scheme Using SVD", IEICE TRANS. COMMUN., VOL.E89–B, NO.5 2006.



Nidhal El Abbadi, received BSc in Chemical Engineering, MSc, and PhD in computer science, worked in industry and many universities, he is general secretary of colleges of computing and informatics society in Iraq,

Member of Editorial board of Journal of Computing and Applications, reviewer for a number of international journals, has many published papers and three published books (Programming with Pascal, C++ from beginning to OOP, Data structures in simple language), his research interests are in image processing, biomedical, and steganography, He's Associate Professor in Computer Science in the University of Kufa – Najaf, IRAQ.



In 1963 born in Najaf city, Iraq .Has MSc in applied mathematics from university of Technology, Baghdad. Has PhD in fractal geometry. He is a viewer of scientific journals and conferences. Member of the ministerial committee for updating the career. Member

of ministerial virtual learning committee. More than 25 papers were published in locally journals and conference.



He graduate from College of education department of mathematics from AL-Mustanseria University/ Baghdad 1994, Worked as a teacher of Mathematics in the secondary schools, currently MSc. student in

university of Kufa, mathematical department.

Iterative Decoding Termination Schemes for Turbo Code Performance Optimization In Mobile Wi-Max Environment

Jagdish D. Kene¹ and Dr. Kishor D. Kulat²

¹ ECE Department, Visvesvaraya National Institute of Technology, Nagpur, Maharashtra, India.

² ECE Department, Visvesvaraya National Institute of Technology, Nagpur, Maharashtra, India.

Abstract

Use of turbo codes is more popular in most of the wireless applications, because of its greater Error control ability. The BER performance reaches to the Shannon's channel capacity limit. Turbo code implementation using SISO decoders with iterative MAP decoding algorithms introduces large time delay to recover the transmitted information bits. This results in increasing Wi-Max system complexity and storage requirement (Memory size). In this paper, the efforts have been made to propose the methods for effective termination of iterations to make the decoder efficient, in terms of reduction in the time delay and the requirement of memory size while maintaining the BER performance. Authors have propose various termination techniques which help in reducing the complexity as compare to conventional MAP decoding algorithm for same BER performance.

Keywords: Turbo code, Iterative decoding, MAP Algorithm, Termination detection schemes

1. Introduction

In advanced wireless communication network system such as Wi-Max, the error control code used for controlling the signal errors is turbo code. Using turbo code one can think of reaching capacity limit defined by the Shannon for low power transmission [1]. Turbo code provides high data rates for low order modulation schemes such as BPSK or QPSK [2]. Turbo code is implemented by employing (i) Parallel concatenated encoder structure and (ii) Soft input soft output iterative decoders [2], [3]. This decoder uses Maximum A Posteriori (MAP) decoding algorithm to generate the soft output. The operation of the turbo decoder is based on iterative decoding which is considered as the main feature of turbo codes. Since the MAP decoding algorithm needs relatively large number of iterations to achieve the expected BER performance at low SNR [4]. This leads to excessive time delay and computational complexity for deciding the system performance. The overall system complexity increases with the number of iterations carried out. Hence to reduce the number of iterations a scheme for fixed number of iteration

is planned to get expected BER at relatively low SNR. The scheme describes the procedure for termination of iterations. When decoder approaches the performance limit of the system, no significant improvement is expected for further iterations, so it is better to stop the decoder operation. Based on this, the performance limit threshold of the system has been decided for a scheme with fixed number of iterations and the decoded output of such scheme send for further processing in the system.

In order to reduce average computational time without degrading the system performance, turbo decoder terminates the iterations for each individual frame immediately after receiving the bits as estimated. The decoder complexity can be reduced by effective termination schemes. The authors have proposed various schemes for effective termination of iterations which are refer to as decoder stopping rule [5]. This will improve the turbo code performance in the field of mobile Wi-Max. The BER vs. SNR characteristics for the system using each stopping rule is simulated and relative comparison is carried out to discuss the system performance.

2. Turbo Decoder

Turbo decoder extracts the systematic bits and recursive bits from the received information. A block diagram of turbo decoder is as shown in figure 1. The input to the turbo decoder is the received sequence (R_k) consisting of systematic and recursive bits. Decoder consists of two soft-input soft-output decoders namely DEC1 and DEC2. DEC1 decodes sequence from recursive systematic encoder1 while DEC2 decodes sequence from recursive encoder2 of turbo encoder. Each of the decoder operating with Maximum A Posteriori (MAP) Algorithm in an iterative decoding process. The DEC1 receives systematic sequence and parity sequence as inputs and generates the soft estimated output called extrinsic data which is label as EXT1. For first iteration this does not contains any information. The output sequence of DEC1 is interleaved and passed as input to second decoder DEC2. The systematic received bits and parity bits also input to DEC2

with the interleaved form of the extrinsic information EXT1 from the first decoder DEC1 [3].

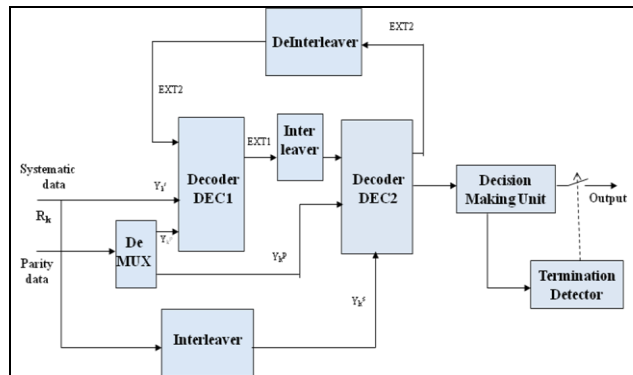


Figure 1: The structure of Turbo Decoder.

The DEC2 produces the soft estimated output called EXT2 which then de-interleaved and feedback to DEC1. This procedure is repeated in an iterative manner and continues until the bit error rate is zero (converges). At the end of decoding process simple threshold operation is performed to carry out hard decision on the soft output of the second decoder DEC2 [4].

This paper modifies the turbo decoder structure by implementing the termination detector and decision making unit. Termination detector is comparing the soft estimated output of decoder DEC2 for individual frame of current iteration to the threshold defined by the termination schemes [5]. Decision maker unit performed hard decision on the soft output of second decoder and stop the process of iterative decoding [7].

2.1 MAP Algorithm

In turbo decoding process, encoded information sequence (X_k) is transmitted over an AWGN channel and a noisy signal (Y_k) is received at the destination. In general, each decoder computes the Log Likelihood Ratio (LLR) to pull out the information data bit from received signal (Y_k). The LLR is calculated for each bit (d_k) of data block length N is defined in equation (1) [1].

$$L(d_k) = \text{Log} \left[\frac{\text{Pr}(d_k=1|Y)}{\text{Pr}(d_k=0|Y)} \right] \quad (1)$$

Where $\text{Pr}((d_k=1)|Y)$ is A Posteriori Probability (APP) of the information input data at time k (d_k). APP is the measure of probability of correct decision that helps the MAP algorithm to minimize the bit error probability [4], [11]. The MAP decoding algorithm can use a recursive technique and estimate the forward, backward and branch metric for each encoder state at time k [4],[12] is defined in equation 2. The Map algorithm calculates the LLR for each received data bit as

$$L(d_k) = \ln \left[\frac{\sum_{S_k} \sum_{S_{k-1}} \gamma_1 S_{k-1}, S_k \alpha S_{k-1} \beta S_k}{\sum_{S_k} \sum_{S_{k-1}} \gamma_0 S_{k-1}, S_k \alpha S_{k-1} \beta S_k} \right] \quad (2)$$

Where α is the forward state metric, β is the backward state metric, γ is the branch metric and S_k is the encoder trellis state at time k [12]. But the original MAP algorithm demands a large number of operations involving multiplication and exponentiation that needs excessive memory and more computational efforts in both forward and backward recursion. The complex mathematical calculations and computations of MAP decoding algorithm can be simplified by performing computation in the logarithmic domain [11] known as Log-MAP algorithm. Thus the MAP decoder provides soft output in terms of logarithms of APP of each data bits. This can be achieved as multiplications turn into additions and additions into maximum detection with an additional correction term [11], the maximum detection operation is defined by equation 3

$$\begin{aligned} \text{Max}^* x, y &= \ln e^x + e^y \\ &= \text{Max} x, y + \ln 1 + e^{-|y-x|} \end{aligned} \quad (3)$$

In log-MAP algorithm, the extrinsic information can be obtained by performing subtraction between input and output of both decoders. The resultant extrinsic sequence is fed back to first decoder through the deinterleaver for further iterations. Turbo decoding algorithm carries out fixed number of iterations per frame or block typically between 6 and 15 [8]. Therefore Log-MAP algorithm requires large memory and more complex computation analysis that leads to extra time delay for decoding the transmitted information. Even though the Log-MAP algorithm produces the decoding delay with complex computation, but the BER performance of this algorithm is unmatched to other existing decoders [3]. So the Log-MAP algorithm uses in the system while the system performance does not degrade. This is possible to reduce the decoding iterations that can be achieved by terminate the operation of decoder before fixed decoding iteration. Author has suggested various iteration termination schemes in this paper.

3. Termination schemes

Iterative decoding is a key feature of turbo codes. As the number of iterations increases (within certain limit), the bit error rate (BER) of the decoder decrease. In this process a fixed number of iterations typically between 6 and 15 are chosen and each frame is decoded for these iterations.

Usually it is set with the worst corrupted frames in the received data [7].

In the received information, it may possible that some of the frames don't have errors or most of the frames content least errors that need lesser number of iterations to converge. It would reduce the average computation without performance degradation. Thus we can improve the average decoding speed and decrease the power consumption of turbo decoder. It is possible if the termination scheme with a variable number of iterations per frame is implemented [6]. The decoder terminated the iterations for each individual frame immediately after the bits are correctly estimated [5], [6]. This is impractical when the transmitted bits are unknown. Various schemes have been proposed to control the operation of termination detector. We refer these methods as Stopping Criterion for iterations in turbo decoder are define as

- i. Cyclic Redundancy Check (CRC).
- ii. Cross Entropy (CE).
- iii. Sign Change Ratio (SCR).

3.1 Cyclic Redundancy check

This is a simple criteria based on hard decisions. In this approach, CRC bits are added to the end of each information frame. Modified frame is sent to the turbo decoder through the channel. At the decoder end, CRC bits are used for checking the errors and decoder makes the hard decision. Decision maker unit compare hard decisions of two successive Iterations. It stops the decoder operation when results of two conjugative frames in the iterations are same (In other words, CRC bits are error free) [7], [14].

This theme uses limited CRC bits for the determination of errors. Large bits are not economical because that degrades the transmission efficiency [14]. Therefore this technique is not applicable for the large frame size.

3.2 Cross Entropy (CE)

The cross entropy (CE) advice by Hagenauer, Offer and Papke is the useful criterion for iterative decoding [11]. This scheme is based on determining the Cross entropy between estimated output of two different SISO decoders. The CE between two estimated output distributions $L_2(d)$ and $L_1(d)$ of a received data sequence $d = \{d_1, d_2, \dots, d_N\}$ is defined as

$$Y_{CE} = E_p \left\{ \text{Log} \frac{L_2(d)}{L_1(d)} \right\} \quad (1)$$

Where E_p denotes the expectation operation perform over estimated output $L_2(d)$. Since d is independent and distributed identically then the equation (1) can rewritten as

$$Y_{CE} = \sum_k E_p \left\{ \text{Log} \frac{L_2(d_k)}{L_1(d_k)} \right\} \quad (2)$$

In turbo decoder, using an iterative decoding algorithm, the CE between estimated output distributions of two decoders at i^{th} iteration can be approximated as

$$T(i) = \sum_k^N \left\{ \frac{\left| \Delta L_{e2}^{(i)} \hat{d}_k \right|^2}{\exp \left| L_1^{(i)} \hat{d}_k \right|} \right\} \quad (3)$$

Where N is the frame length, $\Delta L_{e2}^{(i)} \hat{d}_k$ denoted as difference of extrinsic LLR between present (i) and previous ($i-1$) iteration of detected information bits (d_k) generated at the output of second decoder. Whereas

$L_1^{(i)} \hat{d}_k$ is the extrinsic LLR of received information generated by the first decoder at i^{th} iteration. Cross entropy of estimated LLR $T(i)$ is calculated for each frame in an iteration [6]. The result of each iteration is compared with a threshold in terms of $T(i)$. Iterative decoding stops when the value of $T(i)$ (much smaller) is measure in the range of threshold define as $(10^{-2} \sim 10^{-4})T(1)$. The mean of which two extrinsic distribution are closed enough and result of that decision maker unit stop the further iterations and hard decision send to the output port. CE reduces the average iterations and complexity with performance degradation.

3.3 Sign Change Ratio

A practically simple and computationally effective scheme for CE termination is Sign Change Ratio (SCR) scheme. The scheme directly focuses on number of sign changes of extrinsic LLR of each data bit in an iteration. It gives number of sign changes $s(i)$ for the extrinsic information for two successive iterations. The number of sign changes $s(i)$ decides the threshold for turbo decoder termination of iteration process [6], [12]. The necessary condition to fulfill this is $s(i) \leq Q \cdot N$. Where, $Q \Rightarrow$ Constant, ranges from 0.005 to 0.03 i.e. $0.005 \leq Q \leq 0.03$, and $N \Rightarrow$ Frame size. In our case, it comes out to be 19.2(Upper limit) taking Q as 0.03 and 3.2 (lower limit) for Q as 0.005 at 640 bits frame size. Instead of computing cross entropy the scheme suggested for the count of sign change in extrinsic LLR for successive iterations. Hence SCR more effectively reduces the computational efforts.

4. Simulation Results

The simulation is carried out using MATLAB version 7.10 using turbo codes for following specifications

- Code rates --- 1/3
- Encoder shift register --- 3.
- Generator Polynomials --- 13_{dec} and 15_{dec} .
- Digital Modulations --- BPSK/ QPSK.
- Signal to Noise Ratio --- 0 - 2.0 dB
- Turbo decoding algorithm --- Log-MAP.
- Amount of Information --- 640 bits/ Frame.
- Terminations schemes --- CRC, CE and SCR.

Using the above parameters, the simulation of the Wi-Max system [3] using turbo code is set for desirable (fixed) number of iterations. Turbo decoder uses Log-MAP algorithm. Decoding algorithm is modified by implementing the termination schemes as mentioned above. Termination scheme updates the system performance at lower SNR ranges from 0 to 2 dB. The BER performance of these three methods has been shown in Figure 2. We can observe that performance of CE is more effective towards minimum values of SNR due to soft estimation of received bits. Due to hard estimation, the CRC and SCR provide quite poor performance but relatively better than original decoding algorithm.

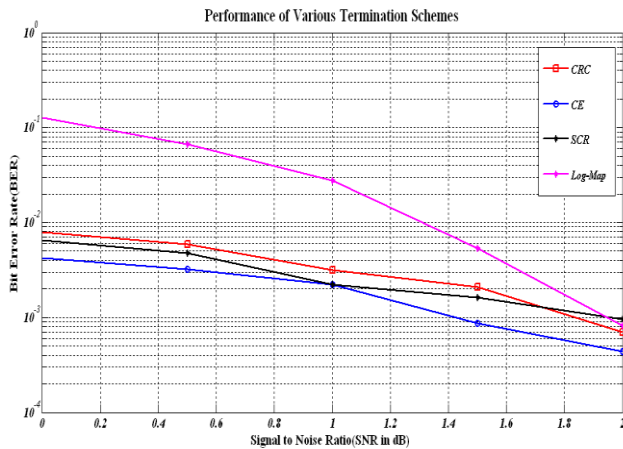


Figure 2: BER Performance of Various Termination Schemes.

The numbers of iterations consumed by each stopping method with respect to error probability are tabulated in the table1. CE operating on soft estimation of received data consumed comparatively less iterations towards lower values of SNR as observed in table1. The Bar diagram in percentage saving of computational analysis due to reduced iterations is shown in figure 3. The CE provides smooth functioning of decoder with soft estimation. These termination methods produced minimum number of decoding Iterations, decoding delay, memory requirement and the hardware complexity provided for optimizing turbo code performance of the Wi-Max system.

Table 1: Iterations consumed for various termination schemes.

SNR\Methods	CRC	CE	SCR
0.1 dB	9	7	8
0.5 dB	4	6	5
1.0 dB	4	4	5
1.5 dB	3	3	4
2.0 dB	3	3	3

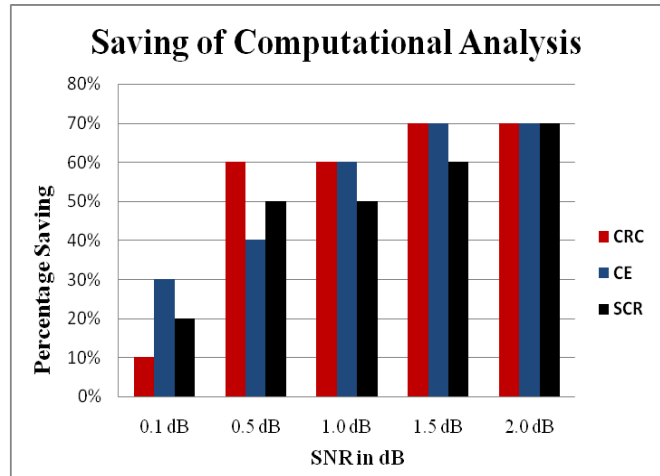


Figure 3: The saving of computational analysis.

4. Conclusions

BER Performance of turbo codes is improved by Iterative Decoder. Map Decoding Algorithm is implemented to get soft estimated output. Turbo MAP decoder achieves better performance with quite higher complexity that reduces the coding efficiency. Various termination schemes are implemented that helps to minimise (i) The number of decoding Iterations required, (ii) Decoding Delay, (iii) Memory requirement and (iv) The hardware complexity. Cross entropy is most effectively performed towards very low SNR with respect to system stability. Turbo codes approaches channel capacity at very low SNR ranges from 0 to 2 dB. Termination schemes speedup the turbo decoder iteration process and promises to minimise the iteration time delay without degrading the Wi-Max system performance.

References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes," *Proceeding of IEEE ICC 93*, PP. 1064-1070.
 [2] Jagdish D. Kene, Kishor D. Kulat, "Performance Evaluation

- of IEEE 802.16e Wi-Max Physical Layer”, IEEE Conference Proceedings' NUICONe, 8-10 December 2011, PP. 1-4.
- [3] Jagdish D. Kene, Kishor D. Kulat, “Performance Optimization of Physical Layer Using Turbo Codes: A Case Study of Wi-Max Mobile Environment”, IEEE Conference Proceedings, ET2ECN, 19-21 December 2012.
- [4] Jagdish D. Kene, Kishor D. Kulat, “Soft Output Decoding Algorithm for Turbo Codes Implementation in Mobile Wi-Max Environment”, Published in the Elsevier Journal Vol. 6, PP. 666-673, 6-8 October 2012.
- [5] Yufei Wu, Brian D. Woerner, William J. Ebel, “A Simple Stopping Criterion for Turbo Decoding”, IEEE Communications Letters, vol. 4, no. 8, AUGUST 2000.
- [6] D. Bokolamulla, T. Aulin, “A New Stopping Criterion for Iterative Decoding”, IEEE International Conference on Communications, 20-24 June 2004, PP. 538 – 541.
- [7] Zhen-chuan, Li Ying, “Research and Improvement on Stopping Criterion of Turbo iterative Decoding”, IEEE Conference Proceedings of International Conference on Wireless Communications Networking and Mobile (WiCOM), 23-25 Sept. 2010.
- [8] A. Taffin, “Generalized stopping criterion for iterative Decoders”, Electronics Letters, 26th June 2003, vol. 39, no. 13.
- [9] Nam Yul Yu, Min Goo Kim, Yong Serk Kim and Sang Uoon Chung, “Efficient stopping criterion for iterative decoding of turbo codes”, Electronics letters, 9th January 2003, vol. 39, no. 1.
- [10] Fan-Min Li, An-Yeu Wu, “On the New Stopping Criteria of Iterative Turbo Decoding by Using Decoding Threshold”, IEEE Transactions on Signal Processing, vol. 55, no. 11, November 2007.
- [11] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” IEEE Transactions on Information Theory, vol. 42, pp. 429–445, March 1996.
- [12] R. Y. Shao, S. Lin, and M. P. C. Fossorier, “Two simple stopping criteria for turbo decoding,” IEEE Transactions on Communications, vol. 47, PP. 1117–1120, August 1999.
- [13] P. Robertson, E. Villebrum, and P. Hoeher, “A Comparison of Optimal and Sub-Optimal MAP Decoding algorithms Operating in the Log-domain”, International Conference on Communications, PP.1009-1013, June 95.
- [14] H.R. Sadjadpour and F. Park, “Maximum A Posteriori decoding algorithms for turbo codes”, In digital wireless communication II, Proc of SPIE, vol. 4045, pp. 73-83, 2000.
- [15] G. Bauch, H. Khorram, J. Hagenauer, “Iterative equalization and Decoding in Mobile Communications Systems”, Second European Personal Mobile Communications Conference (EPMCC '97).
- [16] IEEE 802.16-2006: "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems”.
- [17] IEEE. Standard 802.16-2004. Part16: “Air interface for fixed broadband wireless access systems, October 2004”.

Mr. Jagdish D. Kene completed his bachelor degree in Electronics Engineering in 2001, from Manoharbai Patel Institute of Engineering and Technology, Nagpur University, Nagpur and Master degree in Electronics Engineering in 2005, from Yashwantrao Chawan College of Engineering, Nagpur University, Nagpur, M. S. India. He is currently associated with U. C. O. E. Nagpur, as Assistant Professor in Electronics Engineering Department having total experience



of 11 years. He is currently pursuing P-hd in the field of wireless communication under the supervision of Dr. Kishor D. Kulat at Visvesvaraya National Institute of Technology (VNIT), Nagpur, M.S., India. His research work is related to Performance evaluation and optimization solution of physical layer by implementing various error correction coding techniques in mobile Wi-Max environment. He has published one Journal Paper in Elsevier, three papers in International Conferences in his research area. He also publish 1 paper in International Conference and more than 8 have been published in National Conferences in his academic carrier. He is member of Professional societies like ISTE. He believes that Trust and Honesty is the secrets of success.

Dr. Kishor D. Kulat completed his degree in Electrical Engineering, BE in 1980, from VRCE (at present VNIT) Nagpur and ME degree in 1984 from VJTI, Mumbai, India. He completed his Ph.D. degree in Electronics Engineering, in the year 2003 from VNIT, Nagpur. Having a total experience of more than 25 years, he is currently associated with VNIT, as Professor and Head in the Electronics & Computer Science Department. With his



profound knowledge & experience in his field he is guiding around 15 research scholars for their doctoral degree, 9 have been awarded the Ph. D. degree. He has published around 80 Journal Papers, more than 75 papers in International Conferences & more than 100 have been published in National Conferences. He has worked as Reviewer for many National & International Conferences. He is a member of Board of Studies for Electronics Engineering, Nagpur University for last 10 years. He is member of Professional societies like IETE, IEI and ISTE. With all his faith in God, Dr. K. D. Kulat believes in achieving excellence through the process of continuous upgradation.

DESIGN A SECURE ELECTRONIC VOTING SYSTEM USING FINGERPRINT TECHNIQUE

Sanjay Kumar¹, Manpreet Singh²

¹Computer Science & Engineering Department, Maharishi Markandeshwar University
Mullana, Ambala, Haryana-133203, India

²Computer Science & Engineering Department, Maharishi Markandeshwar University
, Ambala, Haryana-133203, India

Abstract

Fingerprint biometric is the most widely deployed publicized biometrics for identification. This is largely due to its easy and cost effective integration in existing and upcoming technologies. The integration of biometric with electronic voting machine undoubtedly requires less manpower, save much time of voters and personnel, eliminate rigging, ensure accuracy, transparency and fast results in election. In this paper, a framework for electronic voting machine based on biometric verification is proposed and implemented. The proposed framework ensures secured identification and authentication processes for the voters and candidates through the use of fingerprint biometrics.

Keywords: EVM, Fingerprint, Biometric, Fingerprint module.

1. Introduction

1.1 Traditional Voting Process:

Traditional voting process can be divided into different phases [10]:

1. *Authentication:* In this phase, voter authenticates himself or herself by showing his or her voting card, this step is public and verified by the presiding officer. At the end of authentication process, presiding officer give a ballot paper to voter to cast his or her vote.
2. *Vote:* The vote takes place in a protected booth where voter cannot be seen by any person. The voter cast their vote by writing it with a pen on the paper ballot, folds the ballot paper and put into the ballot box where all the votes are mixed.
3. *Vote counting:* At the end of voting time, the presiding officer collect the ballot box containing all ballot papers and submit it to the counting centre. After that with the help of members of the election committee nominated by election commission of India, the ballot boxes are opened and votes are counted and the results are then announced.

4. *Verification:* Various types of verification process are used, most procedure are public and verified by the representative of candidates of competing parties. Recount is also possible if there is any fraud or error.

Conventional voting systems are not efficient due to long period of preparation, bogus voting, include papers, punch cards, mechanical levers, optical-scan machines [1]. These systems are not efficient as they are conducted manually and therefore very often are not accurate. As a consequence, it is obligatory to carry the available voting through an electronic system.

1.2 Requirement of E-Voting:

The requirement in traditional voting process is also applicable for e-voting and some of them are mentioned below [12].

1. *Fairness:* No person can learn the voting outcomes before the tally.
2. *Eligibility:* Only eligible voters are allowed to cast their vote.
3. *Uniqueness:* No voter is allowed to cast their vote more than once.
4. *Privacy:* No person can access the information about the voters vote.
5. *Accuracy:* All the valid votes should be counted correctly.
6. *Efficiency:* The counting of votes can be performed within a minimum amount of time [2].

1.3 Biometric Authentication:

Fingerprint matching is one of the most popular and reliable biometric techniques used in automatic personal identification. There are two main stages during the use of fingerprints authentication: fingerprint verification and fingerprint identification. While the goal of fingerprint verification is to verify the identity of a person, the goal of fingerprint

identification is to establish the identity of a person [13].

In a traditional biometric recognition system, the biometric template is usually stored on a central server during enrolment. The candidate biometric template captured by the biometric device is sent to the server where the processing and matching steps are performed [6].

The objective of voting is to allow voters to exercise their right to express their choices regarding specific issues, pieces of legislation, citizen initiatives, constitutional amendments, recalls and/or to choose their government and political representatives [11]. Technology is being used more and more as a tool to assist voters to cast their votes. To allow the exercise of this right, almost all voting systems around the world include the following steps:

- Voter identification and authentication
- Voting and recording of votes cast
- Vote counting
- Publication of election results

Voter identification is required during two phases of the electoral process: first for voter registration in order to establish the right to vote and afterwards, at voting time, to allow a citizen to exercise their right to vote by verifying if the person satisfies all the requirements needed to vote (authentication) [14].

The field of biometrics was formed and has since expanded on to many types of physical identification. Still, the human fingerprint remains a very common identifier and the biometric method of choice among law enforcement [8]. These concepts of human identification have led to the development of fingerprint scanners that serve to quickly identify individuals and assign access privileges. Fingerprinting recognition, the electronic methods of recording and recognizing an individual finger print, advanced substantially during the last decade of the 21st century [15]. Today, identification can be achieved in a few seconds with reasonable accuracy. As a result, the use of automated fingerprint identification systems (AFIS) that record, store, search, match and identify finger prints is rapidly expanding. AFIS can be integrated with a microcontroller and other peripherals to form an embedded system which is a comprehensive electronic voting machine with fingerprint print identification system.

2. Existing E-Voting System

The category “electronic voting” is potentially broad, referring to several distinct possible stages of electronic usage during the course of an election.

A. Electronic voting: Electronic voting refers to any system where a voter casts his or her ballot using an electronic system, rather than a paper. Once recorded, an electronic vote is stored digitally and

transferred from each electronic voting machine to a counting system [16].

B. Electronic vote counting: Electronic vote counting refers to the system that is used to tabulate ballots and award seats. It would be possible to vote using a non-electronic medium and then convert these votes to an electronic system and award seats through an electronic vote counting system [3].

Electronic Voting Machine is a simple electronic device used to record votes in place of ballot papers and boxes which were used earlier in conventional voting system [4]. It is a simple machine that can be operated easily by both the polling personnel and the voters. Being a standalone machine without any network connectivity, nobody can interfere with its programming and manipulate the result. Keeping the erratic power supply position in many places in the country, the machines have been made to run on batteries. It has mainly two units: Control unit and Ballot unit. The Control Unit is the main unit which stores all data and controls the functioning of EVM. The program which controls the functioning of the control unit is burnt into a micro chip on a “one time programmable basis”. Once burnt it cannot be read, copied out or altered. The EVMs use dynamic coding to enhance security of data transmitted from ballot unit to control unit. The new EVMs have also got real time clock and date-time stamping facility which enables them to record the exact time and date whenever a key is pressed. After the voting is completed and the close button is pressed, the machine does not accept any data or record any vote. Through the press of “total” button, the control unit can display the number of votes recorded till that time which can be cross checked with the register of voters. The display system of the control unit shows the total number of votes polled in a polling station and the candidate-wise votes polled in the machine when the ‘result’ button is pressed by the counting staff in the presence of counting agents at the counting centre. The control unit can also detect any physical tampering made with the connecting cable and indicate the same in the display unit [16].

The security of an EVM can be significantly improved by using biometric authentication system. The main reasons for augmenting a biometric authentication with electronic voting system is that biometrics are traits of a person which can be hardly copied or shared thereby it becomes very difficult to forge the identity of a person [5]. The major biometric-based technologies include finger-scanning, hand geometry, facial recognition, iris scanning, retinal scanning, finger geometry, voice recognition and dynamic signature verification.

In [6], a scheme for a dynamic voter registration, enrolment and voting in an online biometric electronic voting system is proposed. An indexing technique for facilitating the search of a matching identity to an input fingerprint is incorporated. In [7],

a web based secure e-voting system with fingerprint authentication is implemented. A public voting system based on biometric fingerprint method to make the election process transparent and efficient is implemented [8]. In [9], the challenges existing in the conventional electoral system of India are analyzed with the aim of addressing fraudulent electoral prices by the use of biometric authentication based Electronic Voting System.

This research work deals with the design and development of fingerprint recognition based Electronic Voting System. The proposed system also takes into account the essential voting requirements in terms of privacy, uniqueness, completeness, efficiency and fairness.

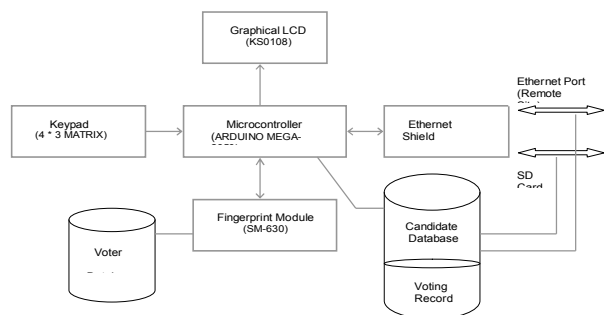


Figure 1: Block Diagram of Proposed Secure E-Voting System

3. System Framework

The system framework for secure voting comprises of a key pad, graphical LCD, microcontroller, finger print module and system interface as shown in Fig 1.

Arduino Mega 2560 is a microcontroller board having a number of facilities for communicating with a computer or other devices. It can be programmed for serial communication on any of the Mega 2560's digital pins. **SM630** fingerprint verification module consists of optical fingerprint sensor, high performance Digital Signal Processor and Flash memory. It boasts of functions such as fingerprint login, deletion, verification, upload and fingerprint download etc. The voter information is stored in fingerprint module, whereas, the candidate database along with voting record is kept in microcontroller flash memory and remote site through Ethernet port. At the lowest level, keyboards are organized in a matrix of rows and columns. The microcontroller accesses both rows and column through ports; therefore, with a port of microcontroller, a 4 x 3 matrix of keys can be connected. The graphical LCD has a display format of 128x64 dots and yellow-green color backlight. It makes the use of **KS0108** controller to execute its internal operations.

4. Working of Proposed Secure Electronic Voting System

The main phases of a voting system are registration, authentication, accessibility, casting and counting. The implementation of all these phases in the proposed system is elaborated in following steps:

//variables used are:

candidate [100] – candidate database
 candidate_tot = 0 – total number of registered candidates
 voter_tot = 0 – total number of registered voters
 voted = 0 – total number of votes
 SECURITY_PIN – stores security pin //

- Step 1:** Display Welcome Screen
- Step 2:** Security Check. If password is correct go to step 3 else repeat 2
- Step 3:** Detect memory card. If memory card found go to step 4 else display No Memory card Detected
- Step 4:** Display main menu options
- Step 5:** Candidate Zone
- Step 6:** Voter Zone
- Step 7:** Vote Now
- Step 8:** Result
- Step 9:** Change Pin
- Step 10:** Exit
- Step 11:** If Keypad input is 1 than go to step 12
- Step 12:** If Keypad input is 2 than go to step 41
- Step 13:** If Keypad input is 3 than go to step 59
- Step 14:** If Keypad input is 4 than go to step 73
- Step 15:** If Keypad input is 5 than go to step 80
- Step 16:** If Keypad input is 6 than go to step 85
- Step 17:** If Keypad input is greater than 6 then print Invalid option. Please try again. Go to step 4
- Step 18:** Display candidate zone options
- Step 19:** New Registration
- Step 20:** Modify Candidate
- Step 21:** Empty Database
- Step 22:** Back to Main Menu
- Step 23:** Exit
- Step 24:** If Keypad input is 1 than go to step 19
- Step 25:** If Keypad input is 2 than go to step 25
- Step 26:** If Keypad input is 3 than go to step 37
- Step 27:** If Keypad input is 4 than go to step 4
- Step 28:** If Keypad input is 5 than go to step 85
- Step 29:** If Keypad input is greater than 5 then display Invalid option. Please try again. Go to step 12
- Step 30:** Enter Candidate code using keypad

- Step 31:** If candidate code exists in candidate database display Already registered. Go to step 12
- Step 32:** candidate_tot++
- Step 33:** candidate[candidate_tot] = New candidate's code and display Candidate has been registered.
- Step 34:** Store candidate information in memory card, EEPROM and remote system. Go to step 12
- Step 35:** If list is candidate database is empty then display List is empty. Go to step 12
- Step 36:** Display option for modification.
- Step 37:** Change Code
- Step 38:** Delete Candidate
- Step 39:** If Keypad input is 1 than go to step 29
- Step 40:** If Keypad input is 2 than go to step 34
- Step 41:** If Keypad input is greater than 2 then print Invalid option. Please try again. Go to step 25
- Step 42:** Enter candidate code for modification using keypad
- Step 43:** Enter new code for candidate
- Step 44:** If code already exists in candidate database then display Already registered. Go to step 12
- Step 45:** Replace old code with new one and display Information has been updated.
- Step 46:** Updated information in memory card, EEPROM and remote system. Go to step 12
- Step 47:** Delete candidate code form database and display Candidate removed.
- Step 48:** Remove candidate information from memory card, EEPROM and remote system.
- Step 49:** candidate_tot--. Go to step 12
- Step 50:** Security Check. If password is correct go to step 38 else repeat 37
- Step 51:** candidate_tot = 0
- Step 52:** Delete complete database and display Database Clear.
- Step 53:** Update candidate information in memory card, EEPROM and remote system. Go to step 12
- Step 54:** Display voter zone options
- Step 55:** Register voter
- Step 56:** Delete Voter
- Step 57:** Empty Database
- Step 58:** Back to Main Menu
- Step 59:** Exit
- Step 60:** If Keypad input is 1 than go to step 48
- Step 61:** If Keypad input is 2 than go to step 52
- Step 62:** If Keypad input is 3 than go to step 56
- Step 63:** If Keypad input is 4 than go to step 4
- Step 64:** If Keypad input is 5 than go to step 85
- Step 65:** If Keypad input is greater than 5 then print Invalid option. Please try again. Go to step 41
- Step 66:** Input voter thumb print using fingerprint module
- Step 67:** If fingerprint matches in database then print Already registered. Go to step 41
- Step 68:** Add fingerprint in voter database.
- Step 69:** voter_tot++. Go to step 41
- Step 70:** Input voter thumb print using fingerprint module
- Step 71:** If fingerprint doesn't matches in database then print Doesn't exist. Go to step 41
- Step 72:** Delete fingerprint from voter database.
- Step 73:** voter_tot--. Go to step 41
- Step 74:** Security Check. If password is correct go to step 57 else repeat 56
- Step 75:** Delete complete database and display Database Clear.
- Step 76:** voter_tot= 0. Go to step 41
- Step 77:** Display vote now options
- Step 78:** Vote
- Step 79:** Back to Main Menu
- Step 80:** If Keypad input is 1 than go to step 63
- Step 81:** If Keypad input is 2 than go to step 71
- Step 82:** If Keypad input is greater than 2 then display Invalid option. Please try again. Go to step 59
- Step 83:** Input voter thumb print using fingerprint module
- Step 84:** If fingerprint doesn't matches in database then display You are Not Eligible. Go to step 59
- Step 85:** Enter candidate code using keypad
- Step 86:** If candidate code doesn't matches in database display invalid candidate code. Go to step 59
- Step 87:** Votes++. Increase the vote of selected candidate by 1.
- Step 88:** Update voted database in memory card, EEPROM and remote system.
- Step 89:** Delete fingerprint from voter's database.
- Step 90:** Display successfully voted. Go to step 59
- Step 91:** Security Check. If password is correct go to step 4 else repeat 71
- Step 92:** Display Winner of election's code with votes
- Step 93:** Display Result section's options

- Step 94:** Full list
- Step 95:** Main menu
- Step 96:** Exit
- Step 97:** If Keypad input is 1 than go to step 78
- Step 98:** If Keypad input is 2 than go to step 4
- Step 99:** If Keypad input is 3 than go to step 85
- Step 100:** If Keypad input is greater than 3 then print Invalid option. Please try again.
- Step 101:** Display whole list of candidates and their respective votes. Go to step 12
- Step 102:** Security Check. If password is correct go to step 80 else repeat 79
- Step 103:** Enter new security pin.
- Step 104:** Confirm security pin.
- Step 105:** If value of step 80 and step 81 matches then go to step 84
- Step 106:** If value of step 80 and step 81 matches print Pin not matched. Go to step 12
- Step 107:** SECURITY_PIN = new security pin. Go to step 12
- Step 108:** Display Thank you for using EVM. Exit.

The interaction among the various entities in the proposed framework is shown in Figure 2.

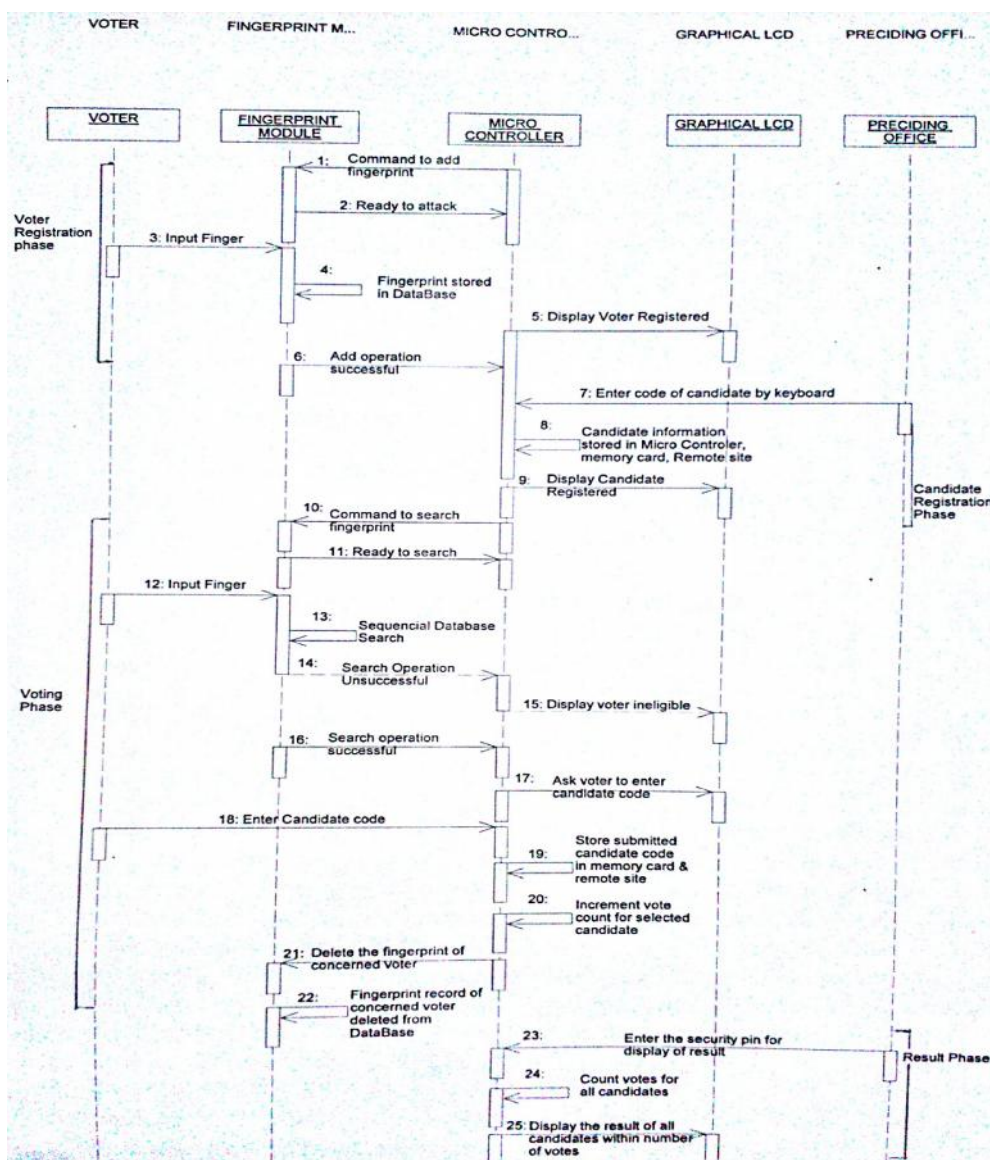


Figure 2: Sequence Diagram showing the Control flow in Proposed System

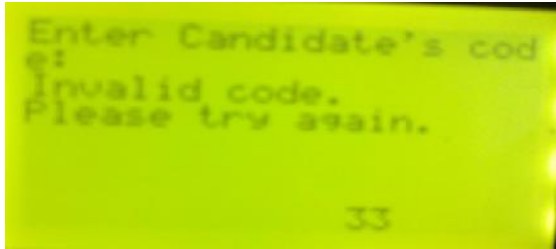


Figure 10: Validation of Candidate's Selection



Figure 11: Information of Winning Candidate

Code	Votes
10	10
21	71
29	29

Figure 12: Voting Result Summary

Some of the salient features of the proposed system are as follow:

- Voters during voting cannot perform the tempering as he is not authorized to scroll any other screen.
- A voter cannot do bogus voting as his fingerprint must match the previously stored data. After voting, the voter's fingerprint record is deleted from the database thereby not allowing him to cast voting more than once.
- The candidate information and voting records are stored at three different places: SD card, EEPROM of microcontroller and remote site through Ethernet port, thereby, improving the availability and reliability of system.
- The replication of voting information at multiple locations reduces the risk of biasing during vote counting.
- The existing Electronic Voting Machine comprises of two separate components: Ballot unit and Control unit; however in the developed system all functionalities are embedded in one module making it as compact and concise.
- The design of currently used voting machines depends on the number of candidates within a

constituency; however this is not a constraint for proposed system and the same model of machine can be used anywhere during voting.

6. Conclusion

Electronic voting system is emerging as significant alternative to the conventional systems in the delivery of reliable and trusted elections. In this paper, a framework for electronic voting system based on fingerprint biometric is proposed and implemented with the objective of eliminating bogus voting and vote repetition, less election expenditure, more transparency and fast results.

REFERENCES

- [1]. M. Byrne, K. Greene, and S. Everett, "Usability of Voting Systems: Baseline Data for Paper, Punch Cards, and Lever Machines," ACM International Conference on Human Factors in Computing Systems, pp. 171-180, 2007.
- [2]. Sussane Caarls, "E-voting Handbook: Key Steps in the Implementation of E-enabled Elections", Council of Europe, 2010.
- [3]. Santin, R. Costa and C. Maziero, "A Three-Ballot-Based Secure Electronic Voting System", IEEE Transaction on Security & Privacy, Vol. 6(3), pp. 14-21, 2008.
- [4]. Kumar, "Electronic voting machine- A review", IEEE International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME), pp. 44-48, 2012.
- [5]. Jain and D. Maltoni, "Handbook of Fingerprint Recognition," Springer-Verlag, New York, USA, 2003.
- [6]. O. Iloanusi and C. Osuagwu, "Framework for a dynamic Fingerprint Indexing Biometric-based Voting System", African Journal of Computing & ICTs, pp. 55-63, Vol. 5(4), 2012.
- [7]. Altun and M. Bilgin, "Web baesd secure e-voting system with fingerprint authentication", Scientific Research and Essays, pp. 2494-2500, Vol. 6(12), 2011.
- [8]. K. Memon, D. Kumar and S. Usman, "Next Generation A secure E-Voting System Based On Biometric Fingerprint Method", International Conference on Information and Intelligent Computing (IPCSIT), pp. 26-32, 2011.
- [9]. S. Yadav and A. Singh, "A Biometric Traits based Authentication system for Indian Voting System", International Journal of Computer Applications, pp. 28-32, Vol. 65(15), 2013.
- [10]. S. Kumar and M. Singh, "Security Enhancement of E-Voting System", Global Journal of Computer Science and Technology, Volume 12 Issue 5 Version 1.0 March 2012.
- [11]. R. Udupa, G. Garg and P. Sharma, "Fast and accurate fingerprint verification", International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 192-197, 2001.

- [12]. R. Haenni, E. Dubuis, and U. Ultes-Nitsche, "Research on e-voting technologies." Bern University of Applied Sciences, Technical Report 5, 2008.
- [13]. M. Min and Y. Thein, "Intelligent Fingerprint Recognition System by using geometry approach", International Conference on the Current Trends in Information Technology (CTIT), pp. 1-5, 2009.
- [14]. M. Khan, "Fingerprint biometric-based self-authentication and deniable authentication schemes for the electronic world." IETE Technical Review, vol. 26(3), pp. 191, 2009.
- [15]. L. O'Gorman, "An overview of fingerprint verification technologies." Information Security Technical Report 3.1, pp. 21-32, 1998.
- [16]. S. Kumar and E. Walia, "Analysis of Electronic Voting system in Various Countries", International Journal on Computer Science and Engineering, vol. 3(5), 1825-1830, 2011.

The Effect of Changing the Speed and the Number of Nodes on Packet Delivery Ratio in MANET

Imad I. Saada, Majdi Z. Rashad, Mohamed A. Abu ElSoud

Department of Computer Science, Faculty of Computers and Information
Mansoura University, Egypt

Abstract

Mobile ad hoc network MANET is wireless network without any infrastructure nor centralized control, MANET may include many nodes that work as a router, number of nodes in MANET may change and each node in the network may change its speed frequently, so the routing in this kind of network is more complex than the routing in the wired network, routing has been discussed in many researches in order to find the routing protocol that provides a high performance. Packet delivery ratio PDR is an important metric in performance of MANET. The concern in this research is to study the effect of changing the number and the speed of nodes on PDR when implementing AODV and DSDV routing protocols, the research will deal with several number of the nodes, in addition it will deal with several values of the speed of the nodes, this study will lead to conclude which routing protocol may be more suitable for each case in terms of PDR.

Keywords: MANET, Routing protocols, DSDV, AODV, packet delivery ratio, PDR.

1. Introduction

Mobile ad hoc network MANET is a set of mobile nodes connected via wireless media without infrastructure, MANET is important network in several environments such as Military environments, civilian environments, emergency fields and Personal area networks, this kind of network has not fixed topology because the nodes can move randomly and rapidly, it is clear that MANET has special characteristics, these characteristics make MANET encounters several challenges such as security threats and routing process, so it must be looking for a routing protocol which can deal with the challenges to get a MANET with high performance.

- At the beginning in this research the routing protocols AODV and DSDV will be discussed.
- The simulation will be performed to clear how PDR may be affected when changing the number and the speed of nodes in case of implementing the mentioned routing protocols.
- The results showed in the graphs of simulation will be discussed.

- The simulation results will be discussed to show the effect of changing the number and the speed of nodes on PDR and to conclude which routing protocol can grant the best value for PDR.

1.1 Packet delivery ratio (PDR):

PDR is a metric indicates the reliability of data packets delivery, this metric can be calculated by getting data packets delivered to destination divided by the number of packets sent.

1.2 Routing Protocols

MANET routing protocols are classified into two types according to the ability of providing a track of routes for all destination:

- Proactive or table- driven routing protocols.
- Reactive or on-demand routing protocols.

Destination-Sequence Distance Vector (DSDV) Protocol: In DSDV the proactive scheme is used, it maintains up-to-date routing information from each node to every other node in the MANET, tables are used to store routing information.

DSDV is based on the proactive routing mechanism Bellman Ford, the nodes in MANET record information of the destinations in routing tables with using sequence number to prevent routing loop, routing table is stored in the nodes, it is used for transmit packets between nodes.

The routing information in the routing tables is periodically updated by nodes in MANET, and this information must be transmitted for each neighboring node, the updating of this information is necessary to detect the dynamically changing topology. This scheme enables each node to continuously have the recent information of each node with low route discovery delay but with high overhead, the routing table that is maintained by each node must include:

- Next hop towards each destination.
- A cost metric for the path to each destination.
- A destination sequence number that is created by the destination itself.

- Sequence numbers used to avoid formation of loops.

Ad Hoc On Demand Distance Vector (AODV) routing Protocol:

AODV is creating a path based on demand, so it is considered as an improvement of DSDV, the source node uses RREQ (route request), when it wants to send a packet to the destination, it sends RREQ to its neighbors, and these neighbors send the RREQ to their neighbors and so on, if any node has a route to the destination, it will send RREP to an intermediate or to destination node, after the source receives RREP, and it will send the packet to the destination using the path that was established when RREQ was sent. The source or the intermediate node selects the fresher route to the destination based on the destination sequence number.

This scheme enables each node to discover the path to destination on demand with low overhead but with high route discovery delay.

The nodes in MANET use hello message to indicate a link failure by exchanging hello messages between neighbors periodically, Alternatively the failure to receive several MAC-level acknowledgement may indicate a link failure.

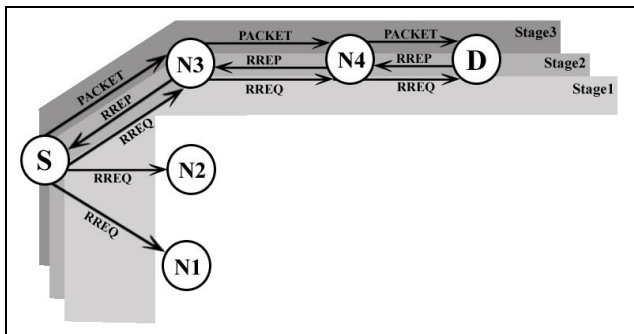


Fig.1 Ad-hoc On-demand Distance Vector routing algorithm

2. Simulation

2.1 Simulation environments:

The seed of simulation equaled 1, terrain dimension 1000x1000 m, selection simulation time was 30 minutes, and the Position of nodes was read from NODE-PLACEMENT-FILE, mobility random-way point with pause time 10s was selected, radio bandwidth was 2000000 and MAC protocol was 802.11.

Simulation one: The metric used in this part was packet delivery ratio with changing the values of number of nodes with implementing AODV and DSDV routing protocols, with minimum speed of 60 m/s to maximum speed of 100 m/s, number of nodes in the area were 20, 50, 70 and 130 nodes, so simulation was done by four scenarios for each routing protocol.

Simulation two: The metric used in this part was packet delivery ratio with changing the values of number of nodes with implementing AODV and DSDV routing protocols, number of nodes in the area were 20 nodes, with average speed 20, 50, 100, 200 m/s, so simulation was done by four scenarios for each routing protocol.

The following table gives the simulation parameters used during the simulation.

Parameter	Value
Simulator	GlomoSim
Routing protocol	AODV, DSDV
The seed	1
Terrain dimension	1000x1000 m
Simulation time	30m
Mobility	NODE-PLACEMENT-FILE
Pause time	10s
Radio bandwidth	2000000
MAC protocol	802.11
Simulation one	
Minimum speed	60 m/s
Maximum speed	100 m/s
Number of nodes	20, 50, 70 and 130 nodes
Simulation two	
Number of nodes	20 nodes
Average speed	20, 50, 100 and 200 m/s

2.2 Simulation result

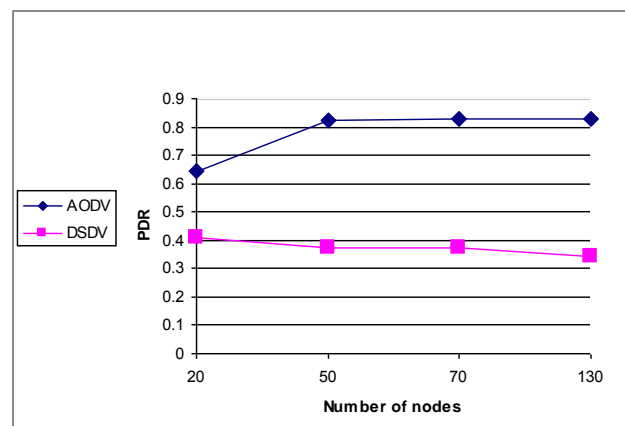


Fig.2 PDR vs. Number of nodes

From Fig.2 of simulation one it is clear that:

- In case of AODV by increasing the number of nodes from 20 nodes to 130 nodes we observe a simple increment of PDR but without a certain behavior.

- In case of DSDV by increasing the number of nodes from 20 to 130, PDR is decreasing continuously.
- When increasing the number of nodes to 130 nodes we observe a considerable difference of PDR for AODV and for DSDV, and we observe that the value of PDR in case of AODV is greater than the value of PDR in case of DSDV regardless of the number of nodes.

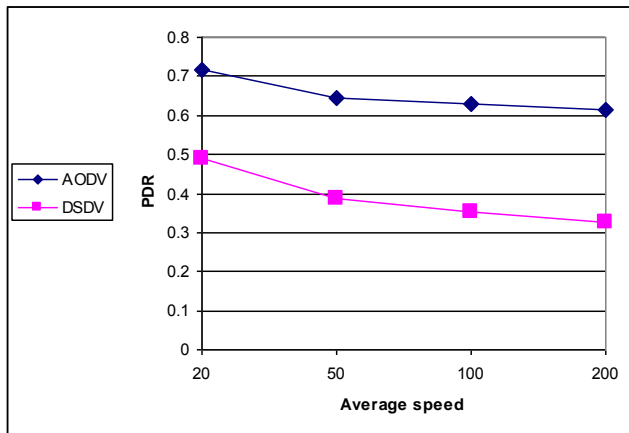


Fig.3 PDR vs Average speed

From Fig.3 of simulation two it is clear that:

- In case of AODV by increasing the average speed of nodes from 20m/s to 200m/s, we observe that the value of PDR is continuously decrementing.
- In case of DSDV by increasing the average speed of nodes from 20m/s to 200m/s we observe that the value of PDR is continuously decrementing.
- When increasing the average speed of nodes to 200m/s we observe a considerable difference of PDR for AODV and for DSDV, and we observe that the value of PDR in case of AODV is greater than the value of PDR in case of DSDV regardless of the average speed.

3. Conclusion:

The simulation showed that the value of PDR decreased by increasing number of nodes when DSDV was implemented, and the value of PDR decreased by increasing the average speed of nodes when AODV and DSDV were implemented.

On the other hand, it can be concluded that the value of PDR when implementing AODV is always higher than the value of PDR when implementing DSDV regardless of the number or the average speed of nodes.

After studying the effect of changing the number and the speed of nodes on PDR, and after comparing PDR in both cases when implementing AODV and DSDV, this research

showed that it is necessary to continue the process of developing a routing protocol that may provide higher PDR, because there was a considerable lost of packets when using AODV and DSDV in MANET.

References

- [1] Panagiotis Papadimitratos, Zygmunt J. Haas, "Secure Routing for Mobile Ad hoc Networks", the SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS), 2002.
- [2] Asis Nasipuri and Samir R. Das, "On-Demand Multipath Routing for Mobile Ad Hoc Networks", ACM volume 6, 2001.
- [3] Sushil Kumar, Dinesh Singh & Mridul Chawla "Performance Comparison of Routing Protocols in MANET Varying Network Size" International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) Volume-1, 2011.
- [4] J. Costa-Requena, T. Vardar, M. Ayyash and R. Kantola, "Scalable Hybrid Ad Hoc Routing Approach", IJCSNS volume 6 no. 8, 2006.
- [5] Manal Abdullah and Helen bakhsh "Agent-based dynamic routing system for MANETs", ICGST-CNIR jornal, volume9, 2009.
- [6] V. Ramasubramanian, Z. J. Haas, E. G'un Siner, "SHARP: A Hybrid Adaptive Routing Protocol for Mobile Ad Hoc Networks", ACM new york, 2003.
- [7] Jing Xie, Luis Girones Quesada and Yuming Jiang, "A Threshold-based Hybrid Routing Protocol for MANET", IEEE ISWCS, 2007.
- [8] H. L. seba "ARPM : an adaptive routing protocol for MANETs", ACS/IEEE International Conference on Pervasive Services, 2006.
- [9] A. H. abd rahman and z. ahmad zukarnain, "Performance Comparison of AODV, DSDV and I-DSDV Routing Protocols in Mobile Ad Hoc Networks", European Journal of Scientific Research Vol.31 No.4, 2009.
- [10] F. aslam and Chi aching ooi, "Mobile Ad Hoc Networks", www.docstoc.com, university of freiburg, 2007.

Imad I. Saada is a PHD student in computer science department in Mansoura University and a member of the academic staff at IT. department in AL-Quds Open University. His subject is in the distributed systems.

Magdy Z. Rashad is an assistant professor and chairman of computer science department in Mansoura University. He is the decision support systems unit coordinator at faculty of computers & information in Mansoura University. He has supervised over 10 PhDs and 21 masters mostly specialized in artificial intelligence and its applications related to real life. As a result of his work he has published over 84 papers. current project is grid computing.

Mohamed A. Abu Eisoud is an assistant professor in computer science department in Mansoura University. He has supervised over 7 PhDs and 13 masters mostly specialized in artificial intelligence and its applications related to real life. As a result of his work he has published over 32 papers.

The development of algorithms for alleviating the problem of discontinuity in speech synthesis from text written in Albanian

Adnan Maxhuni¹, Agni Dika¹, Avni Rexhepi¹ and Dren Imeraj¹

¹Faculty of Electrical and Computer Engineering, University of Prishtina
Prishtina, 10000, Kosovo

Abstract

Efforts to generate speech from written text are different, ranging from mechanical ones of centuries ago, to software solutions of recent decades, known as TTS (Text-to-Speech). The ultimate goal of a speech synthesizer from written text is 'ability' to read any text [2]. Reading should be understandable and natural. A speech synthesizer contains "the part" for the linguistic analysis (NLP - Natural Language Processing) and the digital signal processing (DSP - Digital Signal Processing)[4].

Albanian language as a separate language in Indo-European family of languages, with its own specifics in writing and reading differs from other languages and models for these languages cannot be used for conversion of Albanian texts. We acquired the Concatenation Synthesis model of building the conversion system of text to speech. This technique is based on preliminary recording of the text, from which acoustic segments are cut and stored in a database. Then, we create application which concatenates these acoustic segments, to "create" words and sentences [4].

During concatenation of sequences, due to differences in their amplitudes emerge problem of disconnection. In this paper we present some algorithms for minimizing the effect of discontinuity along concatenation of acoustic segments of the synthesized speech [9].

Keywords: *Sequence, Acoustic Segments, Algorithm, Amplitude.*

1. Introduction

Simple assembly of two acoustic sequences in order to make words causes disconnection/discontinuity that make the generated speech not very clear, especially unnatural. Voice generated in this way is known as "metallic" and intermitted. For these reasons we have developed a set of algorithms in order to minimize these problems. There is no single algorithm which would solve all these problems. Here we address the problem of amplitude differences, caused during reunion of sequences stored in database [10].

The first case is when there is a break after the end of the sequence (sequence at the end of the word). In these cases,

the sequence should be reduced gradually. Change in amplitude should be linear and there should be a certain period of transition [7].

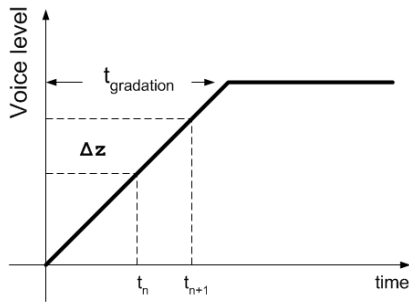
The other case is the connection of two sequences (in the middle of the word). In most cases, the amplitudes of the sequences do not match. This problem is mitigated in two ways: through the amplitude equalization or through gradual amplitude reduction to a certain level [6].

2. The effect of discontinuity during union of acoustic segments

Algorithm for mitigating the effect of discontinuity during union with gradually changing the amplitude of the phoneme, can be realized with multiplying the sequence samples with a constant. In the case of increasing the amplitude from zero to its maximum value, or the reduction from its full value to zero, this "constant value" takes values from 0 to 1, respectively, from 1 to 0. Since the period of change and the starting/ending point are taken as parameters, the expression for the amplitude change should be depending on them [6].

2.1 Algorithm for gradual increase and decrease of the amplitude

Because it is required that this change has to be linear (Fig.1), we must first set a constant Δz (linearity constant) which shows the amplitude change in the unit of time (between samples) [1].



$$\Delta z = \frac{1}{t_{gradation}}$$

Fig.1 Gradual amplitude increase

Pseudo code for amplitude rise/increase from zero to its full value (max. value):

```

dZ = 1 / t_gradation
gradation = 0
t = t_start
WHILE (t < t_start + t_gradation)
    samples[t]=samples[t] * gradation
    gradation = gradation + dZ
    t = t + 1
REPEAT
    
```

After execution of this algorithm, the values of the samples[] vector will be modulated so as to gradually reach maximum amplitude value. Values t_start and t_gradation are input values of the algorithm. Similarly is done gradual amplitude decrease.

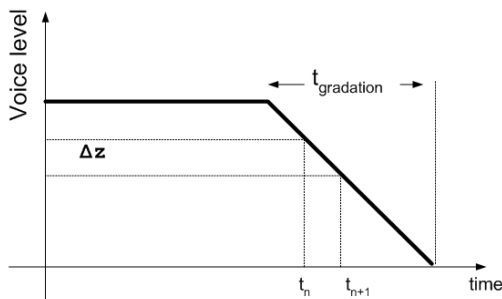


Fig.2 Gradual amplitude decrease

Pseudo code of the algorithm for amplitude decrease to the zero value (Fig. 2), is as follows:

```

dZ = 1 / t_gradation
gradation = 1
WHILE (t < t_start + t_gradation)
    samples[t]=samples[t]*gradation
    gradation = gradation - dZ
    t = t + 1
REPEAT
    
```

Abovementioned algorithms can be applied in two ways: The first way is to apply the algorithm after sequence construction. In this case, the sequence is constructed by copying samples from the database of sequences and then algorithm for amplitude changing is applied. This way is convenient when change of the amplitude must be applied in the middle of the sequence copied from the base of sequences [9].

The second way is the application of change during sequence copying. This method reduces the execution time of the algorithm, as for to apply change, it uses the sequence copying cycle. Depending on the use case, one or the other method may prove more effective [9].

To test the algorithm we take the sequence for letter 'e' in the word 'teksti'. In this word, after the letter 'e' should be a pause and then continues the rest. Because the sequences of the letter 'e' does not end gradually, then the algorithm for gradual reduction of amplitude must be applied [10].

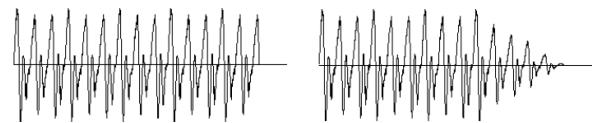


Fig.3 Sequence before and after the application of the algorithm for amplitude decrease/reduction

From Fig. 3 it can be seen that all points of the sequence (not only the amplitudes) are gradually reduced. Indeed gradual extinction of the sequence is the goal of this algorithm. Also, the starting point of the amplitude change can be noticed. The starting point of changing the amplitude is given as a parameter. Here, the amplitude decrease started in 2000 samples before the end of the sequence [3].

In algorithms for copying the sequence and overlapping, is used vector samples[]. This vector contains samples for the sequence. Since the algorithms defined in this section apply PCM, each sample shows the amplitude of sequence in a particular moment. In the PCM technique, samples are equally distant from each other, because there is a frequency of sampling [5].

Each sample is a numeric value that represents the current value of the amplitude. If in a sequence, is used PCM technique with 16 bits per sample, the amplitude values can be in the range [-32 768; 32 767]. By changing values (magnitudes) of the samples, is done the change in the form of sequence.

2.2 Algorithms for equalization of amplitudes

During the union (link) of the sequences repeatedly (e.g. transition from 'e' to 'ek'), usually appears the problem of differences in amplitudes.

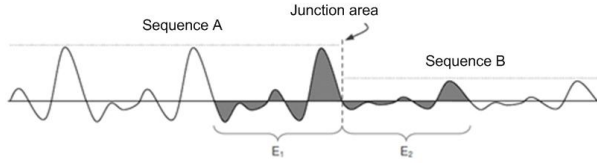


Fig.4 Problem of difference in amplitudes

Since recordings are made at different times, it is natural that the amplitudes differ from sequence to sequence, although their wave shape is almost identical (Fig. 4). Rapid changes in amplitude create harmful effects on comprehensibility and naturality of speech.

To eliminate this difference of amplitudes, we tried to equalize the energy of periods of the sequence B (E2) with energy of periods of the sequence A (E1).

As the wave forms of the two sequences are very similar, but differ in amplitude, then the sequence A can be describes with $a \cdot f(x)$, while the sequence B with function $b \cdot f(x)$. In this way, equation for energy of a discrete signal for sequences A and B will be [1]:

$$E_1 = \frac{1}{N} \sum_{n=n_0}^{n_0+N} |af[n]|^2$$

$$E_2 = \frac{1}{N} \sum_{n=n_0}^{n_0+N} |bf[n]|^2 \dots\dots\dots(1)$$

Equations in (1) can be further extended:

$$E_1 = \frac{1}{N} \sum_{n=n_0}^{n_0+N} |af[n]|^2 = \frac{1}{N} \sum_{n=n_0}^{n_0+N} |a|^2 |f[n]|^2 =$$

$$= \frac{1}{N} \sum_{n=n_0}^{n_0+N} a^2 |f[n]|^2 = a^2 \frac{1}{N} \sum_{n=n_0}^{n_0+N} |f[n]|^2$$

$$E_2 = \frac{1}{N} \sum_{n=n_0}^{n_0+N} |bf[n]|^2 = \frac{1}{N} \sum_{n=n_0}^{n_0+N} |b|^2 |f[n]|^2 =$$

$$= \frac{1}{N} \sum_{n=n_0}^{n_0+N} b^2 |f[n]|^2 = b^2 \frac{1}{N} \sum_{n=n_0}^{n_0+N} |f[n]|^2 \dots\dots\dots(2)$$

If from the last equation (2), we find the ratio of the energies of sequences A and B, we'll have:

$$\frac{E_1}{E_2} = \frac{a^2 \frac{1}{N} \sum_{n=n_0}^{n_0+N} |f[n]|^2}{b^2 \frac{1}{N} \sum_{n=n_0}^{n_0+N} |f[n]|^2} = \frac{a^2}{b^2}$$

$$\frac{a}{b} = \sqrt{\frac{E_1}{E_2}} \dots\dots\dots(3)$$

If we multiply each sample of the sequence B with $\sqrt{E_1/E_2}$, then sequence B ($b \cdot f(x)$) will become:

$$\sqrt{\frac{E_1}{E_2}} b \cdot f(x), \text{ that from (3) can be written as:}$$

$$\sqrt{\frac{E_1}{E_2}} b \cdot f(x) = \frac{a}{b} b \cdot f(x) = a \cdot f(x) \dots\dots\dots(4)$$

So, if each sample of sequence B is multiplied with $\sqrt{E_1/E_2}$, sequence B will have amplitudes equal with sequence A.

To implement the amplitude equalization algorithm it is necessary to know some parameters related sequences. The most important parameter is the length of the period, which must be precise in order to make accurate normalization [5].

According to the algorithm, initially must be calculated energies of the two periods to be joined. Then, each sequence of the second sample should be multiplied by the square root of the quotient of energies. Input parameters of the algorithm for normalization are:

- Point_Junction** - the point where the two sequences join together
 - Length_Period1** - the length of the last period of the first sequence
 - Length_Period2** - the length of the first period of the second sequence
 - Length_Sequence2** - total length of the second sequence
- Below is presented the algorithm that does the equalization of amplitudes after join of the sequences:

```

E1 = 0
t = Point_Junction
WHILE t > PointStart - Length_Period1
    E1 = E1 + samples[t]^2
    t = t - 1
REPEAT
    E1 = E1/Length_Period1
    E2 = 0
    t = Point_Junction
    
```

```

WHILE t < PointStart + Length_Period2
    E2 = E2 + samples[t]^2
    t = t + 1
REPEAT
    E2 = E2/Length_Period2
    FactorNormalization = sqrt(E1/E2)
    t = Point_Junction
WHILE t < GjatesiaSekuenca2
    samples[t]=samples[t]*FactorNormaliza
    tion
    t = t + 1
REPEAT
    
```

After applying the above algorithm, the amplitudes of the second sequence will be equal to those of the first sequence. In this way are eliminated the effects of the immediate backdrop of amplitude at the point of junction. Applying the above algorithm has the effect of reducing the amplitude of the first sequence from the original level to the level of the amplitude of the second sequence (Fig. 5).

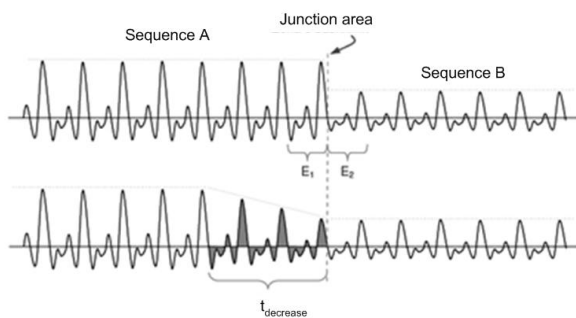


Fig.5 Application of the algorithm for decreasing the amplitude at the junction area

2.3 Algorithm for gradual decrease of the amplitude

Equalization of amplitudes through algorithm for gradual decrease of amplitude, in some cases it is not possible, because in the sequential sequence after the join, cannot be changed all values of the amplitudes. In such cases is applied the algorithm which gradually decreases the amplitudes of the first sequence, so that at the point of junction, the amplitude is approximately as the amplitude of the second sequence. The decrease must be realized during the time tdecrease.

If we analyze the requirement of the latest algorithm, we'll see that there are similarities with the algorithm for the gradual increase of amplitude. The difference between these two algorithms is the level of the last amplitude. While in the algorithm for gradual increase of the amplitude the last level is zero, in the algorithm for equalization of amplitudes after joining of sequences, as

can be seen in the last figure, should be as the level of the first amplitude of the following sequence.

Requirement is that the amplitude drops linearly from A0 (amplitude of sequence A before application of the algorithm) to A1 (amplitude at the beginning of the sequence B) within the time tdecrease. In this case Δz will have the value:

$$\Delta z = \frac{1}{t_{decrease}} \dots \dots \dots (5)$$

If variable ndryshimi with initial value 1 decreases for Δz from pbashkimi – trenja till pbashkimi, at the point pbashkimi will have zero value. Using variable gradation we can write:

$$r = \frac{\sqrt{E_2} \cdot gradation + \sqrt{E_1} \cdot (1 - gradation)}{\sqrt{E_2}} \dots \dots \dots (6)$$

If each sample of the first sequence from p_junction – t_decrease till p_junction is multiplied with r, there will be gradual decrease during the interval.

Algorithm for equalization of amplitudes after joining of the sequences uses division of the square roots of energies, which means that it is necessary to calculate the energies of the periods. After calculating the energies, is calculated ratio r given by equation (6) and each sample of the first sequence from p_junction – t_decrease to p_junction is multiplied by this value (r), (pseudocode below).

```

E1 = 0
t = Point_Junction
WHILE t > Point_Start - Length_Period1
    E1 = E1 + samples[t]^2
    t = t - 1
REPEAT
    E1 = E1/Length_Period1
    E2 = 0
    t = Point_Junction
WHILE t < Point_Start - Length_Period2
    E2 = E2 + samples[t]^2
    t = t + 1
REPEAT
    E2 = E2/Length_Period2
    gradation = 1
    dz = 1/t_decrease
    r = (sqrt(E2)*gradation + sqrt(E1)*(1-gradation))/sqrt(E2)
    t = Point_Junction - t_decrease
WHILE t < Point_Start
    samples[t] = samples[t] * r
    r = r + dz
    t = t + 1
REPEAT
    
```

3. Conclusions

Concatenation Synthesis is the most used technique for speech synthesis from written text. Acoustic segments are created by cutting dyphones from a reading text. These acoustic segments stored in a database. Through software applications, with the union (junction) of these acoustic segments is synthesis the speech. The union of acoustic sequences to form words in speech synthesis from written text occur problems, which causes no meaningful and non-natural speech. This is caused by the change in the amplitude level of the sequences. The algorithms presented in this paper do minimize these effects by improving the quality of the generated speech. These algorithms have to do with increasing or decreasing the amplitude of the sequence at the beginning or end of the word. We also present an algorithm for equalization of the two sequences amplitude in the union area.

Adnan Maxhuni Master of Computer Science (2005) at the University of Prishtina; Teaching and Research Assistant at the University of Prishtina.

Agni Dika PhD degree in Computer Sciences (1989) at the University of Zagreb, Croatia; Full Professor at the University of Prishtina, Faculty of Electrical and Computer Engineering.

Avni Rexhepi PhD degree in Computer Science (2013) at the University of Prishtina; Teaching and Research Assistant at the University of Prishtina.

Dren Imeraj BSc in Computer Science (2012) at the University of Prishtina.

References

- [1] Abramowitz, Milton; Stegun, Irene A., "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", Dover, 1965.
- [2] Dutoit, Thierry, "An introduction to text-to-speech synthesis", Kluwer Academic Publishers, 2001.
- [3] Hayes, M. Horace, "Digital Signal Processing", McGraw Hill, 1999.
- [4] Jurafsky, Daniel & Martin, James H., "Speech and Language Processing", Prentice Hall, 2009.
- [5] Kaiser, J. F., "System Analysis by Digital Computer", Wiley, 1966.
- [6] Lemmetty, Sami., "Review of Speech Synthesis Technology", 1999.
- [7] Agni Dika, Adnan Maxhuni, Avni Rexhepi, "The principles of designing of algorithm for speech synthesis from texts written in Albanian language", IJCSI, May 2012.
- [8] Avni Rexhepi, Agni Dika, Adnan Maxhuni, "Conversion of numbers to text and to speech in Albanian", WSEAS - World Scientific and Engineering Academy and Society, Greece, 2012.
- [9] MBROLA. Multilingual Speech Synthesizer
<http://tcts.fpms.ac.be/synthesis/mbrola.html>, 2008.
- [10] Agni Dika, Mentor Hamiti, "Options for Converting Albanian Written Texts into Speech", BCI, 2009.

A Fuzzy Based Feature Extraction Approach for Handwritten Characters

Mahmood K Jasim¹, Anwar M Al-Saleh² and Alaa Aljanaby¹

¹Department of Mathematical & Physical Sciences, College of Arts & Sciences, University of Nizwa, Oman

² Computer Science Department, College of Sciences, Al-Mustansiriyah University, Iraq

Abstract

This paper describes a technique that can be used to generate fuzzy rules to extract the features of handwritten characters. The feature extraction is a complicated problem as different people write the same character in different ways. The development of a technique that can generate the description of handwritten characters is still a challenge for the handwritten recognition systems. The fuzzy logic offers a good opportunity to build a rule-based feature extraction technique for handwritten characters with low computational cost.

Keywords: Feature Extraction, Handwritten Characters, Fuzzy Logic.

1. Introduction

The term feature extraction consists of two meanings; feature detection, and feature selection [1]. The purpose of feature detection is to obtain those features, which preserve the useful information about the image to the largest extent. The aim of feature selection is to determine those principal feature components depending on a certain classification task in order to achieve an effective classification [1, 2]. The above idea shows that the output of feature detector reflects the information of the image. Feature extraction is responsible for extracting all possible features that are expected to be effective in diagnosing all information of image, without concerning the disadvantages of excessive dimensionality [3].

The feature after selection may not contain enough information about the original image, but it must contain the information that is useful to distinguish different classes for image classification [2, 3]. Figure 1 shows a block diagram of feature extractor for classification system.

In the Handwritten Recognition systems, many tedious tasks can be made more efficient by automating the process of reading handwritten numerals. In such system an optical scanner converts each handwritten numeral to a

digital image, and computer software classifies the image as one of the digits zero through nine. By reducing the need for human interaction, numeral-recognition systems can speed up jobs such as reading income tax returns, sorting inventory, and routing mail. Several steps are necessary to achieve this. A recognition system must first capture digital image of handwritten numerals. Before attempting to classify the numerals, some preprocessing image might be necessary. An algorithm must then classify each handwritten numeral as one of the ten decimal digits [4, 5].

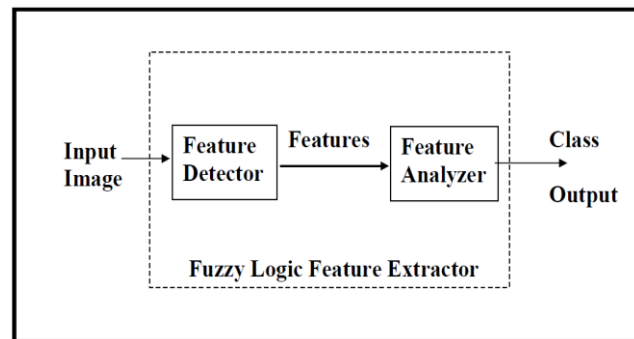


Fig. 1: Block Diagram of Feature Extractor

Although a qualitative description of this process is straight forward, it cannot be easily reduced to a few simple mathematical rules. The difficulty results from the natural variations in human handwritten. A useful recognition system must be robust to alterations in size, shape, orientation, thickness, etc. Closed-form mathematical models tend to be inadequate for such a task because of the many possible representations of the same image. The problem presents certain obstacles that make pattern matching on a pixel-by-pixel basis impractical. For instance, the edge of a character segment can show up in two or more data slices (all the pixels along one column), depending on where the slices overlap. Further, slight variations in printing cause character height and width to vary and misfeeding of the document can skew the imaged character.

Fuzzy logic, which is inherently superior for processing imprecise data, is a natural for this application [1, 7]. However, a data preprocessor is necessary to simplify the problem so that it can be easily described in fuzzy rules. Feature extraction is the crucial phase in numeral identification as each numeral is unique in its own way, thus distinguishing itself from other numerals. Hence, it is very important to extract features in such a way that the recognition of different numerals becomes easier on the basis of the individual features of each numeral [8, 9].

In the present paper, the authors propose the transition calculation and sum of pixels of an image as a feature detector, fuzzy logic technique as a feature analyzer, and extraction the most useful information as outputs for classifier process. Handwritten numerals recognition system has been designed and implemented and a high degree of accuracy has been gained using fuzzy logic.

2. Feature Extraction

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. The literature is replete with high accuracy recognition systems for separated handwritten numerals and characters [10]. However, research into the recognition of characters extracted from cursive and touching handwriting has not had the same measure of success [11]. One of the main problems faced when dealing with segmented, handwritten character recognition is the ambiguity and illegibility of the characters. Figure 2 illustrates the difficulties a programmer encounter when trying to match incoming patterns against an idealized pattern, or template. Each of the three sections of Figure 2 shows twenty data slices of typical read of the character 0.

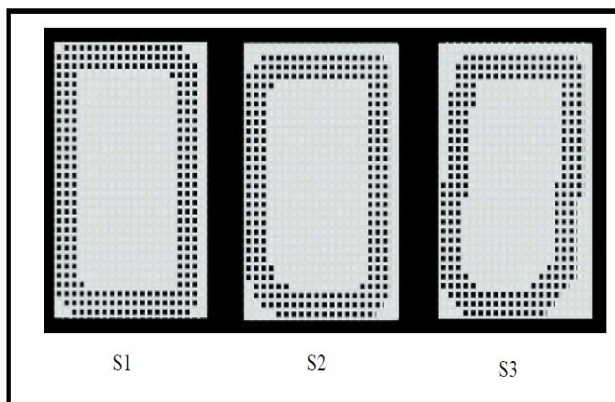


Fig. 2: Different Patterns of Character Zero

The leftmost portion of Figure 2 (S1) represents the pattern associated with an ideal read of a character 0. This portion of the figure can be considered to be a template for the read of a character 0. The center and the right portions of Figure 2 (S2, S3) show some patterns of character 0.

One approach to recognition would have a program compare scanned characters to templates on a pixel-by-pixel basis. Clearly, this procedure could often fail (in this case). For instance, the program would expect a 1 in slice 1, local 3 of a character 0 in pattern S1, and neither S2 nor S3 characters would satisfy the expectation.

Another approach would have the program sum all the pixels in each slice and compare the resulting slice totals to corresponding slice totals from templates. As shown in table 1 below, this approach also cannot produce a match in both S2 and S3 case.

Table 1: Slice Totals for the Three Readings of Fig. 1

Slice	S1	S2	S3
1	27	23	12
2	29	26	23
3	30	28	27
4	7	9	21
5	6	7	10
6	6	7	6
7	6	6	6
8	6	6	6
9	6	6	6
10	6	6	6
11	6	6	6
12	6	6	6
13	6	6	7
14	6	6	8
15	6	6	8
16	7	7	16
17	30	29	26
18	29	28	25
19	27	26	15
20	0	0	0

The data in Table 1 provides a useful insight. It is apparent in all three cases that the magnitude of the slice total increases to a high value of approximately 30, decreases to a low value of approximately 6, increases again to a high value of approximately 30, and then finally decreases to zero. It is possible to locate and quantify these transitions, or changes of magnitude, in slice totals. Quantified transitions will form the input to the handwritten recognition fuzzy system. The fuzzy rules will look something like this: A very large positive transition, followed by a large negative transition, followed by a large positive transition, followed by a very large negative transition, indicates a character zero. A transition is defined as the difference between a current local maximum (or minimum) and the previous local minimum (or maximum). The data preprocessor takes a data slice,

obtains its slice total, and compares the magnitude of the slice total to previous slice total to determine whether it constitutes a new local maximum or minimum.

3. The Transition Concept

Figure 3 below shows the block diagram of the preprocessor for Transition Calculation algorithm which takes in slice data and generates transition outputs. The variables that are updated during preprocessor operation are listed in Figure 3. Preprocessor outputs take the form of a transition number and an associated transition magnitude. For instance, T1=27 means transition number 1 has a magnitude of 27. The algorithm incorporates hysteresis in determining a direction change. In other words, a transition must be of three pixels or greater magnitude to be recognized. For instance, if a current reading produce a slice total of 6 and the previous reading left DIR as (-) and L-min as 4, the current reading would fail the test $CR > PR + 2$, but would pass the test $CR > PR$. Since DIR-, none of the variables are changed. The preprocessor algorithm has no effect on system throughput because it can be run during the delay for integration time.

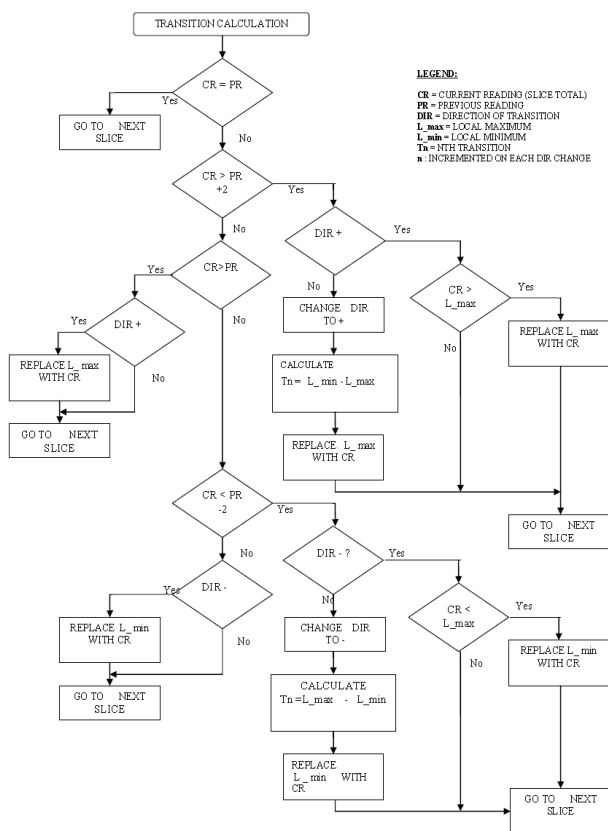


Fig. 3: The Block Diagram of the Preprocessor for Transition Calculation Algorithm

Table 2 shows how variables are updated after each slice. The slice data applied is from the S3 case shown in Figure 2. Prior to entering the routine Transition Calculation, variables are initialized to the values shown in the column labeled Init. Data slice #1 is defined as the first slice with a slice total greater than 2. A final transition number magnitude calculation is forced after the 22nd slice.

Table 2: Translation Calculation for S3 Character Scan

Slice	CR	PR	DIR	T	L_min	L_max	Tmag
Init	0	0	+		0	0	0
1	12	0	+		0	12	0
2	23	12	+		0	23	0
3	27	23	+		0	27	0
4	21	27	-	T1	21	27	27
5	10	21	-		10	27	0
6	6	10	-		6	27	0
7	6	6	-		6	27	0
8	6	6	-		6	27	0
9	6	6	-		6	27	0
10	6	6	-		6	27	0
11	6	6	-		6	27	0
12	6	6	-		6	27	0
13	7	6	-		6	27	0
14	8	7	-		6	27	0
15	8	8	-		6	27	0
16	16	8	+	T2	6	16	-21
17	26	16	+		6	26	0
18	25	26	+		6	26	0
19	15	25	-	T3	15	26	20
20	0	15	-		0	26	0
21	0	0	-		0	26	0
22	0	0	-	T4	0	26	-26

The preprocessor found the following four transitions while reading this character. T1=27, T2=-21, T3=20, T4=-26. For Comparison, the preprocessor would return the following values for both S1 and S2 character 0 (respectively): (S1: T1=30, T2=-24, T3=24, T4=-30) and (S2: T1=28, T2=-22, T3=23, T4=-29). The visual representation of Table 2 can be represented by the following figure 4 below.

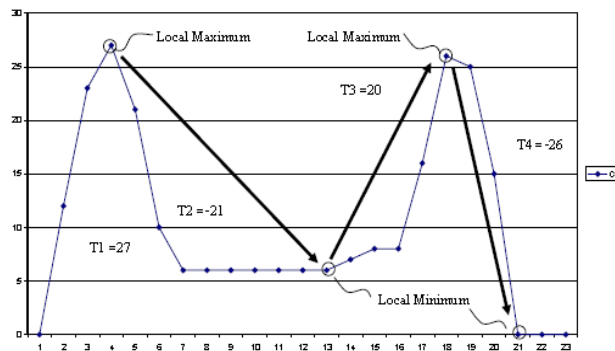


Fig. 4: Visual Representation of Table 2 Transition

4. Fuzzifying Transition Inputs

The transition visualized in figure 3 above make it very easy to write a fuzzy rule that recognizes a character 0: *If T1 is Very_Large_Positive and T2 is Large_Negative and T3 is Large_Positive and T4 is Very_Large_Negative, then character is 0.* A similar visualization of all ten characters is required to write the remaining rules. Table 3 below represents the range of transition magnitudes for all ten characters. This data was obtained by unconstrained images represented handwritten numeral. The number of transitions per character varies from two (characters 1, 3, and 7) to four (characters 2, 4, 5, 6, 8, 9, and 0).

The first step in fuzzifying this data is to establish a universe of discourse that defines the range of possible values for fuzzy input. Once the universe of discourse is defined, fuzzy sets can be created within it. In this case, T1, T2, T3, and T4 are fuzzy inputs. From Table 3 below, transition magnitudes, measured in pixels, vary from -28 to +28, since a slightly over sized character or stray marks on the document can cause more pixels to be counted, the universe of discourse is represented by the range of value from -30 to +30. There are actually two universes of discourse: a positive one associated with T1 and T3, and a negative one associated with T2 and T4. The positive universe of discourse is defined as the range of values from 5 to 28 pixels, and the negative universe of discourse is defined as the range of values from -3 to -28 pixels.

Table 3: Transition Magnitude Ranges for All Ten Characters

Char	T1 (LO,HI)		T2 (LO,HI)		T3 (LO,HI)		T4 (LO,HI)	
0	21	25	-14	-20	14	20	-21	-25
1	13	13	-13	-13	---	---	---	---
2	17	19	-6	-9	5	10	-17	-20
3	23	27	-23	-27	---	---	---	---
4	15	19	-15	-19	12	16	-15	-19
5	16	19	-6	-10	5	9	-14	-20
6	21	24	-12	-16	6	8	-12	-17
7	25	28	-25	-28	---	---	---	---
8	20	22	-7	-11	13	17	-23	-27
9	13	16	-3	-8	17	21	-23	-28

The distribution of transition values across the universe discourse label denotes each character and transition number, putting in graphical representation of transition range, from Table 3 above.

It is obvious from Figure 5 that there is some clumping of transition data could be used to create fuzzy sets that apply all four transition inputs. Based on the natural grouping of data, fuzzy sets are assigned and labeled. At this point, transition inputs and fuzzy sets are defined.

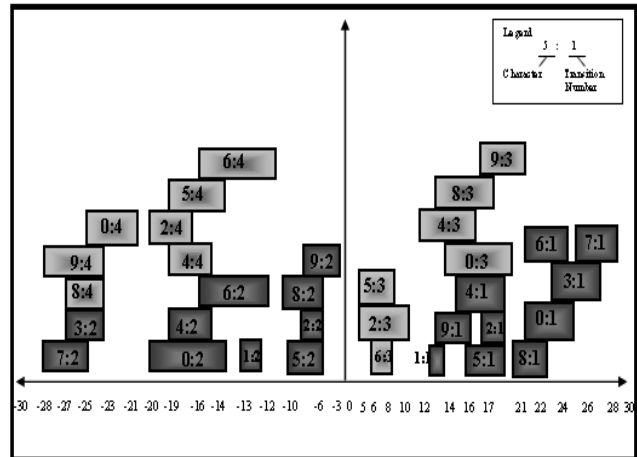


Fig. 5: Distribution of Transitions across the Universe of Discourse

Figures 6, 7, 8 and 9 show the fuzzy set distribution for T1, T2, T3 and T4. Unfortunately, transition inputs alone are not adequate to classify characters. Consider the characters 3 and 7. Each has only two transitions, T1 and T2. The rule for character 3 is: if T1 is Large and T2 is Small then character is 3. The rule for character 7 is: if T1 is large and T2 is Small then character is 7. Thus, from transition data, 3 and 7 are indistinguishable.

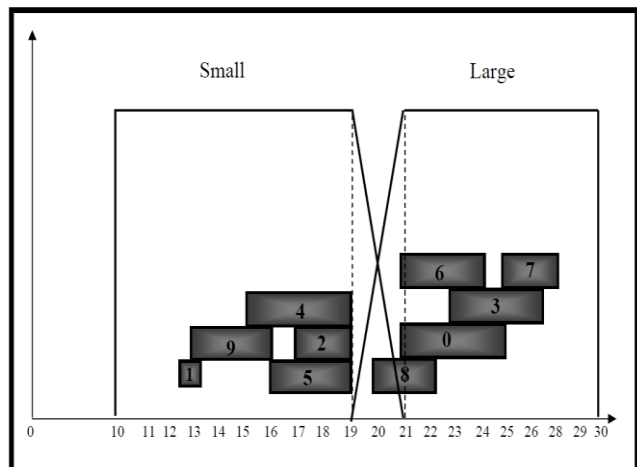


Fig. 6: T1 Fuzzy Set Distribution

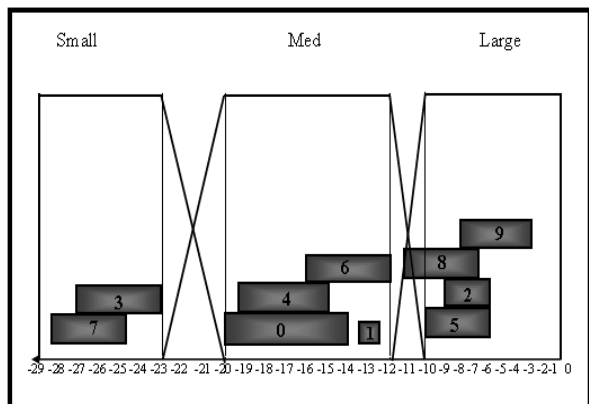


Fig. 7: T2 Fuzzy Set Distribution

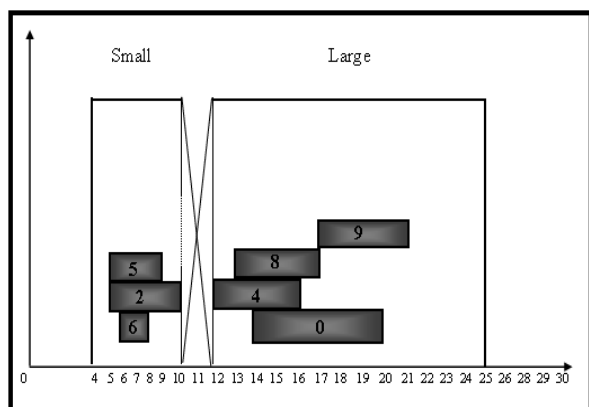


Fig. 8: T3 Fuzzy Set Distribution

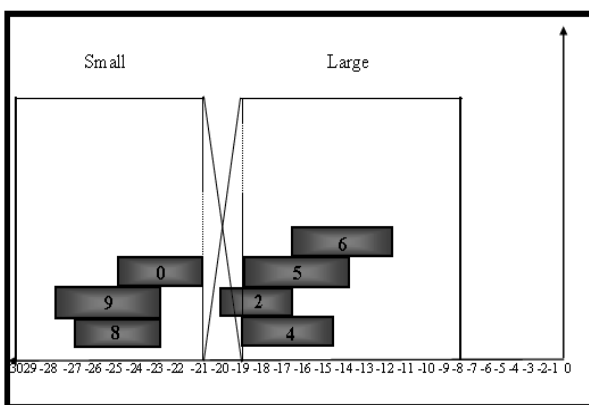


Fig. 9: T4 Fuzzy Set Distribution

Table 4 shows the fuzzy magnitude of each transition presented by character. The conflict column shows the other characters which present the same fuzzy magnitude. Several conflicts occur, so it is necessary to introduce input variables in addition to transition magnitude.

Table 4: Transition Magnitudes for Each Character

Char	T1	T2	T3	T4	Conflict
0	Large	Med	Large	Small	---
1	Small	Med	---	---	4
2	Small	Large	Small	Large	5
3	Large	Small	---	---	7
4	Small	Med	Large	Large	1
5	Small	Large	Small	Large	2
6	Large	Med	Small	Large	---
7	Large	Small	---	---	3
8	Large	Large	Large	Small	---
9	Small	Large	Large	Small	---

5. Sums Of Pixels (SOP)

Some of the conflicts shown in Table 4 can be resolved by considering the total dark area in each character. Total dark area is measured as a sum of pixels of the image, or SOP. The image is divided into four equal quarters and computes the sum of pixels for each quarter. The sub image 1 is the sub image in the top left of the original image (character image), and its size is (15×10), as shown in Figure 10 bellow.



Fig. 10: SOP1 and SOP2

SOP1, represent sum of pixels of sub image 1, will be the fifth fuzz input. Table 5 shows the range of SOP1 values for each character.

Table 5: SOP1 Range for Each Character

Char.	0	1	2	3	4	5	6	7	8	9
SOP1 (low)	56	12	47	50	42	58	53	27	73	47
SOP1 (high)	61	12	65	56	50	84	68	30	80	70
SOP1 (avg.)	59	12	56	53	53	71	61	29	77	59

Figure 11 shows how the conflicts shown in Table 4 are resolved by the addition of the SOP1 input variable. Consider the conflict between characters 1 and 4. The value of SOP1 for the character 1 is always produces a degree of membership of 1 in the fuzzy set SOP1 Small and a degree of membership of 0 in the fuzzy set SOP1 Large, so that a 1 is never recognized as a 4. Conversely, as long as the possible range of SOP1 values for 4 always produces a higher degree of membership in the fuzzy set Large than in the fuzzy set Small, a 4 is recognized as a 4

rather than as a 1. SOP1 for character 4 ranges from 42 to 50 (Table 5). Figure 11 shows that any value of SOP1 from 42 to 50 produces a degree of membership of 1 in the fuzzy set SOP1 Large. The conflict is completely resolved. Each character appears exclusively within a fuzzy set with no overlap.

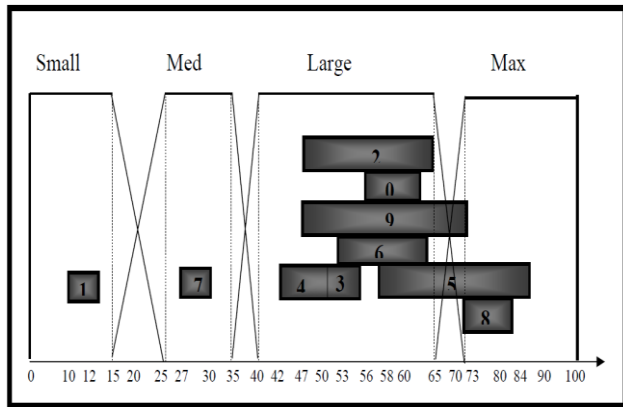


Fig. 11: SOP1 Fuzzy Set Distribution

The input variable SOP1 also resolved the conflict between the two characters 3 and 7. As long as the possible range of SOP1 values for 3 always produces a higher degree of membership in the fuzzy set Large than in the fuzzy set Med, a 3 is recognized as a 3 rather than as a 7. SOP1 for character 3 ranges from 50 to 56 (Table 5). Figure 11 shows that any value of SOP1 from 50 to 56 produces a degree of membership of 1 in the fuzzy set SOP1 Large. Values from 40 to 35 produce declining degrees of membership. Since the minimum value of SOP1 from character 3 is 50, the character 3 always produces some degree of membership in the fuzzy set SOP1 Large and none in the fuzzy set SOP1 Med. Therefore, a 3 is never recognized as a 7. Conversely, the range of SOP1 values for the character 7 is always produces a degree of membership of 1 in the fuzzy set SOP1 Med and a degree of membership of 0 in the fuzzy set SOP1 Large, so that a 7 is never recognized as a 3. The conflict is completely resolved between characters 3 and 7. Each character appears exclusively within a fuzzy set with no overlap.

Notice that the lowest value of SOP1 for 5 produces a higher degree of association with Large than with Max, preventing a resultant 5 for low values of SOP1. Then because of the interaction of its SOP1 ranges with multiple fuzzy sets, SOP1 should not be used as a qualifier for character 5. So the range of SOP1 values for character 5 is not adequately represented by either SOP1 Large or Max. In this case, it is best to leave the SOP1 input variable out of the fuzzy equation for character 5. Therefore an

additional input variable is required. SOP1 is only resolving the conflict between characters 1 and 4 and between characters 3 and 7. However, it makes sense to assign the remaining characters to fuzzy sets in the SOP1 universe of discourse. The minimal additional code that required for these rule additions produces better-qualified results and a more robust classification system.

An additional characteristic that may resolve the remaining conflicts is by considering the sum of pixels in each character sub image 2, or SOP2, where sub image 2 is a sub image of character image with Size (15 x 10) as shown in Figure 6. Table 6 shows the range of SOP2 values for each character. Figure 7 shows the universe of discourse and fuzzy sets for SOP2.

Table 6: SOP2 Range for Each Character

Char	0	1	2	3	4	5	6	7	8	9
SOP2 (low)	59	14	61	18	45	32	78	0	72	30
SOP2 (high)	63	14	70	30	67	52	90	5	82	42
SOP2 (avg.)	61	14	66	24	56	42	84	3	77	36

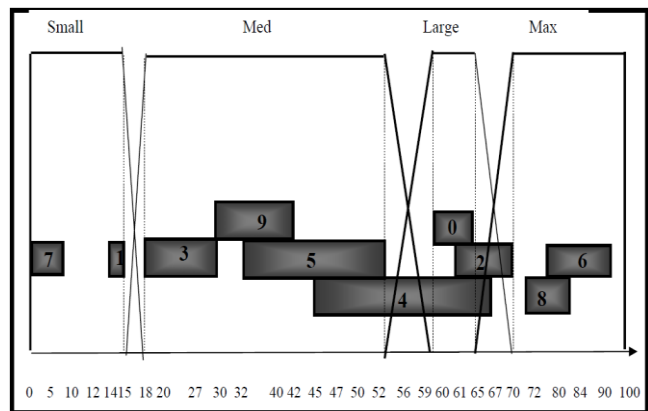


Fig. 12: SOP2 Fuzzy Set Distribution

Figure 12 shows how the conflict between characters 2 and 5 is resolved by the addition of the SOP2 input variable. As long as the possible range of SOP2 values for 2 always produces a degree of membership in the fuzzy set Large than in the fuzzy set Med, a 2 is recognized as a 2 rather than as a 5. Conversely, the range of SOP2 values for the character 5 is always produces a degree of membership of 1 in the fuzzy set SOP2 Med and a degree of membership of 0 in the fuzzy set SOP2 Large, so that a 5 is never recognized as a 2. The conflict is completely resolved

between characters 2 and 5. Each character appears exclusively within a fuzzy set with no overlap. Fuzzy rules create associations between specific inputs and desired outputs. Here, the six input variables are T1, T2, T3, T4, SOP1 and SOP2. Ten valid output variables, one for each of the ten characters to be recognized, are defined. Table 7 shows the relationship between fuzzy input variables and fuzzy output variables.

Table 7: Fuzzy Set Association

Char	T1	T2	T3	T4	SOP1	SOP2
0	Large	Med	Large	Small	Large	Large
1	Small	Med	-----	-----	Small	Small
2	Small	Large	Small	Large	Large	Large
3	Large	Small	-----	-----	Large	Med
4	Small	Med	Large	Large	Large	-----
5	Small	Large	Small	Large	-----	Med
6	Large	Med	Small	Large	Large	Max
7	Large	Small	-----	-----	Med	Small
8	Large	Large	Large	Small	Max	Max
9	Small	Large	Large	Small	Large	Med

With the relationship information summarized in Table 7, the fuzzy rules can be written as:

Rule 1: if T1 is Large and T2 is Med and T3 is Large and T4 is Small and SOP1 is Large and SOP2 is Large **then** Character is *Zero*.

Rule 2: if T1 is Small and T2 is Med and SOP1 is Small and SOP2 is Small **then** Character is *One*.

Rule 3: if T1 is Small and T2 is Large and T3 is Small and T4 is Large and SOP1 is Large and SOP2 is Large **then** Character is *Two*.

Rule 4: if T1 is Large and T2 is Small SOP1 is Large SOP2 is Med **then** Character is *Three*.

Rule 5: if T1 is Small and T2 is Med and T3 is Large and T4 is Large SOP1 is Large **then** Character is *Four*.

Rule 6: if T1 is Small and T2 is Large and T3 is Small and T4 is Large and SOP2 is Med **then** Character is *Five*.

Rule 7: if T1 is Large and T2 is Med and T3 is Small and T4 is Large and SOP1 is Large and SOP2 is Max **then** Character is *Six*.

Rule 8: if T1 is Large and T2 is Smell and SOP1 is Med and SOP2 is Small **then** Character is *Seven*.

Rule 9: if T1 is Large and T2 is Large and T3 is Large and T4 is Small and SOP1 is Max and SOP2 is Max **then** Character is *Eight*.

Rule 10: if T1 is Small and T2 is Large and T3 is Large and T4 is Small and SOP1 is Large and SOP2 is Med **then** Character is *Nine*.

6. Conclusion

The method presented in this paper is a reliable and relatively simple method for generating the fuzzy rule-based description of handwritten characters. Even though this method is not deemed to be the ultimate solution for the recognition of handwritten characters but it is a solution which is extremely simple to implement and use. Testing results will be considered in a forthcoming paper incorporating the presented technique in a complete online handwritten recognition system.

References

- [1] Bandara G.E, Pathirana S.D., Ranawana R. M. "Use of Fuzzy Feature Descriptions to Recognize Handwritten Alphanumeric Characters", 1st Conference on Fuzzy Systems and Knowledge Discovery, Singapore, November 2002.
- [2] Belal K. Elfarra and Ibrahim S. I. Abuhaiba. "New Feature Extraction Method for Mammogram Computer Aided Diagnosis", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 6, No. 1, February, 2013.
- [3] Rajashekaradhy, S. and Ranjan, P. "Zone based feature extraction algorithm for handwritten numeral recognition of kannada script", In Advance Computing Conference, 2009. IACC 2009. IEEE International, pages 525–528, 2009.
- [4] L. M. Lorigo and V. Govindaraju, "Offline Arabic handwriting recognition: a survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 5, pp. 712 - 724, 2006.
- [5] Bandara G.E., Pathirana S.D., Ranawana R. M., "A Short Method for On-line Alphanumeric Character Recognition", NAFIPS – FLINT 2002, New Orleans, USA, June 2002.
- [6] Pal, U., Sharma, N., Wakabayashi, T., and Kimura, F. "Handwritten character recognition of popular south indian scripts", In Doermann, D. and Jaeger, S., editors, Arabic and Chinese Handwriting Recognition, volume 4768 of Lecture Notes in Computer Science, pages 251–264, Springer Berlin / Heidelberg, 2008.

- [7] Jan J. “Tutorial on fuzzy logic”, Technical University of Denmark of Automation, Report No. 98-E868, 19 Aug. 1998.
- [8] Yadana T. and San S. Y. “High Accuracy Myanmar Handwritten Character Recognition using Hybrid approach through MICR and Neural Network”, International Journal of Computer Science Issues (IJCSI), Vol. 7, Issue 6, November 2010.
- [9] Muthumani .I , Uma Kumari C.R. “Online Character Recognition of Handwritten Cursive Script”, International Journal of Computer Science Issues (IJCSI), Vol. 9, Issue 3, No 2, May 2012.
- [10] Lauer, F., Suen, C. Y., and Bloch, G. “A trainable feature extractor for handwritten digit recognition”, Pattern Recognition, 40(6):1816–1824, 2007.
- [11] Pal, U., Sharma, N., Wakabayashi, T., and Kimura, F. “Handwritten character recognition of popular south indian scripts”, In Doermann, D. and Jaeger, S., editors, Arabic and Chinese Handwriting Recognition, volume 4768 of Lecture Notes in Computer Science, pages 251–264, Springer Berlin / Heidelberg, 2008.

Mahmood K Jasim is an Associate Professor and head of the Department of Mathematical and Physical Sciences, College of Arts and Sciences, University of Nizwa, Oman. His research interests are Mathematical Modeling, Mathematical Physics, Differential Equations and its application, Numerical Analysis, Artificial Neural Networks and Fuzzy logic.

Anwar M Al-Saleh is a Lecturer at Computer Science Department, College of Sciences, Al-Mustansiriyah University, Baghdad, IRAQ. Her research interests are Fuzzy Logic and Image processing.

Alaa Aljanaby is an IEEE member, Assistant Professor and head of the Computer Science Section, Department of Mathematical and Physical Sciences, College of Arts and Sciences, University of Nizwa, Oman. His research interests are soft computing, bio-inspired optimization algorithms, combinatorial optimization problems, persuasive technology, opinion mining and image processing.

RBAC Architectural Design Issues in Institutions Collaborative Environment

Muhammad Umar Aftab¹, Amna Nisar², Muhammad Asif³, Adeel Ashraf⁴, Burhan Gill⁵

¹ Computer Science Department, National Textile University,
Faisalabad, Pakistan

² Computer Science Department, National Textile University,
Faisalabad, Pakistan

³ Computer Science Department, National Textile University,
Faisalabad, Pakistan

⁴ Computer Science Department, National Textile University,
Faisalabad, Pakistan

⁵ Computer Science Department, National Textile University,
Faisalabad, Pakistan

Abstract

Institutional collaborative systems focus on providing the fast, and secure connections to students, teaching and non-teaching staff members. Access control is more important in these types of systems because different kind of users access the system on different levels. So a proper architecture must be there for these kinds of systems, for providing an efficient and secure system. As lot of work was done in RBAC like for grouping, securing the system, ease of use, and for enterprise etc. but no one apply all these concepts as a whole on institution level. So, this paper will be a step towards administrative load sharing, securing the system with separation of duty (SOD) and ease of use.

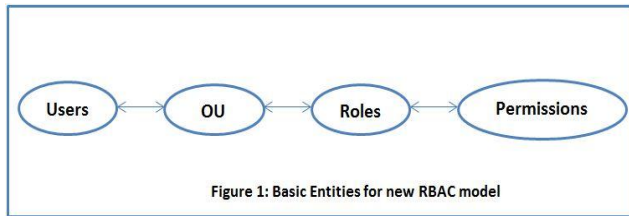
Keywords: *Separation of Duty (SOD), Institutional RBAC Architecture, Architectural Issues, Organizational Unit & RBAC.*

1. Introduction

More secure, efficient and easy to manage access control is a big task for sensitive organizations, especially in institutional collaborative systems. Where different groups of users work jointly, share resources, communicate with each other, on common tasks. The information and resources of such kind of organizations have different kind of sensitivities like unauthorized access, information theft, efficient and secure resource sharing etc. Access control is a mechanism through which organizations can apply restrictions on resources, to protect them from unauthorized access. The access control will allow the

authorized persons to access the granted resources like internet, file sharing, and read/write in a file etc. or restrict the unauthorized persons from accessing the un-granted resources like server access, management of servers and access rights management etc.[1, 2]

One of the most widely used access control model is Role Based Access Control (RBAC). RBAC is a rich technology and a great effort in the field of access control. The basic idea in RBAC is a role that is the central part of the RBAC. In normal practice the permissions are directly assigned to the users but this is the worst case because administrator assigns or revokes permissions one by one to or from users.[3]. This process is time taking, complex and not an efficient approach towards the rights management. Previously, user management was done with groups. Permissions were assigned to groups instead of users one by one. This technique is good for user management but not for the management of permissions.[4] In RBAC, permissions are assigned to the central layer that is role as well as users are also assigned to the roles. If administrator wants to insert or revoke a user then he/she just assigns or revokes the user to/from the role. No need to assign or revoke permissions one by one to/from the user. Through RBAC a more secure, efficient, and easy to manage network can be created.[2] [4, 5]



In this paper, authors try to highlight the RBAC architectural issues in institution collaborative systems. Previously, there was no specific architecture design, defined for the institutions and more important in the context of RBAC as well as divide the working hierarchy with the help of SOD.

In this study, we proposed an architectural design in which we insert another central layer between roles and users that is **organization unit (OU)**. OU is used for directory objects contained inside domains. They are basically the containers used in active directory and we can place groups, computers, users and other OU's in it. The purpose of using OU in system is the efficient users and objects management. [6]. By inserting this layer the work load on the administrator will decrease and management of the access control will become easier. Secondly, we design an architecture design for institutions that is described in hierarchical institution RBAC Architectural design portion.

2. Problem Statement

There exist some problems in the existing systems. To start with the architectural design in context with RBAC in the institutions because there must be a mechanism from which an institution can run its system for the management of its IT resources.[1] On the other hand, there is no defined mechanism for managing users as the permissions like in RBAC. Number of permissions can be managed by assigning these permissions to roles. But how to manage a large number of users like students, teaching and non-teaching staff? Usually, our most concern is about the enterprise but no proper concern with the institutions.

3. Proposed Model Detailed View

RBAC is known as a marvelous model in an institution environment for access control. A role has collection of privileges in RBAC that can be assigned to users.[7] As permissions are assigned to roles, which happens to be a better management of permissions thus, helping administrator in secure and easy management of the network. Before RBAC, permissions were assigned to

users manually and directly without any central layer. That was not a good way towards efficient management of access control. As permission allocation and revocation was very difficult in that approach. On the other hand, organizational unit (OU) and groups' concept can be used for the management of users. Because it is really hard to manage the number of users in an organization without OU.

We proposed a system that used both the concept in a single system, RBAC with OU. The system contains an additional central layer between users and roles that is OU. By combining both concepts, we can easily manage both ends i.e. users and permissions. For managing users, we use organizational units and for managing the permissions, we use roles. The purpose of adding OU is to manage users in an efficient way, as well as decrease the extra burden of administrator. In normal practice, administrator creates, delete and manage users with the other important IT tasks like applying access control, and manage servers etc. The creation and deletion of users is a time taking job and hard to manage with other IT tasks. Sometime user deletion and creation delayed due to some other priority tasks that were not good approach towards user facilitation. We proposed two things regarding user management and load sharing of administrator.

Firstly, we introduce OU concept for the efficient management of users in an institution. As institutions have number of students studying there and after every specific interval, new students are enrolled and previous students get passed from there. So, the user deletion and creation is a separate job for the institutions. By applying OU, users can be easily managed on department level and in specific department, more segregated on specific course with semester/year wise. Secondly, we proposed that the user management or OU creation will be done on department level. Because in every department there are IT coordinators that works under the instructions of administrator or IT manager. If we only assign the user creation and deletion rights to IT coordinator then the work load of administrator will be shared or decreased. As well as, this approach for load sharing is secure because IT coordinator has user creation and deletion rights only. So, these approaches will enhance user facilitation, load sharing of administrator, and efficient management of users in an institution collaborative system.

RBAC is known as a marvelous model in an institution environment for access control. A role has collection of privileges in RBAC that can be assigned to users.[7] As permissions are assigned to roles, which happens to be a better management of permissions thus, helping administrator in secure and easy management of the network. Before RBAC, permissions were assigned to

users manually and directly without any central layer. That was not a good way towards efficient management of access control. As permission allocation and revocation was very difficult in that approach. On the other hand, organizational unit (OU) and groups' concept can be used for the management of users. Because it is really hard to manage the number of users in an organization without OU.

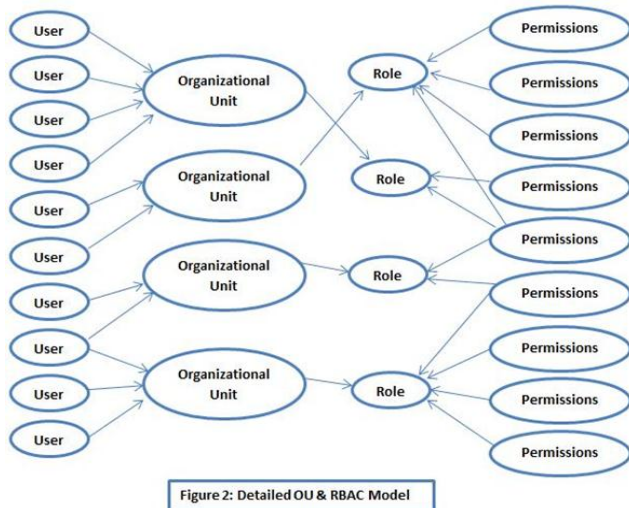


Figure 2: Detailed OU & RBAC Model

We proposed a system that used both the concept in a single system, RBAC with OU. The system contains an additional central layer between users and roles that is OU. By combining both concepts, we can easily manage both ends i.e. users and permissions. For managing users, we use organizational units and for managing the permissions, we use roles. The purpose of adding OU is to manage users in an efficient way, as well as decrease the extra burden of administrator. In normal practice, administrator creates, delete and manage users with the other important IT tasks like applying access control, and manage servers etc. The creation and deletion of users is a time taking job and hard to manage with other IT tasks. Sometime user deletion and creation delayed due to some other priority tasks that were not good approach towards user facilitation. We proposed two things regarding user management and load sharing of administrator.

Firstly, we introduce OU concept for the efficient management of users in an institution. As institutions have number of students studying there and after every specific interval, new students are enrolled and previous students get passed from there. So, the user deletion and creation is a separate job for the institutions. By applying OU, users can be easily managed on department level and in specific department, more segregated on specific course with semester/year wise. Secondly, we proposed that the user management or OU creation will be done on department

level. Because in every department there are IT coordinators that works under the instructions of administrator or IT manager. If we only assign the user creation and deletion rights to IT coordinator then the work load of administrator will be shared or decreased. As well as, this approach for load sharing is secure because IT coordinator has user creation and deletion rights only. So, these approaches will enhance user facilitation, load sharing of administrator, and efficient management of users in an institution collaborative system.

4. Hierarchical Institutional RBAC Architectural Design in Collaborative Environment

As discussed in previous section, the paper is about the architectural design for institution RBAC in collaborative environment. As we introduce the concept of OU for the efficient management of users. So, there must be an architecture that helps the institution and IT manager for the proper deployment of RBAC in an efficient manner, with OU concept.

The Institution model is at the top, it depends upon the particular institution's structure. As well as the permission and roles, implementation and deployment depends on this model. After this the major person that will manage, control and monitor all the activities like permission, and role, assigning, deletion to OUs etc and that person is Admin RBAC Manager. He has all the rights to manage the RBAC and its instances like permissions, roles and users etc.

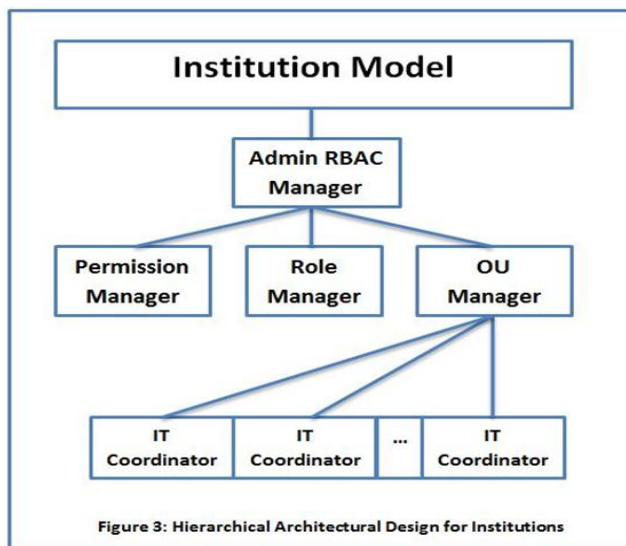
The working of RBAC is further segregated in three managers Permission manager, Role manager and OU manager. This segregation is more secure, efficient and a step towards load sharing. For example, if there is only one person in an institution where 10,000 plus students are studying as well as different courses are offered there. The person is the only one who will manage RBAC. Now, he is doing user creation, deletion, management of users department wise and he is applying permissions according to the requirements of departments, to roles. On the other hand, he is also managing the roles according to permissions and users. This whole scenario will create a hectic working system and this is not possible for a single person. So, this segregation and hierarchy will allow the managers to facilitate their users, system and institution in an efficient way.

The permission manager will manage, create, and revoke permission in the permission portion. He has all the rights related to permission management. The information for

any kind of change, new addition, or deletion in the permissions will be given by RBAC manager.

The role manager will manage, create and revoke roles in the roles portion. He has all the rights necessary for the management of roles. As well as he will assign permissions to roles and OUs to roles. The important information like which permission must be allocated to which role or roles will be described by the RBAC manager.

The OU manager will manage, create and revoke OUs in the OU portion. He has all the rights related to OU management. He communicates with IT coordinators for the proper management of OUs. As, IT coordinators are the persons who directly interact with users like students, teaching and non-teaching staff. IT coordinators have limited rights given by the OU manager. They can create, delete or edit user information only. As well as they assign or delete users in/from the OU and they can also change the OU of any user. So, the OU manager work is just related to OU because the major work load of user management will be shared by IT coordinators. IT coordinators will pass the information of user, necessary rights for users and which rights are required for which user like internet user, system user, and thin client users etc.



Finally, the RBAC manger manages all the mangers working underneath. It manages the whole RBAC process. As well as, he guide and transfer important information to permission, role and OU managers, through which those managers start and end their working.

This whole hierarchical scenario will secure the working of RBAC in an efficient manner. Because three managers are working in parallel, after the combination of all managers work a whole RBAC scenario completes that is more

secure. Like in bank lockers, a locker is open from two keys, one is of the user and second one is of the bank officer that secures the locker as compared to the single key locker. [5] [2, 8]

5. Related Work

Shared and collaborative systems, multi-user apps or groupware help users or groups in cooperating and communicating for common chores. A wide range of these types of applications are available like video/audio conferences, workflow applications, shared editing/writing software and more. These systems have resources and information having different levels of sensitivity. All the applications implemented in such systems make, control, and offer access to a range of secure resources and information. Access control models are most widely used in deciding how the accessibility of information and resources in a system will be supervised and how the group decisions will be expressed. Requirements for access control in collaborative settings have been studied and stated for example where it should be applied, better scalability, must be able in protecting information, constraints for obvious access for authorized users, high degree pattern of access rights, changeable policy facility and bounded costs. Access control models for collaboration have also been discussed which include Access Matrix Model, Role Based Access, Task Based Access, Team Based Access, Spatial Access Control, Context Aware Access Control[9]. Not only the significance of these models been discussed but also the shortcomings and weakness were illustrated and previously combined models were specified on the basis of preliminary work done. Evaluation criteria for these models were expressed on the basis of complexity, understandability, applicability, supporting, policy schemes, and user groups, enforcement of policies, context, and fine-grained control for particular permission.[9].

Enterprise collaborative systems frequently focus on how to build practical and worthy connections amongst users, resources, information and tools. This is where, access control is helpful. Concerns and issues of enterprise access control architecture design were explained. RBAC and TBAC (Task Based Access Control) models have been examined and their benefits and shortcomings were observed in application. A better and improved R & T model (role and task-based access control model), that combines the beneficial properties of both models, has been presented. This model is general and abstract. Object-oriented method was applied to detail out the model policy schemes related to security, and to devise an organization-

based task and resources or tools management method, in particular on dynamic role administration of collaborative system as a virtual organization unit (OU) where TBAC was used. A hierarchical access control architecture that can help out administrators to define, specify and implement security policy in hierarchical approach was presented. [10].

Security is an important issue, when it comes to access policies. Organizations define their own procedures/policies to prevent their information and resources from illegitimate access. Access policies or procedures are principles that state which users can perform which actions in order to implement rules of management control. Existing requirement modeling approaches have issues regarding roles and access procedures. These issues have been addressed and a framework has been proposed that will help to achieve the goals related to security. The significance of a macro-organizational had been studied and analyzed before role or actor definitions were specified in the perspective of modeling of access policies. Current modeling procedures lack this which makes it hard to clearly state the access policies and to improve them into operational constraints. A simple and new way of obtaining roles from the macro-organizational scenario had been proposed. Defining the groupings, the degrees of authority, and the management spheres/areas from which role can be described had also been confirmed. A contribution was made by demonstrating how the access policies satisfy the minimum rights and privileges. [11].

As we also gave the concept of separation of duty (SOD) in our proposed architecture. [2, 5] In the following paper and thesis they also work and describe the benefits of SOD. By dividing the work hierarchy, in more than one person, will secure the system. Because in the case of theft, the identification of one person is more difficult as compared to more than one person. Like in the bank example. A bank voucher contains two or more signatures from the different bank officials, just for the purpose of security and perform the SOD.[8].

6. Conclusion

As security is the main issue for any organization. By deploying RBAC, in an organization especially in our particular domain that is institutions, will provide a more secure and efficient system for the information and resources. As well as, the deployment of RBAC with OU that will facilitate the users in the sense of quick response from IT team members and administrators in sense of management, user facilitation and efficient working etc. In

last, we design an architectural design for the RBAC based institutions that will facilitate the administrators, for the understanding the whole system working and for the proper deployment of OU-RBAC system in an institution. The proposed system will be an excellent approach towards SOD by dividing the RBAC working in different managers and load sharing of administrator, efficient, ease of use and the deployment of system through a proper architectural design.

References

- [1] X. Zu, L. Liu, R. Xu, Access Control Architecture Design Issues in Enterprise Collaborative Environment, IEEE, (2009).
- [2] ANSI, Role Based Access Control, ANSI INCITS, (2004) 56.
- [3] D. Ferraiolo, Kuhn, D. R & Chandramouli, R., Role Based Access control, Artech House 2nd ed (2003).
- [4] Ravi S. Sandhu, E.J. Coyne, H.L.F. and, C.E. Youman, Role-Based Access Control Models, IEEE, 29 (1995) 22.
- [5] M.A. Habib, Secure RBAC with Dynamic, Efficient, & Usable DSD, (2011).
- [6] Microsoft, Organizational units, (2005).
- [7] E.B.a.E. Ferrari, The specification and enforcement of authorization constraints in workflow management systems., ACM, Vol. 2 (1999).
- [8] R.T. Simon, M.E. Zurko, Separation of Duty in Role Based Environments, Foundations of Computer Security Workshop, (1997).
- [9] T. William, A. Gail-Joon, P. Tanusree and H.Seng-Phill, Access control in collaborative systems, ACM, (2005).
- [10] X. Zu, L. Liu, R. Xu, Access Control Architecture Design Issues in Enterprise Collaborative Environment, in: Management and Service Science, 2009. MASS'09. International Conference on, IEEE, 2009, pp. 1-4.
- [11] Crook, D. Ince, B. Nuseibeh, On modelling access policies: Relating roles to their organizational context., IEEE, (2005).

Muhammad Umar Aftab is currently MS (CS) student in National Textile University Faisalabad. He did his BS (IT) Hons from Government College University Faisalabad (GCUF). He also completed his MCITP certification in 2012. Currently He is working as a Research Associate in National Textile University Faisalabad. He has more than one year experience as an EDP Supervisor in Metro Cash & Carry Faisalabad.

Amna Nisar is currently doing MS(CS) from National Textile University Faisalabad. She Done her MCS from Agriculture University Faisalabad.

Muhammad Asif is currently working as an Assistant Professor in department of computer science, National Textile University Faisalabad. He did his PhD candidate in computer science and information management program, school of engineering and

technology, Asian Institute of Technology Bangkok Thailand. He received his master degree in computer science from Quaid-I-Azam University Islamabad, Pakistan.

He worked as a Software Engineer in a project of Air traffic control system for Pakistan air force from May 2006 to July 2007. After wards He joined the Asian Institute of Technology for his masters and PhD studies on a scholarship of Govt. of Pakistan. He was selected for an exchange program with National Institute of Informatics Japan where He carried out his part of research.

Adeel Ashraf is currently doing MS (CS) from National Textile University Faisalabad. He did his MCS from Agriculture University Faisalabad.

Burhan Gill is currently doing MS (CS) from National Textile University Faisalabad. Done BS (IT)Hons from GCUF. Currently working in AA Spinning Mills Faisalabad as a Software Engineer.

An Efficient Approach for Sky Detection

Irfanullah¹, Kamal Haider², Qasim Sattar¹, Sadaqat-ur-Rehman¹, Amjad Ali¹

¹ Sarhad University of Science & Information Technology (SUIT), Peshawar, Pakistan

² Gandhara Institute of Science and Technology, Peshawar

Abstract— Blue color has been proven to be a useful and robust cue for sky detection, localization and for tracking RGB color in different applications of image processing. In this paper a pixel based solution utilizing the sky color information has been proposed for sky detection. The sky color information is extracted through the comparison of RGB values of a pixel. Based on the experimental results on highly complex still images, our approach for sky detection has been proved to be accurate, fast and simple.

Key words: Sky detection, RGB, pixel based sky detection, Color spaces.

1. INTRODUCTION

The sky detection is not a new problem for the researchers in the image processing domain. Due to its vast range of applications in weather forecasting, solar exposure prediction, image acquisition and understanding and in image retrieval and orientation, sky detection has been a keen area for researchers.

Sky detection becomes very difficult under certain circumstances, especially in overcast conditions and different types of clouds makes a real challenge for sky detection algorithms.

Mostly the pixel based approaches are used for sky detection in images/ videos to improve the image/ video quality as it becomes very easy to predict noise in the sky regions in an image because of its smooth appearance.

The color is considered the most robust and accurate feature for the sky detection. In this paper, we proposed a three step pixel based approach to sky detection that incorporates the

RGB values for pixels classification. Simple if else conditions are used for the identification of sky region pixels in an image.

2. RELATED WORK

Due to various applications like video/image quality enhancement, solar exposure prediction and weather forecasting, sky detection has been thoroughly studied and tackled through various approaches. Zafarifar et al. [1] represent a novel approach for sky detection using two different features. This algorithm utilizes the adaptive positioning and color modeling for segmentation and extracting the sky region information in an image/video. The proposed algorithm produced better performance compared to state-of-the-art approaches in sky detection in natural scenes.

Schmitt et al. [2] used color, position and shape as features for the sky detection in their approach. The performance of the proposed algorithm was tested on a number of outdoor images and the based on the analysis of different of the experimental results under different weather and lighting conditions the author claim for highly accurate performance in classifying the sky regions in images taken in clear, overcast and partially clouded weather.

Laungrunthip et al. [3] represent a solar exposure system based on the image processing techniques. Image processing algorithms are used for segmentation of the outdoor images taken under different lighting conditions to segment out the sky regions in a scene. A robust approach composed of canny edge detection algorithm [4] and Morphological closing algorithm [5], was adopted for identifying and separating the sky regions in color images.

Gallagher et al. [6] used two-dimensional polynomial model for sky detection. Their approach is composed of two parts, in 1st part the high confidence blue sky regions are detected while in 2nd phase the other candidate regions for sky are tested through two dimensional polynomial model. The performance of this algorithm is evaluated on 83 different images and produced accurate results.

3. COLOR SPACES DISTRIBUTION FOR SKY COLOR

Computer 3D graphic card and video communication standards have given origin to many color spaces with different properties. As our proposed approach use the color information for sky detection so a short overview of different color spaces and other issues related to color spaces is presented in this section.

3.1. RGB COLOR COMBINATION

RGB color space is based on Cathode rays tube display applications and is composed of three basic colors including: Red, Green and Blue. RGB is widely used in different applications of image/video processing, however the high correlation and significant perceptual non-uniformity are among the major drawbacks of this color space. Apart from this, the chrominance and illumination problems mitigate its significance in color analysis and color based recognition algorithms [7]. As for as, the sky color is concerned, it's a perceptual phenomenon not a physical property of an object. The general perception about the sky color is that it represents blue color; however it depends on the weather forecasting and this mixing of chrominance and luminance is a real challenge in color based sky detection approaches.

3.2. RGB COLOR RATIO

The sky regions in an image invariably contain a large number of pixels having values of blue

color. Using this consideration, most of the values of R/G ratio are used as skin presence indicators as discussed by Warket al. [8]. The significance of R/G ratio in skin pixel identification encouraged us to utilize this concept of RGB color ratio in sky detection.

3.3. NORMALIZED RGB

Normalized RGB is easily obtained by the following three equations

$$r' = R^*/R^*+G^*+B^* \quad (1)$$

$$g' = G^*/R^*+G^*+B^* \quad (2)$$

$$b' = B^*/R^*+G^*+B^* \quad (3)$$

The sum of the above three equations is equal to 1 i.e. ($r'+g'+b'=1$), which is a standard equation for getting the normalized value of RGB. The 3rd component is important factor in our proposed algorithm because its main focus is on blue color in an image. The other two remaining components are called “pure color”.

In the presence of “r” and “g”, the brightness of RGB color is reduced by the normalization to get the required result to point out the sky blue areas in an image.

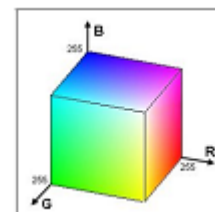


Figure.1: RGB Color Model [9]

3.4. HSI, HSV, HSL

To explain different color ranges with natural values, are based on the idea of artist's tint, saturation and tone. Hue represents the dominant color (such as red, green, purple and yellow) in the specific define area, saturation calculates for the colorfulness of an area in proportion to its

brightness in the different images [10]. The term lightness is related to the color luminance which is naturally presented in nature. The intuitiveness of the color spaces components and explicit discrimination between luminance and chrominance properties make these color spaces very effective in skin color segmentation [11,12, 13]. Here the same concept is used to detect the sky color in an image by targeting the specific pixels in an image.

The polar coordinate system of Hue-Saturation spaces, resulting in nature cycle of the color-space makes it difficult for parametric sky color models that need tight clustering of sky colors for best performance.

3.5. NATURE OF TINT, SATURATION AND LIGHTNESS

The standardized TSL Color space (Tint, Saturation, and Lightness) is a convolution of the standardized RGB into more likely close to hue and saturation and can be explained by the standard equations given below.

$$S = [9/5(r'^2 + g'^2)]^{1/2} \quad (4)$$

$$T = \begin{cases} \arctan(r' / g') / 2\pi + 1/4, & g' > 0 \\ \arctan(r' / g') / 2\pi + 3/4, & g' < 0 \\ 0, & g' = 0 \end{cases} \quad (5)$$

$$L = .299 R + .587 G + .114 B \quad (6)$$

let $r' = r - 1/3$, $g' = g - 1/3$ and r, g are taken from equation (1) and equation (2). Terril-lon et al. [14] have compared nine different color spaces for skin detection with a uni-model Gaussian joint pdf (only chrominance components of the color spaces were used). The proposed algorithm uses this concept for sky color detection with much improved results.

3.6. TECHNIQUE FOR RGB

YCrCb is most widely used technique now-a-days for an encoded nonlinear RGB color signal, which

is mostly used by international television networks worldwide for image compression in different applications. General Color is Classified by luminance (computed as weighted sum from RGB values), and two color difference values C_r and C_b that are formed by subtracting luminance from RGB red and blue components as shown in given equations.

$$Y = .299R + .587G + .114B \quad (7)$$

$$C_r = R - Y \quad (8)$$

$$C_b = B - Y \quad (9)$$

The transformation and separation of luminance and chrominance components makes this color space attractive for skin color modeling which were proofed in different state of art work such as presented by Phung et al. [15], Zarit et al. [11] and Menser et al. [16]. We used that concept in our algorithm for sky color modeling to separate the blue color from the rest of the colors in an image.

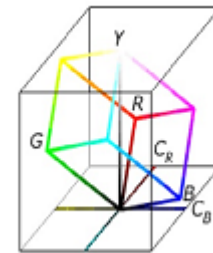


Figure.2: YCrCb Color Model relative to RGB Model

4. IDENTIFYING THE SKY BLUE AREA

After a number of experiments with the sky color images: the following generic rule was extracted:

1. If the red and green values are close to each other, and if the blue value is greater than the red and green values, the color is going to be in the blue range.
2. Identify the blue sky pixels by turning them off i.e. assign a 0.

More specifically, the final algorithm works on the following principle:

if (abs(R - G)<5 && abs(G - B)<5 && B > R
&& B>G && B>50 && B<230)

5. DATASET

Caltech dataset [18] is used for the evaluation of proposed algorithm for sky detection. The dataset contains complex images with cloudy and partially cloudy sky images. We also include some images containing different objects too to test the algorithm to separate the sky regions from the other objects in images.

6. RESULTS AND ANALYSIS

The figures 3 and 4 show the resultant images of some of the experiments. It is clear from the Figure 3 that the proposed algorithm has successfully detected the sky regions in the background of the aero planes. 1st column in Figure 3 shows the original images with clear, clouded and partially clouded sky which are accurately detected by the proposed algorithm.

The Figure 4 shows the experimental results on other types of images with partially clouded and clear sky. The proposed algorithm detected the sky regions efficiently in this case too.

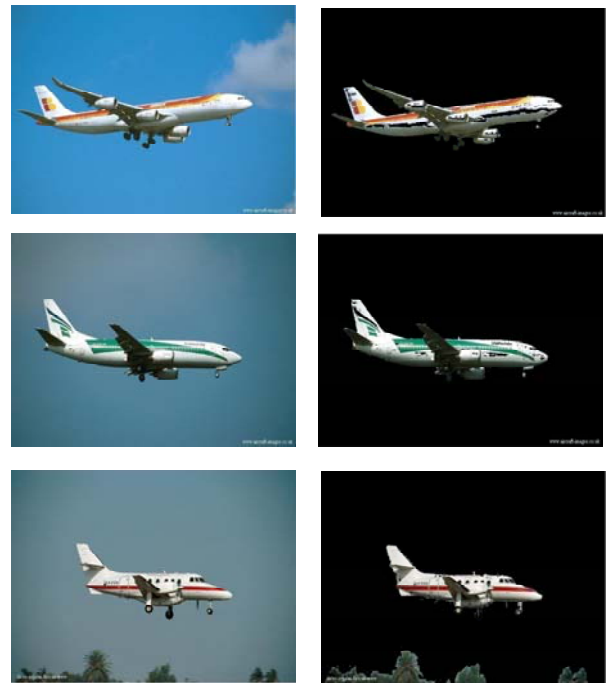


Figure.3: Experimental Results



Figure.4: Experimental results

7. CONCLUSION

Sky detection problem is very versatile in its applications ranging from entertainment (video/image quality enhancement) to weather forecasting and solar exposure prediction. A

robust algorithm is needed to fulfill the demands of its all applications. Based on experimental results on a number of images it can be concluded that RGB values plays an important role in identifying the sky regions in an image and leads to an accurate and robust approach that fulfill all the requirements of sky detection in demanding applications.

References

- [1] B. Zafarifar and P. H. N. de With, "Blue Sky Detection for Content based Television Picture Quality Enhancement," in Proc. IEEE International Conference on Consumer Electronics, Las Vegas, USA, 2007, pp. 437–438.
- [2] Schmitt F., Priese L., "Sky detection in CSC segmented color images". In Proceedings. Computer Vision Theory and Applications (VISAPP), (2009), vol. 2, pp. 101–106.
- [3] Nuchjira Laungrunthip, Alan Mckinnon, Clare Churcher, Keith Unsworth, "Sky Detection in Images for Solar Exposure Prediction", (2008), <http://hdl.handle.net/10182/855>
- [4] Green, B. (2002), "Canny edge detection tutorial", DOI=
http://www.pages.drexel.edu/weg22/can_tut.html
- [5] Fisher, R., Perkin, S., Walker, A. and Wolfart, E. , "Hypermedia Image Processing Reference, (1996).
- [6] Gallegher, A.C, JieboLuo, Wei Hao, "Improved Blue Sky Detection using Polynomial Model Fit", International Conference of Image Processing (ICIP), 2004, Vol. 4, pp.2367-2370.
- [7] Brand, J., Mason, J., "A comparative assessment of three approaches to pixel level human skin-detection". In Proceedings of the International Conference on Pattern Recognition, 200, vol. 1, pp. 1056 -1059.
- [8] Wark, T., Sridharan, S. "A syntactic approach to automatic lip feature extraction for speaker identification". In ICASSP,1998, pp. 3693–3696.
- [9] Vandevenne, L.. Lode's, "Comuter Graphics Tutorial: Light and Color". Retrieved from <http://student.kuleuven.be/~m0216922/CG/color.html>
- [10] POYNTON, C. A.. "Frequently asked questions about colour", (1995) <Inftp://www.inforamp.net/pub/users/poynton/doc/colour/ColorFAQ.ps.gz>
- [11] ZARIT, B. D., S UPER, B. J., AND QUEK, F. K. H., "Comparison of five color models in skin pixel classification". In ICCV'99 Int'l Workshop on recognition, analysis and tracking of faces and gestures in Real-Time systems, 1999, pp. 58-63.
- [12] Mckenna, S., G Oong, S., Raja, Y., "Modeling facial color and identity with Gaussian mixtures". Pattern Recognition, 1998, pp. 1883–1892.
- [13] Jordao, L., P Eerrone, M., C Osteira, J., Santos-Victor, J., "Active face and feature tracking". In Proceedings of the 10th Inter-national Conference on Image Analysis and Processing, pp. 572–577.
- [14] Shirazi, M.N, Fukamachi, H., Akamatsu, S., "Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for Automatic Detection of Human Face in Color Images", IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 54-61.
- [15] Phung, S. L., B Ouzerdoum, A., Chai, D., " A novel skin color model in ycbcr color space and its application to human face detection". In IEEE International Conference on Image Processing (ICIP'2002), vol. 1, 289–292.
- [16] MENSER, B., AND WIEN, M., "Segmentation and tracking of facial regions in color image sequences". In Proc. SPIE Visual Communications and Image Processing 2000, pp. 731–740.
- [17] Jao Schwartz, "Solmetric SunEye", last retrieved on feburay 2013, from: <http://resources.solmetric.com/get/Home-Power-SunEye-review.pdf>
- [18] <http://www.vision.caltech.edu/html-files/archive.html>

Real Time Network Server Monitoring using Smartphone with Dynamic Load Balancing

Dhuha Basheer Abdullah ¹, Zeena Abdulgafar Thanoon ²,

¹ Computer Science Department, Mosul University, Computer Sciences and Mathematics College
Mosul, Iraq

² Computer Science Department, Mosul University, Computer Sciences and Mathematics College
Mosul, Iraq

Abstract

The services provided by network servers are very important, therefore, the monitoring of these servers in real time is required to discover the obstacles in order to improve servers performance. In this paper, a file server system was designed to be monitored by a smartphone in real time through the use of dynamic scheduling algorithm for three main metrics: CPU usage, Memory usage and Free hard disk space. Dynamic load balancing (DLB) was implemented between servers depending on CPU usage of servers. Also, a dynamic checkpoint (DCP) technique was introduced to minimize files download time. The experiments show that the use of DLB reduced the CPU Usage for both servers and the download time for all clients. Also, the use of the proposed DCP technique is more efficient than the traditional static checkpoint technique.

Keywords: Real time, dynamic load balancing, checkpoint, smartphone.

1. Introduction

Applications of Real Time Network Monitoring Systems RTNMS provide a clear view for the metrics that are being monitored in the network, also give a flexible dealing with network problems in a proactive way.

Real time systems can be defined as those systems that consider the time as an important factor. Those systems are categorized depending on time constraints into Hard real time and Soft real time. RTNMS can be considered as a soft real time.

Network administrators are always the first defense line in diagnosing network faults, therefore it is very important to facilitate and support them by monitoring tools and software like smartphones for remote monitoring. One of the important objects to be monitored is the Network Servers [1]. With the rapid growth of both information and users, the need of effectively improve the quality of services provided by servers becomes an urgent problem to be addressed. This situation caused by heavy demand on

the server node which led to slowness in responding to requests. For these reasons load balancing algorithms should be implemented for better performance and to increase the ability to handle more users. So server farm is of wide interest to enterprises, which enables a group of independent servers to be managed as a single system for higher availability, easier manageability and greater scalability [2][3][4]. The main objective of load balancing is to minimize the response time and maximize throughput of the network

2. Contributions

In this work, a real time monitoring system was designed to monitor the performance of a network file server in a server farm environment supported by DLB between the servers. This system offers the following contributions:

- Implementing real time constraints to monitor the most important metric associated with the servers like CPU usage, Memory usage and Free hard disk space.
- Developing a new method associated with the checkpoint technique that gives the clients the ability to resume file download from the stopped point without the need to restart the download operation when changing from the crowded server to less loaded server.
- Developing a DLB algorithm to obtain better performance.

3. Related works

Previous works in this field:

- L. Yucheng, L. Yubin [5] designed a system using B/S mode which enables administrators to view the server-side situation in the performance testing and network maintenance.
- G. Kanagaraj, N. Shanmugasundaram, S. Prakash [6] suggested a method to dynamically load balance using

service queue depending on the summation of CPU utilization, memory utilization and network utilization of web servers.

- N. Bessho, T. Dohi [7] developed a stochastic model to evaluate the expected total recovery overhead for cluster computing system with three well-known checkpoint and rollback recovery schemes, one of which is for central file server checkpointing.

4. Real time system

A real-time system is a computing system which has timing constraints and the accuracy of such a system depends not only on its logical results but also on the time at which the results are available. Real-time systems can be classified as hard real time systems in which the consequences of missing a deadline can be catastrophic and soft real time systems in which the consequences are relatively tolerable. There are two types of Real time scheduling algorithms: fixed priority and dynamic priority. In fixed priority algorithm the priorities are assigned to each task before the activation of all tasks, an example of fixed priority algorithm is the Rate-Monotonic (RM) which assigns priorities to tasks on the basis of their period times. In dynamic priority algorithm the priorities are computed during the execution of the system, an example of dynamic priority algorithm is the Earliest-Deadline-First (EDF) which assign priorities to individual jobs on the basis of their absolute deadline; the shorter the deadline, the higher the priority. [8].

4.1. Real time network monitoring

Real time network monitoring applications are able to aggregate, quantify, and analyze network traffic for a clearer view. Thus, without network monitoring systems, it would be difficult to identify and resolve network problems. Real time network monitoring includes the monitoring of network performance in real time. It is the network administrator's job to ensure that their network and network applications are performing properly. However, advanced tools are needed to help network administrators to intensively and continuously monitor their network performance. Real time network monitoring should be accurate in detecting network disruptions and the cause of these disruptions [1].

5. Load balancing

The main objective behind load balancing is to distribute a set of requests among the different server nodes, to prevent some server nodes from being heavily loaded while others are lightly loaded [2][3]. Load balancing technology can

balance conflicting factors such as cost, performance, and scalability, through a relatively low total cost of the server farm to achieve a strong performance that cannot be achieved by stand-alone system. Load balancing algorithms generally can be classified as either static or dynamic [4]:

A- Static load balancing algorithm:

In this algorithm, load balancing decisions, are made at compile time and takes less time, which doesn't refer to the state of the servers, so it does not need to constantly monitor the nodes for performance statistics [6]. Such a hypothesis may not apply to a distributed environment. Because the static approach cannot respond to a dynamic runtime environment, it may lead to load imbalance on some nodes. However, static algorithms only work well when there is not much variation in the load on the workstations; In addition it is not satisfactory for parallel programs that are of the dynamic and/or unpredictable kind.

B- Dynamic load balancing algorithm (DLB):

DLB algorithm, make changes to the distribution of work among nodes at run-time; they use current or recent load information when making distribution decisions. Despite the higher runtime complexity, dynamic algorithms can potentially provide better performance than static algorithms [9]. DLB system mainly includes two processes: monitoring the load state of servers and assigning requests to the servers. In the dynamic approach, the load balancing decisions are based on the current state of the system; tasks are allowed to move dynamically from an overloaded node to an under-loaded node to receive faster service. This ability to react to change in the system is done through the use of checkpoint technique.

5.1. Checkpointing

One commonly used technique to recover from application failure is the use of checkpoints. During static checkpointing, an application writes its entire state to non-volatile secondary storage so that when the application is interrupted, it can resume its work from the last checkpoint rather than from the beginning. While writing and reading the checkpoint data is a type of overhead that consumes valuable system resources, the savings in rework times due to failure can often outweigh the cost of performing checkpointing in the first place.

One aspect of employing static checkpointing is properly assigning a checkpoint interval, i.e., the time from the beginning of one checkpoint to the beginning of the next. Given a set of jobs and failure parameters, it is possible to assign this interval in such a way that it maximizes application efficiency, i.e., the ratio of time the job spends making forward progress compared to the entire wall-clock time that includes checkpoint, restart, and rework

overload[10]. Here the proposed technique depends on DCP without the need to store fixed points for the downloaded files. In this technique, when an overload occurs on the main server, DCP will start by saving the current point (byte number of the file that is being downloaded) to complete downloading operation on the other server instead of restarting from the starting point.

6. System model

The proposed system was based on a model for a network file server used by the clients for downloading files. The infrastructure of the system has the hardware components shown in figure (1).

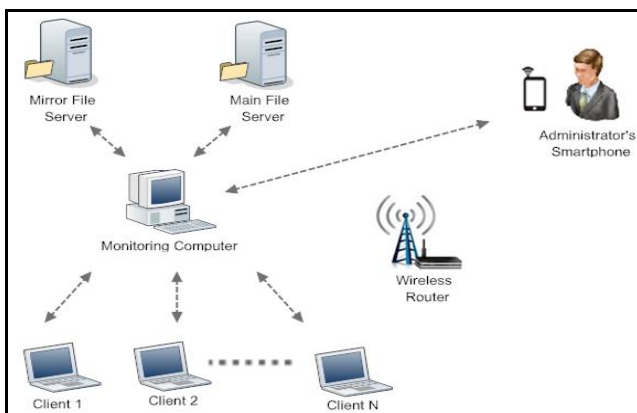


Fig. 1 System model

The monitoring computer was responsible for monitoring the performance of the two servers like CPU usage, Memory Usage and Free hard disk space and it saves the values of these metrics in Google drive. Clients communicate with the monitoring computer to download files; the monitoring computer will send the requests to the server that has less load on its CPU usage using DLB algorithm. The monitoring process will be done depending on the metric of real time constraints. The EDF algorithm was used to schedule the sending of performance metrics. When the administrator wants to use the smartphone for monitoring the metrics of two file servers, the request will be send from smartphone to the monitoring computer, and then the metrics will be displayed periodically in smartphone. If an urgent case occurs in the main server - like CPU usage exceeds the threshold value - the administrator will send an order to the monitoring computer to start re-load balancing between the two servers by implementing the new DCP technique. The connection between the network parts was done using WiFi IEEE 802.11 protocol. Figure (2) illustrate the general behavior of the system.

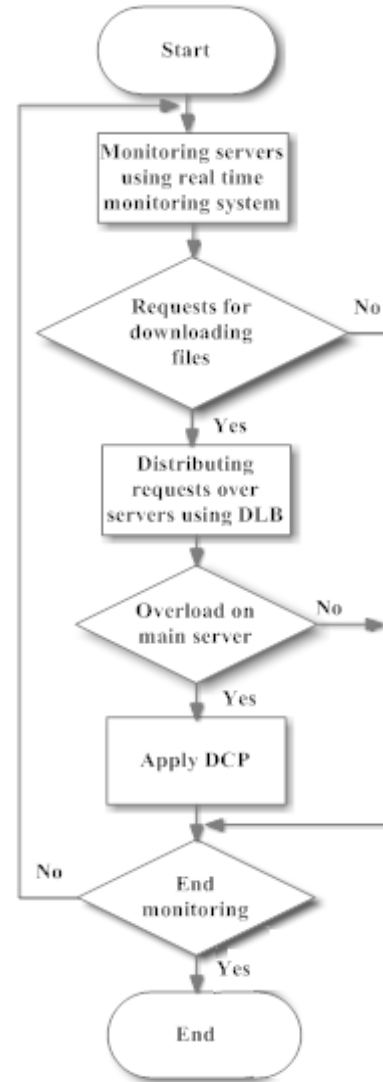


Fig. 2 System general behavior

6.1 Software description

Three algorithms were used in this system; all of them were implemented in the monitoring computer. These algorithms are:

A- Servers real time monitoring algorithm:
 Connections will be open between the monitoring computer and the two servers. Both servers will apply the EDF algorithm to periodically send the performance metrics to the monitoring computer depending on the priority configured by the administrator. The monitoring computer saves the values of those metrics in a text file and uploads them to the cloud drive to archive them for future works. When the administrator wants to use the smartphone for monitoring the metrics of two file servers, a request is sent to the monitoring computer, which will

send these metrics periodically in a real time base. In case of one or all metrics exceed the threshold values; a notification alarm will notify the administrator via the smartphone. Figure (3) shows the flowchart of this algorithm

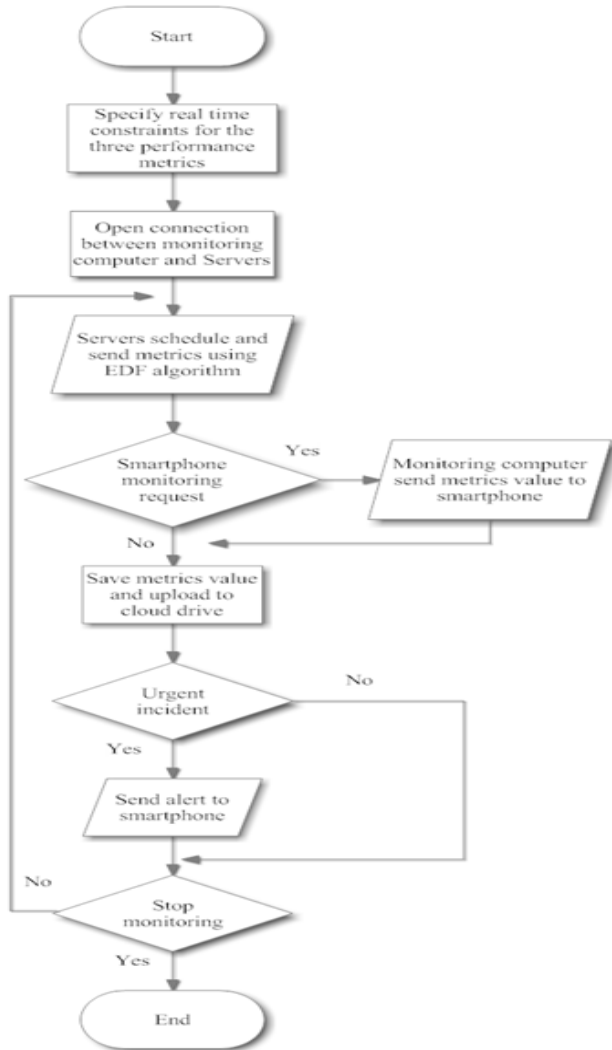


Fig. 3 Monitoring algorithm

B- Servers load balancing algorithm:

When the client wants to download a file, a request is sent to the monitoring computer. The monitoring computer will send a list of all existed files in the main file sever to the client; the client will choose the file from the list. Then the monitoring computer will specify the server with the lowest value of the CPU usage. After that the IP address of the client and the name of the selected file will be send to the specified server to connect with the client and start sending the file. Figure (4) shows the flowchart of this algorithm.



Fig. 4 Load balancing algorithm

C- Servers re-load balancing with DCP algorithm:

An alert will be send to the administrators' smartphone when the CPU usage of the main server exceeds the threshold value. Therefore, the administrator will send a command to the monitoring computer to start the implementation of re-load balancing algorithm. The monitoring computer will send a request to the main server to obtain information concerning the files currently under downloading. The monitoring computer will send an order to all clients to stop download and disconnect the connection with the main file server. The monitoring computer requests these clients to send the name of the file plus the number of the last received byte (that will be a checkpoint for each file) to the mirror server. The mirror server will connect the clients and complete the downloading process by starting from the checkpoints. Figure (5) shows the flowchart of this algorithm.

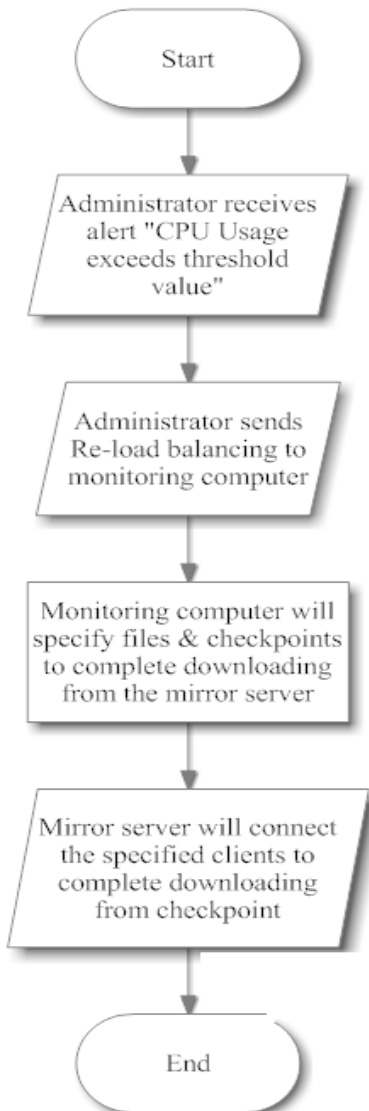


Fig. 5 Re-load balancing with DCP algorithm

7. Results and discussion

In this section, the experimental results for both of DLB and the proposed DCP technique will be explained. Assuming that; there are two servers, one computer acting as load balancer plus monitor, and three clients. Table (1) shows the hardware specification of both servers.

Table 1: Servers' hardware specifications

H/W	Main server	Second sever
CPU	Core i5 2.5 GHZ	Core i3 2.4 GHZ
RAM	4 GB	2 GB

A- DLB (Less CPU usage)

The experimental result presented in table (2) depends on two metrics, CPU Usage and Download Time.

Table 2: Experimental results

File size (MB)	Single server (Without DLB)		Two servers (With DLB)		
	CPU Usage (%)	Average download time (Second)	Serv.1 CPU Usage (%)	Serv. 2 CPU Usage (%)	Average download time (Second)
6	1.89	5.2	1.6	1.59	2
28	4.5	21	3.4	3.51	8.2
65	7.6	31	5.9	5.8	13

An average percentage was calculated for servers' CPU Usage, and also an average download time for clients' requests. The measurements were performed on two cases. The first does not use DLB; all the clients' requests are forwarded to the main server only. The second one uses DLB by distributing clients' requests on both servers. The experiment was performed by assuming that all the clients requested the same file at the same time. The experiment was repeated many times with different file size.

The results show that the use of DLB reduced the CPU Usage for both servers and the download time for all clients. Figure (6) shows the results of the average download time-in the case of the use of DLB and without using it.

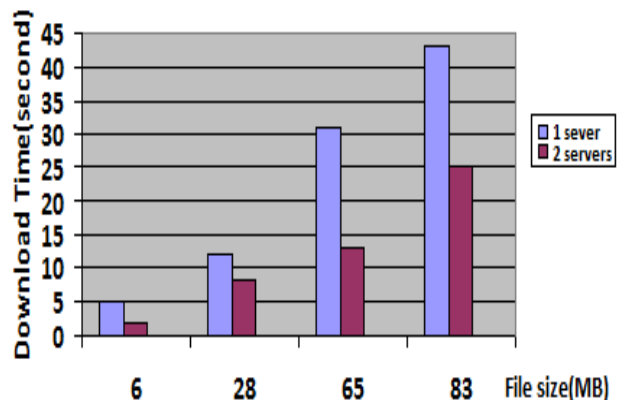


Fig. 6 Download time with/without DLB

B- DCP

The results proved that the use of the proposed DCP technique is more efficient than the traditional static checkpoint technique. In DCP technique, the time needed for switching between servers ranges between (0.5~1) second depending on the speed of the network.

Total download time using DCP technique was calculated depending on the following equation:

$$Total\ Time = T1 + ST + T2$$

Where:

T1: download time before switching

ST: switching time

T2: download time of the remaining size.

While in static checkpointing, assuming that there are three checkpoints, therefore the total download time was calculated depending on the following equation.

$$Total\ Time = T1 + ST + TC + T2$$

Where:

T1: download time before switching

ST: switching time

TC: needed time to go back to the last checkpoint before switching

T2: download time of the remaining size.

Figure (7) shows the results for downloading a specified file with 28MB using static checkpointing technique and the proposed DCP technique. Assuming that the switching time is one second and the interrupt or switching operation occurred in five different locations.

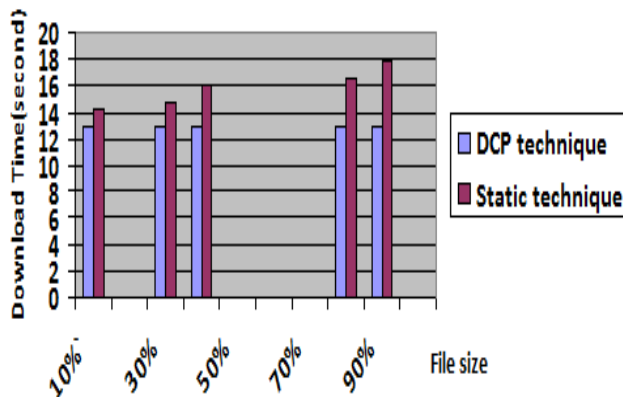


Fig 7. Comparing between DCP & static technique

8. Conclusions

The use of a real time scheduling algorithm in the proposed system maximizes the quality of servers' monitoring operation by giving the administrator the ability to specify the priorities for the monitored metrics. This priority was specified depending on real time constraints. The use of smartphone by the administrator gives flexibility in the job. The DLB proved high

efficiency in load balancing between servers because it takes into consideration the current state of the servers. Also, the use of the new DCP technique achieved a noticeable reduction in download time of files compared with traditional static checkpoint technique.

References

- [1] A. Manasrah, "Real Time Distributed Network Monitoring Platform (RTDNM)", Ph. D thesis, School of computer science, University Sains Malaysia, Malaysia 2009.
- [2] A. Meiappane, and V. Prasanna Venkatesan, et al, "The Pattern as a Reusable Component to Adaptive Framework for Load Balancing Mobile Agent in Internet Banking", IJCSCI, Vol. 9, Issue 2, No. 2, 2012, pp. 332-338.
- [3] Y. Jiao, and W. Wang, "Design and implementation of load balancing of distributed systems based web server", IEEE International Conference on High Performance Computing and Communications, 2010, pp. 337 – 342.
- [4] K. Kungumaraj, and T. Ravichandran, "An Efficient Load Balancing Algorithm for a Distributed Computer System", IJCTA, Vol. 2, No. 6, 2011, pp: 4012-4020.
- [5] L. Yucheng L., and L. Yubin, "A Monitoring System Design Program Based on B/S Mode", IEEE, Vol. 1, 2010, pp. 184-187.
- [6] G. Kanagaraj, N. Shanmugasundaram, and S. Prakash, "Adaptive Load Balancing Algorithm Using Service Queue", 2nd International Conference on Computer Science and Information Technology (ICCSIT) Singapore, 2012, pp. 143-146.
- [7] N. Bessho, and T. Dohi, "Comparing Checkpoint and Rollback Recovery Schemes in a Cluster System", Springer-Verlag Berlin Heidelberg, 2012, pp. 531-545.
- [8] W. S. Jane Liu, "Real-Time Systems", New Jercey, Prentice Hall, 2000.
- [9] S. Kumar, and N. Singhal, "A Study on the Assessment of Load Balancing Algorithms in Grid Based Network", International Journal of Soft Computing and Engineering (IJSC), Vol. 2, Issue: 1, 2012, pp. 402-405.
- [10] W. M. Jones, J. T. Daly., and N. DeBardeleben, "Impact of Sub-optimal Checkpoint Intervals on Application. Efficiency in Computational Clusters", ACM, USA, 2010, pp. 276-279.

¹ **Dhuha Basheer Abdullah** is the head of Computer Science Department, Computer Sciences and Mathematics College, Mosul University. She received her PhD degree in computer sciences in 2004 in the specialty of computer architecture and operating system. She supervised many Master degree students in operating system, computer architecture, dataflow machines, mobile computing, real time, and distributed databases. She has three PhD students in FPGA field, distributed real time systems, and Linux clustering. She also leads and teaches modules at both BSc, MSc, and PhD levels in computer science. Also, she teaches many subjects for PhD and MSc students.

² **Zeena Abdulgafar Thanoon** is an MSc. student in Computer Science Department, Computer Sciences and Mathematics College, Mosul University. She is interested in Computer Networks, Distributed Systems, and Operating System subjects.

Heuristics for Routing and Spiral Run-time Task Mapping in NoC-based Heterogeneous MPSOCs

Abbou El Hassen Benyamina¹, Mohammed kamel Benhaoua^{1,2} and Pierre Boulet²

¹ Department of Computer Science, Faculty of Sciences, University of Oran – Es Senia, Algeria
BP 1524, EL M'Naouer, Oran, Algeria

² University Lille 1, LIFL, CNRS, UMR 8022
F-59650 Villeneuve d'Ascq, France

Abstract

This paper describes a new Spiral Dynamic Task Mapping heuristic for mapping applications onto NoC-based Heterogeneous MPSoC. The heuristic proposed in this paper attempts to map the tasks of an applications that are most related to each other in spiral manner and to find the best possible path load that minimizes the communication overhead. In this context, we have realized a simulation environment for experimental evaluations to map applications with varying number of tasks onto an 8x8 NoC-based Heterogeneous MPSoCs platform, we demonstrate that the new mapping heuristics with the new modified dijkstra routing algorithm proposed are capable of reducing the total execution time and energy consumption of applications when compared to state-of-the-art run-time mapping heuristics reported in the literature.

Keywords: *MultiProcessor System on Chip (MPSoC), Network on Chip (NoC), Heterogeneous architectures, Dynamic mapping Heuristics, Routing Algorithm.*

1. Introduction

Multiprocessor Systems-on-Chip (MPSoCs) is a solution that implements a multiple processing elements (PEs) in the same chip. Advancement in nanometer technology enables that a Future MPSoCs contained thousands of PEs in a single chip by 2015[3], [14], [8]. MPSoCs are being increasingly used in complex embedded applications. The Network-On-Chip (NoC) has been introduced as a power-efficient, scalable inter-communication, interconnection mechanism between PEs [14], [8]. Mapping is important phase in architectural exploration in NoC based MPSOC. The application and architectural platform are represented by processing model, application task graph and architectural graph respectively. Considering the moment when task mapping is executed, approaches can be either static or dynamic. Static mapping defines task placement at design time, having a global view of the MPSoC resources. As it is executed at design time, it may use complex algorithms to better explore the

MPSoC resources, resulting in optimized solutions [24],[9],[16],[7],[12], [23],[19]. However, static mapping is not able to handle a dynamic workload, new tasks or applications loaded at run-time. To cope with this feature of actual MPSoCs, Dynamic (runtime) mapping techniques are required to map them onto the platform resources [4],[21],[13],[5],[6],[17],[15].

The main goal of this paper is to present a new Spiral Dynamic Task Mapping heuristic for run-time mapping applications. The presented heuristics are applied onto NoC-based Heterogeneous MPSoC platform. Two types of PEs are considered instruction set processors (ISPs) and Reconfigurable Areas (RA). Instructions set processors are used to execute software tasks and Reconfigurable Areas for hardware tasks. Heuristic also try to map the tasks of an application in a clustering region to reduce the communication overhead between the communicating tasks. The heuristic proposed in this paper attempts to map the tasks of an applications that are most related to each other in spiral manner and to find the best possible path load that minimizes the communication overhead with using a newly modified dijkstra routing algorithm proposed also in this paper. The new presented heuristic show significant performance improvements when compared to the latest run-time mapping heuristics reported in the literature. The performance metric includes execution time and energy consumption.

The rest of the paper is organized as follows. Section2 provides an overview of related work. Section 3 presents the MPSoC architecture. In Section 4, proposed mapping strategies along with the routing algorithm are presented. Experimental set-up and the results are presented in Section 5 with Section 6 concluding the paper and providing future directions.

2. Related Work

Mapping of tasks onto the MPSoC platform require finding the placement of tasks into the platform in view of

some optimization criteria like reducing energy consumption, reducing total execution time and optimizing occupancy of channels. If the MPSoC platform is heterogeneous, then a task binding process is required before finding the placement for a task. The binding process involves defining a platform resource for each task type like instruction set processors for software tasks and FPGA tiles for hardware tasks. Task mapping is accomplished by static (design-time) or dynamic (run-time) mapping techniques [3].

2.1 Static mapping techniques

The most existing work's in literature to solve the problem of mapping in the NOC platform are Static mapping. Static mapping defines task placement at design time, having a global view of the MPSoC resources and the tasks of applications (tasks graph). As it is executed at design time, it may use complex algorithms to better explore the MPSoC resources, resulting in optimized solutions. Heuristics like Genetic approach and exact methods like Tabu Search and stimulated annealing are presented in [24],[7], [12],[23],[19]. In [9], [16], energy-aware mapping algorithms are presented. These techniques find fixed placement of tasks at design-time with a well known computation and communication behavior. Related works classified for static mapping techniques are shown in Table 1. However, static mapping is not able to handle a dynamic workload, new tasks or applications loaded at runtime. To cope with this feature of actual MPSoCs, Dynamic (run-time) mapping techniques are required to map them onto the platform resources.

2.2 Dynamic mapping techniques

The challenge in the latest work's to solve the problem of mapping in the NoC-based heterogeneous MPSoCs are to present run-time mapping techniques for mapping application's tasks onto them. Wildermann et al. [22] evaluate the benefits of using a runtime mapping heuristic (communication and neighborhood cost functions), which allows decreasing the communication overhead.

Holzspies et al. [20] investigate another run-time spatial mapping technique, considering streaming applications mapped onto heterogeneous MPSoCs, aiming on reducing the energy consumption imposed by such application behaviors. Schranzhofer et al. [2] suggest a dynamic strategy based on pre-computed template mappings (defined at design time), which are used to define newly arriving tasks to the PEs at run-time. Carvalho et al. [11] evaluate pros and cons of using dynamic mapping heuristics (e.g. path load and best neighbor), when compared to static ones (e.g. simulated

annealing and Taboo search). Carvalho's approach was extended by Singh et al. [3], [1], employing a packing strategy, which minimizes the communication overhead in the same NoC-based MPSoC platform. Additionally, Singh's approach was improved to support multitask mapping onto the same PE. Different mapping heuristics were used to evaluate the performance. According to the Authors, the communication overhead of the whole system is reduced, decreasing the energy consumption. Faruque et al. [18] propose a decentralized agent-based mapping approach, targeting larger heterogeneous NoC-based MPSoCs (32x64 system is used as case study).

The most related works classified for Dynamic mapping techniques are shown in Table 2. Mapping heuristics Nearest Neighbor (NN) and Best Neighbor (BN) presented in Carvalho and Moraes [10] and two run-time mapping heuristics presented in Singh et al. [3] are taken for evaluation and performance comparison with our new proposed mapping heuristics.

Table 1: Related work classified for static mapping techniques

REF	Tasks Mapping	Communications Mapping	Tasks Scheduling	Communications Scheduling	Voltage selection link	Voltahe Selection tasks	Objectives	Guarantee hard deadlines
Y.Zhang and al 2002	List Scheduling	----	List Scheduling	----	----	ILP	Min Energy consumption	No
S.Kumar and al 2003	Genetic Algorithm	----	ASAP/ALAP	----	----	----	Max Performance time	No
R.Marcu Iescu and al 2004	List Scheduling	Deterministic	List Scheduling	Deterministic	----	----	-Max performance time -Min Energy consumption	Yes
D.Shin and al 2004	Genetic Algorithm	Genetic Algorithm	List Scheduling	----	List Scheduling	----	Min Energy consumption	Yes
M.Armi n and al 2007	List Priority (TPL,PPL)	----	----	----	----	----	Min Energy consumption	No
C.Chou and al 2009	Heuristic	Heuristic	----	Deterministic	----	----	Max Performance time	Yes
E.Vardi and al 2009	List Priority (TPL1,TPL2 ,ATPL,PPL)	----	----	----	----	----	- Max performance time -Min Energy consumption	No
X.Wang and al 2009	Heuristic	Deterministic	----	----	----	----	Min Energy consumption	Yes

3. Heterogeneous MPSoC Architecture

MPSoC architecture used in this work contains a set of different processing elements which interact via a communication network [14]. Software tasks execute in

instruction set processors (ISPs) and hardware tasks execute in reconfigurable logics (reconfigurable area-RA) or in dedicated IPs.

Authors	Mono / Multi Tasks	Type of Architecture	Type of control	Optimizations Goals
Smit 2005	Mono	heterogeneous	Centralised	Energie Consumption
Ngouanga 2006	Mono	homogeneous	Centralised	Communication volume, Computational Load
Hölezspies 2007/2008	Mono	heterogeneous	Centralised	Energie Consumption
Chou 2007/2008	Mono	homogeneous	Centralised	Energie Consumption
Al Faruque 2008	Mono	heterogeneous	Distributed	Execution time, Mapping time
Mehran 2008	Mono	homogeneous	Centralised	Mapping time, Energie Consumption
Wildermann 2009	Mono	homogeneous	Centralised	communication, latency Energie Consumption
Schranzhofer 2010	Mono	homogeneous	Centralised	Energie Consumption
Carvalho 2010	Mono	heterogeneous	Centralised	Communication volume, Network contention
Singh 2009, 2010	Multi	heterogeneous	Centralised	Energie Consumption, Communication volume, Network contention
Marcelo 2011	Multi	homogeneous	Centralised	Tasks interdependency evaluation, energy consumption
Proposed Work	Mono	heterogeneous	Centralised	Communication volume, Network contention, Execution time, energie consumption

Table 2: Related work classified for Dynamic mapping techniques

One of the processing node is used as the Manager Processor (M) that is responsible for task scheduling, task binding, task placement (mapping), communication routing, resource control and reconfiguration control. The M knows only the initial tasks of the applications. The initial task of each application is started by the M and new communicating tasks are loaded into the MPSoC platform at run-time from the task memory when a communication to them is required and they are not already mapped.

4. Proposed Approach

This section describes our proposed approach. Firstly we describe our dynamic spiral task mapping. Secondly we describe the modified dijkstra routing algorithm. First we introduce some definitions for proper understanding of the proposed approach.

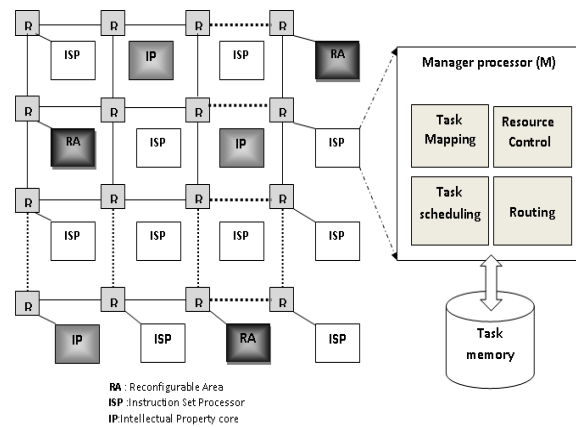


Fig. 1: Conceptual MPSOC architecture

4.1 Definitions:

Definition 1:

An application task graph is represented as an acyclic directed graph $TG = (T, E)$, where T is set of all tasks of an application and E is the set of all edges in the application. Figure 2 (a) describes an application having initial, software and hardware tasks along with the edges (E) connecting these tasks and (b) shows the master-slave pair (communicating tasks). The starting task of an application is the initial task that has no master. E contains all the pair of communicating tasks and is represented as $(mtid, stid, (Vms, Vsm))$, where $mtid$ represents the master task identifier, $stid$ represents the slave task identifier; Vms is the data volumes from master to slave; Vsm is the data volumes from slave to master.

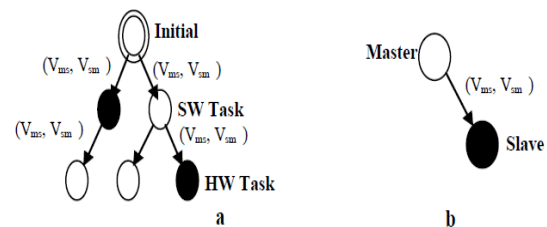


Fig. 2: Application modeling and Master-Slave

Definition 2:

A NoC-based heterogeneous MPSoC architecture is a directed graph $AG = (P, V)$, where P is the set of tiles pi and vi,j presents the physical channel between two tiles pi and pj . A tile pi consists of a router, a network interface, a heterogeneous processing element, local memory and a cache.

Definition 3:

The application mapping is represented by $mpng : ti \rightarrow |pi$ to map the tasks of the application onto the NoC-based heterogeneous MPSoC.

4.2 Reference dynamic mappings heuristics

1) *The First Free (FF) heuristic:* Simply selects the next compatible processor to map a given task, thus walking sequentially through all processors before considering an processor again.

2) *Minimum Maximum Channel load (MMC) heuristic:* Considers all possible mappings for a given task and chooses the one that increases the least the peak load of a channel of the NoC.

3) *Minimum Average Channel load (MAC) heuristic:* Considers all possible mappings for a given task and chooses the one that increases the least the average load of the channels of the NoC

4) *The Nearest Neighbor (NN) heuristic:* Considers only the proximity of an available resource to execute a given task. NN starts searching for a free PE able to execute the target task near the source task. The search tests all n-hop neighbors, n varying between 1 and the NoC limits

5) *The Path Load (PL) heuristic :* Computes the load in each channel used in the communication path. PL computes the cost of the communication path between the source task and each one of the available resources. The selected mapping is the one with minimum cost.

6) *The Best Neighbor (BN) heuristic:* Combines NN search strategy with the PL computation approach. The search method of BN is similar to NN.

4.3 Proposed spiral heuristic based on our Modified Dijkstra routing Algorithm

1) *Spiral heuristic:* To Map the applications, firstly the initials tasks of applications are placed in distributive way the farest possible between them in a middle of the clusters, using a strategy of clusters like shown in Fig. 3. This permit the same tasks of application could be placed in a same region near between them, which reduces the communications costs. The frontiers of clusters are virtual and the common regions could be shared by the tasks of different applications.

After that the initials tasks of each application are placing communicative tasks ask to be placed. To place required task, the master processor (M) try to place it around the processor which has executed the appealed task going from a distance equal 1 (hop) until the limit of NoC. The ressource (the processor according to type of the task) is researched on spiral manner according to sequencement 1, 2, 3, 4, 5, 6, 7, 8 like shown in the figure Fig .3. Explores

spiral neighbors and performs the mapping this prevents the calculation of all possible solutions mapping, as in the PL and calculation of the best neighbor as in the BN heuristic, reducing the execution time for mapping.

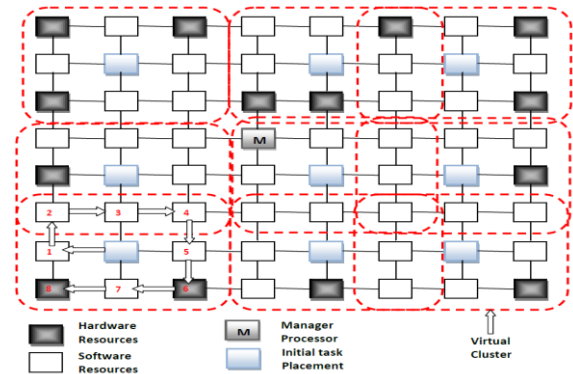


Fig. 3: Initial tasks placement for mapping applications with Spiral packing strategy.

2) *Modified dijkstra routing algorithm:* After mapping the communicating tasks, we need a communication mapping between them. Our new proposed communication mapping tries to search a best path with a high bandwidth. The proposed heuristic reduce the computational time and energy consumption.

Algorithm 1 SPIRAL HEURISTIC

Entrées: NoFreeResources, CurrentProcessor, PE, FreeResources, HopDistance, NocLimit **Sortie:** FreeElement

```

1: HopDistance ← 0
2: HopX ← 0
3: HopY ← 0
4: while (FreeElement = NULL) AND HopX != NocLimit
   AND HopY != NocLimit do
5:   if (CurrentProcessor.left.isFree()) then
6:     FreeElement ← CurrentProcessor.left;
7:   else
8:     HopX- -;
9:   end if
10:  if CurrentProcessor.top.isFree() then
11:    FreeElement ← CurrentProcessor.top;
12:  else
13:    HopY- -;
14:  end if
15:  if CurrentProcessor.right.isFree() then
16:    FreeElement ← CurrentProcessor.right;
17:  else
18:    HopX+ +;
19:  end if
20:  if CurrentProcessor.bottom.isFree() then
21:    FreeElement ← CurrentProcessor.bottom;
22:  else
23:    HopY+ +;
24:  end if
25: end while
    
```

Algorithm 2 MODIFIED DIJKSTRA ROUTING ALGORITHM

```

Entrées: IdSource , IdDestination Sortie: BestTraject
    /*While we did not reach the destination */
2: while (stop=false) do
    /*Got back the neighbor who has the least used link*/
4:   min ← minWeight();
    if min <> idDest then
6:     /*Add the link in the list of the path*/
       Listpath.add(min)
8:   else
       stop ← true
10:  end if
    end while
12: BestTraject ← Listpath
    
```

5. Experimental Set-Up and the Results

For the experimentation we have used a language of programming of high level which the JAVA language.

5.1 Experimental set-up

We have realize a heterogeneous platform simulated which comprises 64 processors of which 14 hardware, 49 software and a processor which is used like processor manager which is responsible of placement of the applications tasks, the configuration and update the platform, the communications routing. This platform uses a network on chip for the communication. We have used XML file for describing graphs of tasks used which are the same used in work of A.K.Singh, tasks (initial,software and hardware). The time processing of tasks depends on the specificity and the capacity of processor. we have fixed parameter software processors needs 40 cycles for an instruction, however hardware processors is fast and needs 20 cycles for one instruction. In a reverse of the consumption of energy or the processors hardware consume more than the processors software that we have fixed to 20 and 10 respectively. The shape of tasks is fixed to a number of instructions. The shape of data changed is 100 parquets. The used scenario is a number of one, three, seven and ten (10) applications witch proceeds between 7 to 9 tasks.

The platform is divided to nine (9) clusters which permits to launch nine (9) applications en parallel beyond this number the others applications which requires to be placed, have to waited in queue. For the placement of the tasks of applications we have implemented our proposed Dynamic task mapping based on Spiral packing strategy and Modified Dijkstra routing method. The spiral method tries to map the tasks in close manner with minimum exploration of the NoC space. The implementation of our method Modified Dijkstra routing which minimize the time processing and energy consumption of the system. For a

comparison study we have implemented the NN and BN dynamic mapping heuristics.

5.2 Experimental results

We have executed the implemented dynamic heuristics the Nearest Neighbor (NN) and the Best Neighbor (BN) for the placement of one, three, seven and ten applications in parallel on the platform simulated of 64 processors whose 14 hardware, 49 software and one for the processor manager (M). For the same we have executed our proposed Heuristics for Routing and Spiral Run-time task Mapping. For the measurements of performances we have calculated the execution time and the energy consumption. The Fig.4 show the optimization brought by our approach in terms of execution time awards the use of the proposed approach. The Fig.5 show the optimization brought by our approach in terms of energy consumption towards the use of the proposed approach.

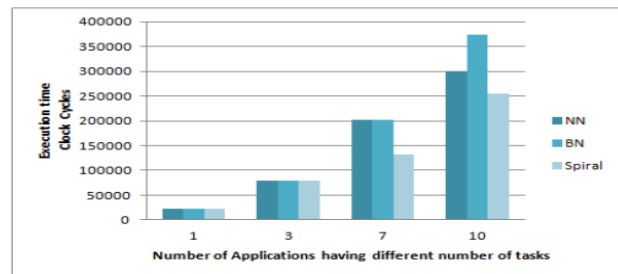


Fig. 4: Execution Time comparison of proposed approach with NN and BN respectively

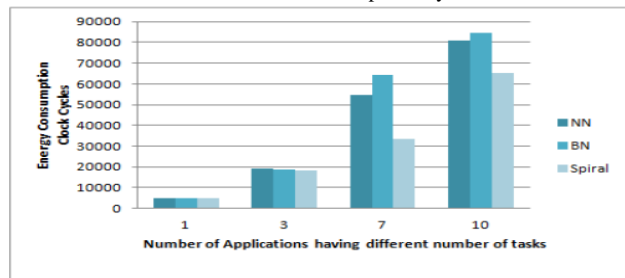


Fig. 5: Energy Consumption comparison of proposed approach with NN and BN respectively

6. Conclusion and Future Directions

A new dynamic task mapping heuristic that try to map the application tasks in close manner to reduce the cost of communications is proposed. To reduce more optimal the cost of the communications a Modified Dijkstra routing is proposed also in this paper. Through the environment of simulation that we have realized the comparative study between the nearest neighbor (NN), the best neighbor (BN) and the proposed dynamic tasks mapping heuristic based on the spiral packing strategy and Modified Dijkstra routing Algorithm. We have showed the offered

optimization by our approach. In our future research works the right challenge is trying to propose others strategies of research other than the one used in the latest works and Proposition of other heuristics of dynamic mapping.

References

[1] A.K.Singh and al. Efficient heuristics for minimizing communication overhead in NoC-based heterogeneous MPSoC platforms. In *Rapid System Prototyping, 2009. RSP '09. IEEE/IFIP International Symposium on*, 2009.

[2] A.Schranzhofer, C.Jian-Jia, L.Santinelli, and L.Thiele. Dynamic and adaptive allocation of applications on MPSoC platforms. In *Design Automation Conference (ASP-DAC), 2010 15th Asia and South Pacific*, 2010.

[3] A.Singh, T.Srikanthan, A.Kumar, and W.Jigang. Communication-aware heuristics for run-time task mapping on NoC-based MPSoC platforms. *Journal of Systems Architecture*, 56 , Issue: 7:242–255, 2010.

[4] Carvalho and al. Evaluation of static and dynamic task mapping algorithms in NoC-based MPSoCs. In *2009 12th Euromicro Conference on Digital System Design, Architectures, Methods and Tools*, 2009.

[5] CL.Chou. Incremental run-time application mapping for homogeneous NoCs with multiple voltage levels. In *Hardware/Software Codesign and System Synthesis (CODES+ISSS), 2007 5th IEEE/ACM/IFIP International Conference on*, 2007.

[6] CL.Chou. User-aware dynamic task allocation in networks-on-chip. In *Design, Automation and Test in Europe, 2008. DATE '08*, 2008.

[7] CL.Chou, C.Wu, and J.Lee. Integrated mapping and scheduling for circuit-switched network-on-chip architectures. In *Electronic Design, Test and Applications, 2008. DELTA 2008. 4th IEEE International Symposium on*, 2008.

[8] D.Bertozzi and L.Benini. A network-on-chip architecture for gigascale systems-on-chip. *Circuits and Systems Magazine, IEEE*, 4 , Issue: 2:18–31, 2004.

[9] D.Shin and J.Kim. Power-aware communication optimization for networks-on-chips with voltage scalable links. In *Hardware/Software Codesign and System Synthesis, 2004. CODES + ISSS 2004. International Conference on*, 2004.

[10] E.Carvalho and F.Moraes. Congestion-aware task mapping in heterogeneous MPSoCs. In *System-on-Chip, 2008. SOC 2008. International Symposium on*, 2008.

[11] E.Carvalho, N.Calazans, and F.Moraes. Dynamic task mapping for MPSoCs. *IEEE Design Test of Computers*, 27 , Issue: 5:26–35, 2010.

[12] F.Vardi, S.Saeidi, and A.Khademzadeh. Crinkle: A heuristic mapping algorithm for network on chip. *IEICE Electronics Express*, 6 , Issue: 24:1737–1744, 2009.

[13] Holzspies. Mapping streaming applications on a reconfigurable MPSoC platform at run-time. In *System-on-Chip, 2007 International Symposium on*, 2007.

[14] L.Benini and G. D. Mecheli. Networks on chips: a new SoC paradigm. *Computer*, 35 , Issue: 1:70–78, 2002.

[15] Marcelo. Multi-task dynamic mapping onto NoC-based MPSoCs. In *SBCCI '11 Proceedings of the 24th symposium on Integrated circuits and systems design*, 2011.

[16] M.Armin, S.Saeidi, and A.Khademzadeh. Spiral : A heuristic mapping algorithm for network on chip. *IEICE Electronic Express*, 4 , Issue: 15:478–484, 2007.

[17] Mehran. Dsm: A heuristic dynamic spiral mapping algorithm for network on chip. *IEICE Electronics*, 5 , Issue: 13:5–13, 2008.

[18] M.Faruque, R.Krist, and J.Henkel. Adam: Run-time agent-based distributed application mapping for on-chip communication. In *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*, 2008.

[19] P.Ghosh, A.Sen, and A.Hall. Energy efficient application mapping to NoC processing elements operating at multiple voltage levels. In *Networks-on-Chip, 2009. NoCS 2009. 3rd ACM/IEEE International Symposium on*, 2009.

[20] P.Holzspies, J.Hurink, J.Kuper, and G.Smit. Run-time spatial mapping of streaming applications to a heterogeneous multi-processor system-on chip (MPSoC). In *Design, Automation and Test in Europe, 2008. DATE '08*, 2008.

[21] Smit and al. Run-time mapping of applications to a heterogeneous SoC. In *Design, Automation and Test in Europe Conference and Exhibition, 2004. Proceedings*, 2004.

[22] S.Wildermann, T.Ziermann, and J.Teichet. Run time mapping of adaptive applications onto homogeneous NoC-based reconfigurable architectures. In *Field-Programmable Technology, 2009. FPT 2009. International Conference on*, 2009.

[23] X.Wang, M.Yang, Y.Jiang, and P.Liu. Power-aware mapping for network-on-chip architectures under bandwidth and latency constraints. *ACM Transactions on Architecture and Code Optimization (TACO)*, 7:1, 2010.

[24] Y.Zhang and al. Task scheduling and voltage selection for energy minimization. In *Design Automation Conference, 2002. Proceedings.39th*, 2002.

Benhaoua Mohammed Kamel : Received his Magister degree in computer science, Oran University, Algeria, in 2009. Currently, he is student towards the completion of his PhD. His research interests include NoC-based MPSoC design, run-time mapping algorithms.

Benyamina Abbou el hassen : received his Ph.D. degree in Computer Science in 2008 from University of Oran (Algeria), He is assistant professor at the university (Algeria). His research works include parallel processing, optimization, design space exploration and Model Driven Engineering with the special focus on real-time and embedded systems.

Pierre Boulet : received his Ph.D. degree in Computer Science in 1996 from University of Lyon, He was assistant professor at the university Lille 1 from September 1998 to August 2002, and then became a researcher at INRIA Futurs from September 2002 to August 2003. Since September 2003, he is a full professor at the university Lille 1. His research works include parallel processing, optimization, design space exploration and Model Driven Engineering with the special focus on real-time and embedded systems.

Broad View of Cryptographic Hash Functions

Mohammad A. AlAhmad¹, Imad Fakhri Alshaikhli²

¹Department of Computer Science, International Islamic University of Malaysia, 53100 Jalan Gombak Kuala Lumpur, Malaysia,

²Department of Computer Science, International Islamic University of Malaysia, 53100 Jalan Gombak Kuala Lumpur, Malaysia

Abstract

Cryptographic hash function is a function that takes an arbitrary length as an input and produces a fixed size of an output. The viability of using cryptographic hash function is to verify data integrity and sender identity or source of information. This paper provides a detailed overview of cryptographic hash functions. It includes the properties, classification, constructions, attacks, applications and an overview of a selected dedicated cryptographic hash functions.

Keywords-cryptographic hash function, construction, attack, classification, SHA-1, SHA-2, SHA-3.

1. INTRODUCTION

A cryptographic hash function H is an algorithm that takes an arbitrary length of message as an input $\{0, 1\}^*$ and produce a fixed length of an output called message digest $\{0, 1\}^n$ (sometimes called an *imprint*, *digital fingerprint*, *hash code*, *hash result*, *hash value*, or *simply hash*). Cryptographic hash functions play a fundamental role in modern cryptography practical applications, for example, digital signature [1,2], digital timestamp [3], message authentication code (or MAC) [4], public key encryption [5], tamper detection of files and many others. This versatility earned them the nickname “Swiss army knife of cryptography”.

2. CLASSIFICATION, PROPERTIES, CONSTRUCTIONS AND ATTACKS OF HASH FUNCTIONS

A. Classification of Hash Functions

Cryptographic hash functions can be classified as unkeyed hash functions and keyed hash functions as Figure 1 shown. Unkeyed hash functions that accepts a variable length message as a single input and produce a fixed hash digest, $H: \{0,1\}^* \rightarrow \{0,1\}^n$. It is also known as modification detection codes (MDCs). Where, keyed hash functions accept a variable length message and a fixed length secret key as two inputs to the hash function design to produce a fixed length hash digest, $H_k: \{0,1\}^k \times \{0,1\}^* \rightarrow \{0,1\}^n$. It is also known as message authentication codes (MACs). Unkeyed hash function is further classified into one-way hash function (OWHF), collision resistant hash function (CRHF), universal one way hash function (UOWHF) [2].

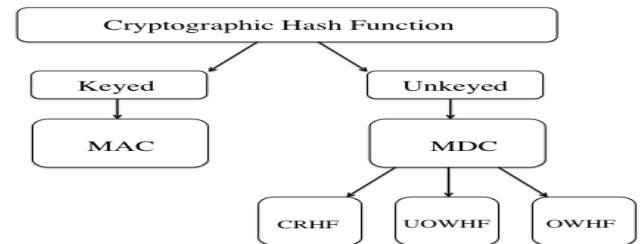


Figure 1. Classification of cryptographic hash function

CRHF usually deals with longer length hash values. An unkeyed hash function is a function for a fixed positive integer n which has, as a minimum, the following two properties:

- 1) Compression: h maps an input x of arbitrary finite bit length, to an output $h(x)$ of fixed bit length n .
- 2) Ease of computation: given h and an input x , $h(x)$ is easy to compute.

Modification Detection Codes are classified as follows:

- 1) One-way hash function is a hash function where finding an input which hashes to a pre-specified hash digest is difficult.
- 2) A collision resistant hash function is a hash function where finding any two inputs having the same hash digest is difficult.
- 3) In a universal one-way hash function, for randomly chosen input x , key k and the function H_k , it is hard to find $y = x$ such that $H_k(x) = H_k(y)$ [2].

A keyed hash function is a function whose specific purpose is called message authentication code (MAC) algorithms as shown in Figure 1. Keyed hash functions should satisfy the following two properties:

- 1) Compression: H_k maps an input x of arbitrary finite bit length, to an output $H_k(x)$ of fixed bit length n .
- 2) Ease of computation: for a known function H_k , given a value k and an input x , $H_k(x)$ is easy to compute. The result is called MAC value [2].

Apart from the classification of keyed and unkeyed hash functions, they can be classified into other ways such as hash function based on block ciphers, hash function based on modular arithmetic and dedicated hash functions.

B. Properties of Hash Functions

Hash functions play a major role in application security today. Hash functions provide different security properties depend on the

security requirements of the application. The basic security properties of hash functions are preimage resistance, second preimage resistance and collision resistance. They are explained below:

1. Preimage resistance: for any given code h , it is computationally infeasible to find x such that $H(x) = h$.
2. Second preimage resistance: for any given input m , it is computationally infeasible to find $y \neq m$ with $H(y) = H(m)$.
3. Collision resistance: it is computationally infeasible to find any pair (m, y) such that $H(m) = H(y)$ [2].

Figure 2 illustrates the definitions of hash function security properties.

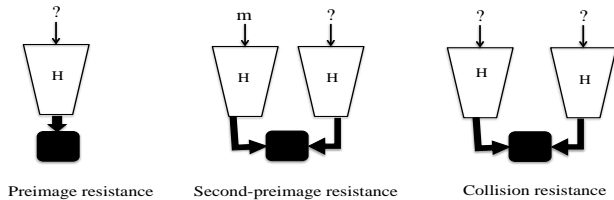


Figure 2. Hash function security properties

The preimage resistance property can be described as the inability to learn about the contents of the input data from its hash digest. The second preimage resistance property can be expressed as the inability to learn about the content of the second preimage from the given first preimage such that both of these preimages have same hash digest. The collision resistance property can be interpreted when two different and separate contents of inputs yield to the same hash digest.

C. Classification of Hash Functions

Building hash functions can be achieved by using various constructions such as Merkle-Damgård or sponge constructions. Merkle-Damgård construction was introduced by R. Merkle's PhD in 1979. It represents a guidance of building dedicated hash functions from compression functions. In 2007, sponge construction introduced in SHA-3 competition by Guido Bertoni, Joan Daemen, MichealPeeter and Gilles Van Assche. It represents the compression function of the new SHA-3 standard (Keccak algorithm).

- The Merkle-Damgård Construction

Merkle-Damgård construction was described by R. Merkle in his Ph.D. thesis [6] in 1979. As Figure 3 shows Merkle-Damgård (MD) construction that iterates sequentially a chaining transformation that takes as input message block and the previous chaining value. The input string x is divided into t equal-sized fixed-length blocks x_i with bit-length r . Bit-length r corresponds to input length of desired compression function f . The algorithm steps of Merkle-Damgård construction as follows:

- 1) Break the input x into blocks x_1, x_2, \dots, x_t .
- 2) Pad the last block x_t with 0-bits if necessary to obtain the multiple length of r .
- 3) Create the length block x_{t+1} with bit length r to hold the right justified binary representation of overall bit-length of x (MD strengthen).

- 4) Inputting x_1, x_2, \dots, x_t to the compression function (iterated processing) to produce an intermediate value of H_i .
- 5) H_i serves as feedback value to f and is processed with x_{i+1} in the next iteration. This implies the need of an initial value (IV) H_0 for the first iteration that is often provided pre-defined with bit-length r .
- 6) After processing all the input blocks, then, function g transforms the preliminary result H_{t+1} of bit-length r to the final hash-value with desired bit-length. Function g is often the identity mapping [6].

The most distinctive and special part of Merkle-Damgård construction is that the problem of designing a collision-resistant hash function reduced to designing a collision-resistant compression function. This means, if the compression function is collision resistant, then, the hash function is collision resistant. So, the properties of the compression function will be transformed to the hash function.

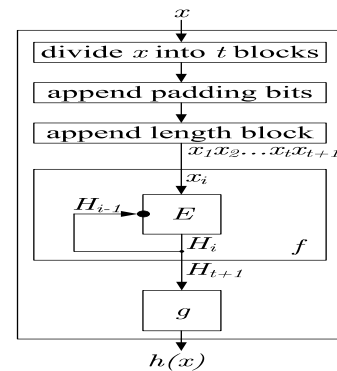


Figure 3. Detailed View of Merkle-Damgård Construction

The well-known Merkle-Damgård construction [6] has determined the basic structure of iterated hash functions. Merkle-Damgård iterates sequentially a chaining of input message blocks and the previous chaining value to produce the final hash digest $h(x)$. Figure 4 shows the Merkle-Damgård strengthen overall design. Padding is an algorithm to extend the input length to become a multiple length of r . Padding is obtained by appending a single '1' bit and '0' bits as many as needed to reach length r . This approach called Merkle-Damgård strengthen or length padding which makes the construction secure.

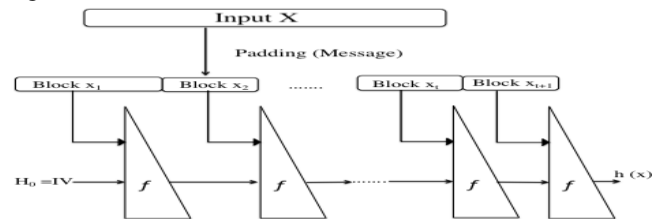


Figure 4. Merkle-Damgård strengthening

The Merkle-Damgård construction used in designing popular hash functions such as MD5, SHA-1 and SHA-2. Also, the Merkle-Damgård construction is quite well studied and several weaknesses (generic attacks) such as multi-collisions [7], long-message second preimage and differentiability [8] have been shown for this construction. Due to the structural weakness founded from these attacks, two intermediate Merkle-Damgård construction versions were developed; wide pipe hash construction and fast wide pipe construction.

- Wide Pipe Hash Construction

Stefanucks [9] introduced the wide pipe hash construction as an intermediate version of Merkle-Damgård to improve the structural weaknesses of Merkle-Damgård design. Figure 5 shows the wide pipe hash construction. The process is similar to Merkle-Damgård algorithm steps except of having a larger internal state size, which means the final hash digest is smaller than the internal state size of bit length.

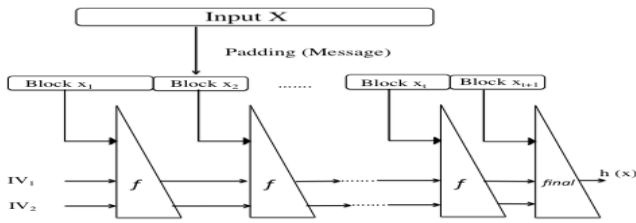


Figure 5. The wide pipe hash construction

Also, the final compression function compresses the internal state length (for ex, 2n-bit) to output a hash digest of n-bit. This simply can be achieved by discarding the last half of 2n-bit output.

• Fast Wide Pipe Construction

Mridul Nandi and Souradyuti Paul proposed the fast wide pipe construction. It is twice faster than the wide pipe construction. Figure 6 shows the fast wide pipe construction. As the Figure shows, the input (IVs) for each compression function is divided into halves.

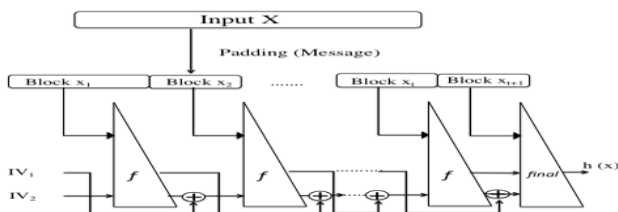


Figure 6. The fast wide pipe hash construction

The first half is inputted in the compression function and the other half is XORed with the output for the same compression function. The feed-forward process makes the overall design faster. Hence, faster process is obtained. The final output of the hash digest can be truncated to the desired digest length using the final compression function.

• The Sponge Construction

Sponge construction is an iterative construction designed by Guido Bertoni, Joan Daemen, MichealPeeter and Gilles Van Assche to replace Merkle-Damgård construction. It is a construction that maps a variable length input to a variable length output. Namely, by using a fixed-length transformation (or permutation) f that operates on a fixed number of $b = r + c$ bits. Where r is called the bitrate and c is called the capacity as Figure 7 shown. First, the input is padded with padding algorithm and cut into blocks of r bits. Then, the b bits of the state are initialized to zero [9]. The sponge construction shown in Figure 7 operates in two phases:

- i. **Absorbing phase:** The r -bit message blocks are XORed with the first r bits of the state of the function F . After processing all the message blocks, the squeezing phase starts.
- ii. **Squeezing phase:** The first r bits of the state are returned as output blocks of the function F . lastly, the number of output blocks is chosen by the user [9].

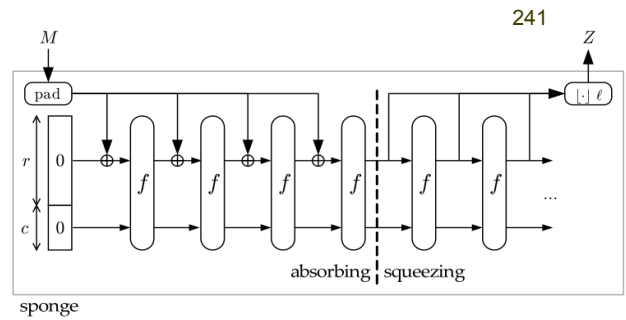


Figure 7. The sponge construction [9]

The sponge construction has been studied by many researchers to prove its security robustness. Bertoni et al. [10] proved that the success probability of any generic attack to a sponge function is upper bound by its success probability for a random oracle plus $N^2/2^{e-1}$ with N the number of queries to f . Aumasson and Meier [11] showed the existence of zero-sum distinguishers for 16 rounds of the underlying permutation f of Keccak hash function. Boura, Canteaut and De Cannière [12] showed the existence of zero-sums on the full permutation (24 rounds).

D. Attacks on Hash Functions

Attacks on hash functions are a technical strategy(s) an adversary may use to defeat the objectives of a hash function. These technical strategies may vary and in many cases attacks applied to the compression function of a hash function. A high-level classification of attacks on hash functions is shown in Figure 8.

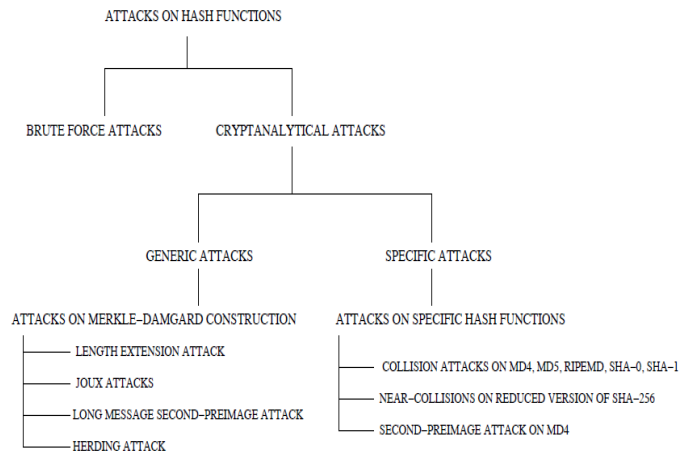


Figure 8. Attacks on hash functions

Attacks on hash functions are mainly classified into two categories: brute force attacks and cryptanalytical attacks.

• Brute force Attacks

Brute force attacks are a particular strategy used to try randomly computed hashes to obtain a specific hash digest. Hence, these attacks do not depend on the structure of the hash function (i.e compression function). The security of any hash function lies on the output hash digests length. Which means, the longer hash digest the more secure hash function. The brute-force attack is a trial and error method to obtain a desired hash function. As an example of a brute-force attack is a dictionary attack which contain a list of dictionary words to try them all in a consecutive manner. These brute-force attacks can always be attempted, however they are not considered as a break unless the required number of evaluations of the hash function is significantly less than both the strength estimated by the designer of the hash

function and that of hash functions of similar parameters with ideal strength [13].

- Cryptanalytical Attacks

A hash function cryptanalysis attempts to attack the properties of hash functions such as a preimage attack, second preimage attack and collision attack. Due to fixed size of the hash values compared to much larger size of the messages, collisions must exist in hash functions. However, for the security of the hash function they must be computationally infeasible to find. Note that collisions in hash functions are much easier to find than preimages or second preimages. Informally, a hash function is said to be “broken” when a reduced number of evaluations of the hash function compared to the brute force attack complexities and the strengths estimated by the designer of the hash function are used to violate at least one of its properties immaterial of the computational feasibility of that effort. For example, assume that it requires 290 evaluations of the hash function to find a collision for a 256-bit hash function. Though it is impractical to generate this amount of computational power today, the hash function is said to be broken as this factor is less than the 2^{128} evaluations of the hash function required by the birthday attack. This theoretical break on the hash function is also termed an “academic break” on the hash function. It should be noted that hash functions are easier to attack practically than encryption schemes because the attacker does not need to assume any secrets and the maximum computational effort required to attack the hash function is only upper bounded by the attacker’s resources not users gullibility. This is not the case with block ciphers where the maximum practical count of executions of the block algorithm is limited by how much computational effort the attacker can get the user to do [14]. As Figure 8 shows that cryptanalytical attacks on hash functions are classified into two categories: generic attacks and specific attacks.

- Generic Attacks (Attacks on Merkle-Damgård construction)

Generic attacks are technical studies used to attack general hash function constructions (i.e Merkle-Damgård construction). The word “generic” means that the attack is not designed for a specific hash function (i.e SHA-2). For example, if the hash function uses a certain block cipher, replacing this block cipher with another should not affect the complexity of a generic attack of that hash function. The generic attacks are classified into four types, discussed in the following sections.

- 1) Length Extension Attacks

An attacker can use the advantage of using the padding scheme for the messages in Merkle-Damgård construction by applying length extension attack (it is also called extension attack). Length extension attack can be used to break secret prefix MAC scheme where the attacker computes the authentication tags without the knowledge of the secret key.

- 2) JouxAttack

Joux attack or Joux multi-collision attack is an attack on Merkle-Damgård hash function, where Antoine Joux shown that finding multiple collisions (more than two messages hashing to the same digest) in a Merkle-Damgård hash function is not much harder than finding single collisions as Figure 9 shown. In his multi-collision attack, Joux assumed access to a machine C that given an initial state, returns two colliding messages [7].

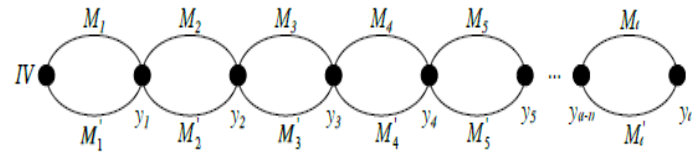


Figure 9. Joux Attack [7]

Also, Joux used his multi-collision in Merkle-Damgård hash function to produce a collision attack in a concatenation of two independent hash functions. Particularly, this attack deemed to be the first spark to look forward to start searching for a new paradigm of mode of operation of hash function other than Merkle-Damgård construction and hence announced SHA-3 competition.

- 3) Long Message second preimage Attacks

In the second preimage attacks, the attacker finds a second preimage S for a given message M, where $M \neq S$ and $H(M) = H(S)$ with an effort less than 2^l computation of H. In the long message second preimage attack, the attacker tries to find a second preimage for a long target message M of 2^{q+1} message blocks. The attacker does this by finding a linking message block M_{link} . Where, the digest of f_{IV} of the linking message block M_{link} matches one of the intermediate states H_i obtained in the hashing of M. The computation cost of this attack is about 2^{q+1} calls to the compression function f .

- 4) Herding Attack

This attack is due to Kelsey and Kohno [15] and is closely related to the multi-collision and second preimage attacks discussed above. A typical scenario where this attack can be used is when an adversary commits to a hash value D (which is not random) that he makes public and claims (falsely) that he possesses knowledge of unknown events (events in the future) and that D is the hash of that knowledge. Later, when the corresponding events occur, the adversary tries to herd the (now publicly known) knowledge of those events to hash to D as he previously claimed [15].

- Specific Attacks (Attacks on Specific Hash Functions)

Specific attacks on hash functions are based on the hash function itself. For example, attacks on MD5 [16], SHA-0 [17] and SHA-1 [18] are called multi-block collision attacks. Multi-Block Collision Attack (MCBA) technique on iterated hash function (i.e Merkle-Damgård construction) finds two colliding messages each at least two blocks on length. In such attack, collisions are found by processing more than one message block. In fact, multi-block collisions attack is applicable and valid on MD5, SHA-0 and SHA-1, since these hash functions use more than a single and collisions are distributed randomly.

As Figure 8 shows the sub-categories of “attacks on specific hash functions” which are (collision attack on MD4, MD5, RIPEMD, SHA-0 and SHA-1; near collisions on reduced version of SHA-256 and second preimage attack on MD4) are only examples of specific attacks on these hash functions. Meaning that, attacks can be customized and applied based on the hash function behavior and architecture.

3. APPLICATIONS OF HASH FUNCTIONS

Hash functions play a major role in application’s security such as certification, data integrity and authentication. The following sections illustrate these applications.

A. Digital Signature

Digital Signature is a mathematical scheme used to validate the authenticity of the sender, message and signer of the document identity as Figure 10 shown. Digital signatures are commonly used in Web-commerce, financial transactions and other cases where it is crucial to detect alteration of a message or a document. It uses private and public keys along with the hash digest to create the signature for a document. Digital Signature provides signer authentication and authorization of a document. It indicates who signed a document, message or record and makes difficult for another person to produce the same without authorization [29].

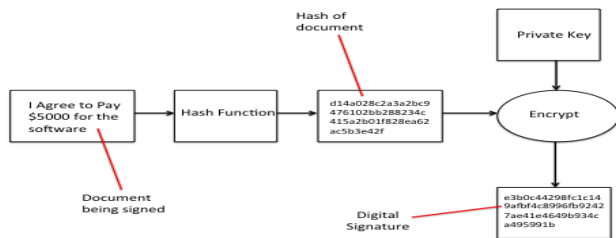


Figure 10. Digitally signed document

B. MAC

A Message Authentication Code (MAC) is similar in usage to a message digest. It is designed especially for applications to detect message tampering and forgery. MAC accepts a shared secret symmetric key (K) as input along with the arbitrary length and outputs MAC (sometimes called tag). Figure 11 shows the MAC algorithm process.

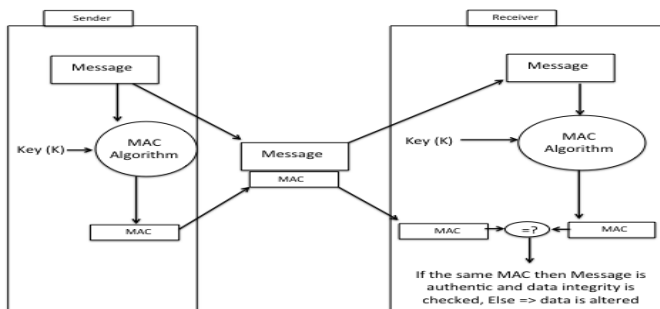


Figure 11. MAC algorithm

As Figure 11 shows the process of the MAC algorithm, sender calculates the MAC by first calculating message digest of the message or document and then applying secret key K to the message digest. Then, the message with the calculated MAC sends to the receiver [29]. Independently, the receiver calculates a new MAC value by using the symmetric secret key (K) and generate new hash digest. If the attached MAC with original message matches the new calculated MAC performed by the receiver then the message is authenticated and integrity verified.

MACs differ from digital signature as MAC uses a symmetric secret key and digital signature uses asymmetric key (public and private keys).

C. HMAC

A popular and specific implementation of message authentication codes is the HMAC (Hash Message Authentication Code). It is a specific construction for calculating a message authentication code which involves a secret key and cryptographic

hash function to ensure secure data transfer over unsecure channels. As computers becoming more powerful, the need arises for complex hash functions. As a result, HMAC is a preferable to use with hash functions other than MAC due to its higher security. Cryptographic hash functions such as SHA-1 or MD5 may be used to calculate HMAC. In this case, the resulting hash function called HMAC-SHA-1 and HMAC-MD5 respectively.

D. Kerberos

Kerberos is network authentication protocol. It is designed to provide strong authentication and improved security for users and between client/server applications. Kerberos was developed at the Massachusetts Institute of Technology (MIT) in 1998. Using Kerberos, a user can request an encrypted “ticket” from an authentication process so it can be used to request a specific service from a server.

E. Key Derivation

A key derivation function (KDF) is an algorithm to derive a key of a given size from a secret value or other known information. That is used to derive keys from a secret value such as a value obtained by Diffie-Hellman key establishment. Keyed cryptographic hash function can be used for key derivation.

F. One Time Password

Cryptographic hash functions are used to compute one time password (OTP). OTP is a password that is valid for a single login or transactions. By using cryptographic hash functions, hashed passwords are saved instead of keeping the password itself. So that, if the file of passwords are revealed then the passwords still protected if the hash function is preimage resistance.

G. Pseudorandom generator

A cryptographic hash function can be used to generate pseudorandom generator (PRG). PRGs are used to generate pseudorandom bits from a short random seed, which can then be used in place of truly random bits that most cryptographic schemes rely on. On the foundational side, PRGs can be used as a building block for more complex cryptographic objects like pseudorandom function (PRF), bit commitment, etc [19].

H. Pretty Good Privacy

Pretty Good Privacy or PGP is a popular program that is used to encrypt/decrypt and authenticate e-mails over the internet. PGP uses a hash function to ensure the integrity of e-mail message.

I. Secure Socket Layer/Transport Layer Security

Secure Socket Layer (SSL) and Transport Layer Security (TLS) protocols are used to authenticate servers and clients over an untrusted network. SSL/TLS can help to secure data transferred using encryption. Also, SSL/TLS can authenticate servers as well as clients through secure communication.

4. AN OVERVIEW OF A SELECTED HASH FUNCTIONS

A. MD4 and MD5

The cryptographic hash function MD4 (Message Digest 4) was introduced by Ronald Rivest in 1990. MD4 was a novel design, which compresses an arbitrary input length and produce 128-bits as a hash digest. Later, other hash algorithms such as MD5, SHA-0, SHA-1 and HAVAL were derived and influenced by MD4. In 1991, hash function Message Digest 5 or MD5 was designed by Ronald Rivest as a strengthened version of MD4. MD5 is widely used algorithm in a variety of security applications.

Working of MD5 is almost similar to MD4 but some changes have been made to MD4. One extra round is added in MD5. MD5 also compresses arbitrary bit-length input into a 128-bit hash value [29].

B. RIPEMD

RIPEMD is a cryptographic hash function developed by Hans Dobbertin, Antoon Bosselaers and Bart Preneel, and first published in 1996. Its design is based on MD4. Which is consists of two parallel versions of the MD4 compression function. RIPEMD produce 160-bits as a hash digest. Dobbertin found a collision attack on two rounds of RIPEMD. Strengthen versions of RIPEMD were developed due the weakness founded in RIPEMD-160 bits [29]. These versions are RIPEMD-128, RIPEMD-256 and RIPEMD-320. RIPEMD produce 128-bits of hash digest. The extended version of RIPEMD-128 is RIPEMD-256, which produce 256-bits as a hash digest. Also, the extended version of RIPEMD-160 is RIPEMD-320, which produce 320 bits as a hash digest.

C. SHA-x Family

• **Secure Hash Algorithm-0**

National Institute of Standard and Technology (NIST) along with National Security Agency (NSA) published the *Secure Hash Algorithm (SHA)* in 1993. At present, SHA is commonly referred to SHA-0. SHA-0 is an algorithm that produces a 160-bits hash digest. SHA-0 was developed to replace MD4 but it was withdrawn shortly after publication due to security issues.

• **Secure Hash Algorithm-1**

SHA-0 was replaced by SHA-1 in 1995. *Secure Hash Algorithm-1* or SHA-1 is a message digests algorithm, which is regarded the world’s most popular hash function, which takes input a message of arbitrary length and produce output a 160 bits “fingerprint” of the input. However, the security level of this standard is limited to a level comparable to an 80-bit block cipher [20]. It is based on the design principle of MD4, and applies the Merkle-Damgård model of compression function.

• **Secure Hash Algorithm-2**

In August, 2002, NIST has published three additional hash functions, SHA-256, SHA-384 and SHA-512. These new hash functions family known as *Secure Hash Algorithm-2* or simply SHA-2. SHA-2 was introduced due to the need of a larger key of a hash function to match the new *Advanced Encryption Standard (AES)* which introduced in 2001. In February 2004, another hash function SHA-224 was added to the SHA-2 family. SHA-224 and SHA-384 are the truncated versions of SHA-256 and SHA-512 respectively. The proposed system architecture of SHA-2 hash family can support efficiently the security needs of modern communication applications such as WLANs, VPNs and firewall [29].

• **Secure Hash Algorithm-3**

In October 2012, NIST announced the winner of SHA-3 competition which started in 2008. Keccak was the winner of NIST competition and become the new SHA-3 standard. Keccak is a cryptographic hash function designed by Guido Bertoni, Joan Daemen, Michael Peeters, and Gilles Van Assche. Keccak has completely different construction than SHA-0, SHA-1 and SHA-2 families. It supports at least four different output lengths $n \in \{224, 256, 384, \text{ and } 512\}$ in a high security levels [21]. According to [22] and [23] the construction of Keccak sponge design is building the compression function from different permutation f operates components in the following:

1. Signify the length of message bitstring by $|M|$, as a sequence of blocks in fixed length x , when calculated the ranges from 0 to $|M|_x - 1$.
2. Pad the message M in a sequence of x -bit blocks to signify by $|M| \text{ pad } [x] (|M|)$. Thus, padding rules have append a bitstring to determined the bit length of M and the block length x .
3. It is a sponge hash functions to construct a function of $[f, \text{pad}, r]$ where the permutation f has different length in input and fixed length of output, a padding rule “pad” and a *bitrate* r . The permutation f operates has seven set of bits, which denoted as Keccak- $f[b]$ where $b = 25w$. Keccak- $f[b]$ is a permutation over, when the bits are figured from 0 to $b-1$. Thus, the different versions of Keccak- f permutation have limited output values in $\{25, 50, 100, 200, 400, 800, 1600\}$ to represent the hypercube of sponge construction of three dimensional array, and $c = b - r$ is the *capacity* of absorbing phase of the compressing state (Bertoni et al. 2011). Figure 12 from depicts the sponge construction of Keccak- $f[r + c]$.

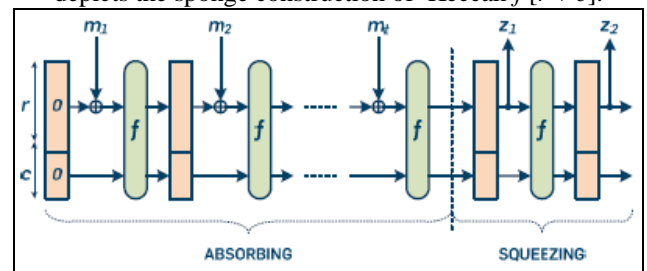


Figure 12. The sponge construction of Keccak [22]

4. The permutation is a sequence of operations on the three-dimensional array of elements of $GF(2)$, specifically $a[5][5][w]$, with $w = 2^l$, where $w \in \{1, 2, 4, 8, 16, 32, 64\}$. The expression $a[x][y][z]$ with $x, y \in \mathbb{Z}_5$ and $z \in \mathbb{Z}_w$, signifies the bit in position (x, y, z) , follows by indexing starts from zero. The mapping between the bits of s and those of a is $s[w(5y + x) + z] = a[x][y][z]$. That terms in the x and y coordinates should be taken modulo 5 and expressions in the z coordinate modulo w . The source state has a fixed value and should never consider as an input [24].

5. CONCLUSION

This paper presented an extensive study of cryptographic hash functions. The presented study surveys cryptographic hash functions from various aspects. It included the properties, classification, constructions, attacks, applications and an overview of a selected dedicated cryptographic hash functions. Practically, MD4, MD5 and SHA-0 considered broken hash functions. Theoretically, SHA-1 considered a broken hash function. But SHA-2 considered secure one. SHA-3 was presented due to the need for a long term security hash function which has a new promising sponge construction.

6. REFERENCES

- [1] S. Goldwasser, S. Micali and R. Rivest. A Digital Signature Scheme Secure Against Adaptive Chosen-Message Attacks, *SIAM Journal of Computing*, vol 17, No. 2, pp. 281-308, April 1998.
- [2] M. Naor and M. Yung. Universal one-way hash functions and their cryptographic applications, *Proceedings of the*

- Twenty First Annual ACM Symposium on Theory of Computing, ACM Press, pp 33-43, 1989.
- [3] S. Haber and W. S. Stornetta. How to timestamping a digital document. *Journal of Cryptology* 3(2), pp. 99-111, 1991.
- [4] H. Krawczyk, M. Bellare and R. Canetti. HMAC: Keyed-Hashing for Message Authentication. Internet RFC 2104, February 1997.
- [5] V. Shoup. Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal of Computing* 33:167-226, 2003.
- [6] I. Damgård. A design principle for hash functions. In G. Brassard, editor, *Advances in Cryptology - CRYPTO '89, 9th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 1989, Proceedings*, volume 435 of *Lecture Notes in Computer Science*, pages 416-427. Springer, 1990.
- [7] Antoine Joux. Multicollisions in Iterated Hash Functions. Application to Cascaded Constructions. In Matt Franklin, editor, *Advances in Cryptology- CRYPTO 2004*, volume 3152 of *Lecture Notes in Computer Science*, pages 306-316. Springer, August 15-19 2004.
- [8] U. Maurer, R. Renner, and C. Holenstein. Indifferentiability, Impossibility Results on Reductions, and Applications to the Random Oracle Methodology. In M. Naor, editor, *Theory of Cryptography, First Theory of Cryptography Conference, TCC 2004, Cambridge, MA, USA, February 19-21, 2004, Proceedings*, volume 2951 of *Lecture Notes in Computer Science*, pages 21-39. Springer, 2004.
- [9] Lucks, S. (2004). Design principles for iterated hash functions, *Cryptology ePrint Archive*, Report 2004/253, 2004, <http://eprint.iacr.org>.
- [10] G. Bertoni, J. Daemen, M. Peeters, and G. Van Assche. On the indifferentiability of the sponge construction. *Advances in Cryptology - EUROCRYPT 2008, 27th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Istanbul, Turkey, April 13-17, 2008. Proceedings*, 4965:181-197, 2008.
- [11] J.-P. Aumasson and W. Meier. Zero-sum distinguishers for reduced Keccak-f and for the core functions of Luffa and Hamsi. *NIST mailing list*, 2009.
- [12] C. Boura, A. Canteaut, and C. D. Cannière. Higher-order differential properties of Keccak and Luffa. *Cryptology ePrint Archive*, Report 2010/589, 2010. <http://eprint.iacr.org/2010/589.pdf>.
- [13] Eli Biham, Rafi Chen, Antoine Joux, Patrick Carribault, Christophe Lemuet, and William Jalby. Collisions of SHA-0 and Reduced SHA-1. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT 2005*, volume 3494 of *Lecture Notes in Computer Science*, pages 36-57. Springer, 2005.
- [14] [24] Yuliang Zheng, Josef Pieprzyk, and Jennifer Seberry, 1993." Haval – A One-Way Hashing Algorithm with Variable Length of Output", *Lecture Notes in Computer Science*, Volume 718, *Advances in Cryptology – Auscrypt '92*, pp. 83-104.
- [15] J. Kelsey and T. Kohno. Herding hash functions and the nostradamus attack. In S. Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*, volume 4004 of *Lecture Notes in Computer Science*, pages 183-200. Springer, 2006.
- [16] Xiaoyun Wang and Hongbo Yu. How to Break MD5 and Other Hash Functions. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT 2005*, volume 3494 of *Lecture Notes in Computer Science*, pages 19-35. Springer, 2005.
- [17] Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Efficient collision search attacks on SHA-0. In Victor Shoup, editor, *Advances in Cryptology— CRYPTO '05*, volume 3621 of *Lecture Notes in Computer Science*, pages 1-16. Springer, 2005, 14-18 August 2005.
- [18] Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding collisions in the full SHA-1. In Victor Shoup, editor, *Advances in Cryptology— CRYPTO '05*, volume 3621 of *Lecture Notes in Computer Science*, pages 17-36. Springer, 2005, 14-18 August 2005.
- [19] Blum, M. and S. Micali (1984). "How to generate cryptographically strong sequences of pseudorandom bits." *SIAM journal on Computing* 13(4): 850-864.
- [20] Vincent Rijmen and Elisabeth Oswald, 2005." Update on SHA-1". In Alfred Menezes, editor, *Topics in Cryptology - CT-RSA 2005, The Cryptographers' Track at the RSA Conference 2005, San Francisco, CA, USA, volume 3376 of LNCS*, pp. 58-71.
- [21] E. Andreeva, B. Mennink, B. Preneel & M. Skrobot (2012), Security Analysis and Comparison of the SHA-3 Finalists BLAKE, Grostl, JH, Keccak, and Skein. from Katholieke Universiteit Leuven, Belgium.
- [22] G. Bertoni, J. Daemen, M. Peeters, & G. V. Assche (2012), Keccak An update. Retrieved March 22-23, 2012, from Third SHA-3 candidate conference, Washington DC.
- [23] E. B. Kavun & T. Yalcin (2012), On the Suitability of SHA-3 Finalists for Lightweight Applications. from Horst Görtz Institute, Ruhr University, Chair of Embedded Security, Germany. Website: http://csrc.nist.gov/groups/ST/hash/sha-3/Round3/March2012/documents/papers/KAVUN_paper.pdf
- [24] Imad Fakhri Al-shaikhli, Mohammad A. Alahmad and Khansaa Munther. The "Comparison and analysis study of sha-3 finalists." *International Conference on Advanced Computer Science Applications and Technologies*(26-28 Nov 2012): 7.
- [25] Mohammad A. Ahmad, I. F. A. S., Hanady Mohammad Ahmad (2012). "Protection of the Texts Using Base64 and MD5." *JACSTR Vol 2, No 1 (2012)*(1): 12.
- [26] Imad F. Alshaikhli, M. A. Ahmad. (2011). "Security Threats of Finger Print Biometric in Network System Environment." *Advanced Computer Science and Technology Research* 1(1): 15.
- [27] Al-Kuwari, S. and Davenport, J.H. and Bradford, R.J. *Cryptographic hash functions: recent design trends and security notions*. Science Press of China, 2010.
- [28] Sobti, R. and G. Geetha (2012). "Cryptographic Hash Functions: A Review." *IJCSI International Journal of Computer Science Issues* 9(2): 461-479.

Heterogeneous Vehicle Routing Problem with profits Dynamic solving by Clustering Genetic Algorithm

Sawsan Amous Kallel¹ and Younes Boujelbene²

¹ UREA : Research Unit in Applied Economics, Faculty of Economics and Management, Airport Road Km 4, Sfax 3000, Tunisia

² Director of the Research Unit UREA Faculty of Economics and Management, Airport Road Km 4, Sfax 3000, Tunisia

Abstract

The transport problem is known as one of the most important combinatorial optimization problems that have drawn the interest of many researchers. Many variants of these problems have been studied in this decade especially the Vehicle Routing Problem. The transport problem has been associated with many variants such as the Heterogeneous Vehicle Routing Problem and dynamic problem.

We propose in this study dynamic performance measures added to HVRP that we call "Heterogeneous Vehicle Routing Problem with Dynamic profits" (HVRPD), and we solve this problem by proposing a new scheme based on a clustering genetic algorithm heuristics that we will specify later.

Computational experiments with the benchmark test instances confirm that our approach produces acceptable quality solutions compared with previous results in similar problems in terms of generated solutions and processing time. Experimental results prove that the method of clustering genetic algorithm heuristics is effective in solving the HVRPD problem and hence has a great potential.

Keywords: *The Heterogeneous Vehicle Routing Problem, dynamic problems, genetic algorithm heuristics, k-means clustering.*

1. Introduction

Within the wide scope of logistics management, transportation plays a central role and is a crucial activity in the delivery of goods and services. The transport problem is one of the mainly essential combinatorial optimization problems that have taken the interest of several researchers. Huge research efforts have been devoted to the study of logistic problems and thousands of papers have been written on many variants of this problem such as Traveling Salesman Problem (TSP), Vehicle Routing Problem (VRP), supply chain management (SCM) and so on [6].

The VRP is one of the most studied combinatorial optimization problems and it consists of the optimal design of routes to be used by a fleet of vehicles to satisfy the demands of customers. In general, the number of vehicles used in VRP is a variable decision.

A variant of VRP has a dynamic nature and can be modelled as Dynamic Combinatorial Optimization Problem (DCOP). This variant has been used in many researches on transportation problems. The results of these studies show that the complex real transportation problems are dynamic and change over time in terms of their objective functions, decision variables and constraints. This implies that the optimal solution might change at any time due to the changes in the transport environment.

Many other related problems are associated with VRP such as the Heterogeneous Fleet Vehicle Routing Problem (HVRP). The HVRP differs from the classical VRP in that it deals with a heterogeneous fleet of vehicles having various capacities and both fixed and variable costs. Therefore, the HVRP involves designing a set of vehicle routes, each starting and ending at the depot, for a heterogeneous fleet of vehicles which services a set of customers with specific demands. Each customer is visited only once, and the total demand of a route should not exceed the loading capacity of the vehicle assigned to it. The routing costs of a vehicle is the sum of its fixed costs and a variable costs incurred proportionately to the travel distance.

The objective is to minimize the total of such routing costs. The number of available vehicles of each type is assumed to be unlimited [5].

In the literature, three HVRP versions have been studied. The first one was introduced by Golden and al. [10], in which variable costs are uniformly given over all vehicle types with the number of available vehicles assumed to be unlimited for each type. This version is also called the Vehicle Fleet Mix (VFM) [19], the Fleet Size and Mix VRP

(1)

[11] or the Fleet Size and Composition VRP [8]. This version that we are consider in this paper.

The second version considers the variable costs, dependent on the vehicle type, which are ignored in the first version. This version is referred to as the HVRP [7], the VFM with variable unit running costs [18].

The third one, called the VRP with a heterogeneous fleet of vehicles [20] or the heterogeneous fixed fleet VRP [21], generalizes the second version by limiting the number of available vehicles of each type.

The remaining part of the paper is organized as follows: the next section introduces the notation used throughout the paper and describes the variants of heterogeneous VRPs studied in the literature then we introduce our problem: heterogeneous vehicle routing problem with dynamic profits associated to the customer. In section 3 we propose a hybrid algorithm that we call “Clustering Genetic Algorithm”. Finally, Section 4 reviews the experimental results.

The remaining part of the paper is organized as follows: the next section introduces the notation used throughout the paper and describes the variants of heterogeneous VRPs studied in the literature then we introduce our problem: heterogeneous vehicle routing problem with dynamic profits associated to the customer. In section 3 we propose a hybrid algorithm that we call “Clustering Genetic Algorithm”. Finally, Section 4 reviews the experimental results.

2. Heterogeneous Vehicle Routing Problem with Dynamic profits associated to the customer priority

In this part, we describe the formulation of HVRP then we explain our problem and we propose a new mathematic formulation for HVRPD.

2.1 The Heterogeneous Vehicle Routing Problem: formulation of the problem

The Heterogeneous Vehicle Routing Problem is defined as follows [2]. Let's $G = (V, A)$ is a directed graph where $V = \{v_0, v_1, \dots, v_n\}$ is the set of $n + 1$ nodes and $A = \{(v_i, v_j) : v_i, v_j \in V, i \neq j\}$ is the arc set. Node v_0 represents a depot on which a fleet of vehicles is based,

while the remaining node set $V' = V \setminus \{0\}$ vertices corresponds to n cities or customers.

Each customer $v_i \in V'$ has a non negative demand q_i . The vehicle fleet is composed of m different vehicle types, with $M = \{1, \dots, m\}$. For each type $k \in M$, m_k vehicles are available at the depot, which have a capacity equal to Q_k . Each vehicle type is also associated with a fixed cost, equal to F_k and a variable cost G_k per distance unit.

The number of vehicles of each type is assumed to be unlimited. With each arc (v_i, v_j) is associated a distance or cost C_{ij} . The HVRP aims at designing a set of vehicle routes, each starting and ending at the depot, visiting each customer only once, limiting the total demand of a route to the loading capacity of the vehicle assigned to it, and minimizing the total cost of the route [7].

All the HVRP variants are NP-hard as they include the classical VRP as a special case. All the presented studies have thus focused on developing heuristic algorithms as a substitute for exact algorithms. They can be generally grouped into two kinds: classical heuristics [11], [18], [16], frequently derived from the classical VRP heuristics, and meta-heuristics like the tabu search method [7], [21]. For more information, see [18] and [21] for a review of HVRP variants.

In our approach, we propose a new problem that we called HVRPD it is to introduce the priority of the customers with dynamic profits in the formulation of HVRP.

2.2 Proposed problem DHVRP

Practically, “dynamism” can be attributed to several factors, such as new customer order, cancellation of old demands... Demands of customers that have been studied by Taillard [20], Gendreau and al. [7], Ismail and Irhamah [12] have already used “dynamism” in their study. Other factors are stochastic like travel times between customers, customers to be visited, locations of customers, capacities of vehicles, and number of vehicles available have been dealt with by Prins [15], Renaud [16]...

In the static vehicle routing problem, information is assumed to be known, including all attributes of the customers such as the geographical location, the service time and all the details about the customer's demand. However, in the dynamic applications, information on the problem is not completely known in advance.

The problem is said dynamic when not all information related to the planning of the routes is known by the planner when the routing process begins. Besides, the information is progressively revealed to the decision maker and it is likely to change according to the dynamic nature of the transport environment.

For the problem under study, we will combine HVRP and dynamic VRP including the priority of customers. We will present in this paper, a mathematical model for HVRPD.

Hence, the mathematical formulation is similar to that of general HVRP [9] with the introduction of the dynamism and priorities by clients.

Note $z_{ir} = 1$ if the client i is at the rank r and 0 otherwise.

The modeling of the problem is similar to standard HVRP [8] and is as follows:

$$\text{Min} \sum_{k \in M} F_k \sum_{j \in V'} x_{0j}^k + \sum_{k \in M} \sum_{i, j \in V'} c_{ij} x_{ij}^k - \sum_{i, r \in V} q_i g_{ir} z_{ir}$$

Subject to

$$\sum_{k \in M} \sum_{i \in V} x_{ij}^k = 1 \quad \forall j \in V' \quad (1)$$

$$\sum_{i \in V} x_{iv}^k - \sum_{i \in V} x_{vj}^k = 0 \quad \forall v \in V', \forall k \in M \quad (2)$$

$$\sum_{j \in V'} x_{0j}^k \leq p_k \quad \forall k \in M \quad (3)$$

$$\sum_{i \in V} y_{ij} - \sum_{i \in V} y_{ji} = q_j \quad \forall j \in V' \quad (4)$$

$$q_j x_{ij}^k \leq y_{ij} \leq (Q^k - q_i) x_{ij}^k \quad \forall i, j \in V, i \neq j, \quad (5)$$

$$y_{ij} \geq 0, \quad \forall i, j \in V, \quad \forall i \neq j \quad (6)$$

$$x_{ij}^k \in \{0, 1\}, \quad \forall i, j \in V, i \neq j, \quad \forall k \in M \quad (7)$$

With the addition of other constraints to introduce the principle of dynamic gains are:

$$\sum_{r=1}^n r z_{ir} + 1 - n(1 - x_{ij}^k) \leq \sum_{r=1}^n r z_{jr} \quad \forall i, j \in V, i \neq j \quad (8)$$

$$\sum_{r=1}^n r z_{jr} \leq \sum_{r=1}^n r z_{ir} + 1 + n(1 - x_{ij}^k) \quad \forall i, j \in V, i \neq j \quad (9)$$

$$\text{And} \quad \sum_{r=1}^n z_{ir} = 1 \quad \forall i \in V \quad (10)$$

Given the random character of the needs q_i , changing continually, and availability of vehicles, we assume a probability distribution according to the history of the customer's request.

In the above formulation, constraint (1) ensures that the customer is visited only once and the constraint (2) ensures the continuity of a tour that means: if a vehicle visits a customer i , it marks the end for this customer and

the start of the next one. The maximum number available for each type of vehicles is imposed by constraint (3). Constraint (4) states that the difference between the quantities of goods that a vehicle carries raised before and after visiting a customer is equal to the demand of this customer. Constraint (5) ensures that the vehicle capacity does never exceed the demand of customers.

The two constraints (6) and (7) show that the quantity transported must be positive and the value of x_{ij}^k can take only 0 or 1.

For the two new constraints (8) and (9), we introduce the profits: the lower the rank of the client is the higher the gains are.

The ranking of client can replace the principle of time window and it makes it easier to implement and reduce the computation time with the same efficiency. It will be beneficial to have a small place.

In this paper, and to solve this problem, we will develop hybrid approaches that incorporate the best features of the metaheuristic "genetic algorithms" and the method of classification "k-means" that we called "Clustering Genetic Algorithm". So, we develop five steps, that we will mention later, that begin by clustering method and finish with finding the optimal tours at the same time, they satisfy the demand of customers.

3. Clustering Genetic Algorithm

The growing importance of combinatorial optimization problems requires the search for efficient algorithms to find optimal and near optimal solutions. Clustering methods are important tool to analyze and manage data. They are used to classify data into homogeneous classes. Inspired by the high performance of Genetic Algorithms (GA); we propose to introduce a clustering process into a genetic algorithm in order to design an efficient tool to deal with complex optimization data. In this domain, some proposed works have appeared recently [4].

To solve the HVRPD, taking into account the new capacity constraints and priorities for customers, a hybrid algorithm will be initialized named Clustering Genetic Algorithm (CGA), which must make appropriate changes on the structural representation of the elements of the problem and reset, and that crossing operations with regard to the specifics of the problem in five steps.

First, client groups within the capacity of the vehicles are formed. They are based on clustering techniques of k -means. Then, a traveling salesman problem is solved in each group by the GA. In the third phase, the vehicles are assigned to tours obtained within the constraints of capacity. They should also aim at minimizing the costs. The fourth phase is to introduce two variables that show the

degree of certainty with which all the requirements must be satisfied by a given vehicle k .

Finally, the fifth phase aims to identify the customers who are satisfied, taking into account the levels of priority, by which groups of customers are aggregated into super-priority nodes.

3.1 Clustering Method

Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data.

A loose definition of clustering could be “the process of organizing objects into groups whose members are similar in some way”. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters [22].

There are many clustering methods available, and each of them may give a different grouping of a dataset. The choice of a particular method will depend on the type of output desired, The known performance of method with particular types of data, the hardware and software facilities available and the size of the dataset.

In general, clustering methods may be divided into two categories based on the cluster structure which they produce. The non-hierarchical or unsupervised methods divide a dataset of N objects into M clusters, with or without overlap and the hierarchical clustering algorithm is based on the union between the two nearest clusters. The beginning condition is realized by setting every datum as a cluster. After a few iterations it reaches the final clusters wanted.

The specific method used in our approach is the k -means. It's one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. k -means consists in partitioning the data into groups, or "clusters". These clusters are obtained by positioning "centroids" in regions of space that have the largest number of population.

Each observation is then assigned to the closest prototype. Each class therefore contains observations that are nearer to a prototype than any other.

Then, the centroids are positioned by an iterative procedure (the centroid of the classes is calculated: the weight of a center of gravity is equal to the sum of the weights of the class of individual forms) which leads progressively to their final stable position.

The weight of a center of gravity aims at minimizing an objective function, in this case a square error function. The objective function [22]:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

where $\|x_i^{(j)} - c_j\|^2$ is a chosen distance measure

between a data point $x_i^{(j)}$ and the centroid c_j , is an indicator of the distance of the n data points from their respective cluster center.

At this state, the final number of classes will be defining. k -means is an algorithm that has been adapted to many problem domains. As we are going to see, it is a good candidate for extension to work with fuzzy feature vectors.

Once the classes are identified in the final partition of the data space, it is possible, to run the Genetic Algorithm cluster by cluster.

3.2 Genetic Algorithm

The Genetic Algorithm (GA) starts from an initial population of candidate solutions or individuals and proceeds for a certain number of iterations until one or more stopping criteria is /are satisfied. This evolution is directed by a fitness measure function that assigns to each solution (represented by a chromosome) a quality value.

Once the population is evaluated, the selection operator chooses which chromosomes in the population will be allowed to reproduce. The stronger an individual is, the greater chance of contributing to the production of new individuals it has.

The new individuals inherit the properties of their parents and they may be created by crossover (the probabilistic exchange of values between chromosomes) or mutation (the random replacement of values in a chromosome). Continuation of this process through a number of generations will result in a group of solutions with better fitness in which optimal or near-optimal solutions can be found [1].

When the size of the problem is becoming wider, genetic algorithms find a difficulty in identifying the optimal solution. Nevertheless, the question is why sacrificing an approach that can have a good opportunity to develop a new area of expertise and keep the problem in touch with reality?

The purpose of the implementation of genetic algorithms is to get a hamiltonian cycle for the clusters identified. This means the genetic algorithms are applied cluster by cluster and the outcome of this step is the tours T_1, T_2, \dots, T_n for the clusters K_1, K_2, \dots, K_n .

3.3 Clustering Genetic Algorithm

In our approach CGA, we have first formed the clusters using *k*-means then we have built the route by solving TSP problems for each tour. So, we have divided our vehicle routing problem (search space) into areas to locate the task of GA. Therefore, we will use a two-phase method. It has shown in many applications [3], [16] of the vehicle routing problem and its generalizations that the set of routes is the most important phase of the algorithm.

The two-phase method uses the principle of "cluster first - route second" [5]. In the case of TSP, such a method comprises, first, splitting the set of vertices (customers) into sub-classes (sets) of customers, and then issuing a routing process on each class to have a solution.

The general outline of the two-phase method is as follows:

General diagram of the two-phase method

step 1 : $i = 0$, initialize $iMax$.

step 2 : Repeat steps 3 through 4 until $i = iMax$.

step 3 : Generate a *Phase1* partition into classes of all customers using a specific method of classification.

step 4 : *Phase2*

Apply a routing procedure on each class of peaks generated by step 3,
 $i++$.

The diversification of class divisions allows the two phase method to provide several solutions to the problem by applying alternatively the classification procedures and routing.

After clustering and genetic algorithm, the next step is to find the optimal tour. Therefore, we must find the complete tour by eliminating some edges and joining clusters. We will take the adjacent centroids and choose the shorts edges that we have to eliminate and repeat the same operation until we have only one tour and ultimately find the minimal cost of the whole tour.

Once the tours have been completed, we will assign the vehicles in a way that makes the fixed costs associated with the trucks the lowest possible.

The next step of our algorithm is to introduce two variables that allow to identify the degree of certainty with which a given vehicle *k* will meet all requirements must be satisfied.

Finally, the fifth phase aims to identify the satisfied customers, taking into account the levels of priorities in

which groups of customers are aggregated into super-priority nodes.

4. Experiment results

In our research, we include the *k*-means method to classify data into identical classes in order to facilitate the procedure of the genetic algorithm and it tackles the travelling salesman problem with many cities. The performances of the heuristics are tested, with a few exceptions, on two sets of benchmark instances: the first one consists of the 20 VFM instances [10], and the second contains 8 instances only with fixed costs [19].

The following table illustrates the results found by comparing the clustering genetic algorithm (Clu.GA) with a classical genetic algorithm (Cla.GA) (two-function point mutation by inversion), including the minimum length of the tour and finally the ratio = average / optimal for different instances of benchmarks from TSP library LIB see table 1.

Table 1: Examples of instances per cluster

Instances	Best know solutions	Minimal tour		Ratio	
		Cla. GA	Clu. GA	Cla. GA	Clu. GA
A 280	2579	4500.447	3954.456	1.744	1.533
Brg 180	1950	3160.621	2267.540	1.620	1.162
Ch 130	6110	15818	8853.430	2.588	1.448
Ch 150	6528	19140	10436	2.931	1.598
Eil 101	629	1954.734	1409.157	3.106	2.240
Eil 76	538	2349.051	1172.204	4.39	2.178
Eil 51	426	1651.796	808.400	3.875	1.890
Gil 262	2378	4460.268	3789.650	1.875	1.563
FI 417	11861	33140	29436	2.931	1.598
FI 1400	20127	34500	25354	1.744	1.833

Results from the experiment in table 1 show the advantages of the proposed algorithm concerning the quality of the solution (the ratio is: Average/optimale).

The proposed algorithm, Clu. GA, is compared with standard genetic algorithms considering a set of instances from the TSP LIB. Our experimental results prove the efficiency of the proposed algorithm in many instances when compared to the best-known solutions of Taillard [19].

We have also tested these instances with Lindo and CPLEX algorithm that gives optimal solutions but only for small instances. Both tests have taken much time to find these solutions.

Even though, our algorithm gives solutions that are not optimal but it has proved to be more efficient than Lindo and CPLEX ones in terms of processing time.

The resolution of the problem shows that our algorithm needs less time to solve big instances.

Let be the following instances with a number of different cities. We compare our algorithm with [12], [19] and [10].

The results are summarized in the following table:

Table 2. Comparable instances

Inst.	n	Osman and Salhi	Taillard	Golden	Our Average Solution	Time
3	20	965	961.03	961.03	965	2.65
4	20	6445	6437.33	6437.33	6536	5.98
5	20	1009	1008.59	1007.05	1016	3.95
6	20	6516	6516.47	6516.47	6596	6.97
13	50	2437	2413.78	2408.62	2785	6.95
14	50	9125	9119.03	9119.28	9150	6.74
15	50	2600	2586.37	2586.37	2960	7.16
16	50	2745	2741.5	2741.5	2830	4.29
17	75	1762	1747.24	1749.5	2086	6.14
18	75	2412	2373.63	2381.43	3360	5.12
19	100	8685	8661.81	8675.16	9642	7.95
20	100	4166	4047.55	4086.76	4213	6.96

The results presented in table 2 show that our algorithm produces solutions that are comparable quality to those of Osman and Salhi[12] and Taillard [19] and it improves the computational time (in secondes). Comparisons with other authors are the only way to evaluate our solutions because it is a new problem in the literature.

Our average results found by Clu.GA are close to those of Osman and Salhi [12], but they are bit lower than those of Taillard [19].

Table 3. Our average results

Ins	n	Veh.	Vehicles nature	Vehicles capacity	Our average value	Fixed costs	Time
3	21	5	5	20-30-40-70-120	965	20-35-50-120-225.	2.65
4	21	5	3	60-80-150	6536	1000-1500-3000.	5.98
5	21	5	5	20-30-40-70-120	1016	20-35-50-120-225.	3.95
6	21	5	3	60-80-150	6596	1000-1500-3000.	6.97
13	51	8	6	20-30-40-70-120-200	2785	20-35-50-120-225-400.	6.95
14	51	8	3	120-160-300	9150	1000-1500-3500.	6.74
15	51	5	3	50-100-160	2960	100-250-450.	7.16
16	51	5	3	40-80-140	2830	100-200-400.	4.29
20	101	8	3	60-140-200	4213	100-300-500.	6.96

Our algorithm has better performance on instances with fixed costs. Here, our average solution values are lower than those of Taillard in most cases, but they allow a good classification of vehicles and their fixed costs and also our algorithm allocate

customers in a better way than that presented by Golden (See table 3).

5. Conclusions

In this paper, we suggest a new mathematic model for HVRPD and we solve this problem by proposing a hybrid metaheuristic based on GA and k-means clustering.

The various experiments give indicate on the sensitivity of the algorithms in different configurations. They can exhibit the advantage of using a definition of compromise proposed to be specifically incorporated into an evolutionary algorithm. In this study, we have investigated a class of dynamic optimization problems. These algorithms need to be evaluated on other classes of problems including other variants and constraints.

We have used the clustering approach as an additional means to control intensification in order to obtain robust algorithms able to produce good solutions.

The results are encouraging especially in relation to processing time if compared with similar problems and similar algorithms.

References

- [1] Amous K, S., Loukil, T. Elaoud, S. and Dhaenens, C. 2008. A New genetic algorithm applied to the traveling salesman problem. International Journal of Pure and Applied Mathematics, Vol. 48, n0 2, pp 151-161, 2008.
- [2] Baldacci, R., Christofides, N., Mingozzi, A. 2007. An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts. Mathematical Programming Ser. A 2007 (<http://dx.doi.org/10.1007/s10107-007-0178-5>)
- [3] Choi, E. and Tcha, D.W. A column generation approach to the heterogeneous fleet vehicle routing problem. Computers & Operations Research; (in press).
- [4] Fisher, M.L. and Jaikumar, R.1981. A generalized assignment heuristic for vehicle routing. Networks 11: 109-124, 1981.
- [5] Gen, M. and Cheng, R. 2000. Genetic algorithms and engineering optimization. John Wiley & Sons, New York, 2000.
- [6] Gendreau, M. Laport, G. Musaraganyi, C. and Taillard, E.D. 1999. A tabu search heuristic for the heterogeneous fleet vehicle routing problem. Computers & Operations Research. 26, 1153-1173, 1999.
- [7] Gheysens, F. Golden, B. Assad, A. 1986. A new heuristic for determining fleet size and composition. Mathematical Programming Study.;26:233-6, 1986.
- [8] Gheysens, F.G. Golden, B.L. and Assad, A. 1984. A comparison of techniques for solving the feet size and mix vehicle routing problem. OR Spectrum, 6(4):207-216, 1984.

- [9] Golden, B. and Assad, A. 1988. *Vehicle Routing: Methods and Studies*. Editors, North-Holland, Amsterdam, pp.359-394, 1988.
- [10] Golden, B. Assad, A. Levy, L. and Gheysens, F. 1984. The fleet size and mix vehicle routing problem. *Computers and Operations Research*;11:49–66, 1984.
- [11] Ismail, Z. and Irhamah. 2008. Solving the Vehicle Routing Problem with Stochastic Demands via Hybrid Genetic Algorithm-Tabu Search. *Journal of Mathematics and Statistics*: 4(3): 161-167, 2008 ISSN: 1549-3644, 2008.
- [12] Osman, IH. and Salhi, S. 1996. Local search strategies for the vehicle fleet mix problem. In V.J. Rayward-Smith, I.H. Osman, C.R. Reeves, and G.D. Smith, editors, *Modern Heuristic Search Methods*, pages 131-153. Wiley: Chichester, 1996.
- [13] Prins C. A simple and effective evolutionary algorithm for the vehicle routing problem, *Computers and Operations Research*, Volume 31, Issue 12, 1985-2002, 2004.
- [14] Prins, C.2002. Efficient heuristics for the heterogeneous fleet multitrip VRP with applications to a large-scale real case. *Journal of Mathematical Modeling and Algorithms*, 1(2):135 - 150, 2002.
- [15] Renaud, J. Boctor, FF. 2002. A sweep-based algorithm for the fleet size and mix vehicle routing problem. *European Journal of Operational Research* ;140:618–28, 2002.
- [16] Roberto, Baldacci, Maria, Battarra and Daniele, Vigo. 2007. *Routing a Heterogeneous Fleet of Vehicles*. DEIS, University Bologna via Venezia 52, 47023 Cesena, Italy, 2007.
- [17] Salhi, S. Sari, M. Saidi, D. and Touati, N. 1992. Adaptation of some vehicle fleet mix heuristics. *Omega*, 20: 653–60, 1992.
- [18] Taillard, ED. 1999. A heuristic column generation method for the heterogeneous fleet VRP. *RAIRO*; 33:1–14, 1999.
- [19] Tarantilis, CD. Kiranoudis, CT. Vassiliadis, VS. 2004. A threshold accepting metaheuristic for the heterogeneous fixed fleet vehicle routing problem. *European Journal of Operational Research*. 152:148–58, 2004.
- [20] Toth, P. & Vigo, D. 2002. Models, relaxations and exact approaches for the capacitated vehicle routing problem. *Discrete Applied Mathematics*. 123, 1-3: 487-512, 2002.
- [21] Wassan, NA. Osman, IH. 2002. Tabu search variants for the mix fleet vehicle routing problem. *Journal of Operational Research Society*;53:768–82, 2002.

An Investigation of the Incidences of Repetitive Strain Injury among computer Users in Nigeria

Olabiysi Olatunde¹, Akingboye Yusuff², Abayomi-Alli Adebayo³, Izilien Fred⁴, Adeleke Iyiola⁵

¹ Computer Science and Engineering Department, Ladoke Akintola University of Technology, Ogbomosho, Oyo State Nigeria.

^{2,4,5} Electrical and Computer Engineering Department, Igbinedion University Okada, Edo State, Nigeria.

³ Computer Science Department, Federal University of Agriculture, Ogun State, Nigeria.

Abstract

Computer has been incorporated into day to day activities of almost every field of human endeavour, offices to different shops. Therefore many people are now working with computer for longer hours of time. There is no doubt that this incorporation of computer has helped users a lot but it also brings problems to the users. One of the problems is Repetitive Strain Injury (RSI).

Five hundred and thirty one (531) questionnaires were personally administered to different categories of people that use computer in various works of life, ranging from banking sector, civil service, educational sector, health sector to private sector. The distribution cut across different professions. A statistical analysis was conducted on the data obtained using frequency distribution, Pearson Correlation and Linear Regression.

The result obtained showed that 94.3% of the respondents suffered pain from one or more parts of the body. 86.8% of the respondents suffered from eyestrain, 63.9% suffered from low back pain, 67.4% with wrist pain, 64.7% finger pain while the least suffered pain was foot pain which only 19% responded positively to it. There are significant relationships between duration of computer usage, type of chair used, type and size of monitor used and the incidence of RSI. RSI modeled was formulated through linear regression which showed that a unit change in computer will result in corresponding 1.76 unit increases in RSI and a unit change in ergonomic deficiency will also result in corresponding 0.66 increases in RSI.

The existence of RSI was established and it was discovered that the more time spent on the computer system, the more the proximity of having strain or pain in one or more part(s) of the body.

Keywords: RSI, Workstation, Computer User, Ergonomic, Computer Task, Ergo Risk Factor.

1. Introduction

In recent times, computer has become a common tool that is being used by almost every individual from various field of human endeavour. This is due to the fact that computer offer a lot of different services and facilities to help the users to perform and complete the tasks more efficient and effective. Injuries due to the usage of computer system has been recognized world-wide and several actions that involved repetitive or forceful movements and the maintenance of constrained or awkward postures have also been associated with musculoskeletal disorder (Repetitive Strain Injury). Poor posture, prolong starring at computer screen, repetitive reaching for mouse, sitting position and type of chairs has been discovered to affect the lower back, eyes, arm of computer users [1].

Repetitive strain injury (RSI), also called cumulative trauma disorder (CTD), occupational overuse syndrome,

or work related upper limb disorder (WRULD), as conditions resulting from continuous use of a tool, e.g. Computer, guitar, knife, etc. over a long period of time or activity that requires repeated movements. It is a syndrome that affects muscles, tendons and nerves in the hands, arms and upper back. The medically accepted condition in which it occurs is when muscles in these areas are kept tense for very long periods of time, due to poor posture and/or repetitive motions as agreed by various researchers [1] [2][3].

Over the years, computer-related injuries (known as repetitive strain injuries or RSIs) have increasingly plagued the modern office workplace, debilitating hundreds of thousands of workers, causing pain, impairment and, in some cases, disability [4]. Several studies linked overuse of computer system with the increase of Repetitive strain injury (Herbert, 2005). With the increase in computer dependency in various works of life, individual who spend more period of time on the computer are prone to a greater risk of developing RSI and it has been observed that individual suffering from RSI tend to be less productive.

RSI is a national tragedy because it has drastically disrupted the lives and livelihoods of many people all over the world [5]. This tragedy is compounded by the fact that it is preventable in most of the cases. However, we need to convince computer users in Nigeria that RSI is a serious public health problem. There is need for proper information on the extent of Repetitive Strain Injury among computer users in Nigeria and the various ways in which it can be prevented in order to aid policy making. In this context, effort has been made to find the root causes of RSI among computer users in Nigeria and provide appropriate awareness to the career threatening syndrome.

1.1 Related Works

Nokubonga Slindele in his study on awkward working postures and precision performance as an example of the relationship between ergonomics and production quality stated that posture adopted has a direct influence on performance variables which are associated with productivity and the quality of product output. Therefore, it provides evidence that ergonomics interventions are relevant and critical in contributing positively to organizational goals simultaneously to taking cognizance of worker needs. And stated further that the results he obtained indicated that, though, precision tasks are

considered to be 'light' tasks, the static muscular contractions required to stabilize individuals in awkward postures are high enough to be a cause for concern with regards to the early onset of fatigue and the precipitation of musculoskeletal injuries. This emphasizes the importance and necessity of workstation design to simultaneously consider performance outcomes and the strain on workers [6].

Adedoyin *et al* did a survey of computer users across six federal university campuses in Nigeria. A frequency analysis was done and his result showed that Low back pain and neck pain were found to be the highest pain complaint with 74% and 73% respectively. 67% of the respondents complained of wrist pain, followed by finger pain (65%), shoulder pain (63%) and general body pain (61%). The knee and foot pains were the least complaints reported with 26% and 25% respectively. In terms of pain severity, low back pain, finger pain, neck pain and shoulder pain are rated to be moderate, while all other joints were said to be of mild pain. His study indicated that low back pain, neck pain and upper limbs are the common disorders complaints among the users. He attributed the cause of the pains to bad ergonomics among the users [7].

Allen E. Akhwa did a survey of computer users in University of Benin, Nigeria on the issue of Occupational Overuse Syndrome (OOS). The data for his study was collected by means of a structured questionnaire administered to respondents (computer users) at the University of Benin campuses comprising staff and students. His results showed that low back pain, neck pain, headache, shoulder pain and eyestrain, are the most prevalent OOS symptoms/pains. He recommended that there is need for computer workplaces to improve on their designs towards finding a lasting solution to the hazardous problem [3].

Sanusi B. also did a survey of incidence of RSI on the students of University of Ibadan, his findings showed that there is correlation and significant relationship between postures maintained by computer users and the incidences of Repetitive Strain Injury and concluded that not keeping a good posture while working on the computer is a major cause of RSI [8].

Many of the studies conducted in Nigeria to investigate the occurrence of RSI focused on its occurrence within the University community but this study survey general computer users from various works of life.

2. Research Methodology

Social survey design was used for this research in order to obtain relevant information from respondents on the incidents of Repetitive Strain Injury (RSI) among computer users in Nigeria. Questionnaire was the instrument used for data collection in this study.

The questionnaire was designed and self-administered to different categories of computer users in various works of life in Nigeria, ranging from banking sector, civil service, educational sector, health sector to private sector.

The distribution was ensured to cut across different field of professionalism. The questionnaire designed was divided into four sections, namely: Bio-data, general question, ergonomic factors, Body strain and ergonomics exercise. This was structured with a checklist of responses.

The data obtained from the questionnaires administered were analyzed using the Statistical Package for the Social Sciences (SPSS) software version 16.0 for windows. The analysis was done in three facets: descriptive analysis through the frequencies procedure which produced frequency tables that displayed both the numbers and percentages of cases for each observed value of variables.

Hypothesis were tested using Pearson Correlation to establish whether or not there is relationship between incidence of RSI and duration of computer usage (years of computer usage and hours of computer usage at a stretch), type of chair used, size and types of monitor used.

2.1 Research Hypothesis

Hypothesis 1

H₀: There is no significant relationship between the duration of computer usage and the incidence of Repetitive Strain Injury.

H_A: There is significant relationship between the duration of computer usage and the incidence of Repetitive Strain Injury.

Hypothesis 2

H₀: There is no significant relationship between the type of chair used when working on the computer and the incidence of Repetitive Strain Injury.

H_A: There is significant relationship between the type of chair used when working on the computer and the incidence of Repetitive Strain Injury.

Hypothesis 3

H₀: There is no significant relationship between the size and type of monitor used and the incidence of eye strain.

H_A: There is significant relationship between the size and type of monitor used and the incidence of eye strain.

2.1.1 Criteria for Rejection and Acceptance of Hypothesis

$$\alpha = 0.01$$

If $p \leq 0.01$ the null hypothesis is rejected

If $p > 0.01$ null hypothesis is accepted

Where α is the level of significant and p-value is the probability that the observed correlation coefficient r was seen by chance.

Linear Regression analysis was done to establish the rate at which ergonomic risk factors explains the incidence of RSI.

2.2 Linear Regression Model Specification

The following models which aided in seeing the causal relationship between the dependent and independent variables are stated below:

$$RSIIndex = f(EF)$$

Where RSIindex – Index of Repetitive Strain Injury

EF – Ergonomic Factors

α 1: coefficient of the independent variables

3. Results

3.1 Demographic Information of Respondents

Five hundred and thirty one questionnaires were administered. The demographic distribution of the respondents according to table 1 indicated the age and gender distribution of the respondents. 59.9% of the respondents were males while 40.1% were females. 1.9% of the respondents were between the age distribution of

The functional form of the model stated above is shown below:

$$RSIIndex = \alpha + \alpha_1 EF$$

A positive relationship is expected between EF and RSIindex because various literature revealed that poor ergonomics has great impact on Repetitive strain Injury. 1-17 years, 47.8% between 18-30 years, and 43.3% between 31-45 years while 7% were 46 years and above.

3.2 Type of Chair Used

Table 2 revealed that 63.7% of the respondents usually sit on fixed chair while working on the computer and the remaining 36.3% sits on adjustable chair throughout their daily computational period.

Table 1: Demographic Information of Respondents

	Frequency	Percent	Valid Percent	Cumulative Percent
Age				
1-17 years	10	1.9	1.9	1.9
18-30 years	254	47.8	47.8	49.7
31-45 years	230	43.3	43.3	93.0
46-65 years	37	7.0	7.0	100.0
	531	100.0	100.0	
Gender				
MALE	318	59.9	59.9	59.9
FEMALE	213	40.1	40.1	100
Total	531	100	100	

Table 2: Type of User's Chair

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fixed	338	63.7	63.7	63.7
	Adjustable	193	36.3	36.3	100.0
	Total	531	100.0	100.0	

3.3 Duration of Computer Usage

Table 3 showed that respondents who used their computer system for 3-4 hours at a stretch takes the highest percentage of 37.5% followed by those that use it for more than 6 hours which is 0.2% greater than

respondents using computer for 4-6 hours with 28.6%. Only 3.8% use computer for 1-2 hours while 1.3% of the respondents rarely use computer. The table also showed that 78.9% of the respondents have been using the computer system for more than 4 years

Table 3: Duration of Computer Usage

		Frequency	Percent	Valid Percent	Cumulative Percent
Years of Computer Usage					
Valid	<1 year	9	1.7	1.7	1.7
	1-3 years	103	19.4	19.4	21.1
	4-7 years	239	45.0	45.0	66.1
	>7 years	180	33.9	33.9	100
	Total	531	100	100	
Hours of Computer Usage at stretch					
Valid	Rarely	7	1.3	1.3	1.3
	1-2 hours	20	3.8	3.8	5.1
	3-4 hours	199	37.5	37.5	42.6
	4-6 hours	152	28.6	28.6	71.2
	>6 hours	153	28.8	28.8	100.0
	Total	531	100.0	100.0	

3.4 Pain Incidences

Respondents were asked to specify the pains experienced and their location. Table 4 above showed the percentage of people that experience different types of pains on parts of the body.

People suffering from eyestrain has the highest percentage with 83.8% followed by Low back pain and wrist pain respectively. Foot pain has the least percentage of people suffering from it followed by ankle pain and elbow pain respectively.

Table 4: Pain Incidences Reported by Respondents

Number of Respondents = 531					
Pains Suffered During Computer use	Never(0)	Mild(1)	Moderate(2)	Severe(3)	**Total Affected
Lowback Pain	30.70%	34.50%	28.80%	6.00%	69.30%
Neck Pain	32.80%	43.30%	20.70%	3.20%	67.20%
Shoulder Pain	55.90%	30.30%	11.30%	2.40%	44.00%
Elbow Pain	75.30%	17.90%	5.80%	0.90%	24.60%
Wrist Pain	32.60%	31.80%	31.10%	4.50%	67.40%
Finger Pain	35.40%	23.20%	36.20%	5.30%	64.70%
Headache	54.80%	28.10%	15.80%	1.30%	45.20%
Arm Pain	40.30%	21.80%	34.80%	3.00%	59.60%
Body Pain	46.00%	31.30%	21.10%	1.70%	54.10%
Hip Pain	68.00%	17.30%	13.90%	0.80%	32.00%
Ankle Pain	81.00%	9.80%	8.70%	0.60%	19.10%
Knee Pain	61.60%	17.10%	19.40%	1.90%	38.40%
Foot Pain	81.00%	9.40%	9.00%	0.60%	19.00%
Eye Strain	16.20%	10.70%	57.10%	16.00%	83.80%
**never excluded					

3.5 Tests for Correlation

3.5.1 Duration of computer Usage and Incidence of RSI

There is a positive and strong correlation between the number of hours spent on the computer at a stretch and the incidence of repetitive strain injury ($r=0.611$, $n=531$). The result showed a significant correlation ($p=0.007$, $p<0.01$) between the number of hours spent on the computer at a stretch and the incidence of repetitive strain injury. The null hypothesis is rejected and alternate hypothesis accepted.

There is also a positive and strong correlation between the number of years spent on the computer and the incidence of repetitive strain injury ($r=0.581$, $n=531$). The result showed a significant correlation ($p=0.001$, $p<0.01$) between the number of years spent on the computer and the incidence of repetitive strain injury. The null hypothesis is rejected and alternate hypothesis accepted (see table 5).

Table 5: Correlations of Duration of Computer Usage and Pain Incidence

		Pain Incidence on Parts of the Body during and after Computer Use		
		Computer Use	at a Stretch	Years of Computer Usage
Pain Incidence on Parts of the Body during and after Computer Use	Pearson Correlation	1		
	Sig. (2-tailed)			
	N	531		
Computer Use Hour at a Stretch	Pearson Correlation	.611**	1	
	Sig. (2-tailed)	.007		
	N	531	531	
Years of Computer Usage	Pearson Correlation	.581**	.607**	1
	Sig. (2-tailed)	.001	.000	
	N	531	531	531

** . Correlation is significant at the 0.01 level (2-tailed).

3.5.2 Chairs Used and Incidence of RSI

The types of chairs used on the computer system was negatively correlated with incidence of low back pain ($r=.538$, $n=532$, $p=.003$ $P<.01$). The correlation is significant at 0.01 levels. This indicate that the more use

of fixed chairs was associated with incidence of repetitive strain injury. The null hypothesis is therefore rejected and the alternate hypothesis accepted (see table 6)

Table 6: Correlations of Types of user's chair and incidence of low back pain

		Incidence of Low back Pain	Type of User's Chair
Incidence of Low back Pain	Pearson Correlation	1	.538**
	Sig. (2-tailed)		.003
	N	531	531
Type of User's Chair	Pearson Correlation	.538**	1
	Sig. (2-tailed)	.003	
	N	531	531

** . Correlation is significant at the 0.01 level (2-tailed).

3.5.3 Type and Size of Monitor Used and Incidence of RSI

There is a positive correlation ($r=0.442$) between the types of monitor used and incidence of Eye Strain. The result is significant at $p=0.004$ ($p<.01$) therefore the null hypothesis is rejected and alternate hypothesis accepted (see table 7). The result showed that there is a significant correlation between monitor screen size and the

incidence of repetitive strain injury, $r= -.519$, $n=452$, $p<.01$, two tails. Monitor screen size is positively associated with the incidence of Repetitive Eye Strain. The null hypothesis is therefore rejected and the alternate hypothesis accepted (see table7)

Table 7: Correlations of Incidence of Eyestrain, Types of Monitor Used and Monitor Screen Size

		Incidence of Eye Strain	Types of Monitor Used	User's Monitor Screen Size
Incidence of Eye Strain	Pearson Correlation	1	.442**	-.519**
	Sig. (2-tailed)		.004	.001
	N	531	452	452
Types of Monitor Used	Pearson Correlation	.442**	1	.381**
	Sig. (2-tailed)	.004		.009
	N	452	452	449
User's Monitor Screen Size	Pearson Correlation	-.519**	.381**	1
	Sig. (2-tailed)	.001	.009	
	N	452	449	452

3.6 Regression Model to Establish Effect of Ergonomic Factors on RSI

RSI index (0 - 1) that is (no strain - strained)

Relation

Dependent: RSI index

Independent: Ergo risk Factors

Analysis

$R=0.808$

The level of correlation among RSIindex (dependent variables) and ErgoFactors (predictor) is 0.907. This result indicates that there is a strong correlation between Ergofactors and RSIindex based on RSIindex at 90.7%. $R^2 = 0.822$.

This showed that the analysis is adjudged accurate at 82.2% and that about 82.2% variation in the dependent variable is being explained by the variation in the independent variables while the remaining 17.2% variation cannot be explained by the variation in the independent variable. (Good).

3.6.1 Test for Regression Parameters

$H_0: \alpha_1 = \alpha_2; H_1: \alpha_1 \neq \alpha_2$

$$F_{cal} = \frac{MS_{Regression}}{MS_{Error(residual)}}$$

$$F_{cal} = \frac{2.093}{0.042} = 69.453$$

$$F_{cal} = 69.453$$

Test statistics F- Test

$$F_{tab} = F_{(V-1) (N-V)}^{\alpha}$$

$$F_{(2)(3)}^{0.05} = P_{value} = 0.000^a$$

Decision Rule

If $F_{cal} < F_{tab}$, accept H_0 (accept the null hypothesis)

If $F_{cal} > F_{tab}$, do not accept H_0 (accept alternative hypothesis)

Comparism

$$F_{cal} = 69.453$$

$$F_{tab} = 0.000$$

Conclusion

While $F_{cal} (69.453) > F_{tab} (0.000)$, we do not accept H_0 , therefore we accept H_1 and conclude that the regression parameter is significant for RSIindex using ErgoFactors.

Model

$$RSIindex = -2.67 + 0.664ErgoFactors$$

Interpretation

1. Sign (+ or -)
The model shows a positive relation between ErgoFactors and RSI index. This implies that ErgoFactor is a corresponding measure of RSI index, that is, increases in non-compliance of ergonomic way of using the computer system and setting up workstation, the more the RSI.
2. Change
A unit change in ErgoFactor will result in correspondent 66.4% increase in RSI. In addition the model has high RSI index based ErgoFactor. This is true because non-compliance with the correct use of computer system, correct way of setting up workstation result in RSI.

4. Discussion

It was found in this study that many people suffered from pains from one part of the body to another during and after working on the computer system, some suffered pain from only one part of the body while majority suffered pains from more than one parts of the body. 95.4% of the respondents said they have incidence of pains while only 5.5% claimed they do not suffered from any form of pain.

The result also showed that more than 80% of the population of computer users in Nigeria lies between the age of 18 and 45 years. Over 63% of the population under study used a fixed chair while working on the computer system while the rest sit on ergonomically designed chairs that are adjustable. This indicates that more people will have pains related to bad sitting position while working for a period of time at a stretch. It was also discovered from the survey that 87.4% of the respondents who suffered from one pain or the other are those that usually work on the computer system for more than three hours at a stretch. This indicates that the more hours one spent on the computer system, the higher the proximity of having RSI from different parts of the body. The same percentage of respondents suffers joint pain while typing on the computer system for up to 4 hours.

5. Conclusion

Our findings have generally linked long duration of computer usage, usage of ergonomically defective chairs, and bad ergonomic work station to incidence of Repetitive Strain injury. We also discovered that with increased computer literacy in the developing countries, many users' workplaces still lack proper design, management strategies, and task design. This showed that proper attention has not been given to the need to minimize or eradicate workplace hazard.

4.1 Recommendations

1. Enlightenment programs should be created especially in the university communities in

- Nigeria to educate students on preventive measures to overcome repetitive strain injury.
2. Establishment of ergonomic programs such as seminars and trainings, these bring awareness and consciousness on the safety measures for the working conditions to all computer users to remedy and prevent workplace injuries.
3. Computer users should cultivate the habit of taking short fixed breaks during their daily computer work to perform different types of ergonomic exercises.
4. The media houses should also create regular awareness on this career threatening syndrome.
5. Workplaces should put proper preventive guidelines in place while those already having ergonomics guidelines should monitor for continual research adherent to it.

Therefore, our findings generally recommend the need to improve computer workplaces in terms of design, management strategies, creation of awareness program and usage of ergonomically designed equipment. When this is done, it will go a long way to avoiding poor posture during computer work and thereby reduce the incidence of RSI.

References

- [1] Pascarelli E. (2004): "Complete Guide to Repetitive Strain Injury", John Willey and Sons, Inc., Hoboken, New Jersey, Pp. 9 and 202.
- [2] Webster J. and Gonzalez M. (2006): "Repetitive Strain Injury" Fact Sheet AHS-06 Utah State University Cooperative Extension, available at extension.usu.edu/agriculturehealthand safety accessed on 04/05/2011
- [3] Allen E. (2007): "Occupational Overuse Syndrome (OOS): a recurring problem in the Computer Industrially developing Countries (IDCs) a case study at University of Benin, Benin City Nigeria." *HFES Journal, Ergonomics Australia* Vol. 21, Number 3 pp. 10-18
- [4] Vivienne G. and Norman J.(2001): "The Typing Injury FAQ, Motion-Based Ergonomics KeyboardRetraining" Available on <http://www.tifaq.org/ergonomics.html> (Accessed on 22/04/2010)
- [5] Nivedita (2007): "The Effect of Stretching and Strengthening exercises as a treatment Program for RSI of Upper Extremity in Computer Professionals", Rajiv Ghadhi University of Health Sciences, Karnakata, Banglore, Unpublished.
- [6] Nokubonga S. (2009): Awkward Working Postures and Precision Performance as an example of the Relationship between Ergonomics and Production Quality, unpublished.
- [7] Adedoyin R. A, Idowu B.O, Adagunodo R. E & Idowu P.A (2005): "Musculoskeletal pain associated with the use of computer systems in Nigeria", IOS Press Amsterdam, The Netherlands, the Netherlands. Vol. 13 issue 2, p125-130.
- [8] Sanusi B, (2011): "investigation of the incidence of repetitive strain injury among computer users" University of Ibadan, Nigeria. Unpublished.

Biography

Olabiyisi, S. O. received his B. Tech., M. Tech and Ph.D degrees in Mathematics from LadokeAkintola University of Technology, Ogbomoso, Nigeria, in 1999, 2002 and 2006 respectively. He also received M.Sc. degree in Computer Science from University of Ibadan, Ibadan, Nigeria in 2003. He is currently an Associate Professor in the Department of Computer Science and Engineering, LadokeAkintola University of Technology, Ogbomoso, Nigeria. He has published in reputable journals and learned conferences. DrOlabiyisi is a full member of Computer Professional (Registration) Council of Nigeria (CPN). His research interests are in computational mathematics, theoretical computer science, information systems and performance modelling and simulation.

Akingboye, A.Y. graduated from the Ladoke Akintola University of Technology, Ogbomoso with a Master of Technology (M.Tech) and Bachelor of Technology (B.Tech) degrees in Computer Science in 2012 and 2005 respectively. His research interests include Microprocessor Systems and Human Computer Interaction. He is presently with the Department of Electrical and Computer Engineering at Igbinedion University Okada.

Abayomi-Alli, A. obtained his B.Tech Degree in Computer Engineering from Ladoke Akintola University of Technology (LAUTECH), Ogbomoso in 2005, MSc Computer Science from the University of Ibadan, Nigeria in the 2009. He started his career at Igbinedion University Okada, Nigeria as a Graduate

Assistant in 2007 before moving to the Federal University of Agriculture Abeokuta (FUNAAB), Nigeria in 2011. His current research interests include microprocessor systems and applications, biometrics, image quality assessment and machine learning.

Izilein, F. A. has a B.Eng and M.Eng Degree in ElectricalElectronics. He presently lectures at the Department ofElectrical and Computer Engineering in Igbinedion University Okada.

Adeleke, I. A. graduated from the Ladoke Akintola University of Technology, Ogbomoso with a Bachelor of Technology (B.Tech) degrees in Computer Engineering in 2005. His research interests include Microprocessor Systems and Electronics. He is presently with the Department of Electrical and Computer Engineering at Igbinedion University Okada.

Improving the methods of email classification based on words ontology

Foruzan Kiamarzpour¹, Rouhollah Dianat², Mohammad bahrani³, Mehdi Sadeghzadeh⁴

¹ Department of Computer Engineering, Science and Research Branch of Bushehr, Islamic Azad University, Bushehr, Iran

² Department of Computer Engineering, Faculty of Engineering University, Qom University, Qom, Iran

³ Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

⁴ Department of Computer Engineering, mahshar branch, Islamic azad University, Mahshahr, Iran

Abstract

The Internet has dramatically changed the relationship among people and their relationships with others people and made the valuable information available for the users. Email is the service, which the Internet provides today for its own users; this service has attracted most of the users' attention due to the low cost. Along with the numerous benefits of Email, one of the weaknesses of this service is that the number of received emails is continually being enhanced, thus the ways are needed to automatically filter these disturbing letters. Most of these filters utilize a combination of several techniques such as the Black or white List, using the keywords and so on in order to identify the spam more accurately. In this paper, we introduce a new method to classify the spam. We are seeking to increase the accuracy of Email classification by combining the output of several decision trees and the concept of ontology.

Keywords: *Data Mining, Email classification, spam detection, decision tree, ontology .*

1 . Introduction

Unwanted spam is sent to the users with different purposes. According to the estimates by Ferris in his research [1], 15 to 20% of emails are spam letters and about half of the users are receiving 10 or more spam every day. Sometimes, some of them receive more than a few hundred unwanted emails. International Data Group [2] has estimated that the traffic of email is more than 60 billion messages daily by 2006.

Nowadays, most of the measures, done on the spam filtering, are based on the techniques such as classifications of Naïve

Bayesian, neural networks, and.... We have provided a framework for efficient spam filtering by using the ontology. Ontology makes sense for the machines to understand the meaning of data [3]. This paper represents an efficient spam filtering method by using the ontology. We have used the SpamBase Dataset[4] and Weka and Jena in order to make the ontology. Using ontology that is specially designed to filter spam, bunch of unsolicited bulk email could be filtered out on the system.

This paper proposes to find an efficient spam email filtering method using ontology. We used Waikato Environment for Knowledge Analysis (Weka) explorer, and Jena to make ontology based on sample dataset.

Ontology is a kind of computational model of the parts in the world and is often shown by the help of semantic networks.

The ontology is in fact an agreement on a common conceptualization and contains the frameworks for modeling the domain knowledge and the agreements in the field of how some of its theories are represented. [5]

Using the meaning of data, the use of electronic technology becomes much easier. There are multiple languages such as XML, RDF, RDF-schema (RDFS), DAML+OIL, and OWL in order to make the ontology and numerous tools for development and implementation of ontology metadata by using these languages .

We will have the following sections in this paper: In Section 2, the research background and related works are presented; in Section 3, the text classification is described; in section 4, the article idea is provided by using the ontology; and in Section 5 the experiments have been presented by using the proposed framework and Finally, in Section 6 we have mentioned the conclusion and future works.

2 . Research background and related works

Some of the conducted studies in the field of Email filtering are described as follows:

Recently the evolution of the Web has attracted interest in defining features, signals for ranking [6] and spam filtering [7, 8, 9, 10, 11]. The earliest results investigate the changes of Web content with the primary interest of keeping a search engine index up-to-date [16, 17].

Many researchers also suggested new systems using other methods such as Bayesian network enhancing the performance of Bayesian classification[12], and WBC (Weighted Bayesian Classifier) that gives weight on some key terms that representing the content's class by SVM(Support Vector Machine)[13]

In this paper we used an ontology for Email Classification ,In additional, j48, ADTree and LADTree was used to classify training dataset.

3. Text Classification

Text classification uses the data mining techniques. Most of the data mining tasks are based on the databases and data structures. Some of these methods are described as follows.

3.1 Support Vector Machines(SVM)

Support Vector Machines are supervised learning methods used for classification, as well as regression. The advantage of Support Vector Machines is that they can make use of certain kernels in order to transform the problem, such that we can apply linear classification techniques to non-linear data. Applying the kernel equations arranges the data instances in such a way within the multi-dimensional space, that there is a hyper-plane that separates data instances of one kind from those of another.

The kernel equations may be any function that transforms the linearly non-separable data in one domain into another domain where the instances become linearly separable. Kernel equations may be linear, quadratic, Gaussian, or anything else that achieves this particular purpose.

3.2 Naïve Bayes Classifier

The Naïve Bayes classifier works on a simple, but comparatively intuitive concept. Also, in some cases it is also seen that Naïve Bayes outperforms many other comparatively complex algorithms. It makes use of the variables contained in the data sample, by observing them individually, independent of each other.

The Naïve Bayes classifier is based on the Bayes rule of conditional probability. It makes use of all the attributes contained in the data, and analyses them individually as though they are equally important and independent of each other.

3.3 J48 Decision Trees

A decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. The internal nodes of a decision tree denote the different attributes, the branches between the nodes tell us the possible values that these attributes can have in the observed samples, while the terminal nodes tell us the final value (classification) of the dependent variable.

The attribute that is to be predicted is known as the dependent variable, since its value depends upon, or is decided by, the values of all the other attributes. The other attributes, which help in predicting the value of the dependent variable, are known as the independent variables in the dataset.

3.4 Alternating Decision Trees(ADTree)

ADTree [14] is a boosted DT. An ADTree consists of prediction nodes and splitter nodes. The splitter nodes are defined by an algorithm test, as, for instance, in $C4.5$, whereas a prediction node is defined by a single value $x \in R^2$. In a standard tree like $C4.5$, a set of attributes will follow a path from the root to a leaf according to the attribute values of the set, with the leaf representing the classification of the set. In an ADTree, the process is similar but there are no leaves. The classification is obtained by the sign of the sum of all prediction nodes existing in the path. Different from standard trees, a path in an ADTree begins at a prediction node and ends in a prediction node.

3.5 LADTree

LADTree [15] produces an ADTree capable of dealing with data sets containing more than two classes. The original formulation of the ADTree restricted it to binary classification problems; the LADTree algorithm extends the ADTree algorithm to the multi-class case by splitting the problem into several two-class problems.

Logical Analysis of Data is the method for classification proposed in optimization literature. It builds a classifier for binary target variable based on learning a logical expression that can distinguish between positive and negative samples in a data set. The basic assumption of LAD model is that a binary point covered by some positive patterns, but not covered by any negative pattern is positive, and similarly, a binary point covered by some negative patterns, but not covered by positive pattern is negative. The construction of Lad model for a given data set typically involves the generation of large set patterns and the selection of a subset of them that satisfies the above assumption such that each pattern in the model satisfies certain requirements in terms of prevalence and homogeneity.

4 . Spam filtering by using the ontology

4.1 General Approach

The first step is to make a smart decision tree, and then we obtain the ontology based on the classification of trees j48, AD and

LAD and use the format RDF based on the object, subject, and prediction for creating the ontology.

The second step is to map the decision tree to the ontology and then get a query from the obtained ontology and give it a test Email and determine whether it is spam or not.

First, we should collect a good database. These data should consider the features of a valid email and junk mail. In this paper, we have used the database spambase [4] which contains 4601 emails of which 39.4% are the spam and 60.6% are the valid emails. We have evaluated a number of decision trees in Weka and have decided to use the trees j48, AD and LAD. The input format of data has the format Arff; we convert the training data into the format Arff and build a decision tree based on the training data. All leaves have the values equal to 0 or 1 in the classification and if it is 0, the email is valid and if is 1, it is the spam. This decision tree is a kind of ontology.

4.2 Architecture and Implementation

The following figure shows our proposed framework for filtering the spam. The training package is a set of emails which provide the result of classification for us. We have classified the training package by using the decision trees j48, AD, and LAD. As a result of data test, a new email entered system, is properly classified and it is determined whether it is spam or not.

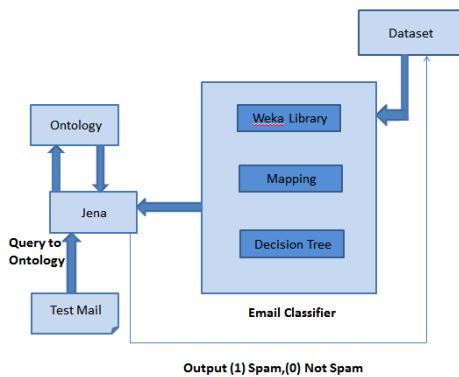


Fig. 1 The architecture of spam filtering by the help of ontology

Therefore, we create the decision trees by using the software Weka and since the decision tree is considered as a kind of ontology, we convert it into the proper format of ontology and ensure its accuracy; we have chosen RDF for creating the ontology (Figure 5)

```
word_freq_our <= 0.71
| word_freq_your <= 1.04: 0 (824.0/29.0)
| word_freq_your > 1.04
| | word_freq_technology <= 0.04
| | | capital_run_length_longest <= 10: 0 (87.0/5.0)
| | | capital_run_length_longest > 10
| | | word_freq_over <= 1.12
| | | | char_freq ( <= 0.064
| | | | | char_freq ! <= 0.154
| | | | | word_freq_will <= 0.95: 0 (8.0)
| | | | | word_freq_will > 0.95
| | | | | word_freq_will <= 1.71: 1 (4.0)
| | | | | word_freq_will > 1.71: 0 (4.0/1.0)
| | | | | char_freq ! > 0.154: 1 (2.0)
| | | | | char_freq ( > 0.064: 0 (16.0)
| | | | | word_freq_over > 1.12: 1 (2.0)
| | | word_freq_technology > 0.04
| | | word_freq_technology <= 0.54: 0 (4.0)
| | | word_freq_technology > 0.54: 1 (3.0)
| word_freq_our > 0.71
```

Fig. 2 Part of j48 Classification Result

Figure 2 shows how we choose the J48 classification filter, which uses the simple c4.5 decision tree for classification.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:eeeg="http://relation/" >
  <rdf:Description rdf:about="http://DecisionTree/Test34">
    <eeeg:isElementOf rdf:resource="http://DecisionTree/Test34/output"/>
    <eeeg:value>0.43</eeeg:value>
    <eeeg:operator>=</eeeg:operator>
    <eeeg:attribute>word_freq_our</eeeg:attribute>
  </rdf:Description>
  <rdf:Description rdf:about="http://DecisionTree/Test85">
    <eeeg:isElementOf rdf:resource="http://DecisionTree/Test85/output"/>
    <eeeg:value>1.17</eeeg:value>
    <eeeg:operator>=</eeeg:operator>
    <eeeg:attribute>word_freq_our</eeeg:attribute>
  </rdf:Description>
  <rdf:Description rdf:about="http://DecisionTree/Test60/output">
    <eeeg:info>(0.0)</eeeg:info>
    <eeeg:decision>28</eeeg:decision>
  </rdf:Description>
  <rdf:Description rdf:about="http://DecisionTree/Test6">
    <eeeg:isElementOf rdf:resource="http://DecisionTree/Test6/output"/>
    <eeeg:value>0.11</eeeg:value>
    <eeeg:operator>=</eeeg:operator>
    <eeeg:attribute>word_freq_our</eeeg:attribute>
  </rdf:Description>
  <rdf:Description rdf:about="http://DecisionTree/Test16">
    <eeeg:isElementOf rdf:resource="http://DecisionTree/Test16/output"/>
    <eeeg:value>0.24</eeeg:value>
    <eeeg:operator>=</eeeg:operator>
  </rdf:Description>
```

Fig. 3 RDF File made for the result of classifying the tree J48

Figure 3 shows the RDF file created based on J48 classification result. The RDF file was used as an input to Jena to create an ontology which will be used to check if the test email is spam or not.

Your RDF document validated successfully.

Triples of the Data Model			
Number	Subject	Predicate	Object
1	http://DecisionTree/Test34	http://relation/isElementOf	http://DecisionTree/Test34/output
2	http://DecisionTree/Test34	http://relation/value	"0.43"
3	http://DecisionTree/Test34	http://relation/operator	"="
4	http://DecisionTree/Test34	http://relation/attribute	"word_freq_our"
5	http://DecisionTree/Test85	http://relation/isElementOf	http://DecisionTree/Test85/output
6	http://DecisionTree/Test85	http://relation/value	"1.17"
7	http://DecisionTree/Test85	http://relation/operator	"="
8	http://DecisionTree/Test85	http://relation/attribute	"word_freq_our"
9	http://DecisionTree/Test60/output	http://relation/info	"(0.0)"
10	http://DecisionTree/Test60/output	http://relation/decision	"28"
11	http://DecisionTree/Test6	http://relation/isElementOf	http://DecisionTree/Test6/output
12	http://DecisionTree/Test6	http://relation/value	"0.11"
13	http://DecisionTree/Test6	http://relation/operator	"="
14	http://DecisionTree/Test6	http://relation/attribute	"word_freq_our"
15	http://DecisionTree/Test16	http://relation/isElementOf	http://DecisionTree/Test16/output
16	http://DecisionTree/Test16	http://relation/value	"0.24"

Fig. 4 Ternaries of the model RDF

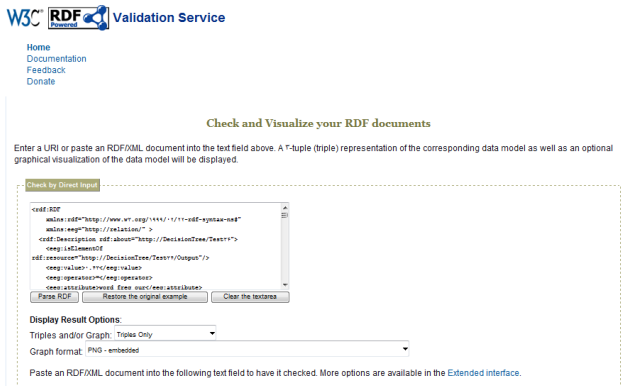


Fig.5 W3C RDF Validation Services

Figure 5 shows RDF validation services. W3C RDF validation services help us to check whether the RDF schema which we are going to give as input to Jena is syntactically correct or not.

5. experiments

We have used 4101 emails in training session and have built the trees by the help of software Weka and converted them to the ontology format. We get the query from this ontology for the test stage and give them 500 test emails for classifying into two groups of spam and valid emails.

The results of processing and all features are shown in the following table; first we investigate the spam class (1) and the valid email class (0):

Definitions of cases mentioned in the next tables and charts:

TP: The spam, which is detected as the spam properly; in other words, the record of test data with the original class 1 is put in the Class 1 by the classifier.

FP: The valid email, which is predicted as the spam by mistake; in other words, the record of test data with the original class 0 is put in the Class 1 by the classifier.

TN: The valid email, which is predicted as the valid email properly; in other words, the record of test data with the original class 0 is put in the Class 0 by the classifier.

FN: The spam, which is predicted as the valid email by mistake; in other words, the record of test data with the original class 1 is put in the Class 0 by the classifier.

$$Precision - Class 1 = \frac{TP}{(TP+FP)} \quad (1)$$

$$Recall - Class 1 = \frac{TP}{(TP+FN)} \quad (2)$$

$$F - Measure = \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

$$Precision - Class 0 = \frac{TN}{(TN+FN)} \quad (4)$$

$$Recall - Class 0 = \frac{TN}{(TN+FP)} \quad (5)$$

Table 1: Results obtained from the class processing of Spam by involving all features

	TP	FP	Precision	Recall	F-Measure
SVM	0.757	0.024	0.959	0.757	0.846
Naive Bayesian	0.963	0.231	0.757	0.963	0.848
ADTree	0.958	0.075	0.957	0.925	0.940
LADTree	0.955	0.112	0.952	0.955	0.924
J48	0.951	0.112	0.948	0.951	0.922
Voting	0.937	0.07	0.937	0.93	0.933

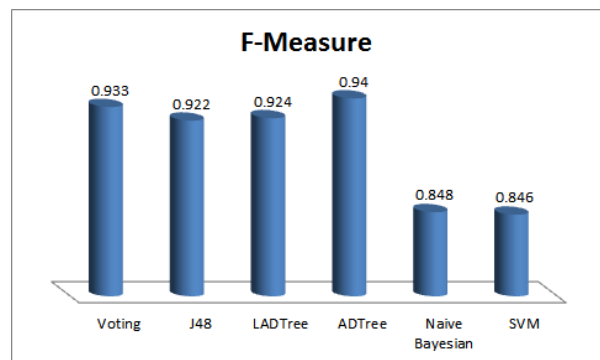


Fig. 6 Comparison of F-Measure, The results obtained from processing the spam class with involving all features

Table 2: Results obtained from processing the class of valid email

	TN	TN	Precision	Recall	F-Measure
SVM	0.976	0.243	0.843	0.976	0.904
Naive Bayesian	0.769	0.037	0.965	0.769	0.856
ADTree	0.925	0.042	0.927	0.958	0.942
LADTree	0.888	0.045	0.895	0.888	0.918
J48	0.888	0.049	0.895	0.888	0.917
Voting	0.941	0.065	0.935	0.941	0.937

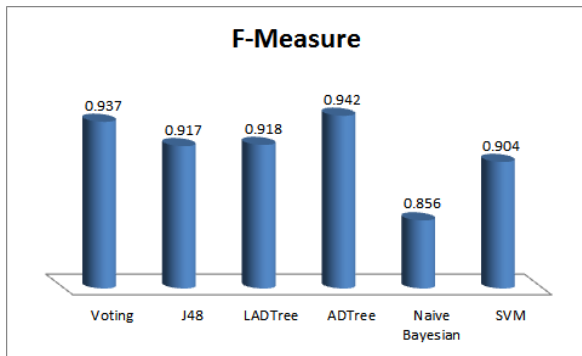


Fig. 7 Comparison of F-Measure obtained from processing the class of valid email with involving all features

As we can see, our proposed method has been able to be more successful than the previous methods in both classes of spam and valid email.

6- Conclusion and Future Work

In this paper, we have Used Weka software to produced the decision trees AD, LAD and J48 and and have converted them into the format of ontology (RDF) and ensured their accuracy, and then the test emails are given to them and we have voted for the predicted results and finally we have compared the obtained results with the results of two methods SVM and Naive Bayesian which are the most common Email classification methods; thus we have found that the results obtained from voting the decision trees between two errors of considering the spam instead of valid email (FN) and a valid email instead of spam (FP) establish a reasonable balance.

Pruning and narrowing the constructed tree as a basis of the proposed ontology can be done as the future work. Furthermore, the classifiers such as Conjunctive-Rule and other non-tree classifiers can be used as well as the tree classifiers in voting step in order to achieve the higher accuracy.

References:

[1] Ferris Research, "Spam Control: Problems & Opportunities", 2003.

[2] International Data Group, "Worldwide email usage 2002 - 2006: Know what's coming your way", 2002.

[3] S.Youn, and D. McLeod, " Ontology Development Toolsfor Ontology-Based Knowledge Management". Encyclopediaof E-Commerce, E-Government and Mobile Commerce,Idea Group Inc., 2006.

[4] Spambase dataset available in:
<http://archive.ics.uci.edu/ml/datasets/Spambase>

[5] A.S. Saffarian, "Service-oriented architecture is rich with meaning", Ms Thesis ,Tehran,Iran, Shahid Beheshti University, 2007.

[6] A. Dong, Y. Chang, Z. Zheng, G. Mishne, J. Bai, K. Buchner, R. Zhang, C. Liao, and F. Diaz." Towards recency ranking in web search". In Proc. WSDM, 2010.

[7] N. Dai, B. D. Davison, and X. Qi." Looking into the past to better classify web spam". In AIRWeb '09: Proceedings of the 5th international workshop on Adversarial information retrieval on the web. ACM Press, 2009.

[8] M. Erdélyi, A. A. Benczúr, J. Masanés, and D. Siklósi. "Web spam filtering in internet archives". In AIRWeb '09: Proceedings of the 5th international workshop on Adversarial information retrieval on the web. ACM Press, 2009.

[9] Y. joo Chung, M. Toyoda, and M. Kitsuregawa. "A study of web spam evolution using a time series of web snapshots". In AIRWeb '09: Proceedings of the 5th international workshop on Adversarial information retrieval on the web. ACM Press, 2009.

[10] Y. Lin, H. Sundaram, Y. Chi, J. Tatemura, and B. Tseng. "Splog detection using content", time and link structures. In 2007 IEEE International Conference on Multimedia and Expo, pages 2030–2033, 2007.

[11] G. Shen, B. Gao, T. Liu, G. Feng, S. Song, and H. Li. "Detecting link spam using temporal information". In ICDM'06., pages 1049–1053, 2006.

[12] M. Sahami,S. Dumais, D. Heckerman,E. and Horvitz, "A Bayesian Approach to Filtering Junk Email". In Learning for Text Categorization , Proc. Of the AAAI Workshop,Madison Wisconsin. AAAI Technical Report WS-98-05, 55-62, 1998.

[13] G. Thomas, A.F and Peter, "Weighted Bayesian Classification based on Support Vector Machine Learning", 207-209, 2001.

[14] Y. Freund and R. E. Schapire."A Decision-Theoretic generalization of on-line learning and an application to Boosting". JCSS, 55:119–139, 1997

[15] G. Holmes, B. Pfahringer, R. Kirkby, E. Frank, M. Hall, "Multiclass Alternating Decision Trees", ECML, 161-172, 2002.

[16] J. Cho and H. Garcia-Molina. "The evolution of the web and implications for an incremental crawler". In The VLDB Journal, pages 200–209, 2000.

[17] J. Cho and H. Garcia-Molina." Synchronizing a database to improve freshness. In Proceedings of the International Conference on Management of Data", pages 117–128, 2000.

“OpenMP” automatic parallelization tools: An Empirical comparative evaluation

Emna KALLEL, Yassine AOUDNI and Mohamed ABID

Sfax University, National school of Engineers of Sfax,
3038, Tunisia,

Abstract

Today, multi-core design has become the trend of enhancing the processor's performance, and most industries have been considering multi-core as the future of development. Thus, a programmer or a compiler explicitly parallelizes the software, which is the key to enhance the performance on multi-core design. Nevertheless, currently, needs an in-depth knowledge of both software and hardware design to develop parallel applications. Automatic parallelization is one of the approaches aiming at a better and easier use of parallel computers. In recent years, several research auto-parallelization tools appeared. However, the automatic parallelization is yet to become a widely adopted industrial practice. This paper presents an empirical comparison between three research tools, namely CETUS, PLUTO and GASPARD. Indeed, we discuss the success of these tools to automatically generate OpenMP parallel codes from serial C codes and compare them using known benchmark C workloads and some evaluation metrics.

Keywords: Multi-core, automatic parallelization, OpenMP, metrics, parallel benchmarks.

1. Introduction

While classical parallel machines serve a relatively small user community, multi-cores aim to capture a mass market, which targets user-oriented, high-productivity programming tools. Furthermore, multi-cores are replacing complex superscalar processors, the parallelism of which was unquestionably exploited by the compiler and underlying architecture [6]. However, while developing a parallel application, multiple challenges, ranging from the extraction of the

application parallelism to the management of the communications between processors, are to be overcome. Dating back to more than 15 years, the automatic parallelization approaches have appeared as one of solutions that aimed at a better and easier use of parallel computers. It consists in taking a program written for a sequential computer (which has only one processor) and to adapt it to a parallel computer. The condition for executing loop iterations in parallel is that the variable usage in the loop does not result in loop-carried dependences. So, the auto-parallelizing compilers try to analyze codes and eliminate these dependences to automatically generate parallel codes. Building an auto-parallelizing compiler has proven to be a very difficult task when producing efficient code in all cases for a wide variety of applications.

Multiple research auto-parallelization projects have been developed to address some of these problems. For example, ROSE [9] is an open source compiler. It builds source-to-source program transformation and analysis tools for large-scale FORTRAN, C, C++, and OpenMP. CETUS [6] is a source-to-source C compiler written in Java and maintained at Purdue University. CETUS provides automatic parallelization, and many other applications have emerged. CETUS 1.2.1 release provides OpenMP to-CUDA package. PAR4ALL [10] is an open-source environment to do source-to-source transformations on C and FORTRAN programs for parallelizing, optimizing, reverse engineering, etc. on various targets, from multicore system-on-chip with some accelerators up to high performance computers and GPU. Also, there are other compilers which can automatically transform serial C/C++ codes to parallel C/C++ codes or to parallel programs by using OpenMP directives [15] or CUDA [5]. In recent years, several commercial compilers that parallelize sequential code also appeared (eg. Intel's[19], IBM's[20] compilers). Despite all those efforts, the automatic parallelization is yet to become a widely adopted industrial practice. The

most important reason is the quality of the generated parallel code, resulting in inefficient hardware utilization and low performance gain.

This paper studies the above outlined problems through an empirical comparison between three available research tools, namely CETUS, PLUTO and GASPARD. The objective of our work is to have an experiment on some auto-parallelization tools, to let us know if these tools are really available, or if they are actually effective when they are used to enhance the performance. Some parallel benchmarks are used through these compilers. To compare their performance some parallel program performance metrics are adopted.

The rest of this paper is organized as follows. Section 2 presents automatic parallelization techniques and some related work. Section 3 describes the analysis model proposed for the experimentation, while section 4 presents the obtained results. Finally conclusions and future work are presented in section 5.

2. Automatic parallelization techniques

Several works in parallel programming area aim to automatically parallelize a sequential code through many analyses programmes. On an academic scale, we find several auto-parallelization tools. In this section, we separated these auto-parallel tools into two types, one is source to source and the other is model to source parallelizer.

2.1 Source-to-Source auto-parallelizer

Many compute-intensive applications often spend most of their execution time in nested loops. The Polyhedral Model provides a powerful abstraction to reason about transformations on such loop nests by viewing a dynamic instance of each statement as an integer point in a well-defined space called the statement's polyhedron [7]. Also, the polyhedral model allows representing easily regular calculations. He allows specifying, examining, analyzing, transforming, and parallelizing affine equations and so generating regular parallel architecture. It can treat programs with static control structure and affine references/loop-bounds. Also, codes with non-affine array access functions or code with dynamic control can be handled, but only with conservative hypothesis on some dependences. Automatic parallelization efforts in the polyhedral model broadly fall into two classes: (1) scheduling/allocation-based, and (2) partitioning based. Griebel [18] (to some extent) fall into the former

class, while Lim/Liao's approach [14] falls into the second class. Griebel [18] presents an integrated framework for optimizing locality and parallelism with space and time tiling, by treating tiling as a post-processing step after a schedule is found. Lim and Liao [14] proposed a framework that identifies outer parallel loops (communication-free space partitions) and permutable loops (time partitions) to maximize the degree of parallelism and minimize the order of synchronization. They employ the same machinery for blocking. Several solutions equivalent in terms of the criterion they optimize for result from their algorithm, and these significantly differ in performance.

No metric is provided to differentiate between these solutions as maximally independent solutions are sought, without using any cost function. As shown through this work, without a cost function, solutions obtained even for simple input may be unsatisfactory with respect to communication cost, locality, and target code complexity. The proposed approach in [7] is closer to the latter class of partitioning-based approaches. However, it is the first to explicitly model tiling in a polyhedral transformation framework, thereby enabling the effective extraction of coarse-grained parallelism along with data locality optimization. At the same time, codes which cannot be tiled or only partially tiled are all handled, and traditional transformations are captured. In fact, it presents an automatic polyhedral source-to-source transformation framework PLUTO that can optimize regular programs for parallelism and locality simultaneously. This framework has been implemented into a tool to automatically generate OpenMP parallel code from C program sections.

However, the front end used in PLUTO accepts only a very little nested loops set. An important limitation of polyhedral parallelizer compilers: there isn't enough information during the compilation to generate a parallel code. For example when the parallelism of the application depends on input data, the compiler is not able to parallelise the program. This problem can be resolved by the taking advantage of the compilation directives which can indicate to the compiler how to decompose the parts of a sequential program for its parallel executing. Automatic parallelization by directives insertion is the set of code transformations at the code source level where the parallelism is expressed by directives given to the compiler. These directives allow the programmer a separation of the preoccupations correction and efficiency [1]: they are a good means to optimize the programmes without questioning the correctness of them. Directives of compilation are used to facilitate the extraction of the parallelism as well as to help in the placement of the calculations and the data on processors. This

methodology is used for example in HPF [4]. In this work, the programmer supplies the directives of partitioning and placement to obtain the placement of the calculations on processors. More than the directives of placement, we distinguish another directive type: the directives of scheduling for shared memory architectures (such as the parallel directive of OpenMP [15]). For example, the CETUS tool [6] provides an infrastructure for research on multicore compiler optimizations that emphasizes automatic parallelization. The CETUS compiler translates OpenMP directives into text strings and stores them in the Intermediate Representation (IR). Then, the OpenMP parser analyzes these text strings to convert them into Annotation, which is a CETUS' map data structure that contains processed OpenMP directive information. For example, "#pragma omp parallel private(a, b)" in the input OpenMP program is transformed into a single text string, which is converted into Annotation map data structure whose (key, values) mappings are (parallel, true) and (private, {a, b}). Compiler analysis passes can easily look up this Annotation map data structure to query the necessary OpenMP information, which is also stored in the CETUS IR. The compiler infrastructure, which targets C programs, supports source-to-source transformations, is user oriented and easy to handle, and provides the most important parallelization passes. HiCUDA [8], a high-level directive-based language for CUDA programming. It allows programmers to perform these tedious tasks in a simpler manner, and directly to the sequential code. Nonetheless, it supports the same programming paradigm already familiar to CUDA programmers. Han [8] prototyped a source-to-source compiler that translates a hiCUDA program to a CUDA program. HiCUDA presents the programmer with a computation model and a data model. The computation model allows the programmer to identify code regions that are intended to be executed on the GPU and to specify how they are to be executed in parallel. The data model allows programmers to allocate and deallocate memory on the GPU and to move data back and forth between the host memory and the GPU memory. The hiCUDA directives are specified using the pragma mechanism provided by the C and C++ standards. The hiCUDA compiler, despite its flexibility, demand to programmer to write correct and optimal programme to obtain preferment results. Lin and Chen [3] introduce a novel compiler based approach for GPGPU programming by providing a high degree of data parallelism. Compiler directives are used to label code fragments that are to be executed on the GPU. The proposed GPGPU compiler, Guru, converts the labelled code fragments into ISO-compliant C code that contains appropriate OpenGL and Cg APIs. A native C compiler can then be used to compile it into the executable code

for GPU. Guru is implemented based on the Open64 compiler infrastructure.

However, for these parallelizer compilers by directives insertion, the process of parallel programming remains ineffective because of absence of the inevitable specific details of programming. The model-based automatic parallelization environments reduce a lot the complexity of development of the parallel programs by their graphic parallel programs modelling in a high level abstraction.

2.2 Model-to-Source auto-parallelizer

The usage of models for the design of multi-processor systems is on its own a great improvement over current practice because it provides a higher abstraction level that especially helps for component reuse and parallel coding. The graphical representation also facilitates the global vision of complex systems and of interactions between the parts of the system. One methodology for the extraction of the parallelism is to partition the original program to diverse independents tasks. It can be realized by using tasks graphs as in ParDT [17]. ParDT is a graphical model-driven development tool suite which supports not only modelling of parallel programs on high abstraction level but the translation of the constructed models into source code skeletons according to the specific runtime environments and libraries. The process of translation, which involves the parsing of graphical models and the generation of source code skeletons aimed at different parallel platforms, are explained in detail. ParDT is implemented based on Eclipse and compatible with its plug-in architecture. The tool suite manages to help programmers relieve the burden of building parallel applications. One important challenge that arises in multicore systems is the ability to dynamically adapt a running application to target architecture in the face of changes in resource availability (e.g., number of cores, available memory or bandwidth). In GASPARD [2], the compilation is a sequence of small and maintainable transformations that allows passing gradually from a high-level description into models closer in abstraction to the final model, which is then converted into code. The specification of the system is done exclusively via UML MARTE [11] models. From the MPSoC model GASPARD provides several transformation chains. As output of a transformation chain, the user expects compilable code which can be used in already available tools. The GASPARD environment permits to select a target into which the SoC should be transformed. The most obvious target is a synthesizable hardware description and application code compilable for this particular hardware. Each chain is a sequence of several model transformations separated by

metamodels and finished by a code generation. It is notable that the majority of auto-parallelization frameworks, like CETUS, use the sequential C code as input and openMP as output. But, despite they have the same objective, they use different transformation techniques. In this paper, we propose an experiment model to evaluate three research auto-parallelization tools, namely PLUTO, CETUS and GASPARD. These tools have the same main goal: automatically generating parallel code with openMP.

3. Measurement model

Experimental Software evaluation is important to discover how some techniques perform, discover its limitations and understand how to improve them [12]. Indeed, in evaluating a system, we need to identify a set of performance metrics that provide adequate information to understand the behavior of the system. Metrics which capture the processor characteristics in terms of the clock speed (MHz), the instruction execution speed (MIPS), the floating point performance (MFLOPS), and the execution time for standard benchmarks (SPEC) have been widely used in modeling uniprocessor performance. A nice property of a uniprocessor system is that given the hardware specifications it is fairly straightforward to predict the performance for any application to be run on the system. However, in a parallel system the hardware specification (which quantifies the available compute power) may never be a true indicator of the performance delivered by the system. This is due to the growth of overheads in the parallel system either because of the application characteristics or certain architectural limitations. So, metrics for parallel system performance evaluation should quantify this gap between available and delivered compute power.

When we wish to evaluate some techniques or process, it is necessary to follow some measurement models that provide the mechanisms to conduct this evaluation. Some mechanisms for defining measurable goals have appeared in the literature like the Goal/Question/Metric Paradigm (GQM) [13] that we choose for defining our measurement model. The GQM paradigm is a mechanism for defining and evaluating a set of operational goals using measurements [13]. A measurement model is defined into three levels: conceptual (goal), operational (question), and quantitative (metric).

In this work, we use the Goal Question Metric (GQM) paradigm for defining our measurement model.

Goal: Evaluate the auto-parallelization tools: CETUS, PLUTO and GASPARD.

Questions:

- What is the parallelization effort of auto-parallelization tools?
- How much performance gain is achieved by parallelizing a given application over a sequential implementation?
- What is its ability to increase performance as number of processors increases?

Metrics to be considered in order to find out answers to these questions are defined as follow:

3.1 Cost/Effort

Cost C reflects the sum of the time that each PE (processor element) spends solving the problem:

$$C = \text{Parallel runtime} \times \text{the no. of PEs used} \quad (1)$$

If $p=1$: The cost of solving a problem on a single PE is the execution time of the fastest known sequential algorithm.

3.2 Speedup

Speedup is a widely used metric for quantifying improvements in parallel system performance as the number of processors is increased. Speedup is defined as the ratio of the time taken by an application of fixed size to execute on one processor to the time taken for executing the same on processors. However, ideal behavior is not achieved because while executing a parallel algorithm, the processors cannot devote 100% of their time to the computations of the algorithm. For Example, part of the time required by the processors to compute the sum of n numbers is spent idling (and communicating in real systems). **Speedup**, S , is the ratio of the time taken to solve a problem on a single PE to the time required to solve the same problem on a parallel computer with p identical PEs.

$$S = T_s / T_p \quad (2)$$

3.3 Efficiency

So, Efficiency is a measure of the fraction of time for which a processor is usefully employed.

Efficiency E is the ratio of *Speedup* S to the number of PEs (p):

$$E = S / p \quad (3)$$

In an ideal parallel system efficiency is equal to one. This means that all processor resources are spent on the task. But, rarely the case because of the overhead associated with coordinating processes. Also some parts of a program (such as I/O) might not parallelized. So, in practice, efficiency is between zero and one.

4. Performance evaluation

The experiments were run on a quad-core Intel Core 2 Quad Q6600 CPU clocked at 2.4 GHz (1066 MHz FSB), with a 32 KB L1 D cache, 8MB of L2 cache (4MB shared per core pair), and 2 GB of DDR2-667 RAM, running Linux kernel version 2.6.22 (x86-64). The used compiler is ICC10.0 [16], which was also used to compile the C codes transformed by the three systems.

Table 1. Benchmarks list for evaluating CETUS, PLUTO and GASPARD systems.

Benchmark	Description
MM (Matrix Multiply)	Computes the product of two matrices.
FT (Fast Fourier Transform: NAS parallel benchmarks)	FT contains the computational kernel of a 3-D fast Fourier Transform (FFT)-based spectral method.
LU (Lower-Upper symmetric Gauss-Seidel: NAS parallel benchmarks)	Solve a synthetic system using three different algorithms involving block tridiagonal, scalar pentadiagonal and symmetric successive over-relaxation (SSOR) solver kernels
MVT (Matrix Vector Transpose)	The MVT kernel is a sequence of two matrix vector transposes. It is found within an outer convergence loop with the Biconjugate gradient algorithm.

The NAS Parallel Benchmark (NPB) suite [23] consists of five kernel benchmarks and three pseudo-applications from the field of computational fluid dynamics. The NPB presents an excellent resource for this study, in that it provides multiple language implementations of each benchmark. Table 1 lists the four used parallel benchmarks code for performance evaluation.

- Runtime and memory usage

One aspect of evaluating a compiler infrastructure is its efficiency in terms of runtime and memory usage when dealing with realistic applications. Runtime consists mainly of parallelization time, of which the dependence analyzer consumes a major portion. Experiments (table 2) show that PLUTO has the best performance than GASPARD and CETUS for the benchmarks running on the same system. Indeed, CETUS and GASPARD are written in Java which is slower and requires more memory than C or C++. Also, memory usage is driven primarily by the complexity of loops analyzed for dependence testing, in terms of their nesting levels and the total number of array accesses they contain. These factors contributed to CETUS and GASPARD taking a noticeably longer time to process its input than, for instance, the PLUTO compiler.

Table 2. Statistics on loops parallelization with PLUTO, CETUS and GASPARD

	Runtime (sec)	Memory usage (Mbytes)
PLUTO	10	70
CETUS	18	110
GASPARD	25	140

- parallelization quality

The performance evaluation of auto-parallelization infrastructure is based on several criteria like the ability to achieve auto-parallelization and to detect parallel and nested loops in the sequential code. In this work, the GASPARD system is tested with only the Matrix Multiply benchmark due to many problems. First, only the Multiply Matrix example is available. Second, in the available GASPARD version, it is not permuted to model our proper parallel application. These preliminary obstacles show the difficulty founded in using a graphical system. But these systems offer an efficient solution to solve the automatic parallelisation problems. They reduce a lot the complexity of development of the parallel programs by their graphic parallel programs modelling in a high level abstraction. Effectively, as shown in Figure 1 GASPARD performs close to or better than PLUTO on MM benchmark, and better than PLUTO in terms of achieve automatic parallelization. Figure 2 show that GASPARD performance can attain to 80% in terms of parallel loops and Nested loop detection, and to 70% in terms of analysis dependence.

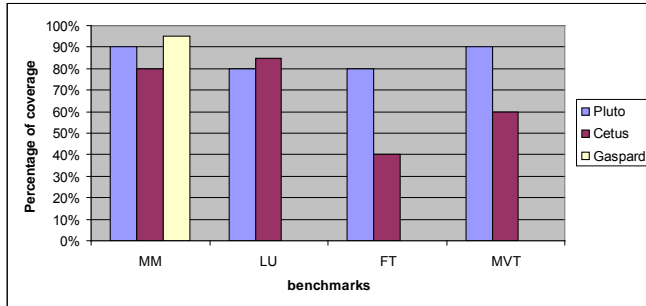


Fig 1. Performance comparison in terms of ability to achieve automatic parallelisation between the different systems

In the case of FT and MVT, CETUS performs poorly than PLUTO. In fact, CETUS is not able to achieve an automatic parallelization. In fact, the codes generated by CETUS need a manual correction to optimally obtain parallel codes. On the other hand, in the LU's case, CETUS performs close to or better than PLUTO. Thus, PLUTO successfully performs the majority of benchmarks by using “-tile” option in the framework execution. In fact, the PLUTO compiler uses the tiling technique which is a key transformation in optimizing for parallelism and data locality.

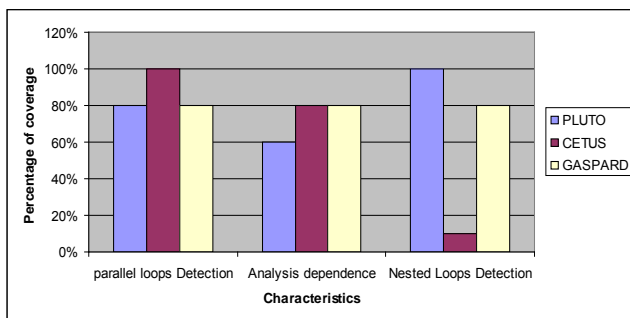


Fig 2. Performance comparison between PLUTO, CETUS and GASPARD systems

As shown in figure 2, PLUTO has a high ability to detect nested loops in the contrary of CETUS which poorly detect such loops. On the other hand, the tests show that CETUS detect parallel loops with 100 percent. Thus, CETUS is powerful in terms of parallel loops detection. But, PLUTO and GASPARD have the same efficiency to detect parallel loops. Moreover, GASPARD and CETUS are better than PLUTO in the terms of dependence analysis. In fact, CETUS enables automatic parallelization by using dependence analysis with the dated Banerjee-Wolfe inequalities, array and scalar privatization, reduction-variable recognition, and induction-variable replacement. These are the

techniques found to be most important for automatically parallelizing compilers.

In GASPARD' case, the designer creates a model of the application with all the information needed for the implementation, that is: without ambiguity and with all the details concerning the realization of a parallel application. This is the main advantage of the model-based systems.

- **Cost, speedup and efficiency**

This section presents a quantitative comparison between the three selected tools to better evaluate their performance and parallelization quality. Indeed, the performance of a compiler is usually measured in terms of the execution efficiency of compiled code [21]. The results presented in this section are based on metrics described in section 3. The experiments are performed using OpenMP standard functions that calculate the wallclock time between two points in the program. For example, OMP GET WTIME(OpenMP) function is intrusive in the code to evaluate the speedup. CPU time does not include the overhead for parallelization.

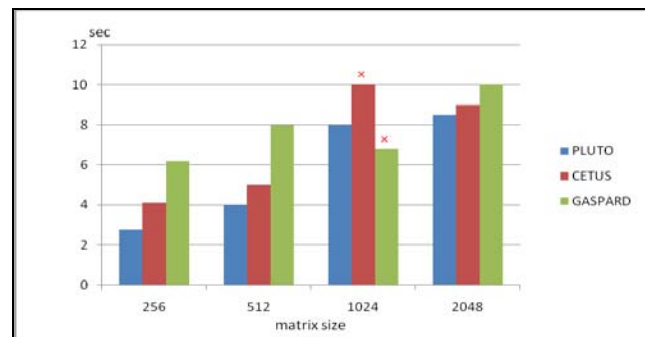


Fig 3. Execution time for parallel Matrix Multiplication

Figure 3 shows the execution time for parallel Matrix Multiplication (size from 256 to 2048) generated by the three tools. We notice that, comparing to PLUTO, CETUS and GASPARD has the higher execution time in all matrix size except in n=1024. Indeed, when we execute the codes transformed by GASPARD and CETUS, we got a run-time error. But, we could only transform successfully by using PLUTO. This is can be explained by insufficient of memory. Indeed, a computer can run out of memory when it is running multiple programs at once or even when running just one or two memory-intensive programs. Running out of available memory causes an error because the computer cannot continue running all of the programs until free memory is available. Execution time is not a serious issue for modestly-sized programs, but it can be a problem for

large benchmarks like LU. Figure 4 shows that when increasing the number of cores (multiplying by 2), the Cost C decrease executing the LU application in the case of both CETUS and PLUTO. So, a super-speedup is achieved.

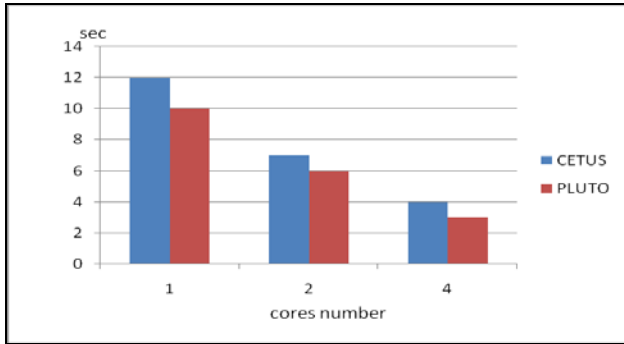


Fig 4. Cost/Effort C running LU application by varying the cores number

Table 3 summarizes the performance of transformed codes using the three tools. The results show a good speedup attained by different cores numbers. Indeed, in the case of MM (n=1024), the three tools have a comparable speedup on dual core design. But, on multi-core design (4 cores), PLUTO has a higher speedup than CETUS and GASPARD. In the case of FT application, PLUTO and CETUS have a comparable speedup on both dual and multi core designs. While, in the case of MVT and LU applications, PLUTO is better than CETUS on both dual and multi core designs.

Figure 5 shows the efficiency of the code generated by the selected tools in the case of multi-core design (4 cores). We notice, firstly, that CETUS performs close to or better than PLUTO on FT application. Secondly, comparing to CETUS and GASPARD, a very high efficiency percentage is attained (97,5%) using PLUTO in the case of MM application. Also, PLUTO is more efficient than CETUS executing LU and MVT application. The final comparison list is presented in table 4.

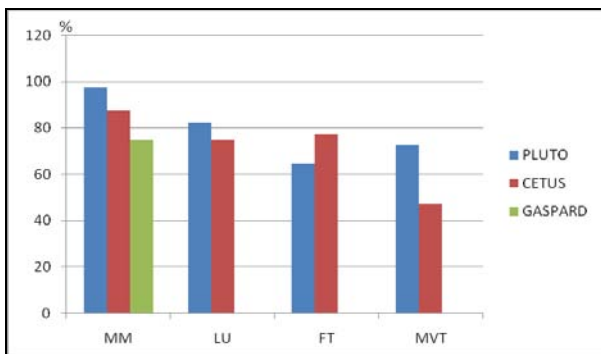


Fig 5. Efficiency percentage

Table 3: speedup comparison

Bench- mark	dual core speedup (2 cores)			Multi-core speedup (4 cores)		
	Pluto	Cetus	Gaspard	Pluto	Cetus	Gaspard
MM	1.9x	1.5x	1.1x	3.9x	3.5x	3x
LU	1.8x	1.5x	-	3.3x	3x	-
FT	1.7x	2.1x	-	2.6x	3.1x	-
MVT	1.1x	0.8x	-	2.9x	1.9x	-

Table 4: comparison list (the smaller is better)

	PLUTO	CETUS	GASPARD
Runtime	1	2	3
Memory usage	1	2	3
Dependency analysis	2	1	1
Parallel loops Detection	2	1	3
Nested loops Detection	1	2	3
Cost/effort	1	2	3
Speed up	1	2	3
Efficiency	1	2	3

From the experimental results in this paper, we find that only PLUTO can transform all benchmarks codes successfully. We conclude also, that CETUS is more efficient than GASPARD. But it gives us, in some cases error results; sometimes we could not transform successfully using CETUS. So, the perfect auto-parallelizing research compiler is yet to be produced. However, there are some cases where auto-parallelization is perfectly suited.

4. Conclusions

Parallel programming is not easy to programmers. So, some frameworks support the auto-parallelization helping them to easily transform sequential codes to parallel codes. In this paper, we presented a comparison between three auto-parallelization tools using a set of criteria. The goal of this assessment was to evaluate the current state of the available automatic parallelization tools when intended to be used by

software designers to develop parallel applications. The comparative study showed that PLUTO is more efficient than the other tools; it gave optimal results for all benchmarks. The advantage of CETUS was his efficiency in terms of dependency analysis and parallel loops detection. But, it gave us error results in detecting and parallelizing nested loops. GASPARD showed the limit of the model-to-source parallelizer comparing to source-to-source parallelizer. It was not flexible and applicable for all benchmarks. But, it gave tolerable results for MM workload.

One common limit of those auto-parallelization tools is the generation of parallel openMP code which depends on the OpenMP API, compiler and OS run time support to realize task partition. However, such support is rarely available in an embedded context where OS is not always present [22]. For future work, we will propose an automatic accelerator generation flow that integrates PLUTO and adapts an application targeting the general purpose processor to an embedded environment. The choice of PLUTO is based on the empirical comparative study presented in this paper.

References

[1] P. Gerner. La sémantique des directives au compilateur: application au parallélisme de données. PhD thesis, Université Louis Pasteur, 2002.

[2] Eric Piel, Philippe Marquet, and Jean-Luc Dekeyser. *Model Transformations for the Compilation of Multi-processor Systems-on-Chip*, GTTSE 2007, LNCS 5235, pp. 459–473, 2008. Springer-Verlag Berlin Heidelberg 2008

[3] *Compiler support for general-purpose computation on GPUs*, Yu-Te Lin · Peng-Sheng Chen Springer Science+Business Media, LLC 2008.

[4] B. Chapman, H. Zima, and P. Mehrotra. Extending HPF for advanced data-parallel applications. *Parallel & Distributed Technology: Systems & Applications*, Jan 1994.

[5]. NVIDIA (2008) NVIDIA CUDA homepage. Website. Online available at <http://developer.nvidia.com/object/cuda.html>

[6] Chirag Dave, Mansang Bae, Seung-jai Min, Seyong Lee, Rudolf Eignmann, Samuel Midkiff “CETUS: A source-to-Source Compiler Infrastructure for Multicores,” *computer*, PLDI’08, vol.42,no.12,pp .36-42, Dec.2009.

[7] Uday Bondhugula, Albert Hartono, J. Ramanujam, P. Sadayappan “A Practical Automatic Polyhedral Parallelizer and Locality Optimizer”, June 7–13, 2008, Tucson, Arizona, USA.

[8] Tianyi David Han, Tarek S. Abdelrahman “hiCUDA: A High-level Directive-based Language for GPU

Programming” GPGPU ’09 March 8, 2009, Washington, D.C. USA

[9] ROSE, <http://www.rosecompiler.org/>

[10] PAR4ALL, <http://www.par4all.org/>

[11] ProMarte partners: UML Profile for MARTE, Beta 1 (August 2007), <http://www.omg.org/cgi-bin/doc?ptc/2007-08.04>.

[12] V. R. Basili. *The role of experimentation in software engineering: past, current, and future*. In ICSE ’96: Proceedings of the 18th international conference on Software engineering, pages 442–449. IEEE Computer Society, 1996.

[13] V. R. Basili, G. Caldiera, and H. D. Rombach. *The goal question metric approach*. Encyclopedia of Software Engineering, 1:528–532, 1995.

[14] A. Lim, S. Liao, and M. Lam. Blocking and array contraction across arbitrarily nested loops using affine partitioning. In ACM SIGPLAN PPOPP, pages 103–112, 2001.

[15] OpenMP Architecture Review Board. OpenMP Application Program Interface, version 3.0. OpenMP Architecture Review Board, May 2008.

[16] <http://software.intel.com/en-us/c-compilers>, 2013

[17] Xu Zhen , Sun Jizhou , Yu Ce, Wu Huabei , Meng Xiaojing , Tang Shanjiang. A Visual Model-Driven Rapid Development Tool suite for Parallel Applications 2009 World Congress on Computer Science and Information Engineering.

[18] M. Griebl. *Automatic Parallelization of Loop Programs for Distributed Memory Architectures*. University of Passau, 2004. Habilitation thesis.

[19] <http://software.intel.com/en-us/intel-parallel-studio-xe>, 2013.

[20] <http://www.ibm.com/developerworks/wikis/display/hpccentral/IBM+Parallel+Environment+Developer+Edition>, 2013.

[21] CHING, W.-M., NELSON, Rick, and SHI, Nungjane. An empirical study of the performance of the APL370 compiler. ACM, 1989.

[22] G. Tian, G Hammami, O. “Performance measurements of synchronization mechanisms on 16PE NOC based multi-core with dedicated synchronization and data NOC”. In: International Conference on Electronics, Circuits, and Systems (ICECS’09), 2009, pp. 988 – 991.

[23] NAS Parallel Benchmarks - <http://www.nas.nasa.gov/Software/NPB/>, 2013

First Author Emna KALLEL: received his Dipl.-Ing. degree in 2006 from the National School of Engineers of Sfax (ENIS) Tunisia, where he is currently working toward the Ph.D. degree with research focused on automatic code generation for

embedded systems. She has been with Enis School, as Research Assistant.

Here further research interests include parallel computing, parallel architectures, automatic parallelisation, Rapid prototyping and video processing and object-oriented methods for hardware generation.

Second Author Yassine Aoudni: Studied Electrical Engineering and Computer software Engineering at the National School of Engineers of Sfax (ENIS) Tunisia. He received Dipl.-Ing. from the National School of Engineers of Sfax in 2002 and Dr.-Ing. from the University of South Brittany, France in 2010. From 2008 till 2011, he worked as a Research Assistant at National School of Engineers of Sfax (ENIS) Tunisia conducting research in FPGA prototyping. Since 2011 he is Assistant Professor at the National School of Engineers of Sfax. He acts as a member of several technical program committees, as a reviewer for different journals. His research interests include joint source FPGA prototyping, signal processing, system high level design, parser design, information theory, as well as multiprocessor architecture

Third Author Mohamed Abid: He received Dipl.-Ing. from the National School of Engineers of Sfax (ENIS) in 1985 and the phddegree from the National Institution of applied Science, Toulouse, France. In 2000, he received his doctorale degree in Electrical and Computer Engineering at National Engineering School of Tunis. He is currently Professor at the Electrical Department of ENIS. Since 2006, he has been on the Head of the research laboratory «Computer Embedded System » CES-ENIS. He is responsible for research projects in the area of automatic signal and image processing, wireless networks and information systems. He has been on the Head of Federator Research Project since 2009. He has authored or co-authored more than 120 international conference papers, and he has written more than 20 technical contributions to various international standardization projects. He is a member of the Scientific and Program Committees of several international conferences and workshops. He is the Co-coordinator of several Nationals and Internationals projects with universities and industries like DGRSRT, CNRS, INRIA, CMCU, training for research, PNM, Temptra, etc.

Ontology based data warehouses federation management system

Naoual MOUHNI¹, Abderrafiaa EL KALAY²

¹ Department of mathematics and computer sciences , University Cadi Ayyad, Faculty of sciences and technologies Marrakesh, 40000, Morocco

² Department of mathematics and computer sciences , University Cadi Ayyad, Faculty of sciences and technologies Marrakesh, 40000, Morocco

Abstract

Data warehouses are nowadays an important component in every competitive system, it's one of the main components on which business intelligence is based. We can even say that many companies are climbing to the next level and use a set of Data warehouses to provide the complete information or it's generally due to fusion of two or many companies. these Data warehouses can be heterogeneous and geographically separated , this structure is what we call federation, and even if the components are physically separated, they are logically seen as a single component. generally, these items are heterogeneous which make it difficult to create the logical federation schema ,and the execution of user queries a complicated mission. In this paper, we will fill this gap by proposing an extension of an existent algorithm in order to treat different schema types (star , snow flack) including the treatment of hierarchies dimension using ontology

Keywords: Data warehouse Federation, Ontology, Hierarchical dimension, Schema Integration..

1. Introduction

A Data Warehouse represents the enterprise-wide "single source of truth" and corporate memory of all business process data [3], it is "a subject oriented, non-volatile, integrated, time variant collection of data in support of management's decisions." as defined by Bill Inmon in 1990, the father of data warehouses.

In some cases, one data warehouse is not sufficient to provide a complete information about a fact, which makes grouping multiple data warehouses the only solution. e.g. in the context of a hotel chain that is geographically distributed in many countries, it may have several heterogeneous warehouses to store and analyse data about customers reservations.

this set of warehouses is what we call " a data warehouse federation".

Federated data warehouses are different than distributed Data warehouses, in order that distributed data warehouses can refer different subjects and there is a strict rule in data distribution (horizontal, vertical...) which make it easy to integrate the query results by using join or sum operations [7].

In federated system, the user send his query without having an idea about the location of data or its structure, the set of data warehouses is seen as a whole and the result is the combination of data warehouses components results.

The components in FDWS (Federated data warehouse system) can differ in aspects such as : data model, query language and data semantic [9].

So, a FDWS must contains the following elements [8] :

- An integration procedure of the schemas of the component warehouses giving the logical schema of the federation.
- A query language for user who does not need to know the schemas of the component warehouses.
- A procedure which enables decomposition of user queries to the federation into sub-queries which are sent to the component warehouses

the warehouse federation system management is first based on a logical schema called the federation schema , which integrates all the components schemas. to create this schema, we must integrate all the other local schemas, without losing information. During this integration, it may be difficult to decide keeping or not an information by using the procedure shows in [8], which compare every measure to the one in the existing federation schema, if it exists only the location of this measure which is

characterized by the couple (D_0^i, b_name) is added, where :

D_0^i : represent the fact table in the data warehouse i.

b_name : represent the name of the measure

else, a new measure is added to the schema .

An algorithm is implemented to integrate dimensions attributes, respecting the same logic.

In fact, this algorithm present its limits in case we have a measures or a dimension attributes that refer to the same subject, and represented by two different terms in data warehouses local schemas and it doesn't treat the relationship that could been between attributes and the case of hierarchical dimensions.

Our approach consist of using an application ontology defined in [10] as "a description of knowledge necessary to achieve a particular task and that allows to use the same programming language as the application programming language ", to fill this gap instead of using only Meta data that does not fully represent the semantic relationship between local schema measures and dimension attributes, and those of the federation schema.

In this article, we propose an ontology based data warehouses federation management system to solve the problem of semantic heterogeneity during federation schema creation, based on hotel chain data warehouse sources.

Moreover, in our knowledge, there is no studies that used ontology in a federation context to solve this problem, which justify our choice.

Then in section 2; we present and analysis in summary a set of related works.

2. Related works

In all domain research, It is always worth considering the others work , discuss it and check if we can refine and extend it for our particular purpose.

In computer sciences, reusing existing sources is one of the reasons that made the development of this domain possible.

Warehouses federation according to Sheth and Larson [9], and that appears in [7] and [4], is a set of data warehouses that are heterogeneous, autonomous and dispersed. Every component can continue its local operations and at the same time participate in federation.

It's for the better that all the integration operations be done without interrupting the process of component data warehouses.

There are no many studies on the data warehouse federation, however, R. Kern, K. Ryk, and Ngoc Thanh Nguyen, proposed a framework for building logical

schema and query decomposition in data warehouse federations [7], they developed an algorithm to integrate component schemas into one global logical federation schema.

But this algorithm presents some limits in order to treat the case of warehouses with star schema only, and it doesn't consider the hierarchical dimensions and all the heterogeneity types, which are described in [9] as the difference in structure, where different data models provides two different structural primitives. then , differences in constraints ,differences in query languages and semantic heterogeneity.

Semantic heterogeneity, is one of the biggest problem that faces information integration nowadays, it occurs when two synonym terms from two different sources describe the same subject [1] (e.g: schedule and timetable are synonyms but we have to show it to the system) .

one of the solutions to fill this gap is using ontology, which is according to [6] " ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). An ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins."

According to their use, we distinguish many types of ontologies, Generic Ontology, Domain ontology, Application ontology, Representation ontology, The ontology of methods, tasks and resolution of problems, Light ontology and rich ontology[2].

Even if using ontology may resolve the heterogeneity problem in federated data warehouses, it is not yet used in this context, and all the solutions proposed are based on Meta data repositories, which solve the problem of structure definition but not the semantic issues.

3. Our contribution

3.1 Presentation of the solution

Our work is an extension to[7] algorithm to create the global logical federation schema .

R. Kern, K. Ryk, and N. Nguyen, proposed an algorithm of integration of component schemas into a federated logical schema. They assume that all warehouses are with

star schema, so they do not deal with hierarchies in dimensions.

In fact, even with a star schema the hierarchy for dimension are stored are stored in the dimensional table itself.

Whereas, in a snow flack schema, a dimension table have more or more parent tables, and hierarchies are broken into separate tables in snow flake schema . These hierarchies helps to drill down the data from topmost hierarchies to the lowermost hierarchies[5].

Our objective is an improvement of this integration algorithm to cover heterogeneous schemas (snow flack or star schema). And use ontology as a tool to solve the semantic heterogeneity problem instead of using meta data only.

We propose a federation data warehouse management system (FDWS), which cover :

- Improved algorithm for schemas integration using application ontology
- A query analysis and decomposition tool .
- An ontology-based integration Algorithm for query results.

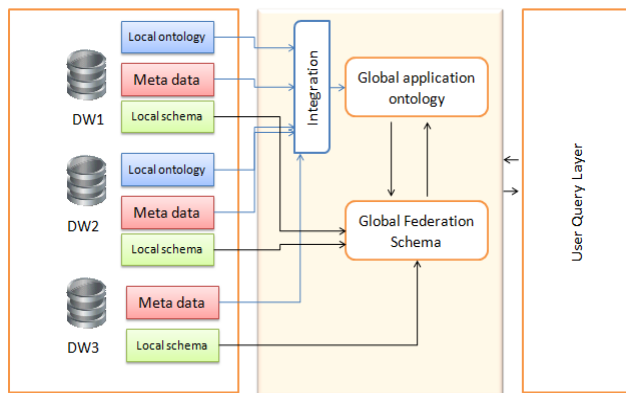


Fig. 1. Structure of the proposed Data warehouses federation management system

1. Every federation component may or not have its own local application ontology, which is written in OWL language describing the semantic of every attribute and measure, and describe the relationship between items and hierarchies of dimensions by using *is_a* and *parentOf* relations.
2. This local ontologies are exported to the logical layer ontology repository, besides that a meta data xml file is loaded into the federation system to describe data structure.
3. The user query is analyzed by the FDWS, decomposed, executed on the selected components

4. The query results are integrated using ontology to solve the heterogeneity problem .

3.2 Integration schema's algorithm

In our case, the input can be with different schemas types (star , snow flack), so to treat the dimension hierarchies we propose the following algorithm:

Annotation

We use the same notation as [7].

Input.

P_i^j as the set of parents of a dimension defined by

$$P_i^j = (D_1, \dots, D_n)$$

H_p a Data warehouse schema defined as

$$H_i = (D_0^i, D_1^i, \dots, D_{\alpha i}^i)$$

F an existing federation schema defined by

$$F = (D_0, D_1, \dots, D_m)$$

Output.

F the federation after integration with H_p .

Other notations are used:

a_name : name of attribute a

b_name : name of measure b

$D_x \sim D_y$: D_x is similar to D_y (based on ontology and meta data OR expert's decision)

$a_x \equiv a_y$: a_x is similar to a_y (based on ontology and meta data OR expert's decision)

$b_x \Leftrightarrow b_y$: similar measures (based on ontology and meta data OR expert's decision)

Recall of the Measure integration algorithm.

R. Kern, K. Ryk, and N. Nguyen in [7], defined a measure integration algorithm as follow:

For each measure from input data warehouse try to find corresponding measure in federation schema. If such a measure exists in federation schema add a mapping between them. If none of the federation measures corresponds to the current one add it to the federation and make a mapping between new measure and the current

one.

Dimension Integration

In every iteration of the algorithm, the global schema is being updated by integrating parents of dimensions, then integrating the dimension it self.

1. For each dimension from a component schema, using ontology, we extract the set of this dimension parents, this set can be equal to \emptyset or contains one or many items.
 - a. For each parents item, we look for similarity in F , if it contains a similar structure, we compare its attributes with the existing one, in case two attributes are similar, we add a new location to the attributes inventory represented by the couple (D_y^p, a_name) , else, we add the attribute as a new one to the dimension. In case that the attribute doesn't exist in the target dimension, we add a new attribute.
 - b. After integrating all the dimension parents, we integrate using the same operations the dimension it self.

```

foreach  $D_y^p$  in  $H_p, y=1,2,\dots,\alpha p$ 
  if  $P_y^p \neq \emptyset$ 
    foreach dimension  $D_i$  in  $P_y^p$ 
      if  $\exists D_t \in F : D_t \sim D_i$ 
        foreach attribute  $a'$  in  $D_i$ 
          if  $\exists a \in D_t : a \equiv a' \wedge a$  is characterized
            by  $(a\_name, list)$ 
              list = list  $\cup \{(D_i, a'_name)\}$ 
            else
               $D_t = D_t \cup \{(a'_name, \{(D_i, a'_name)\})\}$ 
          endif
        endforeach
      else
         $D_t = \emptyset$ 
      foreach  $a''$  in  $D_i$ 
         $D_t = D_t \cup \{(a''_name, \{(D_i, a''_name)\})\}$ 
    
```

```

endforeach
 $F = F \cup \{D_t\}$ 
endif
if  $\exists D_t \in F : D_t \sim D_y^p$ 
  foreach attribute  $a'$  in  $D_y^p$ 
    if  $\exists a \in D_t : a \equiv a' \wedge a$  is characterized
      by  $(a\_name, list)$ 
        list = list  $\cup \{(D_y^p, a'_name)\}$ 
      else
         $D_t = D_t \cup \{(a'_name, \{(D_y^p, a'_name)\})\}$ 
      endif
    endforeach
  else
     $D_t = \emptyset$ 
    foreach  $a'''$  in  $D_y^p$ 
       $D_t = D_t \cup \{(a'''_name, \{(D_y^p, a'''_name)\})\}$ 
    endforeach
   $F = F \cup \{D_t\}$ 
endif
endforeach

```

4.Example

We consider that we have two data warehouses which represent the sources of our federation system. The first component is with start schema, so hierarchies dimension are represented in dimension itself. e.g. the hierarchy $Country \rightarrow Region \rightarrow City$.

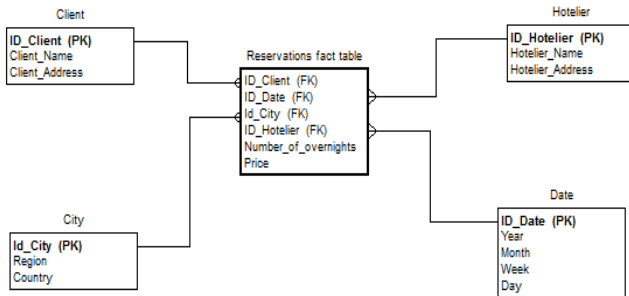


Fig. 2. A star schema of hotel reservations

The second component, is a snow flack schema representing Hotel reservations. this schema contains some hierarchies of dimensions.

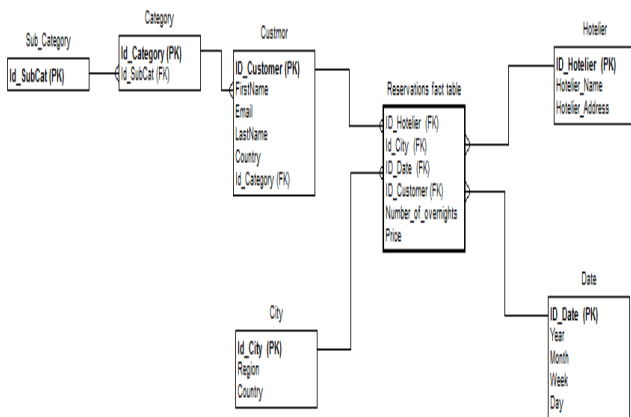


Fig. 3. A snow flack schema for hotel reservations

After applying the proposed integration algorithm we get the global schema as follow:

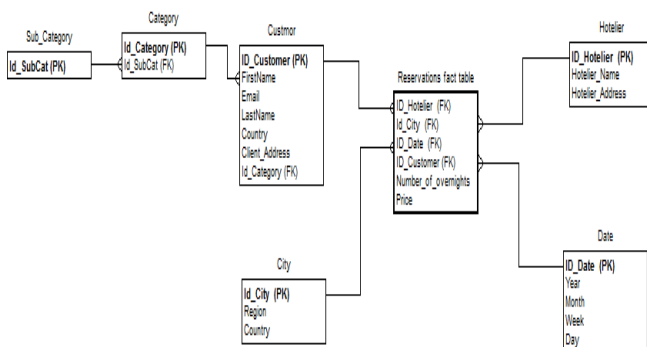


Fig. 4. The result of components schemas integration

Let consider two data ware houses, the first one (Fig2) with a star schema and the second one (Fig3) is a snow

flack schema related to a reservation management in a hotel chain.

1. We first extract ontologies and metadata files from different nodes, in the integration layer of the FDWS, then include new entries into the global ontology repository.
2. Then we integrate fact tables by testing the existence of this table in the global federation schema, if it exists, we compare its measures to the existing ones referring to the ontology repository.
3. next step is to integrate dimensions and hierarchies dimension, e.g: we first integrate the client dimension from DW1 into the global schema, then when we try to include Customer dimension, which is a synonym of client dimension, so referring to the ontology repository we don't add it as a new dimension, and we compare its attributes with clients attributes.

Based on *parentof* relationship mentioned in ontology files, between Customer/client and Category and sub_Category we integrate this hierarchy.

5. Implementation

The integration schema algorithm was implemented using Java API Jena, to manipulate RDF language from java application. We are using two data warehouses; the first one with a star schema and has an ontology written in OWL/RDF, the second data warehouse with a snow flack schema and has no local ontology. Metadata files and OWL/RDF files are mapped into xml file and transferred into the network to the Federated data warehouses management system.

6. Conclusion

In this article, we have presented a part of our data warehouses federation management system. In particular the process of creating the federation schema based on the integration of local schemas using application ontology. Which makes possible to treat the hierarchies of dimensions by analyzing the parentOf relationships, and make the it easy to automate the integration process in federation context.

References

- [1] D. AWAD, H. T., V. COURBOULAY, A. REVEL 2012. Ontology-based Solution for Data

- Warehousing in Genetic Neurological Disease
Proceeding of the world congress on Engineering
- [2] GAYO DIALLO, L. D. T. D. L. I., DE LA MODÉLISATION ET DE LA COGNITION 2006. Une Architecture à Base d'Ontologies pour la Gestion Unifiée des Données Structurées et non Structurées. *TIMC-IMAG — UMR*, 5525.
- [3] INMON, W. H. 1999. *The Operational Data Store*.
- [4] JINDAL, R., ACHARYA, A 2003. Federated Data Warehouse Architecture. *White paper, Wipro Technologies*.
- [5] LLC, P. B. S. 2012. *Learn data modeling* [Online]. Available:
<http://www.learn-datamodeling.com/star.php>.
- [6] MCGUINNESS, N. F. N. A. D. L. 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*.
- [7] R. KERN, K. R., N. NGUYEN 2011. A Framework for Building Logical Schema and Query Decomposition in Data Warehouse Federations. *Computational Collective Intelligence. Technologies and Applications, Lecture Notes in Computer Science Volume 6922*, pp 612-622.
- [8] R. KERN, T. S., N. NGUYEN 2013. A formal framework for query decomposition and knowledge integration in data warehouse federations. *Expert Systems with Applications*, 40, 2592–2606.
- [9] SHETH, A. P., LARSON, J.A. September 1990. Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Computing Surveys* Vol. 22, No. 3.
- [10] VAN HEIJST, G., SCHREIBER, A. ET WIELINGA, B. 1997. Using explicit ontologies in kbs development. *Int. J. of Human-Computer Studies*, 46(2/3), 183 292.

An Effective Web Service Ranking Algorithm based on Quality of Experience of Users

Vandan Tewari¹, Nirmal Dagdee² and Aruna Tiwari³

¹ Dept. of Comp. Tech. & Appl. , S.G.S.I.T.S
Indore- 452003, MadhyaPradesh, India

² Dept. of Computer Engg., S.D. Bansal College of Technology, Umaria
Indore, MadhyaPradesh, India

³ Dept. of Computer Sc., IIT, Indore,
Indore, MadhyaPradesh, India

Abstract

With the increasing number of Web services, discovering and selecting best services for a client is becoming very significant. While discovering a user can benefit from experiences of other users. This can actually be exhibited through a collaborative filtering mechanism where a user is able to rate a service based on his experiences. A user can be offered services based on the Quality of Experience (QoE) of all the users which have used the given services in past. The service ratings given by all the users can be aggregated into a single list to prepare the overall service ranking which can be rendered to a client to help him in selection of better service. Further if a user wants to see the service ranking on other aspects such as its popularity etc., an aggregate ranking of services is presented using different ranking parameters. This paper presents a client oriented approach of Service rating and rank aggregation based on user oriented QoE based rating as well as popularity.

Keywords: *Web Service, Quality of Experience, Service Rating, Service Ranking.*

1. Introduction

In the age of globalization, day by day business to business and business to consumer operations are finding huge importance in internet computation around the world. Web services [1] are one means by which we can fulfill all these demands in an easy and efficient way.

Web Services are based on Service Oriented Architecture[2] which enables application-to-application communication over the internet and easy accessibility to heterogeneous applications and devices. As web services become more popular model for Internet computing, the issues of effective and appropriate service discovery become of utmost importance. The web service search using search techniques supported by existing UDDI[3]

APIs may not result in the search results that are appropriate to service requestor's needs.

Current proposals for web service discovery presents the same search results to all clients for the similar query. However evidently the different users have different needs and an objective for web service discovery and therefore it is essential that these differences are accounted for while discovering services for a client. Therefore there is an urgent need of identifying the needs of a client for discovery for rendering him the services which he actually desires.

There will be a large pool of discovered services which fulfills the functional requirements of a user. However to select an appropriate service from this pool is still an issue. To help a user in finding a 'good' service, the past experiences of other users might be used. The users may be asked to give their feedbacks in terms of service's overall behavior such as 'value for money', 'satisfaction level', 'trustworthiness' etc which actually represent its Quality of Service (QoS) behavior. These parameters collectively can be thought as Quality of Experience (QoE) and can be used for ranking a service in its pool.

In this paper an Effective Web Service Ranking Algorithm based on Quality of Experience of Web Service users has been proposed.

2. Service Rating and Ranking

A ranking list of n services is just a vector of permutations of integers 1 through n. In contrast a rating of services is assigning a numerical score to each service. A sorted rating list creates a ranking list. The rank of a web services is its relative importance

to the other web services in the set. A service ranking model is a method of determining a way in which the ranks are assigned to services in a group. Typically a ranking model uses information available to determine a rating for each service. Once we have the ratings the assignment of the service ranks can be as trivial as sorting the services in the descending order of the corresponding ratings.

When multiple services with the similar functionality are discovered for a user, it is difficult for the user to choose any one of them. To ease of this selection, the services can be ranked on various parameters such as User feedback, Popularity, Cost etc. Evidently these requirements vary from user to user. As discussed earlier the services can be rated those users who have already experienced these services. These users can rate services as per their quality of experience despite being aware of QoS parameters. The services can also be ranked based on the cost which is an important parameter for a client while choosing a service.

3. Related Work

A noteworthy contribution has been made in field of web service ranking by various researchers. In [4], Le-Hung Vu et al. have presented present a QoS-based semantic web service selection and ranking solution with the application of a trust and reputation management method to address the problem of false QoS rating claimed by providers. The UX architecture presented in [5] proposes use of dedicated servers to collect feedback of consumers and then predict the future performance of published services. In [6], a proposal has been given where services are allowed to vote for quality and trustworthiness of each other and the service discovery engine utilizes the concept of distinct sum count in sketch theory to compute the QoS reputation for every service. However, these reputation management techniques are still simple and can suffer from problems of cheating behaviors. Authors have proposed a model of reputation-enhanced QoS based Web services discovery which uses a reputation manager to assign reputation scores to the services based on customer feedback of their performance in [7]. A discovery agent facilitates QoS-based service discovery using the reputation scores in a service matching, ranking and selection algorithm. In [8] E. Al Masri et al. have introduced the Web Service Relevancy Function (WsRF) used for measuring the relevancy ranking of a particular

Web service based on client's preferences, and QoS metrics. The proposal in [9] aims at ranking different Web services published by different cloud platforms, taking advantage of PageRank principle. However the service rating based on collaborative feedback service rating has not been explored for service ranking.

3. Proposed Service Ranking Algorithm

Different users can use different scales for rating services. To aggregate them into a single rating, an algorithm is presented in this section. For a naïve user there may be subtle difficulty in rating a service on Quality of Service using objective values and therefore it is very unlikely for him to provide a feedback. However if the parameters are mapped on natural language descriptions i.e. are subjective, a user may be suitably motivated to provide feedback as rating and therefore for service rating an appropriate GUI has been designed as shown in the figure. The users are simply asked to choose to tag a service in five intervals of quality of experience {Best, very good, good, fair, poor} which is relatively simpler to comprehend. However sometimes a user may not be that confident about his capability of rating and therefore user is also asked to judge his confidence in his ratings. This can actually affect the contribution of rating in overall rating. Figure 1 presents the design of the GUI:

Fig. 1 : GUI for User Rating

SERVICE RATING AND REVIEWS	
VALUE FOR MONEY	Excellent <input type="range"/> Poor
SATISFACTION LEVEL	Excellent <input type="range"/> Poor
TRUSTWORTHINESS	Excellent <input type="range"/> Poor
HOW CONFIDENT ARE YOU IN YOUR RECOMMENDATION (Pl. enter a number between 1-10)	<input type="text" value="8.0"/>
OVERALL EXPERIENCE	★ ★ ★ ★ ☆

3.1 Calculating Overall Ranking from the Five User Selected Parameters:

The users can be asked to provide the rating and reviews of used services using above shown GUI. Here the five

parameters can have following ranges: Value of Money, Satisfaction Level Trustworthiness, can be mapped upon (Best, Very Good, Good, Fair, Poor) Where ‘Excellent’ represents the best rating while poor represents the worst and therefore Best can be mapped on 5 on a scale of 1-5, while Poor can be treated as 1.

In addition the ‘Confidence of Rating’ gives us a weight parameter for consideration while calculating the overall rating. The overall experience also can be mapped upon 1-5 where this parameter can represent how well the service has met the functional requirement of a client. For the shown example in Fig 1: GUI, the corresponding overall rating may be calculated as : $R = [(3.5+3.5+4.5+4.0)/4] \times 0.8 = (19.5/4) \times 0.8 = 3.9 \sim V. Good$

3.2 Calculating Overall Service Rating using Collaborative Filtering

Once an overall rating is calculated for a user, a final rating list has to be aggregated for all the users who have given feedback for services in terms of rating. In this section a method has been given for rating aggregation so as to produce a final ranking for candidate web services using the Offense-Defense rating and ranking method [10].

3.2.1 The Offence-Defense Model of Rating and Ranking

A natural approach in rating is to first rate individual attributes of each participant service. Each service in the chosen pool can have two strengths: first an offensive strength representing how many times it has been rated better over similar services in its pool and a defensive rating representing how strong (or weak) the offence of other services. The Offense-Defense Method has been proposed by Anjela Y. Govan et al. in [10] for ranking sports teams based on their play statistics. It is a model for rating the overall strength of each team relative to the others. While there are numerous factors that might be taken into account, the approach is to characterize “strength” by combining each team’s relative offensive and defensive prowess in a non-linear fashion. To compute offensive and defensive ratings authors start with the assumption that if offensive ratings are large then team has greater offensive strength, i.e. the increased capability of winning in a matchup. On the other hand, smaller defensive ratings will correspond to greater defensive strength, i.e., a low defensive rating indicates that it is hard for the opposition to win in a matchup.

However it can very well be employed for ranking services in a pool. The Offence-Defense Method (OD) iteratively produces two ratings for a team i ; an offensive o_i and a defensive d_i . A simple way to combine them to produce an overall rating is $r_i = o_i/d_i$.

The convergence of the method is guaranteed and is fast. If each web service is treated as a team having a matchup with all potential services as other teams in selection. The method can be effectively applied for ranking. The method of OD rating is as follows: If teams i and j compete, then let

$A = [a_{ij}]$ be such that a_{ij} is the score that team j generated against team i (set $a_{ij} = 0$ if the two teams did not play each other). Alternately, a_{ij} can be thought of as the number of points that team i held team j to. When applying for service rating this matrix represents how much better a service has been rated from its opponent service. In other words, depending on how it is viewed, a_{ij} simultaneously reflects relative offensive and defensive capability. To utilize this feature authors have defined the offensive rating of team j to be the combination

$$o_j = a_{1j}(1/d_1) + \dots + a_{nj}(1/d_n)$$

where d_i is the defensive rating of team i that is defined to be

$$d_i = a_{1i}(1/o_1) + \dots + a_{ni}(1/o_n)$$

Since o_j 's and d_i 's are interdependent, these values will have to be determined by a successive refinement technique. Intuitively given $A = [a_{ij}]$:

$$o = A^T \frac{1}{d} ; \quad d = A \frac{1}{o}$$

In other words row sums of normalized average distance vector would give us the measure of offensive output while column sum would give us the defensive output.

Rank aggregation is a function which uses several ratings (or ranks) obtained using various models as an input to produce a single rating (or rank) of each team as an output. The simplest aggregation function that can be applied to the Offense-Defense model is $r_i = o_i/d_i$; i.e., the overall rating score of team i is its offensive rating divided by its defensive rating.

3.2.2 Application of OD model for Web Service Ranking

As an example: suppose the following (overall) ratings have been given by a group of four clients to a type (say weather forecast) of services and Overall ratings received by the services are shown in Table 1 & Table 2:

They should be numbered consecutively throughout the text. Equation numbers should be enclosed in parentheses and flushed right. Equations should be referred to as Eq. (X) in the text where X is the equation number. In multiple-line equations, the number should be given on the last line.

Table 1: Ratings Given by Users

Services	User1	User2	User3	User4
S1	Excellent	Very Good	Excellent	-
S4	Fair	-	-	Good
S6	Good	-	Very Good	Very Good
S8	Good	Good	-	Good
S9	-	Good	Excellent	-
S15	-	Poor	Very Good	Excellent

Table 2: Overall Ratings for Services

Service	Ratings
S1	Excellent, Very Good, Excellent
S4	Fair, Good
S6	Good, Very Good, Very Good
S8	Good, good, good
S9	Good, Excellent
S15	Poor, Very Good, Excellent

The above presented table cannot depict correctly the ratings of various users and also some users may be friendly while rating however some may be strict with ratings .To aggregate these differences in scales, a rating aggregation method [11] may be used using normalized rating matrices for all the users \bar{U} : Calculating \bar{U} matrices for all users:

Table 3: Normalized matrix for user 1

Service	S1	S4	S6	S8	S9	S15
S1	0	3/9	2/9	2/9	0	0
S4	0	0	0	0	0	0
S6	0	1/9	0	0	0	0
S8	0	1/9	0	0	0	0
S9	0	0	0	0	0	0
S15	0	0	0	0	0	0

Here numerator of (s1,s4) entry means user 1 rated a service 1 three positions above three positions above service 4 . The denominator 9 is the cumulative rating differentials of all 4C_2 possible matchups between services rated by user1.e.g. difference in ratings between (s1-s4,s1-s6,s1-s8,s4-s6,s4-s8,s6-s8)=(3+2+2+1+1+0)=9 . Similar

$\bar{U}_2, \bar{U}_3, \bar{U}_4$ rating matrices can be calculated.

and then calculating the average distance matrix \bar{U}_{ave} can be calculated :

Table 4: Average Distance Matrix

Service	S1	S4	S6	S8	S9	S15
S1	0	0.0833	0.1181	0.0833	0.0278	0.1458
S4	0	0	0	0	0	0
S6	0	0.0635	0	0	0.0357	0
S8	0	0.0278	0	0	0	0.0556
S9	0	0	0.0625	0	0	0.1181
S15	0	0.0714	0.0357	0	0	0

The oValue and dValue of the services is as follows:

Here **oValue** represents the offensive vector calculated as sum of row values for each service i.e.
 $oValue = \bar{R}_{avg} E$ where E is the vector of all ones.
 The **dValue** represents the defensive vector calculated as sum of column values for each service i.e.
 $dValue = E^T \bar{R}_{avg}$. Now the rank of each service can be calculated as $R = oValue/dValue$ Using R The final rating list may be calculated as:

Table 5: O-D Values and Overall Rank

Service	oValue	dValue	R=oValue/dValue	Rank
S1	0.4583	0	∞	1
S4	0	0.246	0	6
S6	0.0992	0.216	0.459	4
S8	0.0834	0.0833	1.001	3
S9	0.1806	0.0635	2.844	2
S15	0.1071	0.3195	0.335	5

3.3 Rank Aggregation for Web Service Discovery

While selecting service, a user must have the freedom of choosing the attributes on which the services should be ranked and therefore a scheme has been presented in this section based on various parameters of service ranking. The services may be ranked not only on service rating but on Popularity i.e. invocation frequency as well as cost to be paid while using the service.

3.3.1 Borda Count Method for Rank Aggregation

Borda's method [12] is a "positional" method, it assigns a score corresponding to the positions in which a candidate appears within each voter's ranked list of preferences, and the candidates are sorted by their total score. Given k full lists, Borda's method assigns a k-element position vector to each candidate (the positions of the candidate in the k lists), and sorting the candidates by the L_1 norm of these vectors. A primary advantage of positional methods is that they are computationally simple and can be implemented in linear time on a RAM (i.e., in $O(nk)$ time for k full lists of n candidates). As an example if there are three ranked lists on popularity, collaborative rating and QoS Rating

where 1 represents the best candidate, then Borda Rank is calculated as :

Table 6: Aggregated Rank of Services

Service	OD Rank	Popularity	QoS of Service	Borda Count	Borda Rank
S1	1	2	1	4	1 st
S4	6	6	6	18	6 th
S6	4	5	5	14	5 th
S8	3	3	2	8	3 rd
S9	2	1	3	6	2 nd
S15	5	4	4	13	4 th

3.4 Algorithm for Client Oriented Rank Aggregation

Algorithm : Client Oriented Rank Aggregation

//The presented algorithm calculates the aggregated //ranks as per user's chosen parameters for web //service ranking

Input: User Chosen Parameters for Ranking

Output: Aggregated Ranked List of Services

BEGIN

Step 1. Take Choice of ranking parameters from user

Step 2. As per the chosen parameters, run the ranking algorithms to generate individual ranking lists.

Step 3. Calculate the Borda Count method with respect to given ranked lists.

Step 4. Provide the aggregated ranked list.

STOP

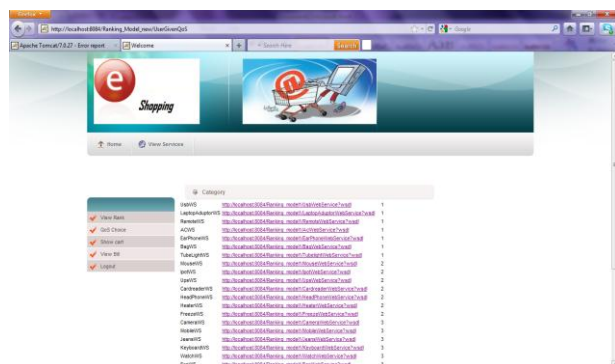
4. Results & Test Cases

For testing the method of service ranking based on Borda Count, a simple test domain for e-shopping had been taken where e- shopping services were ranked based on popularity, QoS of Offered service and User feedback. The aggregated ranked list has been shown in the Table 7 below. It has been observed that Borda Count method suffices well for a small data set of services and is computationally feasible in web service environment for ranking.

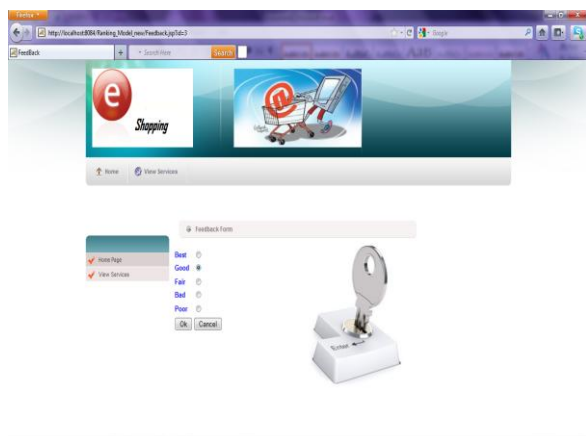
Table 7 : Overall Results for Rank Aggregation

Web Service Name	Rank By Cost	Rank By Feedback (OD Rank)	Rank By Popularity	Rank By Cost & Popularity	Rank By Feedback & Popularity	Rank By Cost, Feedback & Popularity
Ipadws	2	3	4	2	5	2
LaptopWS1	6	2	6	5	6	6
PCWS1	5	1	2	3	2	1
EarphoneWS	12	1	5	10	4	10
SmarttabWS1	6	2	3	5	3	4
IpodWS	8	2	13	7	10	12
UpsWS	7	2	8	6	7	9
HeadphoneWS	11	5	13	11	11	15
LaptopWS3	4	3	7	4	7	7
CameraWS	3	2	10	2	9	8
LaptopWS2	1	2	9	1	8	5
MobileWS	8	1	1	6	1	3
PalmtopWS	9	3	12	9	10	13
KeyboardWS	10	1	11	8	9	11
PCWS2	7	5	14	9	12	14

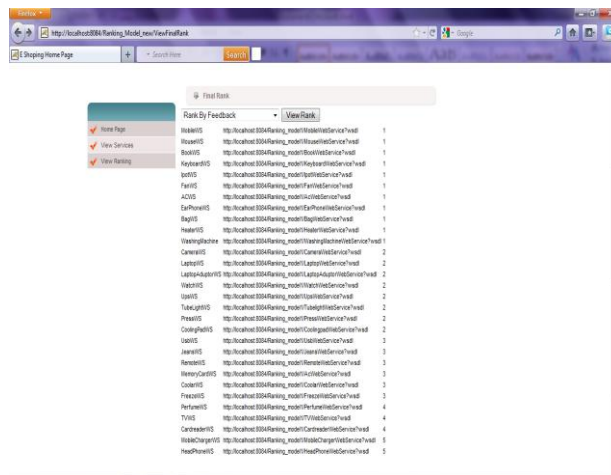
Test Case 1: Ranking based on User Given Quality of Service



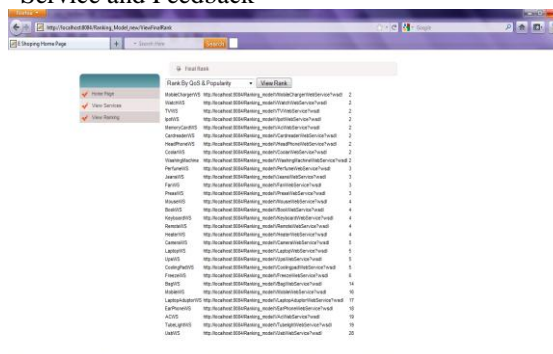
Test Case 2: User feedback of Web Services



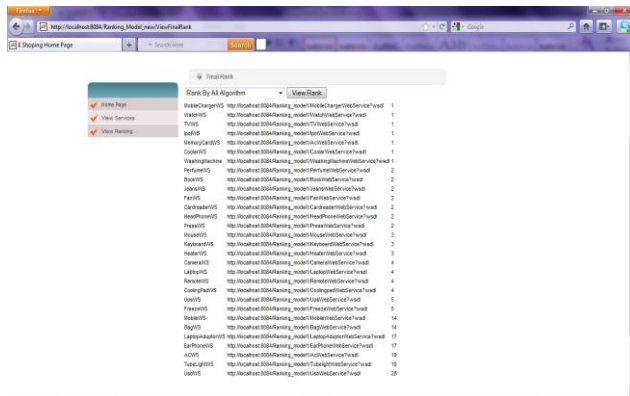
Test Case 3: Rank of Web Services on user Feedback



Test Case 4: Rank of Web Services using Quality of Service and Feedback



Test Case 5: Rank of Web Services by Quality of Service attribute, Popularity and User Feedback



5. Conclusions

An aggregate ranking algorithm based on user feedbacks/service rating has been proposed in this paper which not only calculates an overall rating given by a user but also calculates the overall service ranking for the candidate set of services chosen for a user. This way a user has been benefited from the experiences of his peers. Further there can be other parameters on which a service may be ranked such as Service usages i.e. popularity and QoS parameters. Hence an aggregate ranking algorithm has been proposed for overall ranking based on combination of these parameters which takes into account the fact that user should be able to choose the parameters on which the services are being ranked. This Service Rating & Ranking algorithm has been incorporated in the any service discovery architecture. A Service rating approach has been presented which is user friendly and encourages to a user to give rating for a used service. In this paper the rating and ranking issues that were handled are:

- To calculate Overall service rating based on subjective ratings supplied by a user through a given interface.
- To do the rate aggregation of rating lists provided by group of users.
- To rank the services based on aggregated rate list using rating vectors.
- To present a method for aggregating ranking lists based on different parameters to provide a final ranking list to a client.

The user benefits significantly from usage of these techniques as he is offered better services which are more appropriate to him in a ranked fashion. This in term allows him a better judgment while choosing the best service out of the pool of services which are functionally viable for him.

References

- [1] Web Services
www.w3schools.com/webservices/ws_intro.asp.
- [2] Service Oriented Architecture
<http://java.sun.com/developer/Books/j2ee/jwsa/>
- [3] Universal Description Discovery and Integration (UDDI) version 3.0.2 UDDI Spec Technical Committee Draft, Dated 20041019, http://www.uddi.org/pubs/uddi_v3.htm,
- [4] Le-Hung Vu; Porto, F.; Aberer, K.; Hauswirth, M., "An Extensible and Personalized Approach to QoS-enabled Service Discovery," In proceedings of 11th International Database Engineering and Applications Symposium, IDEAS 2007, pp.37-45, doi:10.1109/IDEAS.2007.4318087
<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4318087&isnumber=4318075>
- [5] Z. Chen, C. Liang-Tien, B. Silverajan, L. Bu-Sung, "UX - An Architecture Providing QoS-Aware and Federated Support for UDDI", *Proceedings of ICWS'03*.
- [6] F. Emekci, O. D. Sahin, D. Agrawal, A. E. Abbadi, "A Peer- to-Peer Framework for Web Service Discovery with Ranking", *Proceedings of ICWS'04*.
- [7] Z. Xu, P. Martin, W. Powley, and F. Zulkernine, "Reputation-Enhanced QoS-based Web Services Discovery," in IEEE International Conference on Web Services (ICWS 2007), pp. 9-13
- [8] Eyhab Al-Masri, Qusay H. Mahmoud, "QoS based Discovery and Ranking of Web Services", IEEE, 2007 pp 529-534.
- [9] WSRank: A Method for Web Service Ranking in Cloud Environment 2011 Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing
- [10] Anjela Y. Govan, Amy N. Langville and Carl D. Meyer. Offense-defense approach to ranking team sports. *Journal of Quantitative Analysis in Sports*. 2009, 5(1), pp. 1-17, 2009
- [11] Amy N. Langville, Carl D. D. Meyer, Who's #1?: The Science of Rating and Ranking ,Princeton University Press, 2012, pp.176-181.
- [12] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. 2001. Rank aggregation methods for the Web. In Proceedings of the 10th international conference on World Wide Web (WWW '01). ACM, New York, NY, USA, 613-622. <http://doi.acm.org/10.1145/371920.372165>

Vandan Tewari is currently pursuing Ph.D. program in Computer engineering in RGPV University, M.P., India. She is presently working as Asst. Prof. in Dept. of C.T.A at S.G.S.I.T.S., Indore. Her area of research is Web Services, Personalization, Data Mining and, Databases

Dr. Nirmal Dagde is a professor of computer science and has about 25 years of teaching experience in various engineering colleges. His areas of active research are data security, SOA and soft computing. He has authored several research papers that are published in reputed journals and conference proceedings. Currently, he is director of S. D. Bansal College Of Technology, Indore, India.

Dr. Aruna Tiwari is working as Assistant professor at Dept. Of Computer Sc. & Engg. at IIT, Indore. She has published more than 25 papers in Journals of repute. Her major area of research are Soft Computing, Data Mining and Databases

Energy Efficient Topology Control Approach for Mobile Ad hoc Networks

T.S.Asha¹ and Dr.N.J.R.Muniraj³

¹ Research Scholar, Karpagam University
Coimbatore

² Principal, Tejaa Shakthi Institute of Technology
Coimbatore

Abstract

In MANET, energy consumption and network connectivity are the two very important issues. Due to the mobility of the nodes, the network partition occurs unlimitedly. To avoid this, several researches concentrated on this issue. But it is not focused on constantly. In this work, we developed Energy Efficient Topology Control Approach (EETCA) is developed to attain both network connectivity and energy consumption. It consists of three main parts. Network and Interference model is introduced to make sure the network connectivity. Energy based topology control is developed to ensure more energy efficiency. Here the power consumption is also determined and validated in each and every route. Energy level of node is equally maintained in both route discovery and route maintenance phase. Packet format of EETCA is proposed which consists of power consumption and link availability. If any link is broken more power consumption will be occupied. So the status of link availability is kept on monitoring during this phase. By using the extensive simulation results using Network Simulator (NS2), the proposed scheme EETCA achieves better network lifetime, packet delivery ratio, less overhead and end to end delay than the existing schemes like NCTC and DM.

Keywords: *MANET, Interference, network connectivity, network lifetime, packet delivery ratio, end to end delay, overhead.*

1. Introduction

A. Mobile Ad Hoc Networks (MANET)

Mobile ad-hoc network is an independent system of mobile nodes connected by wireless links forming a short, live, on-the-fly network even when access to the Internet is unavailable. Nodes in MANETs generally operate on low power battery devices. These nodes can function both as hosts and as routers. As a host, nodes function as a source and destination in the network and as a router, nodes act as intermediate bridges between the source and the destination giving store-and-forward services to all the neighbouring nodes in the network. Easy deployment, speed of development, and decreased dependency on the infrastructure are the main reasons to use ad-hoc network.

B. The topology control problem in MANET

In mobile ad hoc wireless communication, each node of the network has a potential of varying the topology through the adjustment of its power transmission in relation to other nodes in the neighborhood. In contrast, wired networks have fixed established pre-configured infrastructure with centralized network management system structure in place. Therefore, the fundamental reason for the topology control scheme in MANET is to provide a control mechanism that maintains the network connectivity and performance optimization by prolonging network lifetime and maximizing network throughput. A MANET topology can depend on uncontrollable factors such as node mobility, weather, interference, noise as well as controllable factors such as transmission power, directional antennas and multi-channel communications.

A bad topology can impact negatively on the network capacity by limiting spatial reuse capability of the communication channel and also can greatly undermine the robustness of the network. Network capacity means that the bandwidth and ability for it to be used for communication. A network partitioning can occur in a situation where the network topology becomes too sparse. Similarly, a network which is too dense is prone to interference at the medium access (MAC) layer, the physical layer of the network. So the network should neither be too dense nor too sparse for efficient communication amongst nodes to take place.

C. Problem Definition

The problem identified in contemporary research literature pertaining to topology control in MANET is that most of the topology control algorithms do not achieve reliable and guaranteed network connectivity.

2. Related work

Dalu et.al [1] proposed topology control algorithm to maintain the topology without any control message. If any node goes out of range, communication would not get affected. The communication range is higher than the maximum allowable distance. The algorithm controls the movement of node with respect to a target node to make more connectivity of the network through the topology maintenance. Here there is no need

to change routing table as the connectivity of the network is maintained throughout the communication phase.

Manvi and Hurakadli [2] proposed agent based model to address the aspect of topology discovery and routing. In this model, three agents are used. Manager Agent handles the activities of route discovery and routing agency. Monitoring Agent is deployed to monitor resources like transmit power, battery life, bandwidth and reliability. Discovery and Routing Agents discover the links between the mobile nodes, perform routing information fusion and build pre-computed paths.

S.Muthuramalingam and R.Rajaram, [3] proposed clustering algorithm which reduces the number of clusters and optimize the load balancing factor. Network lifetime is also improved. This algorithm does topology management by the usage of coverage area of each node and power management based on mean transmission power within the context of wireless ad-hoc networks. By reducing the transmission range of the nodes, energy consumed by each node is decreased and topology is formed.

Ngo duc Thuan et.al [4] proposed Local Tree based Reliable Topology to update the effects on network connectivity. It preserves edge connectivity which is more reliable. This scheme considered to be the scalable and applicable scheme that is used in MANET. By experimental results, the proposed topology achieves high transmission range and more network connectivity than that of existing scheme.

Fujian Qin [5] proposed algorithm to construct topology that can meet the QoS requirements and decrease the total transmission power in the network. At physical layer, it is adopted that cooperative communication which combines partial signals to obtain complete information. At network layer, the whole topology information can be collected when it is not required to perform packet forwarding. Energy Efficient Topology Control and QoS formulation are proposed to achieve more network connectivity.

Bharathi and Saranya [6] proposed secure adaptive distributed control algorithm which aims at topology control and performs secure self organization. It consists of four phases like Anti node detection, Cluster formation, Key distribution and key renewal. In anti node detection, both encryption and decryption is performed to find the anti node. In cluster formation, cluster head is selected to ensure more information about network. Both key distribution and key renewal phase are developed to provide more integrity of data.

Jie Wu and Fei Dai [7] proposed mobility sensitive topology control method that consists of local view consistency based on synchronous and asynchronous method. A local view consists of locations of 1-hop neighbors within a normal transmission range. It is collected via exchanging "Hello" messages among neighbors and used to select logical neighbors at each node. The weak consistency is introduced to reduce the maintenance cost without any synchronization among neighbours.

Karunakaran and Thangaraj [8] proposed topology control algorithm based on power level. In this technique, topology control is maintained within each cluster. Initially, the cluster-head is selected based on factors power level, stability and connectivity. After the cluster-head selection, the connectivity of each node of the cluster with the cluster-head is checked. If the connectivity is low then the connectivity is increased by increasing the power level. In a cluster, cluster-head which is incharge of a dense area will experience low inter node interference. If there are any unidirectional links in the network then the cluster-head will form bidirectional link with it. If there are no unidirectional links, then the cluster-head will start linking up with the nodes that are not its direct neighbors. If the connectivity is higher than a threshold, then the cluster-head reduces the power level.

Suchismita Rout et.al [9] proposed Distance based Sleep Scheduling to deal with topology control problem at network layer and also help to reduce total energy consumption of network to maximize network lifetime. In this protocol, it takes farthest node in its transmission range for routing. That node is geographically closer to the destination. The number of nodes in packet transferring is less. Energy conservation is done by utilizing energy of small set of nodes. Sleep based approaches for the other idle mode node minimizes energy consumption of the network.

Uma and Shantharajah [10] proposed an energy efficiency analysis topology control algorithm. Our algorithm dynamically adjusts transmission power of mobile nodes to construct new topology which can meet bandwidth and end-to-end delay constraints as well as minimize the total energy consumption in network. This model has been compared with AODV and DSDV protocols in CBR traffic model and the simulation results show that the proposed algorithm has a better performance.

Aron et.al [11] proposed the notion of energy management in the context of heterogeneous wireless mesh networks was introduced. The objective was to develop a minimum-energy distributed topology control that ensures a reduction in the amount of energy consumed per node during transmissions and without loss of connectivity. A three phased topology control algorithm is proposed that executes distributively per node. A node uses only the locally available information to determine the nodes that should be its logical neighbours at any given time. The execution of the algorithm is asynchronous from node to the next till convergence in the transmission power per node is achieved thus runs in one pass thereby reducing concerns on control overheads.

Xiang-Yang Li et.al [12] proposed power assignment and routing protocols and performed extensive simulations to study the performance of our unicast routing protocols. When there is only one common source node, we show that our power assignment and routing are optimal. It is also presented a multicast routing protocol whose energy consumption is no more than two times of the minimum in a one-to-one communication model.

Joao B.D. Cabrea et.al [13] explored theoretic investigation of topology control in MANET. The Local information no topology algorithm is proposed to ensure more node degree. In this work

stability is maintained well but the node connectivity is not well improved. Maximum power is achieved at each level.

Quansheng Guan [14] explored Capacity-Optimized Cooperative (COCO) topology control scheme to improve the network capacity in MANETs by jointly considering both upper layer network capacity and physical layer cooperative communications. They have introduced physical layer cooperative communications, topology control, and network capacity in MANETs. The network capacity of MANETs is improved with cooperative communications.

Atsushi yoshinari et.al [15] proposed mechanism with an adopted topology control technique, based on a localized algorithm, can maintain local connectivity which results in keeping global network connectivity although the network is dynamic. In the proposed topology update mechanism, the update interval in each node is determined based on the transmission range and mobility information of its adjacent nodes so that the network connectivity is guaranteed.

Srinivas Rao et.al [16] proposed power management schemes looks into two directions. First is to balance power consumption during data transfer and secondly to reduce the power consumed in case of a route failure. By balancing power consumption we can avoid the death of some critical nodes caused by excessive power consumption. Reducing power consumption intends to prolong the lifetime of each node which in turn extends the lifetime of the entire network. Each approach proposed in the following sections improves the network's performance either by balancing the power across the network or by reducing the power consumed by the nodes across the network.

The paper is organized as follows. The Section 1 describes introduction about MANET, topology control problem in MANET. Section 2 deals with the previous work which is related to the topology control. Section 3 is devoted for the implementation of Energy Efficient Topology Control. Section 4 describes the performance analysis and the last section concludes the work.

3. Implementation of Energy Efficient Topology Control Approach

In the proposed Energy Efficient Topology Control Approach, there are three phases involved. In first phase, we aim to propose network model. In second phase, we propose the energy based topology control approach which ensures the more energy efficiency of the node. In third phase, new packet format is proposed which contains the status of energy and interference level.

3.1 Network and Interference model

Let G be the collection of mobile nodes K denotes the graph on G in which there is an edge from node m to n . Let τ be the

topology status which depends on energy consumption, throughput, load balancing, network connectivity and mobility.

In a centralized model of mobile networks, the connected topology is constructed that minimizes the maximum interference. It is also introduced that centralized and localized methods for reducing link interference with spanning ratio. In this algorithm, edges are sorted by their weights in ascending order. Starting from the edge with minimum weight, in each iteration of the algorithm an edge mn is processed. If nodes m and n are already connected in the induced graph, the edge mn is just ignored and otherwise it will be added to the topology. The algorithm continues till a connected graph is constructed. The time complexity of this approach is $O(m \log m + hn)$ where h is the number of links in the final structure H . The illustration of Energy Efficient topology Model is shown in fig.1.

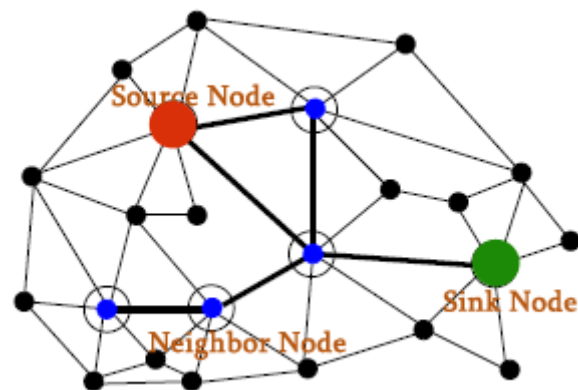


Fig.1. Topology Model of EETCA

A. Energy Efficient Topology Control

In this scheme, link based topology information is used to maintain a connected topology. If a route update indicates that a link failure has occurred such that the network is no longer connected, the appropriate nodes increase their transmit power until it is connected. This technique depends heavily on routing protocol performance, because changes in network connectivity can trigger further routing updates.

The main issue of minimum energy consumption model is to minimize the total energy consumed in forwarding a packet from source to destination mobile nodes. It can exploit path loss and packet loss by forwarding traffic using a sequence of low power transmissions rather than a single direct transmission.

The signal to noise plus interference ratio (SNIR) for successful transmission at the receive node must be greater than some threshold, which depends on the bit error rate. In a basic path loss model, received signal strength decreases exponentially with distance. The measurement data presented earlier show that it is necessary to account for energy consumed in both transmitting and receiving when evaluating the energy cost of a path. The former depends on the transmit power used at each hop, while the latter is roughly constant. If a relay node is added to a minimum hop-count path, the energy saved though reduced transmit power must compensate for the energy consumed by the overhead of the extra transmit and receive operations.

The cost of transmitting a packet over minimum energy consumption link is given as,

$$C(N, t) = \max_k N_k(t) \quad (1)$$

$$N_k(t) = \min_q E_q(t) / P_{kq}(t) \quad (2)$$

Where t is time $E_q(t)$ is the remaining energy of the node q assumed to be known from hardware. $P_{kq}(t)$ is the transmit power of node k in route q as stored in the received packet.

Route discovery and route maintenance are the main components of Energy based Topology model. Initially the route discovery is initiated when the route cache at the source node does not have any entry for the destination node. The source node broadcasts a route request message.

A node that receives the request can do one of two things;

- Forward the request after appending its own id if it's not the destination, or reply using its cached routes.
- The destination would reply and reply messages propagate back to the source. A node ignores a request if it has already processed it.
- It uses the route with maximum remaining lifetime. Remaining lifetime of a node in a route is defined as remaining node energy divided by power required to transmit packet to the next node in the route.
- Remaining lifetime of a route is then minimum of remaining life of nodes in the route. Following the notation used.

Route maintenance is achieved by using Medium Access Control layer acknowledgments to confirm retrieval of packet information about a broken link is propagated back along the route. Nodes invalidate all routes containing the broken link. The source then tries to find the next route in cache. If it is node, the route discovery is initiated.

More precisely, nodes upon receiving of the packets can calculate minimal energy necessary to reach their single hop neighbors (these associations are stored in power table) using the following formula:

$$P_{\min} = P_{tx} - P_{recv} + P_{\text{threshold}} + P_{\text{margin}} - P_{\text{drop}}$$

Where P_{\min} is the minimum required power of the sender to destination node. P_{tx} and P_{recv} are the transmitter and received power. $P_{\text{threshold}}$ is threshold power. P_{margin} is the margin to safeguard against such as channel fluctuation and mobility. The following steps are proposed to verify link availability and power consumption level.

Input : An instance K of MANET. Let f be the value of the maximum speed times the unit time interval, and let P_{maximum} be the common transmission power.

Output: The power assignments to nodes $\{p_1, p_2, \dots, p_n\}$

Where K is the movements connected.

Steps: Each node C_m independently. The following steps are used to estimate that packet transmissions utilizes P_{max} before P_i is computed.

1. $R'_k = R_k - 2d$, where R_k is the maximum transmission range of C_m .
2. Each time the transmission range of C_m is updated to R'_k .
3. Sort C_m 1-hop neighbors in to the list $L_k = (C_1, C_2, \dots, C_m)$ in ascending order by the received signal strength $\{C_{mk}\}$.
4. Broadcast List L_k to every node in L'_k .
5. For $C_{mk} \in L_k$ in ascending order do
 If $\{ \exists_k > l : C_{mk'} \in L_k, C_{nk} = C_{nk'}, |C_{nk}| < |C'_{nk}| \}$,
 then eliminate c_{nk} from L_k ; else break out of For the Loop.
6. Let d_m be the distance between C_k and the first node in L_k .
7. $p_k \leftarrow PL(d_m + 2r)$ where Power consumption threshold level maps distance to power according to the communication radio model.

3.2 Packet format of EETCA

Source/ Destination	Next hop	Link Availability	Hop count	Power consumption	CRC
4	1	4	2	4	1

Fig.2.Proposed Packet format

In fig 2, first field occupies source and destination address. It occupies 4 bytes. In second field next hop is determined to achieve shortest path. In third, link availability occupies 4 bytes which is monitored to ensure network connectivity. Power consumption of all the nodes are determined and send back to source node. It occupies 4 bytes. Cyclic Redundancy Check is in the last field for error detection and error correction. It occupies 1 byte. It can be used by upper level routing algorithm to find a least weighted path. The network topology under EETCA is all the nodes in L and their individually perceived logical neighbor relations. The flow chart of EETCA is shown in fig.3.

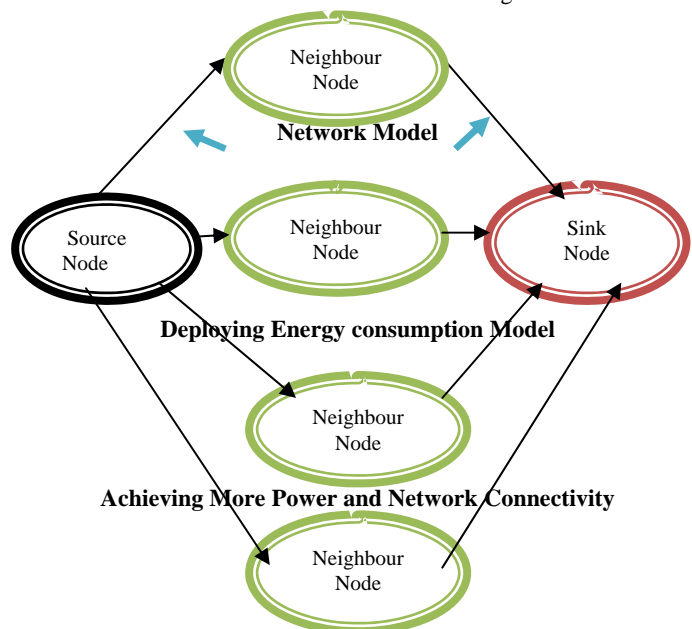


Fig.3.Flowchart of EETCA

4. Performance Analysis

Network Simulator (NS) is an event driven network simulator developed at UC Berkeley that simulates variety of IP networks.

It implements network protocols such as TCP and UDP, traffic source behavior such as FTP, Telnet, Web, CBR and VBR, router queue management mechanism such as Drop Tail, RED and CBQ, routing algorithms such as Dijkstra, and more. NS also implements multicasting and some of the MAC layer protocols for LAN simulations. Currently, NS (version 2) written in C++ and OTel (Tel script language with Object-oriented extensions developed at MIT) is available.

We use NS2 to simulate our proposed algorithm. In our simulation, 200 mobile nodes move in a 1200 meter x 1200 meter square region for 80 seconds simulation time. All nodes have the same transmission range of 300 meters. The simulated traffic is Constant Bit Rate (CBR). Our simulation settings and parameters are summarized in table 1.

Table 1. Simulation settings and parameters of EETCA

No. of Nodes	200
Area Size	1200 X 1200
Mac	802.11
Radio Range	300m
Simulation Time	50 sec
Traffic Source	CBR
Packet Size	512 bytes
Mobility Model	Random Way Point
Protocol	DSR
Packet rate	6pkts/sec

Performance Metrics

We evaluate mainly the performance according to the following metrics.

Control overhead: The control overhead is defined as the total number of routing control packets normalized by the total number of received data packets.

Packet Delivery Ratio: The packet delivery ratio (PDR) of a network is defined as the ratio of total number of data packets actually received and total number of data packets transmitted by senders.

Node degree: It is the important metric to evaluate the performance of topology control algorithms. If the node degree is higher, it indicates that higher collision will be. So value of node degree should be kept small.

Network connectivity ratio: It determines the nodes are connected in the intermediate region. It should be kept small while varying the average speed.

End-to-End Delay: The End-to-End delay is defined as the difference between two time instances: one when packet is generated at the sender and the other, when packet is received by the receiving application.

The simulation results are presented in the next part. We compare our proposed scheme EETCA with NCTC and DM [17] in presence of topology control environment. Figure 4 shows the results of connectivity ratio for varying the mobility from 5 to 25. From the results, we can see that EETCA scheme has slightly lower connectivity ratio than the NCTC and DM method because of light weight calculations.

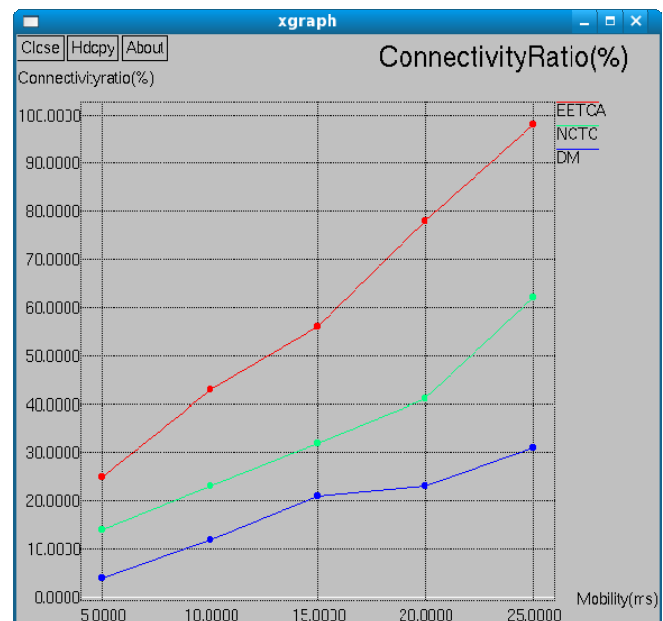


Fig.4. Mobility Vs Connectivity Ratio

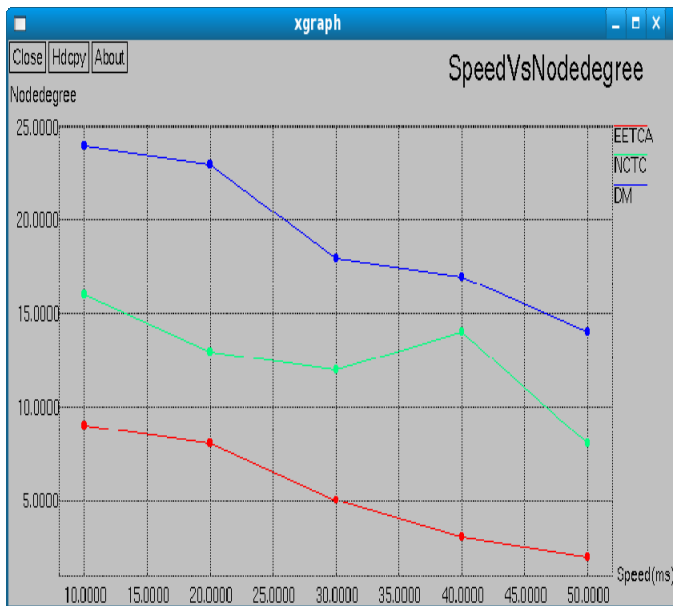


Fig. 5. Speed Vs Node degree

Fig. 5, presents the comparison of node degree. It is clearly shown that the node degree of EETCA has low overhead than the NCTC and DM.

Figure 6 shows the results of Time Vs End to end delay. From the results, we can see that EETCA scheme has slightly lower delay than the NCTC and DM scheme because of stable routines.

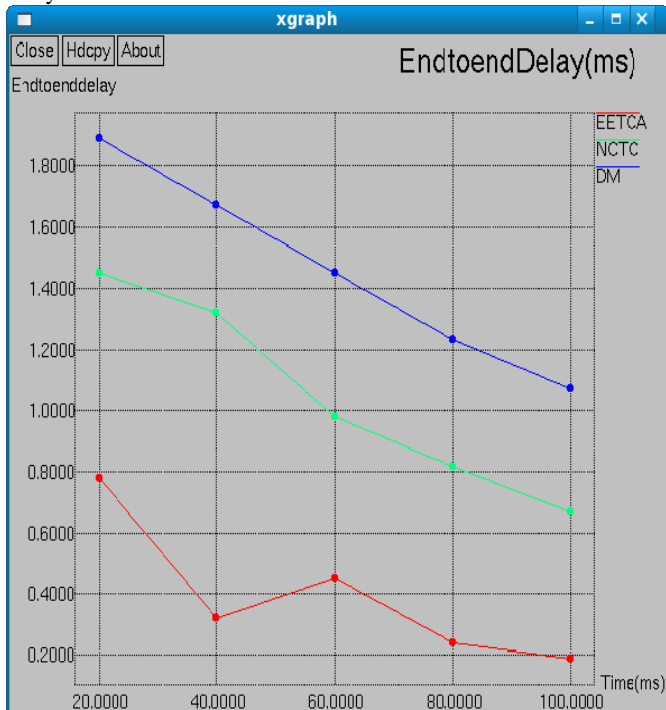


Fig. 6. Time Vs End to end delay

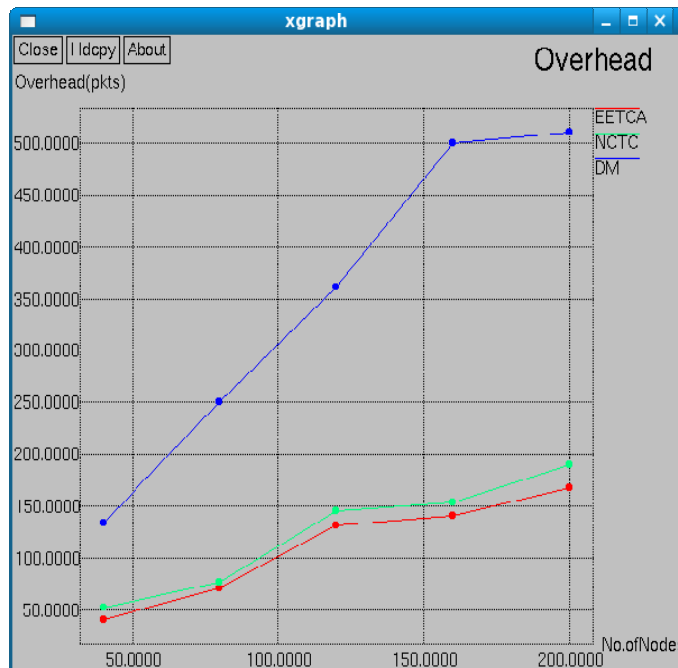


Fig.7.No. of Nodes Vs Overhead

Fig. 7, presents the comparison of overhead while varying the nodes from 0 to 200. It is clearly shown that the overhead of EETCA has low overhead than the NCTC and DM method.

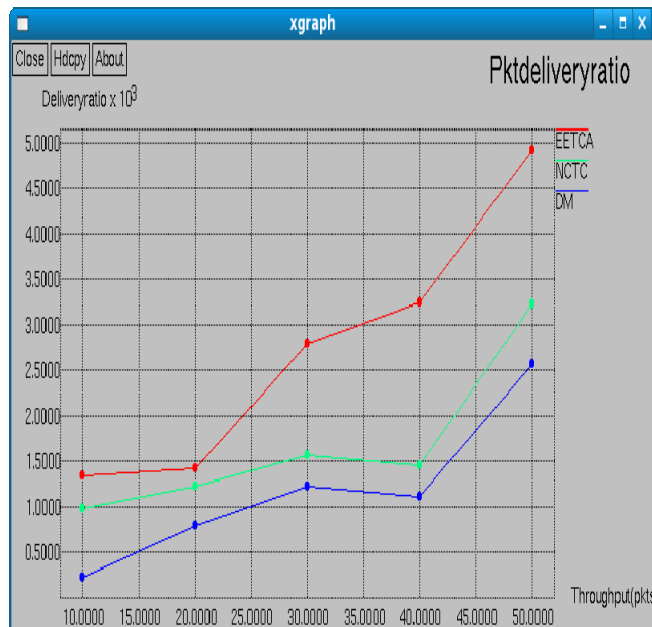


Fig.8. Throughput Vs Packet Delivery Ratio

Figure 8 shows the results of average packet delivery ratio for the throughput 10, 20...50 for the 200 nodes scenario. Clearly our EETCA scheme achieves more delivery ratio than the NCTC and DM scheme since it has topology control features.

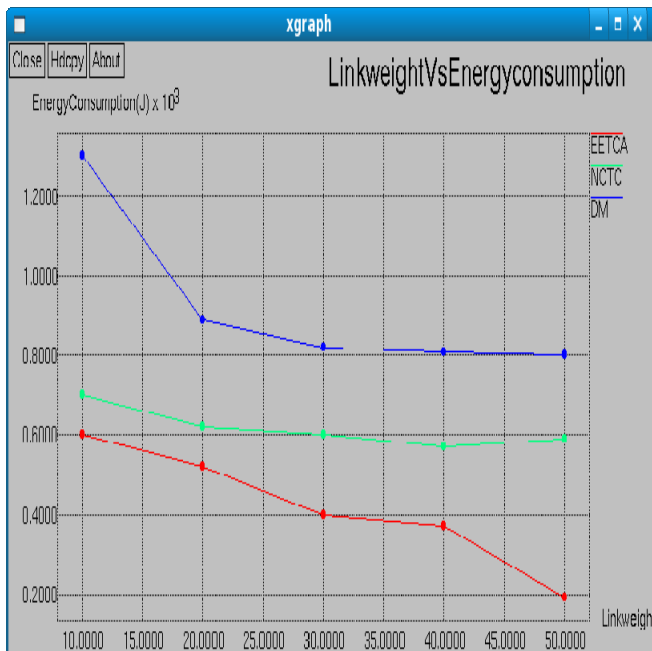


Fig.9. LinkWeight Vs Energy Consumption

Figure 9 show the results of energy consumption for the link weight 10, 20...50 packets for the 200 nodes scenario. Clearly our EETCA scheme achieves consumes less energy than the NCTC and DM scheme since it has topology control features.

5. Conclusions

Mobile nodes are communicating without any access point in MANETs. Due to the uncontrolled topologies, the more interference and more energy consumption is introduced in the networks which degrades the performance of network connectivity. In this paper, we have introduced the Energy Efficient Topology Control Approach to make the correct balance between the energy efficiency and network connectivity. In first phase, we have achieved low interference using based on the recommendation of neighbor nodes. In second phase, the energy based efficient topology control is introduced to extend the network lifetime and energy efficiency of MANET. Packet format is designed and integrated among the network to keep on monitoring the power consumption and link availability. By simulation results we have shown that EETCA achieves good packet delivery ratio, better network lifetime while attaining low delay, overhead, while varying the number of nodes, node velocity and mobility than our previous scheme NCTC and existing scheme DM.

6. References

[1] S. S. Dalu M. K. Naskar and C. K. Sarkar, "Implementation of a Topology Control Algorithm for Mobile Ad hoc Networks Using Pursue Mobility Model", *Journal of Applied Computer Science*, Vol.2, no.2, 2008, pp.14-20

[2] S.S.Manvi and J.J.Hurakadli, "An agent-based model for topology discovery and routing information fusion

in mobile ad-hoc networks", *Journal of Indian Institute of Science*, Vol.86, pp.553-568.

[3] S.Muthuramalingam and R.Rajaram, "A Transmission Range based Clustering Algorithm for Topology Control MANET", *International journal on applications of graph theory in wireless ad hoc networks and sensor networks (GRAPH-HOC)* Vol.2, No.3, 2010, pp68-76.

[4] Ngo Duc Thuan, Hiroki Nishiyama, Nirwan Ansari and Nei Kato, "On Performance Evaluation of Reliable Topology Control Algorithms in Mobile Ad Hoc Networks", *IEEE Conferences*, 2010, pp.1-5.

[5] Fujian Qin, "QoS Topology Control with Energy Efficiency for MANET", *Journal of Convergence Information Technology*, Volume6, Number 6, 2011, pp.300-307.

[6] R. Bharathi, Saranya. A.V, "High Throughput Analysis Using Topological Control & Authentication Scheme in MANET", *International Journal of Emerging Technology and Advanced Engineering*, Volume 3, Issue 3, 2013, pp.857-861.

[7] Jie Wu and Fei Dai, "Mobility-Sensitive Topology Control in Mobile Ad Hoc Networks", *Proceedings of the 18th International Parallel and Distributed Processing Symposium*, 2004, pp.1-29.

[8] S. Karunakaran and P. Thangaraj, "Topology Control Using Efficient Power Management", *Journal of Computer Science*, Vol.7, No.4, 2011, pp.561-567.

[9] Suchismita Rout, Ashok Kumar Turuk and Bibhudatta Sahoo, "Energy Aware Routing Protocol in MANET using Power Efficient Topology Control Method", *International Journal of Computer Applications* (0975 – 8887) Volume 43, No.5, 2012, pp.33-42.

[10] S. Uma and S.P.Shantharajah, "MANET Energy efficiency analysis using MST based topology control algorithm and comparison with reactive and proactive protocols", *International Journal of Advances in Computer Networks and its Security*, Volume 2, Issue 2, pp.23-27.

[11] F. O. Aron, T. O. Olwal, A. Kurien, and M. O. Odhiambo, "A Distributed Topology Control Algorithm to Conserve Energy in Heterogeneous Wireless Mesh Networks", *World Academy of Science, Engineering and Technology*, 2008, pp.530-536.

[12] Xiang-Yang Li, Yu Wang, Haiming Chen, Xiaowen Chu, Yanwei Wu and Yong Qi, "Reliable and Energy-Efficient Routing for Static Wireless Ad Hoc Networks with Unreliable Links", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 20, No. 10, October 2009, pp.1408-1421.

[13] Joao B. D. Cabrera, Ram Ramanathan, Carlos Gutierrez, and Raman K. Mehra, "Stable Topology Control for Mobile Ad-Hoc Networks", *IEEE Communications Letters*, Vol. 11, No. 7, 2007, pp.574-576.

[14] Quansheng Guan, F.RichardYu, Shengming Jiang and Victor C.M.Lung, "Topology Control in Mobile Ad

- hoc Networks With Cooperative Communications”,
IEEE Wireless communications, 2012, pp.2-7.
- [15] Atsushi Yoshinari, Hiroki Nishiyama, Nei Kato, and Dan Keun Sung, "Dynamic Topology Update Mechanism in Local Tree-based Reliable Topology (LTRT) based MANETs," IEEE International Conference on Communications (ICC 2012), Ottawa, Canada, Jun. 2012, pp.1-6.
- [16] Dr D Srinivas Rao, Jooby Ann Thomas and Sake Pothalaiah, "QoS through Power Control in MANETs using AODV Protocol", International Journal of Computer Applications (0975 – 8887) , Volume 31– No.9, October 2011, pp.23-29.
- [17] Hiroki Nishiyama, Thuan Ngo, Nirwan Ansari, and Nei Kato, "On Minimizing the Impact of Mobility on Topology Control in Mobile Ad Hoc Networks," IEEE Transactions on Wireless Communications, vol.11, no.3, pp.1158-1166, Mar. 2012.

Mrs. T.S. Asha is presently working as the Associate professor in N S S College of Engineering, Palakkad. She has 20 years of teaching experiences and published one international journal and five national papers. She had taken her B.E Degree from The Indian Engineering College, Vadakkankulam in 1990. She has her M Tech degree in Opto Electronics and Laser technology from Cochin university of Science and Technology. Her research area include Wireless mobile communication and Optical communication.

Dr. N.J.R.Muniraj is presently working as a Principal of Tejaa Shakthi institute of Technology, Coimbatore. He has more than 23 years of teaching and five years of industrial experience. He has presented more than 40 National and International papers and published fifteen international journal papers. His research area includes VLSI Signal Processing, Neural Networks, Image Processing and MEMS. He is also heading the Tejaa Shakthi Innovation centre.

A High-Speed Residue-to-Binary Converter for Three-Moduli $\{2^{2n+2} - 1, 2^{2n+1} - 1, 2^n\}$ set

Amani Goniemat, Andraws Swidan
Computer Engineering Department, The University of Jordan
Amman, Jordan

Abstract

Residue-to-binary conversion is the crucial step for residue arithmetic. The traditional methods are the Chinese Remainder Theorem and the Mixed Radix Conversion. Both approaches have some well known long standing difficulties, new Chinese remainder theorem used to overcome those difficulties. In this paper presents, and a new converter for specific moduli set $\{2^{2n+2} - 1, 2^{2n+1} - 1, 2^n\}$ was proposed. Our proposed converter is based on the CRTI. A detailed comparative analysis of the proposed converter was carried out. The analysis showed that our proposed converter overcomes other converters by speed and at the same time it is comparable to them by hardware requirements.

Keywords: Residue Number System, Reverse Converters, New Chinese Remainder Theorem.

1. Introduction

Residue number system (RNS) is a non-weighted system and uses residues of a number in particular modulus for its representation.

Arithmetic operations on residues can be performed in each moduli in parallel without carry propagation between them, thus one of the important characteristic of using residue arithmetic is the carry-free property which increases the calculation speed and decrease the consumed power.

In addition instead of performing arithmetic operations on large number, calculations are done on its corresponding residues in parallel. Hence the hardware requirement is reduced and speed operations is improved moreover all tasks are performed parallel.

Considering the characteristics of RNS, it has been applied on many arithmetic operations such as fast number theoretic transforms, discrete Fourier transforms and many other areas. Also it received a considerable attention since

1950 in image processing, digital filters and digital signal processing computation algorithms (DSP).

Unlike conventional number system RNS bear the extra cost of conversion step that is user to interface the RNS with the external world either for convert the binary to residue representation for forward conversion or the inverse in reverse conversion to produce the binary equivalent of residues. Other critical issue concerning the use of RNS is the choice of moduli set as the form of the moduli set and the number of moduli that chosen for RNS processor affects on dynamic range, speed and its VLSI implementation.

Up to now, many moduli sets have been presented with various dynamic ranges either $3n$, $4n$, $5n$ bits with three, four, five and even six moduli sets but always the trend is to offer a moduli set that meet high performance needs with a large dynamic range and parallelism.

In this paper a new three modules $\{2^{2n+2} - 1, 2^{2n+1} - 1, 2^n\}$ with $(5n)$ -bits dynamic range is proposed and a high speed and low complex reverse converter is designed based on new CRT algorithm.

2. Related background

The Residue Number System is defined in terms of a relatively-prime modulus set $\{P_1, P_1, \dots, P_n\}$ where $\gcd(P_i, P_j) = 1$ for $i \neq j$, and $\gcd(a, b)$ denotes the greatest common divisor of a and b . A weighted number X can be represented as $X = (x_1, x_2, \dots, x_n)$, where this representation is unique for any integer X in range $[0, M - 1]$, where M is the Dynamic range of the modulus set and defined as $M = P_1 P_2 \dots P_n$.

In order to convert from binary to residue numbers and vice-versa, a binary to residue (forward converter) is required in the front end of the system and a residue to binary in the bank end of the system. Reverse conversion involves a significant degree of complexity; hence an

efficient design of reverse converter greatly simplifies the operations in RNS.

The algorithms of residue to binary conversion are mainly based on the Chinese remainder theorem (CRT) and mixed radix conversion (MRC), and recently a new implementation of (CRT) is proposed defined as (CRT-I) and (CRT-I I). The New CRTs have potentiality to create higher performance reverse converters than CRT and MRC particularly for some special moduli sets. Hence, many researchers have been done in the recent years to discover efficient moduli sets which can be fitted with properties of New CRTs.

Chinese Remainder Theorem: Let m_1, m_2, \dots, m_n be pairwise relatively prime positive integers, i.e. $\gcd(m_i, m_j) = 1$ for $i \neq j$. The system

$$\begin{aligned} x &\equiv a_1 \pmod{m_1} \\ x &\equiv a_2 \pmod{m_2} \end{aligned}$$

.

.

.

$$x \equiv a_n \pmod{m_n}$$

has a unique solution modulo $m = m_1 m_2 \dots m_n$, i.e., there is a unique solution x with $0 \leq x < m$. Furthermore, all solutions are congruent modulo m .

We can construct a solution as follows.

1. Let $m = m_1 m_2 \dots m_n$.
 2. Let $M_k = \frac{m}{m_k}$ for all $k = 1, 2, \dots, n$.
 3. For all $k = 1, 2, \dots, n$ find integers y_k such $M_k y_k \equiv 1 \pmod{m_k}$
- Since $\gcd(M_k, m_k) = 1$, we know that y_k exists. Euclid's extended algorithm can be used to Find y_k .
- The integer $a_1 M_1 y_1 + a_2 M_2 y_2 + \dots + a_n M_n y_n$ is a solution of the system. The integer
- $$x = (a_1 M_1 y_1 + a_2 M_2 y_2 + \dots + a_n M_n y_n) \pmod{m} \quad (1)$$
- is the unique solution with $0 \leq x < m$.

The Mixed Radix Conversion: The residue to binary converter can be implemented using the MRC as follow

$$X = V_n \prod_{i=1}^n P_i + \dots + V_3 P_2 P_1 + V_2 P_1 + V_1 \quad (2)$$

The coefficients $V_i P$ can be obtained from residues by:

$$V_1 = x_1 \quad (3)$$

$$V_2 = |(x_2 - x_1) |P_1^{-1}|_{P_2} |_{P_2} \quad (4)$$

$$V_3 = |((x_3 - x_1) |P_1^{-1}|_{P_3} - V_2) |P_2^{-1}|_{P_3} |_{P_3} \quad (5)$$

In general case we have

$$V_n = (((x_n - V_1) |P_1^{-1}|_{P_n} - V_2) |P_2^{-1}|_{P_n} - \dots - V_{n-1}) |P_{n-1}^{-1}|_{P_n} \quad (6)$$

New Chinese Remainder Theorem 1: by New CRT-I with the 3-moduli set $\{P_1, P_2, P_3\}$ the number X can be computed from its corresponding residues (x_1, x_2, x_3) using the following equations

$$Z = x_1 + m_1 |K_1 (X_2 - X_1) + K_2 M_2 (X_3 - X_2) |_{m_2 m_3} \quad (7)$$

Where

$$|k_1 m_1 |_{m_2 m_3} = 1 \quad (8)$$

$$|k_2 m_1 m_2 |_{m_3} = 1 \quad (9)$$

Where k_1, k_2 are multiplicative inverses.

3. Design of Reverse Converter

For the design of reverse converter I use the new CRT theorem, the following lemmas and properties are needed for the derivation of the conversion algorithm

Theorem 1: modules are pairwise relatively prime.

Proof:

Using the Euclid's algorithm to find the greatest common divisor:

Euclid (a,b)

if b=0

return a

else

return Euclid (b,a mod b)

if a was 1 then we conclude that numbers are prime.

begin with

$$\gcd(2^{2n+2} - 1, 2^{2n+1} - 1) = \gcd(2^{2n+2} - 1, 1) = 1$$

so $2^{2n+2} - 1, 2^{2n+1} - 1$ are coprime.

For the moduli $\{2^{2n+1} - 1, 2^n\}$ and $\{2^{2n+2} - 1, 2^n\}$, either you can use the previous steps or noting that $2^{2n+2} - 1, 2^{2n+1} - 1$ are both odd numbers while 2^n is even so it is clear that 2^n is relatively prime with both $2^{2n+2} - 1$ and $2^{2n+1} - 1$

Lemma 1:

The multiplicative inverse of $|m_1 |_{m_2 m_3}$ is

$$K_1 = |(2^{2n+2} - 1)^{-1}|_{(2^{2n+1}-1)2^n} \quad (10)$$

$$K_1 = 2^{3n+1} + 2^{2n+2} - 2^n - 1 \quad (11)$$

Proof:

$$|(2^{3n+1} + 2^{2n+2} - 2^n - 1)(2^{2n+2} - 1)|_{2^{2n+1}-1, 2^n=1} \quad (12)$$

Lemma 2:

The multiplicative inverse of $|m_1 m_2|_{m_3}$ is

$$K_2 = |(2^{2n+2} - 1)(2^{2n+1} - 1)^{-1}|_{(2^{2n+1}-1)2^n} \quad (13)$$

$$K_2 = 1$$

Proof:

$$|T_1 + T_2 + T_3|_{2^n}, \quad x_2 \geq x_1 \quad (14a)$$

$$|T_1 + T_2 + T_3|_{2^n}, \quad x_2 < x_1 \quad (14b)$$

$$\begin{aligned} & |(2^{2n+2} - \\ & 1)(2^{2n+1} - \\ & 1)|_{2^n=1} \end{aligned}$$

the moduli set $\{2^{2n+2} - 1, 2^{2n+1} - 1, 2^n\}$ using the new CRT I:

$$X = x_1 + m_1 |K_1(x_2 - x_1) + K_2 m_2(x_3 - x_2)|_{m_2 m_3} \quad (18)$$

$$\begin{aligned} X = x_1 + (2^{2n+2} \\ - 1) | (2^{3n+1} + 2^{2n+2} - 2^n \\ - 1)(x_2 - x_1) \\ + (2^{2n+1} - 1)(x_3 \\ - x_2) |_{m_2 m_3} \end{aligned} \quad (19)$$

Where $m_2 = (2^{2n+1} - 1)$
 And $m_3 = 2^n$

The binary number $X = (x_1, x_2, x_3)$ is given by

$$X = x_1 + (2^{2n+2} - 1)Z \quad (20)$$

Where

$$Z = (2^{2n+1} - 1)Y + |x_2 - x_1|_{(2^{2n+1}-1)} \quad (21)$$

$$Y = |x_3 + x_2(2^n + 1) - 2x_1|_{2^n}$$

Where

$$T_1 = (x_{3,n-1} x_{3,n-2} \dots \dots \dots x_{3,1} x_{3,0}) \quad (22)$$

$$T_2 = (x_{2,n-1} x_{2,n-2} \dots \dots \dots x_{2,1} x_{2,0}) \quad (23)$$

$$T_{31} = (\overline{x_{1,n-2}} \dots \dots \dots \overline{x_{1,1}} \overline{x_{1,0}} 1) \quad (24)$$

$$T_3 = T_{31} + 1 \quad (25)$$

Proof:

The binary vectors x_1, x_2 and x_3 can be represented in bit-level as

$$x_1 = \underbrace{(x_{1,2n+1} x_{1,2n} \dots \dots \dots x_{1,1} x_{1,0})}$$

4. Conversion theorem

In this section I propose a theorem to convert the residue number (x_1, x_2, x_3) into binary representation for

$$x_2 = \underbrace{(x_{2,2n}x_{2,2n-1} \dots \dots \dots x_{2,1}x_{2,0})}_{2n + 1 \text{ bits}}$$

$$x_3 = \underbrace{(x_{3,n-1}x_{3,n-2} \dots \dots \dots x_{3,1}x_{3,0})}_{n \text{ bits}}$$

Using the new CRT I:

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & - 1 \right] (2^{3n+1} + 2^{2n+2} - 2^n \\ & - 1) (x_2 - x_1) \\ & + (2^{2n+1} - 1) (x_3 - x_2) \end{aligned} \right]_{m_2 m_3} \quad (26)$$

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & ((2^{n+1} - 1)(2^n + 2) + 1) (x_2 \\ & - x_1) + (2^{2n+1} - 1) (x_3 - x_2) \end{aligned} \right]_{m_2 m_3}$$

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & (x_2 - x_1) \\ & + (2^{2n+1} - 1) [(2^n + 2)(x_2 - x_1) \\ & + (x_3 - x_2)] \end{aligned} \right]_{m_2 m_3}$$

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & (x_2 - x_1) \\ & + (2^{2n+1} - 1) [2^n x_2 + 2x_2 - 2^n x_1 - 2x_1 \\ & + x_3 - x_2] \end{aligned} \right]_{m_2 m_3}$$

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & (x_2 - x_1) \\ & + (2^{2n+1} - 1) [x_2(2^n + 1) - x_1(2^n + 2) \\ & + x_3] \end{aligned} \right]_{m_2 m_3}$$

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & (x_2 - x_1) \\ & + (2^{2n+1} - 1) [x_2(2^n + 1) - x_1(2^n + 2) \\ & + x_3] \end{aligned} \right]_{(2^{2n+1}-1)(2^n)} \\ -x_1 = |2^n - x_1|_{2^n} \quad (27)$$

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & (x_2 - x_1) \\ & + (2^{2n+1} - 1) [x_2(2^n + 1) + x_1(2^n - 2^n - 2) \\ & + x_3] \end{aligned} \right]_{(2^{2n+1}-1)(2^n)}$$

$$X = x_1 + (2^{2n+2} - 1) \left[\begin{aligned} & (x_2 - x_1) \\ & + (2^{2n+1} - 1) [x_2(2^n + 1) + x_1(-2) \\ & + x_3] \end{aligned} \right]_{(2^{2n+1}-1)(2^n)}$$

$$X = x_1 + (2^{2n+2} - 1) Z$$

$$Z = [(x_2 - x_1) + (2^{2n+1} - 1) [x_2(2^n + 1) - 2x_1 + x_3]]_{(2^{2n+1}-1)(2^n)}$$

When $x_2 \geq x_1$

$$x_2 - x_1 = |x_2 - x_1|_{(2^{2n+1}-1)} \quad (28)$$

$$Z = (2^{2n+1} - 1) |x_3 + x_2(2^n + 1) - 2x_1|_{2^n} + |x_2 - x_1|_{(2^{2n+1}-1)}$$

$$Z = (2^{2n+1} - 1) Y + |x_2 - x_1|_{(2^{2n+1}-1)}$$

$$Y = |x_3 + x_2(2^n + 1) - 2x_1|_{2^n}$$

$$T_1 = |x_3|_{2^n} = (x_{3,n-1}x_{3,n-2} \dots \dots \dots x_{3,1}x_{3,0})$$

$$T_2 = |x_2(2^n + 1)|_{2^n} = |(x_2)(x_2)|_{2^n}$$

$$T_2 = (x_{2,n-1}x_{2,n-2} \dots \dots \dots x_{2,1}x_{2,0})$$

$$T_3 = |-2x_1|_{2^n}$$

$$T_3 = |-2(x_{1,n-1} \dots \dots \dots x_{1,1}x_{1,0})|_{2^n}$$

$$|2x_1|_{2^n} = (x_{1,n-2} \dots \dots \dots x_{1,1}x_{1,0}0)$$

$$T_{31} = (\overline{x_{1,n-2}} \dots \dots \dots \overline{x_{1,1}x_{1,0}1})$$

$$T_3 = T_{31} + 1$$

$$Y = |T_1 + T_2 + T_3|_{2^n}$$

When $x_2 < x_1$

$$x_2 - x_1 = |x_2 - x_1|_{2^{2n+1}-1} - 2^{2n+1} - 1$$

$$Z = [(2^{2n+1} - 1)(x_3 + x_2(2^n + 1) - 2x_1 - 1) + |x_2 - x_1|_{2^{2n+1}-1}]_{(2^{2n+1}-1)(2^n)}$$

$$Z = (2^{2n+1} - 1) |x_3 + x_2(2^n + 1) - 2x_1 - 1|_{2^n} + |x_2 - x_1|_{(2^{2n+1}-1)}$$

$$Z = (2^{2n+1} - 1) Y + |x_2 - x_1|_{(2^{2n+1}-1)}$$

$$Y = |x_3 + x_2(2^n + 1) - 2x_1 - 1|_{2^n}$$

$$T_1 = |x_3|_{2^n} = (x_{3,n-1}x_{3,n-2} \dots \dots \dots x_{3,1}x_{3,0})$$

$$T_2 = (x_{2,n-1}x_{2,n-2} \dots \dots \dots x_{2,1}x_{2,0})$$

$$T_3 = |-2x_1 - 1|_{2^n}$$

$$T_3 = T_{31} = (\overline{x_{1,n-2}} \dots \dots \dots \overline{x_{1,1}x_{1,0}}1)$$

$$Y = |T_1 + T_2 + +T_{31}|_{2^n}$$

5. Example

Given n=4, then

$$m_1 = 1023$$

$$m_2 = 511$$

$$m_3 = 16$$

And suppose that

$$x_1 = 2 = \langle 00 \rangle \langle 0000 \rangle \langle 0010 \rangle$$

$$x_2 = 3 = \langle 0 \rangle \langle 0000 \rangle \langle 0011 \rangle$$

$$x_3 = 1 = \langle 0001 \rangle$$

Then

$$T_1 = |X_3|_{2^n} = \langle 0001 \rangle$$

$$T_2 = (x_{2,n-1}x_{2,n-2} \dots \dots \dots x_{2,1}x_{2,0}) = \langle 0011 \rangle$$

And noting that $x_2 \geq x_1$

$$T_{31} = (\overline{x_{1,n-2}} \dots \dots \dots \overline{x_{1,1}x_{1,0}}1) = \langle 1011 \rangle$$

$$T_3 = T_{31} + 1 = \langle 1100 \rangle$$

$$Y = |1 + 3 + +12|_{2^n} = \langle 0000 \rangle$$

$$Z = (2^{2n+1} - 1)Y + |X_2 - X_1|_{(2^{2n+1}-1)}$$

$$Z = \langle 0000 \rangle \langle 0 \ 0000 \ 0001 \rangle$$

└──────────────────┘
3n+1 bits

$$X = X_1 + (2^{2n+2} - 1) Z$$

$$\begin{aligned} & X_1 + 2^{2n+2} Z \\ = & \langle 0000 \rangle \langle 0 \ 0000 \ 0001 \rangle \langle 00 \rangle \langle 0000 \rangle \langle 0010 \rangle \\ & + \\ - & Z \{ Z \text{ 2's complement} \} \\ = & \langle 1111 \rangle \langle 1 \ 1111 \ 1111 \rangle \langle 11 \rangle \langle 1111 \rangle \langle 1111 \rangle \end{aligned}$$

$$\begin{aligned} X = & \langle 0000 \rangle \langle 0 \ 0000 \ 0001 \rangle \langle 00 \rangle \langle 0000 \rangle \langle 0001 \rangle \\ = & 1025 \end{aligned}$$

Another example if we take

$$x_1 = 931 = \langle 11 \rangle \langle 1010 \rangle \langle 0011 \rangle$$

$$x_2 = 423 = \langle 1 \rangle \langle 1010 \rangle \langle 0111 \rangle$$

$$x_3 = 0 = \langle 0000 \rangle$$

Noting that $x_2 < x_1$

$$T_1 = |X_3|_{2^n} = \langle 0000 \rangle$$

$$T_2 = (x_{2,n-1}x_{2,n-2} \dots \dots \dots x_{2,1}x_{2,0}) = \langle 0111 \rangle$$

$$T_{31} = (\overline{x_{1,n-2}} \dots \dots \dots \overline{x_{1,1}x_{1,0}}1) = \langle 1001 \rangle$$

$$Y = |7 + 9 + +0|_{2^n} = \langle 0000 \rangle$$

$$Z = (2^{2n+1} - 1)Y + |X_2 - X_1|_{(2^{2n+1}-1)}$$

$$Z = \langle 0000 \rangle \langle 0 \ 0000 \ 0011 \rangle$$

$$X = X_1 + (2^{2n+2} - 1) Z$$

$$\begin{aligned} = & \langle 0000 \rangle \langle 0 \ 0000 \ 0011 \rangle \langle 11 \rangle \langle 1010 \rangle \langle 0011 \rangle + \\ & \langle 1111 \rangle \langle 1 \ 1111 \ 1111 \rangle \langle 11 \rangle \langle 1111 \rangle \langle 1101 \rangle + \\ = & \langle 0000 \rangle \langle 0 \ 0000 \ 0011 \rangle \langle 11 \rangle \langle 1010 \rangle \langle 0000 \rangle \\ = & 4000 \end{aligned}$$

6. Hardware Implementation

The proposed reverse converter is based on Eq. (13), Eq. (14), Eq.(15 a) and Eq.(15 b). The hardware implementation of Eq.(15) requires n-bits carry save adder to reduce these three input numbers into two numbers namely, sum and carry then n-bits module adder is used to add the sum and carry together to generate Y.

Let us denote the module that computes Y as Module Y, which employ one n-bits CSA the first used to compute then n-bits carry propagate adder (CPA) with end around carry (EAC) act as n-bits module adders, as shown in Fig. 1

Figure 2 show the execution of Eq. (14) we need a subtractor for $x_2 - x_1$ which can be implemented by (2n+1) bits CSA with EAC, this EAC bit can indicate whether $x_2 \geq x_1$ or not so can be used as control bit for n-bits 2-to-1 multiplexer that decide the correct Y ;
 $Y = |T_1 + T_2 + +T_3|_{2^n}$ when $x_2 \geq x_1$ and $Y = |T_1 + T_2 + +T_3|_{2^n}$ when $x_2 < x_1$.

To complete our implementation of Eq. (14) we simply concatenate Y with SUB1 output into a (3n+1) bits number then another subtractor SUB2 is used to generate Z

Finally for the implementation of Eq. (13) a (5n+3) bits number is generated by concatenation of Z and X_1 and using SUB3 to subtracts Z we can get the final X as shown in Fig. 2.

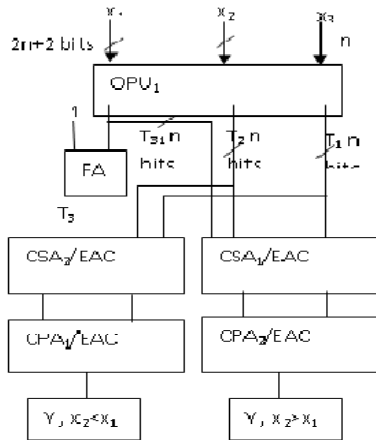


Fig. 1 Module Y: detailed architecture

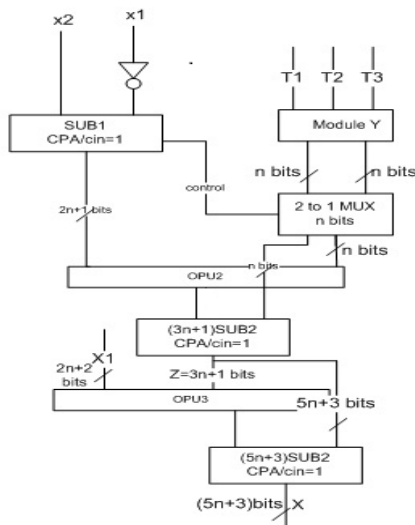


Fig. 2 Proposed Converter

The operands preparation unit of Fig. 1 contains (n) not gates for doing the inversion of x_1 to find its one's complement.

It is important to know that some parts of equation 13 and 14 can be implemented simply by concatenation without the use of any calculative hardware in Fig.3 we represent this by using OPU2, OPU3

Note that SUB_2 is (3n+1)-bit binary subtractor, which employ (3n+1) FA's and n-bit not gates to find the inversion of Y to compute subtraction.

Also since we have (2n+1) bits of 1's, (2n+1) FA's in SUB_2 can be reduced to a pairs of XNOR/OR gates. Again for the implementation of SUB_3 a (5n+3) regular binary subtractor is used where the (3n+1) of the (5n+3) FA's can be replaced by a pairs of XNOR/OR gates.

Performance evaluation

The primary digital characteristics of any digital design are the speed, area and power. The speed can be computed by throughput, latency and timing; the latency is the time between data input and the processing data outputs while timing is the logical delay between elements, when a design doesn't meet timing it means that the delay of the critical path is larger than the target clock period.

So to optimize the performance efficiently in your design you have to reduce the delay in critical path one way to do this by considering the amount of parallelism between entities and reduce the dependencies among them. Parallelism was implemented in the new CRT theorem where in the conversion process; the weighted number can be retrieved faster because the operations are done in parallel, without depending on other results.

In this section, hardware requirement and speed of the proposed reverse converter based on our moduli set is studied.

Firstly we must calculate the hardware requirement and delay of the proposed converter. Then compare the result with other converters from both hardware cost and delay viewpoints.

In Table 1 the complexity and delay introduced by different adders and gates used in the proposed converter are listed.

Table 1: Complexity and Delay of Various Components

Parts	FA	NOT	XNOR/OR pairs	Delay
OPU1		(n-1)		t_{not}
Adder	1			t_{FA}
CSA1	n			t_{FA}
CSA2	n			t_{FA}
CPA1	n			nt_{FA}
CPA2	n			nt_{FA}
SUB1	2n+1	2n+2		$(2n+1)t_{FA} + t_{not}$
MUX	n			t_{FA}
SUB2	n	n	2n+1	$(3n+1)t_{FA} + t_{not}$
SUB3	3n+1	3n+1	2n+2	$(5n+3)t_{FA} + t_{not}$
Total	11n+3	7n+2	4n+3	$(12n+9)t_{FA} + 4t_{not}$

To improve the throughput rate, pipelining is usually applied in real implementation, the delay of Module Y is smaller than that of SUB_1 , and hence the delay of the converter depends on the delay of the critical path consisting of SUB_1 , MUXs, SUB_2 , and SUB_3 . The delay of a CSA or a MUX is the same as that of an FA, namely, t_{FA} . The delay of each of the modulo adders in Module Y is nt_{FA} and that of the modulo subtractor SUB_1 is $(2n+1)t_{FA} + t_{not}$, while the delay of the binary subtractor SUB_2 is

$(3n+1)t_{FA} + t_{not}$, and that of SUB_3 is $(5n+3)t_{FA} + t_{not}$. Thus, the converter has a total delay of

$$\begin{aligned} \text{Delay} = & (2n + 1)t_{FA} + t_{not}SUB_1 + t_{FA}MUX \\ & + (3n + 1)t_{FA} + t_{not}SUB_2 \\ & + (5n + 3)t_{FA} + t_{not}SUB_3 \end{aligned}$$

$$\text{Total Delay} = (10n + 6)t_{FA} + 3t_{not}$$

So as described above doing the computations simultaneously in parallel improves the critical path delay. Table II describe the improvement in delay that was achieved using the new CRT for the conversion process in comparison with converters have the same or less dynamic range.

Table 2: Delay Comparison between the Proposed Reverse Converter and Related Works

Converter	DR(bits);n=4	Delay(t_{FA})
[12]	16	16n+22
[22]	22	14n+8
[28]	19	18n+17
proposed	22	10n+6

Hence, converters based the New CRT's require no big size modulo adders. In many cases, only one modulo operation is needed. The numbers involved in the conversion are smaller than the numbers in the CRT. This will gain speed, since binary arithmetic speed is often bounded by the size of the numbers. But, New CRTs are hardware intensive as they require many inverse modulus operators, modulus operators, multipliers and dividers. Dividers and inverse modulus operators in turn needs many half and full adders and subtractors. but since hardware cost has been driven low nowadays, illustration of the idea that there is always room for better performance.

In order to obtain comprehensive view of the improvement in implementation we extend Table I for different values of n as shown in Table III. since converter of [13] has 4n bits dynamic range we exclude it from comparison.

Table 3: Performance Comparison Using Different Values of n

N	[22]		[28]		Proposed	
	DR (bits)	Delay (t_{FA})	DR (bits)	Delay (t_{FA})	DR (bits)	delay (t_{FA})
4	22	64	19	89	22	46
8	62	120	39	161	62	86
12	82	176	59	233	82	126
16	102	232	79	305	102	166
20	122	288	99	377	122	206

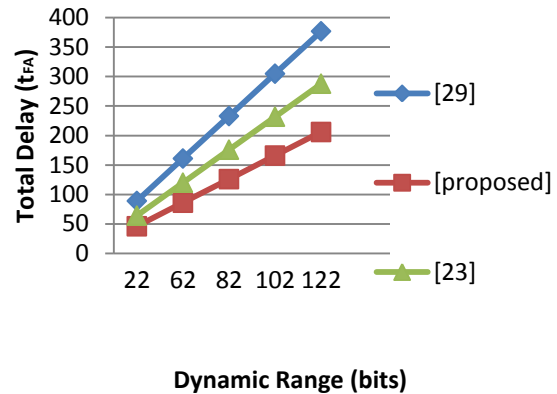


Fig. 3 comparison of delay for different converters

Based on the specific sets $M_i - 8, M_i - 16, M_i - 32$ and $M_i - 64$, the corresponding n that represent each converter found and delay of each moduli set for all converters $C_i - 8, C_i - 16, C_i - 32$ and $C_i - 64$ is computed. It is assumed that $C_1 = \{2^{2n+2} - 1, 2^{2n+1} - 1, 2^n\}$ based on MRC and $C_2 = \{2^n, 2^{n-1} - 1, 2^{n+1} - 1, 2^n - 1, 2^n + 1\}$, finally C_3 is the proposed converter which is for $\{2^{2n+2} - 1, 2^{2n+1} - 1, 2^n\}$ based on CRTI, for 64-bit converter n=13, for both.

Table 4: Specific seytsm_I-8, M_I-16, M_I-32, (I-1, ..., 4)

Converter	8-bit $M_i - 8$	16-bit $M_i - 16$	32-bit $M_i - 32$
$C_{1.1}$	{3,4,5,7,1}	{15,16,17,31,7}	{127,128,129,255,63}
$C_{1.2}$	{63,31,4}	{225,127,8}	{16383,8191,64}

Table 5: Delay for Specific Dynamic Range

Converter	$C_i - 8$	$C_i - 16$	$C_i - 32$	$C_i - 64$
$C_{1.1}$	53/n=2	89/n=4	143/n=7	251/n=13
$C_{1.2}$	36/n=2	50/n=3	120/n=8	190/n=13
$C_{1.3}$	26/n=2	36/n=3	86/n=8	136/n=13

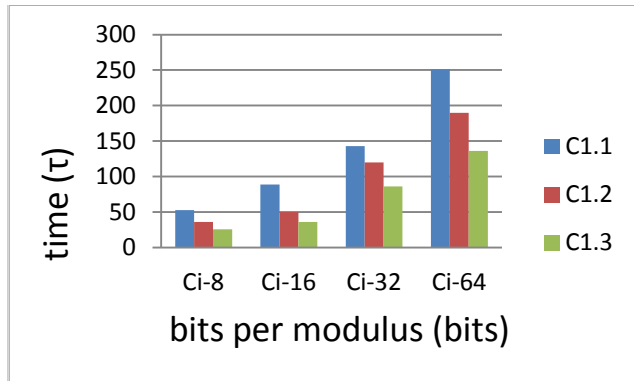


Fig. 4 Time Performance Comparison vs. dynamic range

Conclusion

A new converter for specific moduli set $\{2^{2n+2} - 1, 2^{2n+1} - 1, 2^n\}$ was proposed using New Chinese Remainder Theorem I. The design is compact and provides higher speed of conversion compared to other implementations that operate on the same set. The hardware requirements for the proposed converter are comparable to similar converters. The manipulation technique presented in this paper can serve as a guideline for similar design procedure and will open up many doors for further RNS research. It is expected more efficient arithmetic algorithms can be developed based on it, and many converters as the one proposed here can be implemented.

References

[1] [1] Bajard, Claude, and Plantard(2004) "RNS bases and conversions". proceedings of SPIE , 2004,Vol. 5559,Page:61-75.

[2] Bhardwaj, Srikanthan,and Clarke.(1999)"A reverse converter for the 4-moduli superset $\{2n-1, 2n, 2n+ 1, 2n+1 +1\}$." Computer arithmetic proceedings of 14th IEEE Symposium , 1999,Pages:168-175.

[3] Hiasat, and Sweidan(2003) "Residue number system to binary converter for the moduli set $(2n- 1, 2n - 1, 2 n+ 1)$ ". Journal of systems architecture ,2003,Vol.49,Pages: 53- 58.

[4] Hiasat, and Zohdy (1998)"Residue-to-binary arithmetic converter for the moduli set $(2k, 2k-1, 2 k-1-1)$ ". Circuits and Systems II: Analog and Digital Signal Processing of IEEE Transactions, 1998, Vol. 45,Pages: 204-209.

[5] Hariri, Arash, Navi, and Rastegar(2008)"A new high dynamic range moduli set with efficient reverse converter" .Computers & mathematics with applications , 2008, Vol.55, Pages: 660-668.

[6] Jameii, Mahdi, Taghipour, and Azad(2011) "Using both Binary and Residue Representations for Achieving Fast Converters in RNS".journal of advances in computer research ,2011, Pages: 91-104.

[7] Jenkins (1978)"Techniques for residue-to-analog conversion for residue-encoded digital filters". Circuits and Systems, IEEE Transactions , 1978,Vol. 25,Pages: 555-562.

[8] Lewis, Michael, Jon Mellott, and Taylot(2004) "An efficient residue to analog converter", IEEE International Conference , 2004 ,Vol. 5 ,Pages:157-60.

[9] Modiri, Samira, Movaghar, and Barati(2012) "Study of error control capability for the new moduli set $\{22n+ 1+ 2n-1, 22n+ 1-1, 2n-1, 23n, 23n+ 1-1\}$ ". Journal of Advanced Computer Science & Technology ,2012, Vol. 1, Pages: 176-186.

[10] Mohan,and Ananda(2002) "Residue number systems: algorithms and architectures". Springer, 2002,vol.1,page:1-268.

[11] Mohan, and Ananda.(2007)"RNS-to-binary converter for a new three-moduli set $\{2^{n+1}- 1, 2^n, 2^{n-1}\}$ ". Circuits and Systems II: Express Briefs, IEEE Transactions ,2007, Vol.54, Pages: 775-779.

[12] Mohan, Ananda, and Premkumar(2007) "RNS-to-binary converters for two four-moduli sets $\{2^{n-1}, 2^n, 2^{n+1}, 2^{n+1}- 1\}$ and $\{2^{n-1}, 2^n, 2^{n+1}, 2^{n+1}+ 1\}$." Circuits and Systems I: Regular Papers, IEEE Transactions ,2007, Vol. 54, Pages: 1245-1254.

[13] Molahosseini, and Sabbagh (2011)"Efficient Residue to Binary Conversion Based on a Modified Flexible Moduli Set" . Proceedings of AIP Conference ,2011, Vol. 1389,Pages: 1066-1069.

[14] Molahosseini, Sabbagh, Dadkhah, and Navi(2009) "A new five-moduli set for efficient hardware implementation of the reverse converter" .IEICE Electronics Express ,2009, Vol.6, Pages: 1006-1012.

[15] Navi, Keivan, Molahosseini, and Esmaeildoust (2011)"How to teach residue number system to computer scientists and engineers." IEEE Transactions ,2011,Vol.54, Pages: 156-163.

- [16] Omondi, Amos, and Premkumar.(2007) "Residue number systems: theory and implementation". Imperial College Press, 2007. Systems II: Analog and Digital Signal Processing, IEEE Transactions , 2000,Vol.47,Pages: 1576-1581.
- [17] Pham, Premkumar, and Madhukumar(2010)"Reduced complexity analogue-to-residue conversion employing folding number system" .Circuits, Devices & Systems,2010,Vol. 4,Pages: 30-41.
- [18] Premkumar,Ang, and Lai(2006) "Improved memoryless RNS forward converter based on the periodicity of residues". Circuits and Systems II: Express Briefs, IEEE Transactions,2006 ,Vol. 6, Pages: 133-137.
- [19] Radhakrishnan, and Preethy (1998) "A new approach to data conversion: direct analog-to-residue converter". Acoustics, Speech and Signal Processing, IEEE International Conference,1998,Vol. 5, Pages:3013 – 3016.
- [20] Radhakrishnan and Preethy (1999) "A parallel approach to direct analog-to-residue conversion".Information processing letter,1999,Vol. 69, Pages: 249-252.
- [21] Riazi, Sahel, Hassanpour, and Hosseinzadeh. (2011)"An efficient architecture of residue to binary converter for new four-moduli set".International Journal ,2011,Vol. 2,Pages: 709-715.
- [22] Samiria,Moraghar and Barati.(2012)"An Efficient Reverse Converter for the New Modulus Set{ $2^{2n+2}-1$, $2^{2n+1}-1$, 2^n }". International Journal of Advanced Research in Computer Science and Software Engineering,2012,Vol.2Pages:447-452
- [23] Shende, Radha, and Zode.(2012)"Efficient design $2k-1$ binary to residue converter." Devices, Circuits and Systems (ICDCS), International Conference IEEE, 2012, Pages: 482 - 485
- [24] Taheri, MohammadReza, Pirhoseinlo,Esmaeildoust,and Navi. (2012)"High speed reverse converter for high dynamic range moduli set." International Journal of Advances in Engineering & Technology,2012,vol.3.
- [25] Taheri, Reza, Pirhoseinlo,Esmaeildoust,and Navi.(2012)"Efficient reverse converter design for five moduli set { 2^n , $2^{2n+1}-1$, $2^{n/2}-1$, $2^{n/2}+1$, 2^{n+1} }". Journal of Computations & Modelling ,2012, Vol.2,Pages: 93-108.
- [26] Wang.(2003) " study of the residue-to-binary converters for the three-moduli sets" .Circuits and Systems I: Fundamental Theory and Applications, IEEE, 2003,Vol. 50, Pages: 235-243.
- [27] Wang,Swamy, and Wang . (2000)"A high-speed residue-to-binary converter for three-moduli ($2k$, $2k-1$, $2k-1-1$) RNS and a scheme for its VLSI implementation". Circuits and

An Energy Efficient Data Redundancy Reduction Approach for Data Aggregation in WSN

Bharathi M. A¹, B.P. Vijaya Kumar², Manjaiah D H³

¹Assoc. Prof: Dept of CSE
Reva ITM, Kattaganahalli
Bangalore, India

²HOD & Prof.: Dept. of ISE
MSRIT
Bangalore, India

³Prof. Dept of CSE,
Mangalore University
Bangalore, India

WSN has the potentiality to join the physical world with the virtual world by creating a network of sensor nodes. Here, sensor nodes are usually battery-operated devices, and hence energy reduction of sensor nodes is a major design issue. To extend the network's lifetime, minimization of energy consumption should be used. In cluster-based routing, cluster heads shape a wireless stamina to the sink. Each cluster heads collects data starting from the sensors belonging to its cluster and relay it to the sink. Here, the cluster head position rotates, i.e., each node works as a cluster head for a restricted period of time. Energy saving in BFOA approaches can be done by cluster formation, cluster-head election, data collection at the cluster-head nodes to reduce data redundancy and thus save energy and also to improve energy efficiency of homogeneous WSN. It also defines Bacterial Foraging Optimization Algorithm (BFOA) as an algorithm for selecting best cluster head selection for WSN. The simulation results enhanced performance of BFA based on total energy dissipation and no. of .alive nodes of the network when compared with LEACH.

Keywords: BFOA, Chemotaxis, Swarming, Tumbling

1. Introduction

Wireless Sensor Networks (WSNs) consists of networked sensors that work together in hundreds of thousands of numbers for collaborative signal processing, monitoring, sensing and control tasks. WSNs offer extensive benefits and versatility to low-power and low-cost rapid deployment for many applications that can be automated without any human supervision. Many of the applications include disaster recovery, military surveillance, health administration, environmental & habitat monitoring, target tracking. WSNs can self organize and self-configure independently by inter-node communications possible through multi-hop wireless paths. While the communications of each node is possible via the transceiver unit, each sensor node also consists of a sensing unit, processing unit, and a power basis unit. Each of these units is included together with the help of ICs with integrated signal processing and micro-sensing components. These nodes that form the WSN can be sited far from the actual occurrence, and can still be used for data aggregation and collection from a remote location far away from the point of event-occurrence. This scheme uses bacteria foraging algorithm in wireless sensor network to improve the energy efficiency of each sensor nodes, whereas in previous schemes it used in control system. Nowadays Bacteria Foraging technique is gaining importance in the optimization problems. Because a) Philosophy says, Biology provides highly automated, robust and effective organism, b) Search strategy of bacteria is salutary (like common fish) in nature c) Bacteria can sense, make a decision and act so adopts social foraging (foraging in groups). To perform social foraging an animal needs

communication capabilities and it gains advantages that can exploit essentially the sensing capabilities of the group, so that the group can gang-up on larger prey, persons can obtain safety from predators while in a group, and in a certain sense the group can forage a kind of intelligence. BFA is based on the foraging performance of Escherichia Coli (E. coli) bacteria present in the person intestine.

2. Related Work

Efren Mezura-Montes and Betania Hernandez-Ocana [1] has explored another swarm-intelligence-based model: Bacterial Foraging Optimization Algorithm (BFOA), inspired in the behavior of bacteria E. coli in its search for food. Three behaviors were modeled by Passino in his original proposal (1) Chemotaxis, (2) reproduction and (3) elimination-dispersal.

Swagatam Das et al [2] has presented, Bacterial foraging optimization algorithm (BFOA) has been widely accepted as a global optimization algorithm of current interest for distributed optimization and control. BFOA is inspired by the social foraging behavior of Escherichia coli. BFOA has already drawn the attention of researchers because of its efficiency in solving real-world optimization problems arising in several application domains. The underlying biology behind the foraging strategy of E.coli is emulated in an extraordinary manner and used as a simple optimization algorithm.

R. Vijay [3] has presented an overview of the biology of bacterial foraging and the pseudo-code that models this process also explained. This paper presents a novel BFO to solve Economic Load Dispatch (ELD) problems. The results

are obtained for a test system with three and thirteen generating units. In this paper the performance of the BFO is compared with Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). The results clearly show that the proposed method gives better optimal solution as compared to the other methods.

Ahmed Y. Saber et al [4] has proposed a novel modified bacterial foraging technique (BFT) to solve economic load dispatch (ELD) problems. BFT is already used for optimization problems, and performance of basic BFT for small problems with moderate dimension and searching space is satisfactory. Search space and complexity grow exponentially in scalable ELD problems, and the basic BFT is not suitable to solve the high dimensional ELD problems, as cells move randomly in basic BFT, and swarming is not sufficiently achieved by cell-to-cell attraction and repelling effects for ELD.

Om Prakash Acharya et al [5] has demonstrated, the task of fault finding in antenna arrays was approached as an optimization problem and was solved using BFO technique. BFO was used to find the amplitude excitations from the pattern of the defected array which was then compared with the excitations of the original array to find the location and level of fault in the defected array. Partial as well as complete fault cases were considered and located successfully. Although a linear Chebyshev array was taken as the test antenna in this work, but the same methodology is applicable for any type of array.

Dong Hwa Kim [6] has proposed method using four test functions and the performance of the algorithm is studied with an emphasis on mutation, crossover, variation of step sizes, chemo tactic steps, and the lifetime of the bacteria. The proposed algorithm is then used to tune a PID controller of an automatic voltage regulator (AVR). Simulation results clearly illustrate that the proposed approach is very efficient and could easily be extended for other global optimization problems.

H. Nouri et al [7] has proposed, an attempt is made to tuning chemo tactic and swarming steps parameters meanwhile taking into consideration bacteria foraging optimization algorithm convergence speed and performance. The factorial designed experiment is suggested to create treatments of experiment. The adequacy of the proposed model is analyzed based on some commonly statistical criteria.

Rajasekhar Anguluri et al [8] presented, a new hybrid algorithm combining the features of BFOA and Particle Swarm Optimization (PSO) for tuning a Fractional order speed controller in a Permanent Magnet Synchronous Motor (PMSM) Drive. Computer simulations illustrate the effectiveness of the proposed approach compared to that of basic versions of PSO and BFO.

Hsuan-Ming Feng et al [9] has proposes a novel bacterial foraging swarm-based intelligent algorithm called the bacterial foraging particle swarm optimization (BFPSO) algorithm to design vector quantization (VQ)-based fuzzy-image compression systems. It improves compressed image quality when processing many image patterns. The BFPSO algorithm is an efficient evolutionary learning

algorithm that manages complex global optimal codebook generation problems.

Wael Korani et al [10] has illustrated, a new algorithm Bacteria Foraging oriented by PSO (BF-PSO). The new algorithm is proposed to combines both algorithms' advantages in order to get better optimization values. The proposed algorithm is applied to the problem of PID controller tuning and is compared with conveniently Bacterial Foraging algorithm and Particle swarm optimization.

Yudong ZHANG et al [11] have proposed BFO is a novel and powerful global search technique, and it can find the weights/biases of the neural network quickly and accurately. Experiments indicate that the proposed BFO-NN is superior to GA-NN with respect to convergence speed and forecast accuracy.

A. N. K. Nasir et al [12] has illustrated, a hybrid optimization algorithm referred to as Hybrid spiral dynamics bacterial foraging (HSDBF). The algorithm synergizes spiral adaptive simplified bacterial foraging algorithm (BFA) and spiral dynamics inspired optimization algorithm (SDA).

E. Ben George et al [13] has demonstrated an innovative approach to extract the textural features from the segmented magnetic resonance image to classify the tumors into benign, malignant or normal. Textural analysis methods such as Spatial Gray Level Dependency Matrix (SGLDM) and Surrounding Region Dependency Matrix (SRDM) are used to extract the fourteen Haralick features from the segmented image.

3. Evolution of Bacterial Foraging

During foraging of the real bacteria, locomotion is achieved by a set of tensile flagella. Flagella help an E.coli bacterium to tumble or swim, which are two basic operations performed by a bacterium at the time of foraging. When they rotate the flagella in the clockwise direction, each flagellum pulls on the cell. That results in the moving of flagella independently and finally the bacterium tumbles with lesser number of tumbling whereas in a harmful place it tumbles frequently to find a nutrient gradient. Moving the flagella in the counterclockwise direction helps the bacterium to swim at a very fast rate. In the above-mentioned algorithm the bacteria undergoes chemotaxis, where they like to move towards a nutrient gradient and avoid noxious environment. Generally the bacteria move for a longer distance in a friendly environment. Fig. 1 depicts how clockwise and counter clockwise movement of a bacterium take place in a nutrient solution.

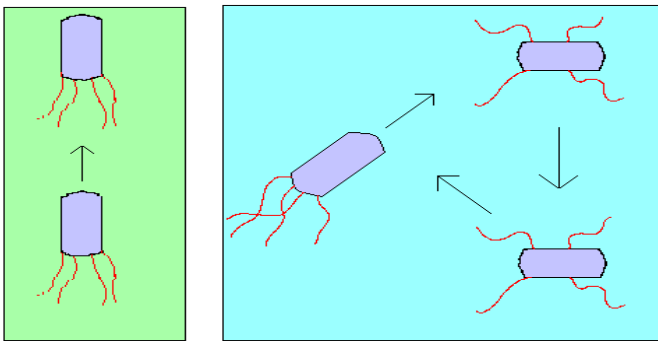


Fig.1 Swim and tumble of a bacterium

When they get food in sufficient, they are increased in length and in presence of suitable temperature they break in the middle to form an exact replica of itself. This phenomenon inspired Passino to introduce an event of reproduction in BFOA. Due to the occurrence of sudden environmental changes or attack, the chemotactic progress may be destroyed and a group of bacteria may move to some other places or some other may be introduced in the swarm of concern. This constitutes the event of elimination-dispersal in the real bacterial population, where all the bacteria in a region are killed or a group is dispersed into a new part of the environment.

4. Proposed System

In this scheme, BFOA (Bacteria Foraging Optimization Algorithm) is used to enhance the energy efficiency in wireless sensor networks. The proposed algorithm aims at balancing the energy consumption among the nodes in every cluster and reducing the energy dissipation of the cluster-heads. Like conventional LEACH protocol [14], this runs many rounds in the lifetime of the network. It includes four phases [2] during the working process: a cluster formation phase, a cluster-head adjustment phase, cooperative nodes selection phase and a steady phase. Algorithm output produced the best cell value where it can be a cluster head. At initial step cell number, problem sizes are given to find the initial population. Based on this population the high energy node is calculated. There are four basic steps[2], such as chemotaxis, swarming, reproduction, removal and dispersal.

$$P(j,k,l) = \{\theta_i(j,k,l) \mid i=1,2,K,S\} \quad (1)$$

Let j be the chemotactic step, k be the reproduction step, l be the elimination and dispersal event, s be the bacteria. Chemotactic is a process of finding nearby nodes. Let N_c be the distance end to end of the lifetime of the bacteria as deliberate by the number of chemotactic steps they take during their life.

Let $C(i) \lambda_0, i, \lambda_1, 2, K, S$, denote a basic chemotactic step size is to define the lengths of steps during runs. Tumbling is a process of moving to nearby nodes based on the

specific set of directions. To signify a tumble, a unit length arbitrary direction, say $\lambda(j)$, is generated; unit length will be used to define the path of movement after a tumble. In particular, we let,

$$\lambda_i(j_1,k,l) \lambda_i(j,k,l) C(i) \lambda(j) \quad (2)$$

So that $C(i)$ is the size of the step taken in the arbitrary direction specified by the tumble. Swimming is same as tumbling but it doesn't have a specific set of directions only there is a little displacement. If at $\theta_i(j+1,k,l)$ the cost $J(i, j+1,k,l)$ is better (lower) than at $\theta_i(j,k,l)$, then another step of size $C(i)$ in this same direction will be taken, and again, if that pace resulted in a position with a better cost value than at the previous step, another step is taken. This swim is continued as long as it continues to reduce the cost, but only up to a utmost number of steps, N_s . This represents that the cell will be apt to keep moving if it is headed in the path of increasingly favorable environments.

The proposed algorithm is discussed under following phases:

Cluster Formation Phase: Every node sends its position information to the base station. A certain number of nodes are elected to act as the auxiliary cluster-heads with a certain probability. According to the number of auxiliary cluster-heads, the network is divided into the same number of clusters evenly. If no node dies, the number of auxiliary cluster-heads is fixed, and also the number of members in every cluster is same means fixed. To construct the cluster network, the base station starts with the furthest node from it as the first auxiliary cluster head, which chooses the nearest fixed number of nodes as its member nodes. Then the base station chooses the furthest node from it as the second auxiliary cluster-head from the rest nodes, which chooses the nearest fixed number of nodes as its members. This process continues until all the auxiliary cluster heads and members are chosen out. The base station sends notifications to all the nodes in the network. According to this procedure, if no sensor node dies in the network, this cluster construction is stable. The network owns the same auxiliary cluster-heads and every auxiliary cluster-head owns the same members.

4.1 Cluster Adjustment Phase

The auxiliary cluster-heads are not the final cluster-heads, and BFOA algorithm is used for the adjustment. The algorithm is employed by every node. Auxiliary cluster heads then decide the final cluster-heads by BFOA after every member node sends its position and residual energy information to its auxiliary cluster-head. Each auxiliary cluster-head computes the position of the final cluster-head, and sends notifications to the final cluster-head and other members in the cluster. In this phase, the BFOA algorithm is used by the auxiliary cluster-head for the adjustment. We can use a fitness function which is based upon the relative distance of each node from the auxiliary cluster head and the residual energy of each node.

On the basis of this fitness function and at the end of all chemotaxis steps we can find suitable positions for each sensor node (by following BFO algorithm). At last the bacterium or all sensor nodes in the cluster send their relative position and residual energy to the auxiliary cluster head, then the auxiliary cluster head compute the final cluster head position. Now suppose before starting iteration, all the bacteria are present at their real position and after iteration all the bacterias are present at their suitable position. Then the most suitable position is mapped into one of the real positions of the nodes in the cluster. Auxiliary cluster head do this mapping according to:
 $D_{min} = \min \{ ||P_b - P_1||, ||P_b - P_2||, ||P_b - P_3||, \dots, ||P_b - P_n|| \}$
 Hear P_b = real position,
 P_n = suitable position

Then the real position of a certain node with d_{min} will be chosen to the position of final cluster head, which means that the nearest node from the P_b will act as the final cluster-head.

4.2 Cooperative node Selection Phase

After the cluster formation, each cluster head will select J cooperative sending and receiving nodes for wireless sensor network communication with each of its neighboring cluster head (j is fixed). Nodes with higher energy close to the cluster head will be elected as sending and receiving cooperative nodes for the cluster. The cluster head will broadcast a cooperative request message, which contains the ID of the cluster itself, the ID of the neighboring cluster head y , the ID of the transmitting and receiving cooperative nodes and the index of cooperative nodes in the cooperative node set of each cluster head to each cooperative node. The cooperative node on receiving the cooperative request message, stores the cluster head ID and sends back a cooperate-acknowledgement message to the cluster head. In this wireless sensor network cluster head itself act as a sender or receiver means cluster head also take part in cooperatively sending and receiving. Such technique is seen implemented on LEACH protocol, but in that it suffers from lot of draw backs, But in the proposed scheme these draw backs are can be removed by considering a good energy balancing fitness function.

4.3 Data Transmission Phase

During this phase, the data sensed by sensor nodes are transmitted to the cluster head and forwarded to the sink using wireless sensor network scheme according to the routing table.

4.4 Intra Cluster Transmission

In this phase, the non-cluster head nodes send their data frames to the cluster head during their allocated time slot.

4.5 Inter Cluster Transmission

After a cluster head receives data frames from its cluster members, it performs data aggregation and broadcasts the

data to J cooperative wireless sensor network sending nodes. When each cooperative sending node receives the data packet, they including cluster head encode the data using space time block code and transmit the data cooperatively. The advantage of using space time block code technique is that they provide diversity gain in both transmits and receive operation in wireless sensor network system. The diversity gain therefore provides reliable and energy efficient transmission. The receiving cooperative nodes use channel state information to decode the space time coded data. The cooperative node relays the decoded data to the neighboring cluster head node and forwards the data packet to the target cluster head.

5. Algorithm Design

5.1 Chemo taxis

This process in the control system is achieved through swimming and tumbling via Flagella. Each flagellum is a left-handed coil configured so that as the support of the flagellum (i.e., where it is connected to the cell) rotates counterclockwise, as viewed from the free end of the flagellum looking in the direction of the cell, it produces a force against the bacterium so it pushes the cell. On the other hand, if they rotate clockwise, each flagellum pulls on the cell, and the net result is that each flagellum operates relatively independently of others, and so the bacterium tumbles about. Therefore, an E. coli bacterium can go in two different ways; it can run (swim for a period of time) or it can tumble, and exchange flank by these two modes of operation in the entire lifetime. To represent a tumble, a unit distance end to end arbitrary direction, say(j), is generated; this will be used to define the way of movement after a tumble[2]. In particular

$$J_{cc}(\theta, P(j, k, l)) = \sum_{i=1}^S J_{cc}^i(\theta, \theta^i(j, k, l))$$

$$= \sum_{i=1}^S [-d_{attractant} \exp(-w_{attractant} \sum_{m=1}^p (\theta_m - \theta_m^i)^2)]$$

$$+ \sum_{i=1}^S [h_{repellent} \exp(-w_{repellent} \sum_{m=1}^p (\theta_m - \theta_m^i)^2)] \quad (3)$$

Where $J_{cc}(\theta, P(j, k, l))$ is the cost function value to be added to the real cost function to be minimized to present a time changeable cost function, S is the total number of bacteria, P is the numeral of parameters to be optimized which are present in each bacterium, and $d_{attract}$, $w_{attract}$, $h_{repellent}$, $w_{repellent}$ are different coefficients that are to be chosen properly.

5.2 Reproduction

The minimum healthy bacteria die and the other healthier bacteria each split into two bacteria, which are located in

the same location. This makes the inhabitants of bacteria constant.

5.3 Elimination and Dispersal

It is probable that in the local environment, the lives of a population of bacteria changes either step by step (e.g., via consumption of nutrients) or unexpectedly due to some other influence. Actions can occur such that all the bacteria in an area are killed or a group is isolated into a new part of the environment. They have the effect of perhaps destroying the chemotactic progress, but they also have the effect of support in chemotaxis, since spreading may place bacteria near good food sources. From a wide perspective, elimination and dispersal are parts of the population-level long-distance motile behavior.

6 System Design

To explain how chemo taxis motions are generated, we must simply explain how the bacteria decides how long to run, since from the above discussion we know what happens during a tumble or run. First, note with the intention of if an E. coli is in some substance that is neutral in the sense that it does not have food or noxious substances, and if it is in this medium for a long time (e.g., more than 1 min), then the flagella will simultaneously alternate between moving clockwise and counterclockwise so that the bacterium will alternately tumble and run. This alternation between the two modes will move the bacterium, but in arbitrary directions, and this enables it to “search” for nutrients. For instance, in the isotropic homogeneous environment described above, the bacterium alternately tumbles and runs with the mean tumble and run lengths given above and at the speed that was given. If the bacteria are placed in a homogeneous concentration of serine (i.e., one with a nutrient but no gradients), then a variety of changes occurs in the characteristics of their motile behavior. For instance, mean run length and mean speed increase and mean tumble time decreases. They do still produce, however, a basic type of searching behavior; even though the bacterium has some food, it persistently searches for more. Suppose that we call this its baseline behavior. As an example of tumbles and runs in the isotropic homogeneous medium described above, in one trial motility experiment lasting 29.5 s there were 26 runs, the maximum run length was 3.6 s, and the mean speed was about 21 Next, suppose that the bacterium happens to meet a nutrient gradient (e.g., serine). The change in the concentration of the food triggers a reaction such that the bacterium will spend large time swimming and less time tumbling.

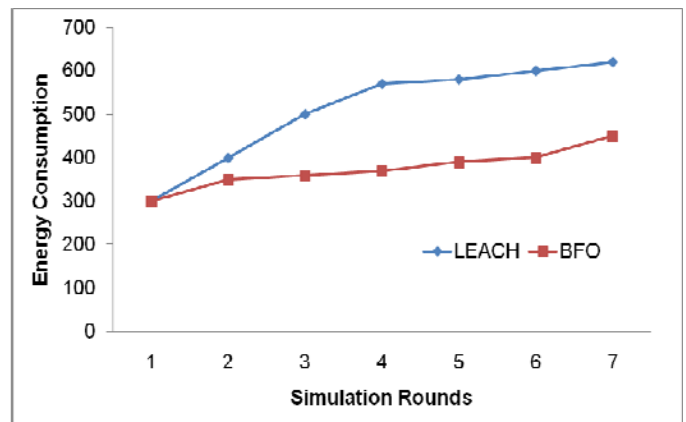


Fig. 4.1 Simulation Time VS Energy Consumption

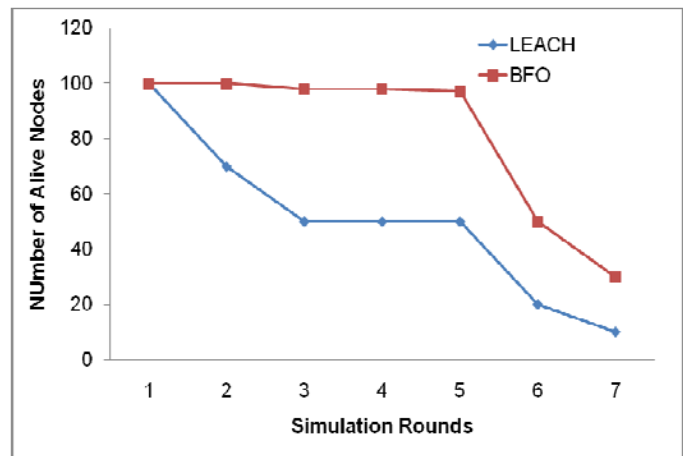


Fig. 4.2 Simulation Time VS No. of. Alive Nodes

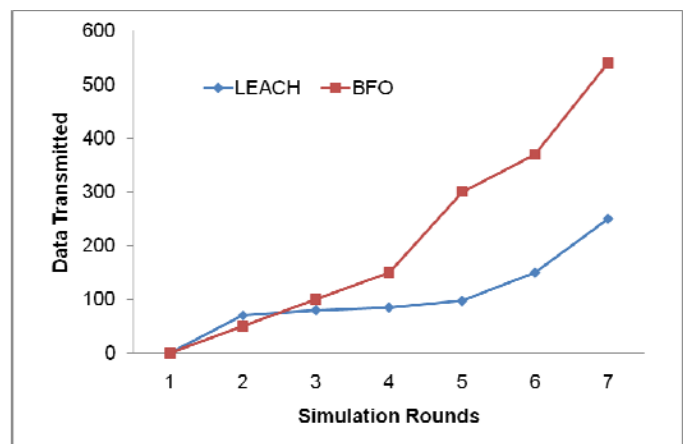


Fig. 4.3 Simulation Time VS Data Transmission

7 Conclusion

The first contribution of the scheme is related to use of bacteria foraging algorithm firstly for WSNS for enhancing network lifetime of sensor nodes. To validate the algorithm, simulations had been carried out using Matlab. simulation results showed better performance of BFA as compared to other clustering protocols like leach, in terms of

performance metrics like number of alive nodes and total energy dissipation in the system. BFA provides better lifetime for nodes compared to LEACH. It is also seen that BFA is able to provide 100% live nodes for maximum duration. leach provides a considerably higher lifetime compared to k-means clustering.

References

- [1] Mezura-Monte, E & Hernandez-Ocaña. B.(2008). Bacterial Foraging for Engineering Design Problems: Preliminary Results, "Memorias del 4o Congreso Nacional de Computacion Evolutiva (COMCEV'2008), CIMAT, Gto. Mexico.
- [2] Das, S., Biswas,A., Dasgupta,S & Abraham,A. (2009).Bacterial Foraging Optimization algorithm: Theoretical Foundations, Analysis, and Applications, Computational Intelligence Volume 203, 2009, pp 23-55
- [3] Vijay, R. (2012). Intelligent Bacterial Foraging Optimization Technique to Economic Load Dispatch Problem, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-2
- [4] Saber, A.Y., Venayagamoorthy, G.K. (2008). Economic Load Dispatch using Bacterial Foraging Technique with Particle Swarm Optimization Biased Evolution, IEEE Swarm Intelligence Symposium St. Louis MO USA, September 21-23
- [5] Acharya, O.P., Patnaik, A & Choudhary, B.(2011). Fault Finding in Antenna Arrays using Bacteria Foraging Optimization Technique, Communications (NCC), 2011 National Conference, Page(s): 1 - 5
- [6] Kim, D.H., Abraham, A & Cho, J.H.(2007). A hybrid genetic algorithm and bacterial foraging approach for global optimization, Information Science, 3918-3937
- [7] Nouri, H., Tang S. H., Ariffin,M.K.A.,Tuah,B.T.H & Samin,R. (2013). Development and Tuning of Bacteria Foraging Optimization Algorithm on Cell Formation in Cellular Manufacturing System, Journal of Automation and Control Engineering, Vol. 1, No. 1
- [8] Anguluri, R., Abraham,A & Snaset, V. (2011). A Hybrid Bacterial Foraging - PSO Algorithm Based Tuning of Optimal FOPI Speed Controller, Acta Montanistica Slovaca,
- [9] Feng,H.M., Horg, J.H., Jou, S.M. (2012). Bacterial Foraging Particle Swarm Optimization Algorithm Based Fuzzy-VQ Compression Systems, Journal of Information Hiding and Multimedia Signal ProcessingISSN 2073-4212 Ubiquitous International Volume 3, Number 3
- [10] Korani, W. (2008). Bacterial Foraging Oriented by Particle Swarm Optimization Strategy for PID Tuning, GECCOACM978-1-60558-131-6/08/07.
- [11] ZHANG, Y., WU,L & Wang, S. (2010). Bacterial Foraging Optimization Based Neural Network for Short-term Load Forecasting, Journal of Computational Information Systems 6:7(2010) 2099-2105
- [12] Nasir, A.N.K., Tokhi,M.O., Ghani,N.M & Ahmad,M.A. (2012).A Novel Hybrid Spiral-Dynamics Bacterial-Foraging Algorithm for Global Optimization with Application to Control Design, Computational Intelligence (UKCI), 2012 12th UK Workshop, Page(s): 1 - 7
- [13] George, E.B., Karnan, M. (2012). Feature Extraction and Classification of Brain Tumor using Bacteria Foraging Optimization Algorithm and Back Propagation Neural Networks, European Journal of Scientific Research ISSN 1450-216X Vol. 88 No 3 October, 2012, pp.327-333
- [14] Wendi B. Heinzelman, An Application-Specific Protocol Architecture for Wireless Microsensor Networks, IEEE Transactions on wireless communications, Vol. 1, No. 4, October 2002

An online collaborative framework for orthography system development

Sook-Kuan Chin¹, Alvin W. Yeo², Nadianatra Musa³

¹Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak,
Sarawak, Malaysia.

²Institute of Social Informatics and Technological Innovations,
Universiti Malaysia Sarawak,
Sarawak, Malaysia

³Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak,
Sarawak, Malaysia.

Abstract

This paper describes a collaborative project for orthography system development involving the Daro community and linguists. We proposed a new model which integrated with Information and Communication Technologies (ICTs) to simplify the orthography system development. Besides, collaborative work also introduced for expanding the relationship between the linguists and community in orthography system development. Currently, our project focuses on Daro community, but we anticipate that this project will extend to other indigenous communities in Sarawak as well.

This study investigates how community able to contribute their efforts in orthography system development. Hence, Daro community required to use the online platform to upload data. Their response towards the online platform was observed.

Keywords: *Orthography system development, collaborative work*

1. Introduction

Many endangered languages in Sarawak, Malaysia are not written yet. Thus, the endangered languages need to be written down to provide a standard spelling guide for the community. Orthography system is the initial step to write down the endangered languages. An orthography system is a writing system with respect to a set of symbols and writing rules [1]. However, orthography development largely difficult as it requires many costs and time from linguists. Thus, a new method was proposed to solve the current situation. The new method is applied together with Information and Communication Technologies (ICTs) to provide an online workspace for linguists and community

to conduct the orthography system development collaboratively. One of the most developed cases of the use of the Internet to preserve and strengthen an indigenous language has occurred in Hawai'i [2]. Since ICTs is an important tool to preserve the indigenous languages, thus we applied it in orthography system development to improve on the manual method.

In this study, Melanau Matu-Daro language was chosen as the development of orthography system for this language does not exist yet. Melanau Matu-Daro is a language spoken in Matu River from North Channel of Rejang River to the sea, Daro and Matu areas [3]. This study is located at Daro, Sarawak.

Melanau Matu-Daro is mostly used within the community, mainly at home or some community activities. They are able to converse with their family members or relatives but not familiar in writing.

Furthermore, so far this language does not have standard orthography. Words are pronounced and spelled based on individual's background and preferences. Based on the criteria of this language, this language had been identified as a suitable case study language to be used in this project.

Therefore, the purpose of this study is to determine that the community can contribute towards orthography system development.

This paper is organised as follows. Section 2 introduces the project framework. In Section 3 we explained the process to implement the online platform. After this, Section 4

discusses the experimental results and analysis. Finally the paper concludes in Section 5.

2. The Project Framework

Currently, linguists need to travel to the rural places to conduct the data collection and conduct the orthography development. Hence, the manual method does not cost effective because much time and costs were consumed to complete the orthography development. The linguists may spend a life time to produce the linguistic materials (dictionary, publication and etc) and thus some technology is necessary to help produce some useful materials [4].

To the authors' knowledge, so far there is no existing model which allows linguists to work collaboratively with the community in online platform.

Hence, we are working toward a model which will support collaborative work aimed at simplifying the orthography development. The framework for this project is applied together with Computer Supported Cooperative Work (CSCW). CSCW is a groupware that brings the people work together with the helping of computer technology (Rodden, 1991). In this context, it defines that we use computer technology to develop orthography system even the participants has different background and they are located at different places.

The framework proposes that linguists work collaboratively with the community in orthography system development using ICTs. ICTs can preserve the knowledge for the coming generation and develop a platform to access the indigenous knowledge [5]. As language represents the culture of the community, therefore it provides the important knowledge for the community. It is important to preserve the knowledge in order to pass to the next generation.

The goal of developing the framework aimed at bringing the speakers to share their knowledge about their languages. Besides, this project provided the environment for expanding a relationship between community and linguists. Linguists and community work together, each contributing their expertise to collaborate in orthography system development.

We believed that ICTs can help to increase the performance of orthography system development. Instead of travelling to the rural places, the linguists can conduct the work using online method with the community. Online communication will facilitate the exchange of information

among each others. This provides the opportunity for the local community to share the information with others.

3. Implementation of CODES

CODES had been implemented to provide the online working space for linguists and community to develop orthography system. The online platform implemented with PHP in Windows environment using Macromedia Dreamweaver. PHP is a widely used general purpose scripting language that is especially suited for Web development and can be embedded into HTML [7].

The online platform (CODES) consists of few major components, which include of the data collection module, IPA and spelling identification module and community evaluation module.

Data collection module allows community to upload their data once the data collected from other community members. Next, linguists use the IPA and spelling identification module to propose suitable spelling for the community. Lastly, community evaluation module allows community to express their opinion towards linguists' proposed spelling.

After the online platform being built, then it tested with different stakeholders for validating the model. The stakeholders who involved in this project are linguists and community. They used different modules in CODES to perform their action respectively.

4. Method

In this section, the experiment had been carried out to determine that community can contribute towards orthography system development. To realise an online working environment for linguists and community, uploading data module in CODES was used in the study. The purpose of the study was to investigate the possibility of community to contribute in orthography system development by collecting the data for linguists.

A. Participants

The participants were two females and two males who spanned the ages of 20-80. All participants were briefed the study's purpose, method and procedure. Table 1 show the profile and information of participants who involved in the study. F1-F2 represents the younger generation who familiar with computer while P1-P2 represents the old folks who provides the pronunciation of the words.

Table 1: Participants' Information

ID	Gender	Age	Occupation	Highest Educational Background	Purpose of experiment
F1	Female	27	-	Secondary school	Collect and upload data
F2	Female	21	-	Secondary school	Collect and upload data
P1	Male	75	Retired Teacher	Primary school	Collect and upload data
P2	Female	64	Housewife	Primary school	Collect and upload data

B. Experimental Setup and Equipment

In this study, the 10-item Swadesh list is used. The ten words were written in Bahasa Malaysia (BM). The words written in BM is because the participants are more familiar with the Malay words. Swadesh list is a classic compilation of basic concepts for the purposes of historical-comparative linguistics [8].

Table 1 below shows the 10-item Swadesh list which used in this experiment.

Table.2: Word List

Word List	
1	I
2	wind
3	fire
4	true
5	flower
6	bite
7	fish
8	road
9	small
10	because

The equipments in this study are includes with sound recorder, timer and computer. Additionally, Internet also required in the study. Sound recorder was used to record the voices of the participants while timer used to calculate the time usage during the experimentation process. On the other hand, computer and Internet connection are necessary for the participants to access the online platform.

C. Experimental Procedure

The experiment involved tests with community to collect and upload data to online platform. Before beginning the experiment, we informed the participants about the

purpose of this study and experiment process.

The subjects were asked to translate the Malay words into Melanau Matu-Daro and asked the older to pronounce the word accordingly. This is mainly due to old folks of the village provides more accurate pronunciation of words. The accurate pronunciation of the words helps the linguists to identify the correct IPA and spelling. Each of the pronunciation of the words was recorded with the sound recorder. The recorded audio files were saved in computer. After this, the uploading module in CODES was used to upload the audio files. The uploaded audio files will be processed by the linguists later on. Participants had to fill out a questionnaire to observe their capability to use the online platform.

D. Results

Table 2 summarised that the time taken to record and upload an item less than 5 minutes. The results showed that participants help linguists to save some time in travelling because they assist in data collection process.

Table 2: Time used by the facilitator

Task	Total time used(minutes)
To record an item	< 5 minutes
To upload an item	< 5 minutes

E. Questionnaires

Table 3 shows the results of questionnaires. The results obtained reflected that participants felt satisfied with the online platform for data uploading process. The rating scales are from 1-5 which shows the agreement of the statements. 1 indicates that subjects very agree with the statement while 5 indicate that subject did not agree with the statement.

Table 3: Questionnaire results

1. Overall, I feel easy to use the online platform	2	1	1
2. Information in online platform is easy to understand.	2	1	1
3. I feel easy upload file to online platform	2	1	1
4. After explanation given, I can easily upload the file.	3	1	1
5. I do not face any problem whenever upload file to online platform.	3	1	1
6. I believe this online platform able to preserve the indigenous languages.	1	1	1
7. Do you will use this online platform in future time?	Ya	Ya	Ya
8. Will you recommend this online platform to others?	Ya	Ya	Ya

F. Discussion

In general, data collection process had been carried out smoothly with the active participation of community members. The purpose of this study was to find out how community members collect and upload data by themselves. Community used the provided word list to collect and upload words using CODES. Thus, linguists able to access the data and do not need to travel to the rural places.

Besides, the results showed that the time taken is very short for them to upload data. Therefore, linguists save the time and energy to collect data while they depend on the helping from the community. In addition, they also can concentrate to create IPA and spelling for the community.

In summary, the results collected from questionnaire showed that community satisfied with the functionality of CODES. Overall, the online platform help the community to keep their language resources alive but linguists also get the benefits from it because they able to access the language resources without travelling to the field.

5. Conclusion

In this paper, online collaborative framework for simplifying the orthography system development is presented. The proposed model is intended to improve the performance of orthography system development. The experiment was carried out to determine that community collect data themselves help to reduce the burdens of linguists. Hence, both the linguists and community can gain a better experience in orthography system development.

Acknowledgments

The authors would like to express the gratitude to Exploratory Research Grant Scheme from Ministry of Higher Education. And also thanks to Universiti Malaysia Sarawak (UNIMAS) for providing computer facilities as well.

Besides, the participation of Daro community members during the pilot study is greatly appreciated.

References

1. Seifart, F., *Orthography development*. Essentials of language documentation, 2006: p. 275-299.
2. Warschauer, M. and I. De Florio-Hansen, *Multilingualism, identity, and the Internet*. Multiple identity and multilingualism, 2003: p. 155-179.
3. Lewis, M.P., *Ethnologue: Languages of the world*. Vol. 9. 2009: SIL international Dallas, TX.

4. Czaykowska-Higgins, E., *Research Models, Community Engagement, and Linguistic Fieldwork: Reflections on Working within Canadian Indigenous Communities*. Language documentation and conversation Vol 3, No1 2009.
5. Adam, L., T. James, and A. Wanjira, *Frequently asked questions about multi-stakeholder partnerships in ICTs for development*. A guide for national ICT policy animators. Published by the Association for Progressive Communications (APC), 2007.

Sook-Kuan Chin received B.E. degree in Software Engineering at Universiti Malaysia Sarawak (UNIMAS), Malaysia in 2009. Now she is pursuing her Master degree of Computer Science at UNIMAS. Her research interests cover the indigenous communities and language preservation method.

Alvin W. Yeo received the PHD degree in Computer Science at University of Waikato, New Zealand in 2002. He is now a director of Institute of Social Informatics and Technological Innovations (ISITI), UNIMAS. He has expertise in the area of Information and Communications Technology for Rural Development (ICT4RD).

Nadianatra Musa received the PHD degree in Computer Science at University of Tasmania, Australia in 2012. She is now a lecturer at Faculty of Computer Science and Information Technology, UNIMAS. She has expertise in IS/IT security governance, IS/IT Security Internal Controls.

Optimal Transmission Power of Target Tracking with Quantized Measurement in WSN

Osama M. El-Ghandour¹ and Amr Lotfy Elewa M.²

¹ Helwan University, Cairo, Egypt

² Helwan University, Electronic Research Institute, Egypt

Abstract

One of the important applications of the quality of monitoring (QoM) on target tracking in wireless sensor networks is reducing the overall power consumption in the monitoring / tracking procedures. We present, an optimal (sensor) transmission power problem is analytically formulated and its optimal solution is found such that a given constraint on the QoM is satisfied. Next, an optimum quantization system for the noise-corrupted sensor observations (measurements) is presented. In this scheme, sensor observations are first quantized into binary levels, and then transmitted to a fusion center where a final decision is made. The significant impact of optimizing the sensor's transmission power and quantizing its observations, is to provide for high QoM while reducing the overall power consumption in the monitoring / tracking procedures.

Numerical validation results show that, our suggested methods decreases energy consumptions in the sensor/fusion communication phases by the constraint binary message transmissions. This is well motivated by the bandwidth limitation of the communication links, and by the limited power budget of local sensors. On the other hand, the energy consumed in target tracking is minimized to an analytically optimal level while the target QoM level is satisfied all the time.

Key words: Wireless Sensor Networks (WSN); optimal transmission power; optimal message quantization; Power adaptation scheme; Quality of Monitoring (QoM).

1- Introduction

This paper presents first: an optimal sensing power that can guarantee, theoretically, error-free communications in WSNs. In traditional transmission scenarios, the system operating points lay in the feasible signal-to-noise ratio (SNR) regions. The objective of the scheme being presented is to improve the global probability of bit-error by compensating for the effect of fading along the communication channels through updating the effective sensor SNR required to optimize the detection performance [1], second; due to bandwidth and power

limitations, each sensor node quantizes its observations into b_i -bits message, and transmits its locally processed data to the fusion center. Then, the fusion node estimates the state vector of the object based on the quantized observations.

A significantly important aspect of the sensor's power optimization scheme is that, it helps in reducing the number of Participating monitor nodes in the target tracking problem (QoM) [1] and also, in node selection procedures, aiming to select the most informative sensors in order to minimize the energy consumptions of monitoring and tracking.

2- Related work

Energy efficiency is another critical design factor in WSNs, because the sensor nodes are usually of low cost and are designed with strict restrictions on their power Consumptions Previous research works on WSNs range from general theoretic analysis, to proposing optimization solutions for the detection process [2], [6]. However, these works mostly neglect the effect of fading over the communication channels, which are an important issue in real environment and, ignoring it, may cause significant degradation of the performance of the detection process. For the purpose of energy conservation, it was shown in [7] that, when the network is subjected to a joint power constraint, having identical sensor nodes (i.e. all nodes using the same transmission scheme), is asymptotically optimal for binary decentralized detection. Efficient node power allocation to achieve a given performance has been considered by [8], [9]–[11]. In [11], the optimal power assignment problem was addressed with amplify-and forward processing at local sensor nodes. It was shown that, such an analog forwarding scheme is optimal in the single sensor case by Shannon's separation principle. It was also shown that, optimal power scheduling improves the mean squared error performance by a large margin compared to that achieved by uniform power allocation scheme. The minimum energy, decentralized estimation with correlated data was addressed in [10]. They exploited knowledge of the noise covariance matrix to optimal quantization levels at sensor nodes the

the power, while meeting a given target mean-squared error.

3- Saving Power Consumption in the Target Tracking

3.1- Power Assignment Algorithm

Assume that the received signal strength at the fusion node [1] is given by,

$$U = |R|\sqrt{a} + n \quad (1)$$

Where, \sqrt{a} denotes the transmitted power, R is the path gain (fading amplitude) between the sensor and the fusion node with n as an additive white Gaussian noise having standard deviation σ . The SNR, at the fusion node is therefore,

$$SNR = \gamma = \frac{|R|^2 a}{\sigma^2} \quad (2)$$

3.1.1- Optimal sensing power.

In the following, we derive an optimal sensing power that minimizes the power consumed by the sensor subjected to constraints on the performance metrics (α, β) (i.e., $Q_0M(\alpha, \beta)$). This, in effect, is a constrained optimization problem, can be formulated as follows,

$$\begin{cases} \text{Min } \sqrt{a}, & \text{such that} \\ Q_0M(\alpha, \beta) \leq Q \left(\left| \frac{\log(\eta)\sigma}{R\sqrt{a}} - \frac{R\sqrt{a}}{2\sigma} \right| \right), \\ \sqrt{a} \geq 0, \end{cases} \quad (3)$$

The inequality in equation (3) above can be rewritten as follows,

$$\varphi \leq \left| \frac{\log(\eta)\sigma}{R\sqrt{a}} - \frac{R\sqrt{a}}{2\sigma} \right|, \quad (4)$$

Where we defined, $\varphi = Q^{-1}(Q_0M(\alpha, \beta))$, hence, the optimization problem (3) can be rewritten as follows,

$$\begin{cases} \text{Min. } \sqrt{a}, & \text{such that} \\ \varphi \leq \left(\left| \frac{\log(\eta)\sigma}{R\sqrt{a}} - \frac{R\sqrt{a}}{2\sigma} \right| \right), \\ \sqrt{a} \geq 0, \end{cases} \quad (5)$$

The optimization problem (5) can thus be reformulated using the Lagrange optimization scheme [12]-[13] as follows, then,

Assume the following objective function, F ,

$$F = \sqrt{a} + \Gamma \left(\varphi \leq \left| \frac{\log(\eta)\sigma}{R\sqrt{a}} - \frac{R\sqrt{a}}{2\sigma} \right| \right) \quad (6)$$

Where: Γ is the Lagrange multiplier, $\sqrt{a} = \frac{\sqrt{a_0}}{d}$. The optimal solution for the problem (6) is given by,

$$\sqrt{a_{opt}} = \frac{d \sqrt{(\sigma \varphi)^2 + 2 \sigma \eta - d \sigma \varphi}}{R} \quad (7)$$

Where $\eta = \log(\eta) \sigma$, (Prove is in appendix),

Equation (7) gives the minimum (sensor) transmission power necessary to balance the effects of channel fading and noise. Substituting equation (7) into (2), gives the target SNR,

$$\sqrt{SNR_{trg}} = d \sqrt{\varphi^2 + 2 \eta / \sigma} - d \varphi \quad (8)$$

$$\text{Where, } SNR_{trg} = \frac{|R|^2 a_{opt}}{\sigma^2}$$

At this point, we can set a sensor selection strategy based on the following procedures:

- A target SNR of the link between the sensor and the fusion center is computed using equation (8), based on the sensor location d ,
- Based on the received SNR at each sensor. It turns itself into active/inactive (participating/nonparticipating) in the target detection process,
- Certainly, such a self sensor activation /deactivation procedure would leads to a significant reduction of the sensor energy along its life time.

3.1.2- Numerical Results

In this section, the performance of the proposed coverage and SNR assignment is validated through numerical examples.

As shown in Figure (1), the sensor's coverage d increases with the SNR according to Equation (8). For instance, the sensing range, d , is about 5m if $\alpha = 2\%$, $\beta' = 93\%$, $SNR = 7.5$ dB. However, at $SNR = 9$ dB, the sensing range covers up to 10m. Figure (1) depicts the sensing ranges for different detection metrics (α, β').

In figure (2), we show an illustrative example on the optimum transmission power for the sensor-to-fusion communication link. In this example, the fading coefficient R , is set to unity (constant). This is done in order to highlight the effects of different communication metric values (α, β') on the power assignment process. As expected, the higher the values of (α, β'), the higher is the transmission power necessary to compensate for the effects of the path losses over the communication range (i.e., twice the sensing range) as given by Equation (7). In Figure (3), we show the effect of channel fading on the optimum power assignment. As expected, in order to maintain a given target values of (α, β'). higher transmission power assignment becomes necessary.

order to compensate for the effects both the fading and the path losses.

Figure (4), presents a practical implementation of the proposed (optimal) power assignment strategy. Assume that each sensor knows it's (discrete) relative location with respect to the fusion node. Assume further that, each sensor sends a pilot signal to the fusion node. Upon receiving the pilot signal from the sensor and, based on the measured channel characteristics, the fusion center performs an estimation of the optimum transmission power necessary to achieve the target (α, β') values and sends it as an update to the sensor. This way, our power assignment strategy would guarantee that the network operational point will always lies in the optimal SNR region.

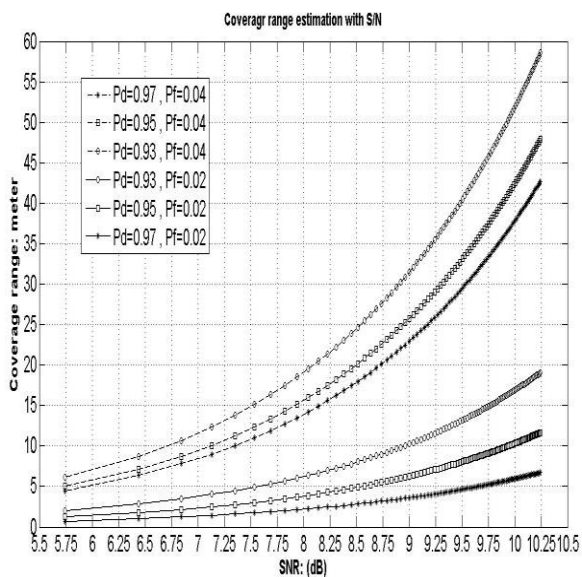


Figure (1): Coverage range estimation

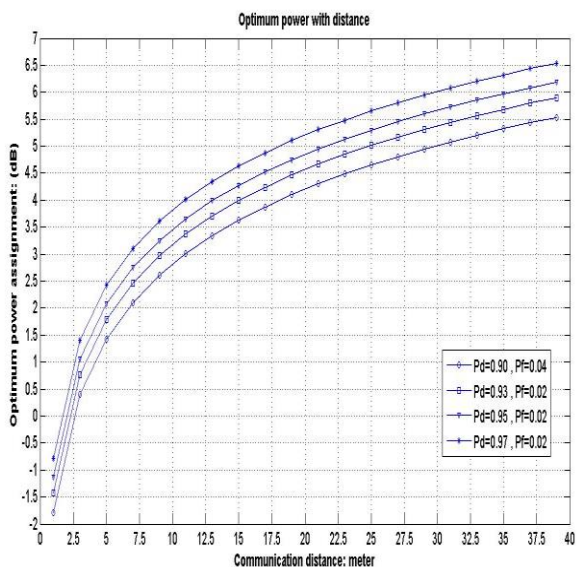


Figure (2): Optimum power assignment

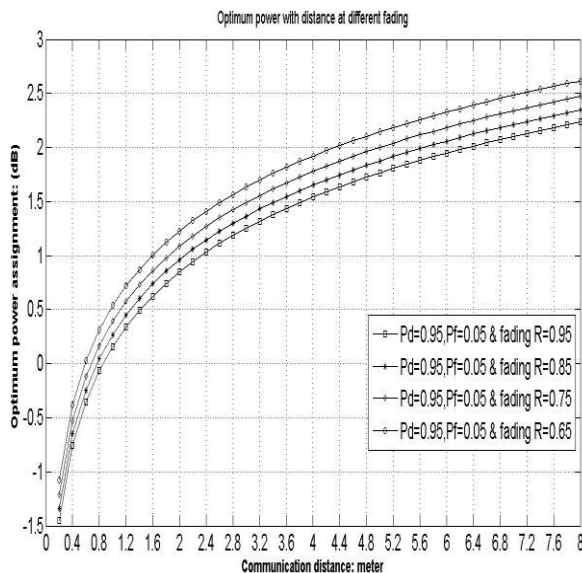


Figure (3): Effect of fading on the power assignment adding model

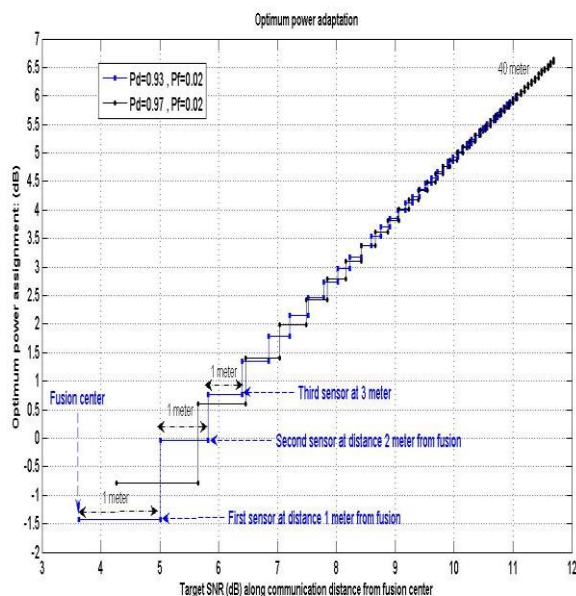


Figure (4): Target SNR(dB) along communication distance from fusion center

3.2- Optimal Message Quantization

Let the quantized message from the i -th sensor to fusion center [1] at time k be modeled as,

$$y_i(k) + q_i \tag{9}$$

Where: q_i is a zero mean quantization error with variance less than $\frac{W^2}{(2^{b_i} - 1)^2}$ [14], and $(\frac{W}{2}, \frac{-W}{2})$ is the available signal amplitude range common to all sensors, b_i is the number of bits, to be determined later, and $L_i = 2^{b_i}$, is the i -th quantization points, these points are uniformly spaced and it follows,

$$\Delta = \frac{2W}{b_i - 1}$$

The quantization model in equation (9) and the uniform quantization error assumption are widely used in the literature due to their analytical tractability.

Assuming that, the channel noise is v_i , quantization noise q_i , are mutually independent. Therefore, the signal of i -th sensor can be express as,

$$u_i(k) = y_i(k) + q_i + v_i \quad (11)$$

Let the noise $n_i = v_i + q_i$ is comprised of uncorrelated components and having zero means with variance,

$$\sigma_{n_i}^2 = \sigma_{v_i}^2 + \sigma_{q_i}^2 \quad (12)$$

The covariance of quantization noise is,

$$\sigma_{q_i}^2 \leq \frac{W^2}{(2^{b_i} - 1)^2} \quad (13)$$

It is easy to see that, the accuracy of the quantized messages is better if the variance of the quantization noise is small which is equivalent to using larger number of bits. That is, we can make its upper bound small which, in turns, means more bandwidth is need. However, in WSNs, both the sensor power and the transmission bandwidth are limited. Hence, it is important to find the optimal quantization bits necessary to achieve a given performance measure such that, a constraint on the sensor's energy/power are satisfied.

3.2.1- Bit Assignment

In this section, we consider the quantization bit assignment problem, assuming that the channel between the i -th sensor and the fusion node experiences a path loss proportional to d_i^m , where d_i is the transmission distance between the i -th sensor and the fusion node. The energy consumed in the i -th sensor is,

$$E_i = \omega_i (2^{b_i} - 1) \quad (14)$$

Where: ω_i is the energy density, in which $\omega_i = \rho' d_i^m \ln(\frac{2}{P_b})$, ρ' depends on the actual noise distribution [23], and P_b is the target bit error rate, assumed common to all sensor links.

At this point, our goal is to minimize the mean square transmission power while meeting a given total power consumption. This goal can be represented by the following optimization problem,

$$\begin{cases} \min \sum_{i=1}^N \omega_i^2 B_i^2 \\ \text{s.t } \sum_{i=1}^N (\sigma_{v_i}^2 + \sigma_{q_i}^2) \leq D \\ D \geq 0 \end{cases} \quad (15)$$

Where: $D > 0$ is a given targeted upper bound on the noise variance, where

$$B_i^2 = (2^{b_i} - 1)^2.$$

3.2.2- The Optimal Solution

In order to facilitate the analysis, we relax the integer b_i to be a real positive number. As we did in previous section, the problem in equation (15) can be reformulated as a Lagrangian convex optimization,

$$F(\Gamma, B_i) = \sum_{i=1}^N \omega_i^2 B_i^2 + \Gamma^* \{ \sum_{i=1}^N (\sigma_{v_i}^2 + \frac{W^2}{B_i^2}) - D \} \quad (16)$$

Letting $\partial F / \partial b_i = 0$ for all i ,

$$2\omega_i^2 B_{i_{opt}} - \Gamma^* \frac{2 B_{i_{opt}} W^2}{B_{i_{opt}}^3} = 0 \quad (17)$$

$$2\omega_i^2 B_{i_{opt}} - \Gamma^* \frac{2W^2}{B_{i_{opt}}^3} = 0 \quad (18)$$

And at the optimum solution, we should have,

$$D - \sum_{i=1}^N \left(\sigma_{v_i}^2 + \frac{W^2}{B_{i_{opt}}^2} \right) = 0 \quad (19)$$

Combining equations (18) and (19) yields,

$$b_{i_{opt}} = \log_2 \left[1 + \frac{W}{\omega_i} \sqrt{\frac{\Gamma^*}{2}} \right] \quad (20)$$

$$\text{Where: } \Gamma^* = \frac{2(W \sum_{i=1}^N \omega_i)^2}{(D - \sum_{i=1}^N \sigma_{v_i}^2)^2}$$

Once the optimal, real-valued $b_{i_{opt}}$ is computed, the associated bit loads can be obtained through simple upper integer rounding. Recall from equation (20) that the energy consumption of each sensor is proportional to the path loss d_i^m . Hence, larger energy consumptions correspond to sensors deployed far away from its fusion node.

3.2.3- Effect of Channel Fading on Quantization Bit Assignment

The relationship between the original signal of i -th sensor and the data received by the fusion node with fading is depicted in [1]. Therefore, equation (11) becomes,

$$u_i(k) = R y_i(k) + q_i + v_i \quad (21)$$

and from equation (14), the energy consumed in the i -th sensor under fading is,

$$E_i = R \omega_i (2^{b_i} - 1) \quad (22)$$

Where: R is the fading gain, ω_i is the energy density, $\omega_i = \rho' d_i^m \ln\left(\frac{2}{P_b}\right)$.

3.2.4- Numerical Results

Recall from equation (11), that the energy consumption of each sensor is proportional to the path loss. Hence, large value of the energy consumptions correspond to sensors deployed far away from the fusion node. In light of this point, the optimal quantization bit assignment is intuitively attractive. Figure (5) illustrates the (optimally) assigned bits versus the path loss of the channel in terms of the coverage distance. As can be seen, the optimal number of bits is proportional to the expected path loss. This is intuitively reasonable since sensors with bad link conditions, should be allocated with more bits in order to improve the received message accuracy at the fusion center. Clearly, the same talking applies well for the channel noise. This is illustrated in Figure (6), in terms of the noise variance. Finally, Figure (7) shows the optimum bit assignments taking into account the bit error caused by the channel. To transmit, the binary bits, we must insure that a given probability of bit error is achieved at the fusion node.

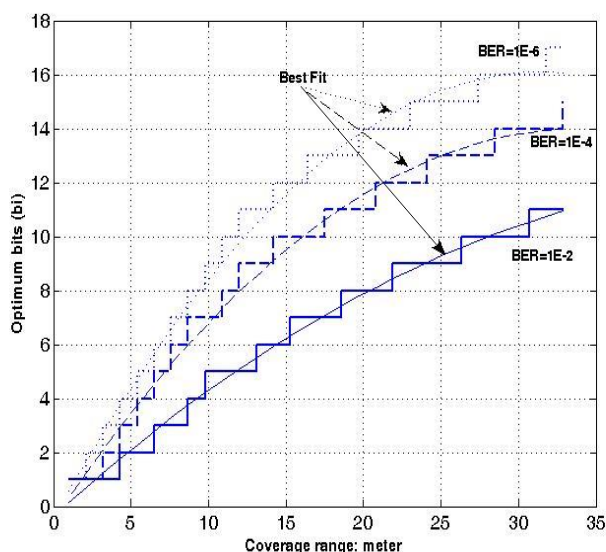


Figure (5): Optimum bit assignment with coverage range

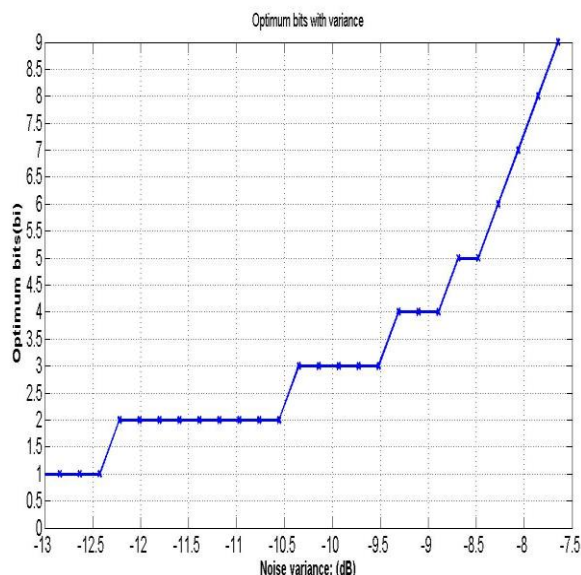


Figure (6): Effect of noise on optimum bit assignment

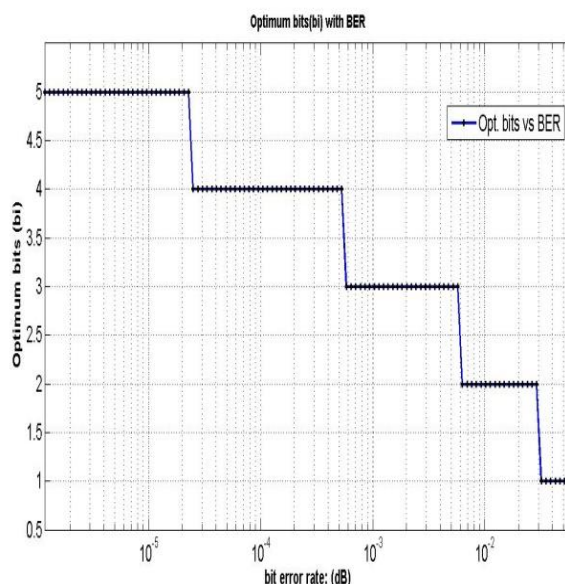


Figure (7): Effect of BER on optimum bit assignment

4- CONCLUSION

These days, energy saving in the monitoring (QoM) of mobile target tracking is considered as one of the important applications of wireless sensor networks.

We considered the optimal (sensor) transmission power problem such that a given constraint on the QoM is satisfied. The significant impact of optimizing the sensor transmission power is, to provide for high QoM while reducing the overall power consumption in the WSN. The scheme is designed with number of objectives: first, the moving target should be covered with predefined QoM level, at optimal transmission power; second, channel quality is below a (computable) SNR threshold.

the corresponding sensor will be completely shut off to save energy. In contrast, when the channel quality is good and the observation noise is low, the corresponding sensor will be active. Hence, the potential duty sensor(s) is the one who can receives a pre computed SNR level. As such, only some sensors will be eligible to participating in the target tracking routine, while others will have to abstain.

Along the same, energy saving line, we presented an optimal bit assignment scheme for the noise-corrupted sensor observations (measurements). In this scheme, sensor observations are first quantized into binary levels, and then transmitted to the fusion center where a final decision is made.

In very broad terms, we claim to have elaborated on the moving target tracking problem, but from different viewpoints. The objective has always been, to challenge the long held paradigm that high tracking quality (low tracking error) necessarily requires high power consumptions.

Numerical validation results show that, our suggested methods decreases energy consumptions in the sensor/fusion communication phases by the constraint binary message transmissions. This is well motivated by the bandwidth limitation of the communication links, and by the limited power budget of local sensors.

On the other hand, the energy consumed in target tracking is minimized to an analytically optimal level while the target QoM level is satisfied all the time.

Appendices

A constrained optimization problem, can be formulated as follows,

$$\begin{cases} \text{Min. } \sqrt{a} \text{ , such that} \\ \left| \frac{\log(\eta)\sigma}{R\sqrt{a}} - \frac{R\sqrt{a}}{2\sigma} \right| \leq -\varphi \\ \sqrt{a} \geq 0 \end{cases}$$

First we convert to the form

$$\begin{cases} \text{Min. } \sqrt{a} \text{ , such that} \\ \left| \frac{\log(\eta)\sigma}{R\sqrt{a}} - \frac{R\sqrt{a}}{2\sigma} \right| \leq -\varphi \\ \sqrt{a} \geq 0 \end{cases}$$

Using Lagrange: the object function is

$$F = \sqrt{a} + \Gamma \left(\left| \frac{\log(\eta)\sigma}{R\sqrt{a}} - \frac{R\sqrt{a}}{2\sigma} \right| \leq -\varphi \right)$$

Put $\sqrt{a} = T$

So we need $\frac{\partial F}{\partial T} = 0$, then after differentiations

$$1 - \Gamma \left(- \left(\frac{-\log(\eta)\sigma}{RT^2} - \frac{R}{2\sigma} \right) \right) = 0$$

$$\left(\frac{\log(\eta)\sigma}{RT} - \frac{RT}{2\sigma} \right) \leq -\varphi$$

$$\Gamma \geq 0 \tag{23}$$

From equation (23)

$$-2\sigma \log(\eta) \sigma + R^2(T_{\text{opt}})^2 = 2\varphi R(T_{\text{opt}}) \sigma \text{ , then}$$

$$R^2(T_{\text{opt}})^2 + 2R(T_{\text{opt}}) \varphi \sigma - 2\sigma \log(\eta) \sigma = 0$$

Put $\hat{\eta} = \log(\eta) \sigma$

$$R^2(T_{\text{opt}})^2 + 2\varphi R(T_{\text{opt}}) \sigma - 2\sigma \hat{\eta} = 0$$

Then $T_{\text{opt}} = \frac{\sqrt{(\sigma\varphi)^2 + 2\sigma\hat{\eta}} - \sigma\varphi}{R}$,

Where:

$$R^2 \frac{(\sigma\varphi)^2 + 2\sigma\hat{\eta} - 2\sigma\varphi\sqrt{(\sigma\varphi)^2 + 2\sigma\hat{\eta}} + (\sigma\varphi)^2}{R^2} + 2R\sigma\varphi \frac{\sqrt{(\sigma\varphi)^2 + 2\sigma\hat{\eta}} - \sigma\varphi}{R} - 2\sigma\hat{\eta} = 0$$

Then

$$\sqrt{a_{\text{opt}}} = \frac{\sqrt{(\sigma\varphi)^2 + 2\sigma\hat{\eta}} - \sigma\varphi}{R}$$

$$\sqrt{a_{\text{opt}}} = \frac{d\sqrt{(\sigma\varphi)^2 + 2\sigma\hat{\eta}} - d\sigma\varphi}{R} \text{ , (prove of equation (7))}$$

Where:


$$T_{\text{opt}} = \sqrt{a_{\text{opt}}} \text{ , } \sqrt{a_{\text{opt}}} = \frac{\sqrt{a_{\text{opt}}}}{d}$$

$$\sqrt{\text{SNR}_{\text{trg}}} = d\sqrt{\varphi^2 + 2\hat{\eta}/\sigma} - d\varphi = \frac{R\sqrt{a_{\text{opt}}}}{\sigma} \text{ , (prove of equation (8))}$$

ACKNOWLEDGMENT

The Authors would like to thank anonymous Reviewers for their valuable comments and suggestions that improve the presentation of these papers.

REFERENCES

- [1] Amr Lotfy Elewa M. and Osama M. El Ghandour, " Performance analysis of target tracking via wireless sensor network," Journal of Computing, Volume 4, Issue 8, August 2012, ISSN (Online) 2151-9617
- [2] Wei-Ho Chung and Kung Yao, " Decision Fusion in Sensor Networks for Spectrum Sensing based on Likelihood Ratio Tests ," Advanced Signal Processing Algorithms, Architectures, and Implementations XVIII, edited by Franklin T. Luk, Proc. of SPIE Vol. 7074, 70740H, 2008
- [3] Gongbo Zhou , Zhencai Zhu, Guangzhu Chen, Lijuan Zhou, " Decision fusion rules based on multi-bit knowledge of local sensors in wireless sensor networks," , Information Fusion 12 (2011) 187–193, Elsevier
- [4] RuixinNiu, Member, Biao C PramodK.Varshney " Fusion of  www.IJCSI.org Transmitted Over Rayleigh Fading Channels in

- Wireless Sensor Networks," IEEE Transactions on Signal Processing, vol. 54, NO. 3, march2006
- [5] P. Varshney. "Distributed Detection and Data Fusion," Springer, 1996.
- [6] X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill., "Integrated coverage and connectivity configuration in wireless sensor networks,". In ACM Sensys'03, Nov., 2003
- [7] E. Masazade, R. Rajagopalan, P. Varshney, C. Mohan, G. Sendur, and M. Keskinoz, "A multiobjective optimization approach to obtain decision thresholds for distributed detection in wireless sensor networks," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 40, no. 2, pp. 444–457, April 2010.
- [8] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," IEEE Communications Magazine, vol. 40, no. 8, pp. 102–114, August 2002.
- [9] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection in wireless sensor networks," IEEE Trans. Signal Processing., Jan. 2007, Submitted.
- [10] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises," EURASIP Journal on Wireless Commun. and Networking, vol. 2005, no. 4, pp. 473–482, 2005.
- [11] J. Xiao, S. Cui, and A. J. Goldsmith, "Power-efficient analog forwarding transmission in an inhomogeneous gaussian sensor network," in 6th Workshop on IEEE Signal Processing Advances in Wireless Commun., June 2005, pp. 121–125
- [12] Kris Hauser, "Constrained Optimization, Lagrange Multipliers, and KKT Conditions", B553 Lecture 7, February 2, 2012
- [13] R. TYRRELL ROCKAFELLAR, "LAGRANGE MULTIPLIERS AND OPTIMALITY ", Vol. 35, No. 2, June 1993, Society for Industrial and Applied Mathematics
- [14] Yan Zhou, Student Member, IEEE, Jianxun Li, Member, IEEE, " Collaborative Maneuvering Target Tracking in Wireless Sensor Network with Quantized Measurements ", Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA - October 2009
- [15] S.J. Orfanidis, "Introduction to Signal Processing," Englewood Cliffs, NJ: Prentice-Hall, 1996
- [16] JJ. Xiao, S. Cui, Z. Q. Luo, and AJ. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," IEEE Trans. Signal Process., vol. 54, no. 2, pp. 413–421, Feb. 2006.
- [17] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," IEEE Trans. Information Theory, vol. 51, pp. 2210–2219, 2005
- [18] Kai Lin, "Research on adaptive target tracking in vehicle sensor networks," Journal of Network and Computer Applications, 2012 Elsevier
- [19] M. AmacGuvensan, A. GokhanYavuz, "On coverage issues in directional sensor networks: A survey", ADHOC 592, 28 February 2011, 2011 Published by Elsevier
- [20] Loredana Arienzo, Maurizio Longo, "Loredana Arienzo, Maurizio Longo, "Energy-efficient tracking strategy for wireless sensor networks", 2008 IEEE
- [21] Loredana Arienzo, "Efficient Particle Filter for Target Tracking in Wireless Sensor Networks", ACM SIGCOMM. Poster Session (Kyoto, Japan), August 27- 31 2007.
- [22] Guillermo Molina, Enrique Alba, "Location Discovery in Wireless Sensor Networks Using Metaheuristics," Applied Soft Computing Journal 2008, Elsevier
- [23] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," IEEE Trans. Signal Process., vol. 52, no. 12, pp. 3454–3458, Dec. 2004.
- [24] J.J. Xiao, S. Cui, Z. Q. Luo, and A.J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," IEEE Trans. Signal Process., vol. 54, no. 2, pp. 413–421, Feb. 2006.

Osama M. El Ghandour received the B.Sc degree from Helwan university in 1982 and M.Sc degree from polytechnic university, New York, in 1986, and D.Sc degree from George Washington university Washington D.C in 1990. Currently he is an Associate professor at Helwan university, Cairo, Egypt, he conducted research in theoretical limits in communication with practical constraints, information-theoretical models for cellular mobile systems, combined modulation and coding, resource allocation management methods for wireless systems, routing protocols in Mobile Ad-hoc systems. E-mail: osamaelghandour90@gmail.com

Amr Lotfy Elewa M. received the B.Sc degree in communications and electronics from Helwan Faculty of Engineering, Helwan University, in 2003, he is a Senior Network Engineer at computers & systems department, Electronics Research Institute (ERI) since 2006. E-mail: amrlotfy55@yahoo.com; amr@eri.sci.eg

Dimension Reduction in Intrusion Detection Features Using Discriminative Machine Learning Approach

Karan Bajaj¹, Amit Arora²

¹ Computer Science and Engineering Department, Chitkara University
Baddi, Himachal Pradesh, India

² Computer Science and Engineering Department, Chitkara University
Baddi, Himachal Pradesh, India

Abstract

With the growing need of internet in daily life and the dependence on the world wide system of computer networks, the network security is becoming a necessary requirement of our world to secure the confidential information available on the networks. Efficient intrusion detection is needed as a defence of the network system to detect the attacks over the network. Using feature selection, we reduce the dimensions of NSL-KDD data set. By feature reduction and machine learning approach, we are able to build Intrusion detection model to find attacks on system and improve the intrusion detection using the captured data.

Keywords: Feature Selection, Weka, NSL-KDD Data Set, Accuracy.

1. Introduction

As the growing need of internet in our daily life and our dependence on the world wide system of computer networks, the network security is becoming a necessary requirement of our world to secure the confidential information available on the networks. The precious information is always prone to maximum attacks over the network. Intrusion may occur due to system vulnerabilities or security breaches, such as system misconfiguration, user misuse or program defects. Attackers can also combine multiple security vulnerabilities into an intelligent intrusion. Intrusion detection plays an important role over the large network system. In a big network system there are large number of servers and on-line services running in the system while such networks may lure more attackers. Efficient intrusion detection model is needed as a defense of the network system [1].

2. Data Set

To test the Intrusion Detection, we use NSL-KDD [2] data set which consists of selected records of the complete KDD'99 [3] data set. In NSL-KDD [2] data set redundancy of the data in the original KDD'99 [3] data set is reduced, which makes the data more realistic for attack detection. The data set contains 41 features and 1 class labeled as Normal and Anomaly.

Data set is divided into separate Train and Test data on which evaluation is performed, with a total of 24 training attack types in the train set [8], with an additional 14 types [8] in test data set. This makes the detection more realistic because now the model is also checked for the unknown attacks.

3. Related Research

The Lightweight Network Intrusion Detection System, LNID, is proposed system for intrusion detection. The filtering scheme proposed consists of two packet filters: Tcpdump Filter and LNID Filter. The former one processes initial packet filtering with tcpdump tool, extracting TCP packets towards Telnet servers of internal local area network [1]. In [4], the authors purpose Intrusion detection using several Decisions Trees and Decision Rules. The prediction accuracy of classifiers was evaluated using 10-fold cross validation, due to cross validation the obtained accuracy was only for the known attacks. Extended security for intrusion detection system using data cleaning in large database [5], this process works on matching policies in database with anomalous information. So, it works well when the policy is matched, therefore technique is good for known attacks whose policies are already defined. Light weight agents for

intrusion detection, this approach is designed and implemented for intrusion detection system (IDS) prototype based on mobile agents [6], but limited for only mobile agents. IP Flow-Based Intrusion Detection [7], this approach find the attack contents by monitoring every packet. However, packet inspection cannot easily be performed at high-speeds. Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani [8], demonstrated the use of multiple machine learning algorithms on their proposed NSL-KDD [2] data set which was free from the redundant data which was in KDDCUP'99 [3]. They use separate training and test set which makes the detection more accurate for unknown attacks.

Literature survey showed that, for the practical, most of the researchers had used KDDCUP'99[3] which suffers from drawback of redundant data, which leads to the biasing in detection of attacks which are more frequent in data sets like DOS and PROBE attacks. Some researchers had applied single algorithm to detect all the attack types or they had used cross validation on data set which is good only for the detection of already known attacks. The researchers that have used the NSL-KDD [2] data set with multiple machine algorithms [8] did not try further any attribute selection measures to improve the accuracy. This motivated us for our assumption that using NSL-KDD [2] data set with different training and test set separately and attribute selection with different machine learning algorithms will yield good performance and improve prediction for detection of attacks including unknown attacks as well.

4. Attribute Selection Measures

For the attribute selection, different feature selection algorithms are used they find the contribution of the 41 features in NSL-KDD [2] data set in intrusion detection. Feature selection reduces the features from the data set without affecting the effective indicators of system attacks.

4.1. Information Gain Attribute Evaluation:

Information Gain Attribute Evaluation evaluates the worth of an attribute by measuring the information gain with respect to the class [9].

$$\text{Info}(G) = -\sum_{i=1}^m p_i \log_2(p_i)$$

Here Information gain G is calculated by calculating the probability of occurrence of class over total classes in data set.

4.2. Gain Ratio Attribute Evaluation:

It uses an extension to the information gain uses the gain ratio [9]

$$\text{Gain Ratio}(A) = \text{Gain}(A) / \text{Split Info}(A)$$

This value represents the potential information generated by splitting the training data set.

4.3. Correlation Attribute Evaluation:

Correlation specifies dependence of feature on each other. It represents the linear relationship between the variables or features.

$$r_{A,B} = \frac{\sum_{i=1}^N (a_i - \bar{A})(b_i - \bar{B})}{N\sigma_A\sigma_B} = \frac{\sum_{i=1}^N (a_i b_i) - N\bar{A}\bar{B}}{N\sigma_A\sigma_B}$$

Here N is the number of tuples, a_i and b_i is the respective values of A and B in tuple i, \bar{A} and \bar{B} are the respective mean values of A and B, σ_A and σ_B are the respective standard deviations of A and B [9]. The value of $r_{A,B}$ lies between -1 and 1. If A and B are completely correlated, $r_{A,B}$ takes the value of 1, if A and B are inversely correlated then $r_{A,B}$ takes value of -1 and if A and B are totally independent then $r_{A,B}$ is zero.

5. Implementation Setup & Methodology

From feature selection and machine learning algorithms we will be able to collect the result data through which we can identify and predict the machine learning techniques that helps to distinguish between alerts, attacks and normal data. Our purpose is to suggest a learning model to reduce the false alarms and improves detection of attacks.

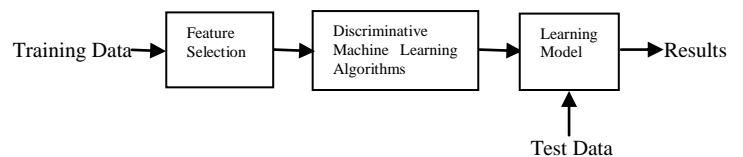


Fig.1 Implementation Setup Model

5.1. Feature Selection

For the attribute selection, we use different feature selection methods.

5.1.1. Information Gain Attribute Evaluation, we process the NSL-KDD [2] train set and retrieves the results. This algorithm use rankers method on features and evaluate the feature by ranking them from most important to least important.

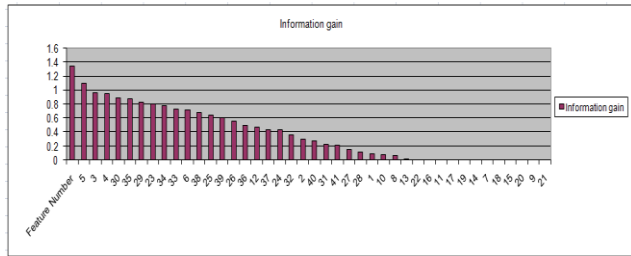


Fig.2 Information Gain Attribute Evaluation

Developed By: Karan Bajaj Date:24/04/2013 Aid:Weka(3.6.9)

5.1.2. We use training data and apply Gain Ratio Attribute Evaluation algorithm on data, this algorithm use rankers method on features and evaluate the feature by ranking them.

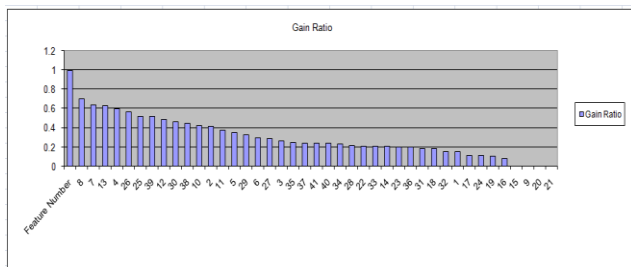


Fig.3 Gain Ratio Attribute Evaluation

Developed By: Karan Bajaj Date:24/04/2013 Aid:Weka(3.6.9)

Information gain measure is biased towards tests with many outcomes. Gain Ratio prefers to select attributes having a large number of values. It uses an extension to the information gain.

5.1.3. Correlation Attribute Evaluation, this algorithm rank the features in NSL-KDD [2] train set based on their correlation with each other, correlation specify dependence of feature on each other.

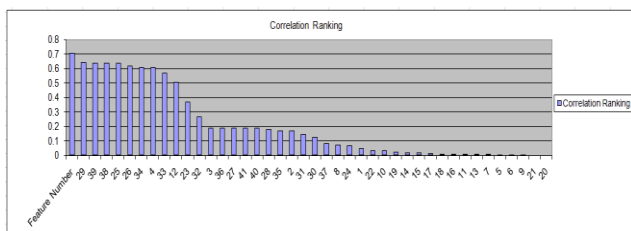


Fig.4 Correlation Attribute Evaluation

Developed By: Karan Bajaj Date:24/04/2013 Aid:Weka(3.7.9)

5.2. Dimension Reduction:

From the three feature selection methods applied on NSL-KDD [2] training data set (Fig 2,3 and 4), we come to find

that feature number 9, 20 and 21 (urgent, num_outbound_cmds and is_host_login) have no role in detection of any attack and further 15, 17, 19, 32 and 40 (su_attempted, num_file_creations, num_access_files, dst_host_count and dst_host_error_rate) have minimum role in detection of attack.

On the basis of analysis of results of feature selection, we reduce the NSL-KDD [2] data set. In both train and test data set the dimensions of data set is reduced by removing the feature numbers 9, 15, 17, 19, 20, 21, 32 and 40.

5.3. Discriminative Machine Learning Algorithms

On the reduced data set, we applied several discriminative machine learning algorithms, now training sets are given to train the machine learning algorithms and test data set is given separately. Using separate train and test set give us advantage to check the accuracy of detection of attacks even on unknown attacks, because training set contain 24 attack types [8] and test set contain additional 14 attacks with previous 24 attacks [8]. This makes the detection more accurate because now the model is also checked for the unknown attacks.

5.3.1. NaïveBayes:

The NaiveBayes [10] classifier provides an approach to represent and learn the probabilistic knowledge [11].

5.3.2. J48:

Is a tree classifier in Weka Tool [12], it is a version of C4.5 algorithm which was developed by Quinlan [13].

5.3.3. NB Tree:

NB Tree [14] builds a naive Bayes classifier on each leaf node of the built decision tree, which just integrates the advantages of the decision tree classifiers and the Naive Bayes classifiers [15].

5.3.4. Multilayer Perception:

It is a neural network classification algorithm [11].

5.3.5. LibSVM:

Support Vector Machines are supervised learning models with algorithms that analyze the data and recognize the patterns. LibSVM is integrated software for support vector classification, regression and distribution estimation. It supports multi-class classification [16].

5.3.6. SimpleCart:

Cart stands for classification and regression. Cart has the ability to generate the regression trees. It enables users to provide prior probability distribution [17].

6. Results & Performance Comparison

Table1 is representing the results, without feature selection on NSL-KDD [2] data set with 41 features and 1 class for Labels. The results are given in terms of accuracy of detection for various learning algorithms from previous benchmarks [8], except SimpleCart algorithm that is not used in previous benchmark paper.

Table 1 : Detection accuracy on NSS-KDD Test Data set

Classifier (Discriminative Machine Learning Algorithms)	Detection Accuracy (%)	Incorrectly Classified Instances
J48 [8]	81.05	**
Naïve Bayes [8]	76.56	**
NB Tree [8]	82.02	**
Multi-layer Perception [8]	77.41	**
SVM [8]	69.52	**
SimpleCart	80.3229	19.6771

** Indicates information not provided by the author in their respective paper.

Table 2 is representing the results, after feature selection on NSL-KDD [2] data set. Now the feature is reduced from 41 features to 33 features and 1 class for Labels. The results shown in Table 2 are compared with previous benchmarks shown in Table 1.

Table 2 : Detection Accuracy on reduced Data set after dimension reduction

Classifier (Discriminative Machine Learning Algorithms)	Detection Accuracy (%)	Incorrectly Classified Instances
J48	81.9375	18.0625
Naïve Bayes	75.7851	24.2149
NB Tree	80.6778	19.3222
Multi-layer Perception	73.5495	26.4505
LibSVM	71.0211	28.9789
SimpleCart	82.3235	17.6765

Result Analysis:

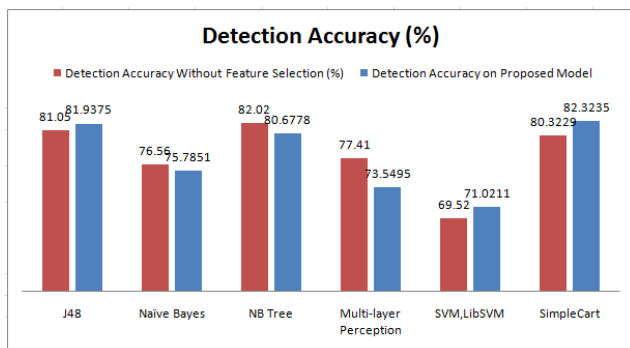


Fig.5 Comparison of Result

Fig.5 is representing the comparative analysis in terms of detection accuracy without feature selection and proposed model.

7. Conclusions

In this paper, we propose model for intrusion detection, which suggests that for the detection of intrusion, it is not necessary to perform the test on all the 41 features of NSL-KDD [2] data set. First by using feature selection the features are reduced to 33 features and further by removing them, the biasing of learning algorithms towards the frequent and easily detectable records in the data set is reduced. And the suggested machine learning algorithm after selection process is SimpleCart for the intrusion detection that leads to improve the computer security alerts from computer security incidents using machine learning techniques.

References

- [1]Chia-Mei Chen, Ya-Lin Chen, Hsiao-Chung Lin, “An efficient network intrusion detection”, Elsevier, Vol. 33, No. 4, 2010, pp. 477- 484.
- [2]“Nsl-kdd data set for network-based intrusion detection systems. Available on: <http://nsl.cs.umb.ca/NSL-KDD/>, March 2009.
- [3]KDD Cup (1999). Available on: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>
- [4]G. Meera Gandhi, Kumaravel Appavoo , S.K. Srivatsa, “Effective Network Intrusion Detection using Classifiers Decision Trees and Decision rules” Int. J. Advanced Networking and Applications, Vol. 2, No. 3, 2010, pp. 686.
- [5] R. Aarthy and P. Marikkannu, “Extended security for intrusion detection system using data cleaning in large database” International Journal of Communications and Engineering, Vol. 2, No. 2, 2012, pp. 56-60.
- [6]Guy Helmer, Johnny S.K. Wong, Vasant Honvar, Les Miller, Yanxin Wang, “Lightweight agents for intrusion detection”, Journal of systems and Software. Elsevier, Vol. 67, No. 2, 2010, pp.109-122.
- [7]Anna Sperotto, Gregor Schaffrath, Ramin Sadre, Cristian Morariu, Aiko Pras and Burkhard Stiller, “An Overview of IP Flow-Based Intrusion Detection” IEEE communications surveys & tutorials, Vol. 12, No. 3, 2010, pp. 343.
- [8]M. Tavallae, E. Bagheri, W. Lu, and Ali A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set” IEEE Symposium on computational intelligence in security and defence application, 2009.
- [9]Jiawei Han and Micheline kamber, Data Mining Concepts and Techniques ,Publisher Elsevier, pp.67-69, 296-301, 2001.
- [10] G. H. John, P. Langley, “Estimating Continuous Distributions in Bayesian Classifiers” In Proc. Of 11th Conference on Uncertainty in Artificial Intelligence, 1995.
- [11] Huy Anh Nguyen and Deokjai Choi, “Application of Data Mining to Network Intrusion Detection: Classifier Selection Model”, Springer-Verlag Berlin Heidelberg, LNCS 5297, 2008, pp. 399–408.

- [12] “Waikato environment for knowledge analysis (weka) version 3.6.9. and 3.7.9” Available on :<http://www.cs.waikato.ac.nz/ml/weka/>
- [13] J. Quinlan, : C4.5: Programs for Machine Learning, Publisher Morgan Kaufmann, San Mateo, 1993.
- [14] R. Kohavi, “Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid,” ser. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press, 1996, pp. 202–207.
- [15] Liangxiao Jiang and Chaoqun Li, “Scaling Up the Accuracy of Decision-Tree Classifiers: A Naive-Bayes Combination,” Journal of Computers, Vol. 6, No. 4, 2011, pp.1325-1331.
- [16] Chih-Chung Chang and Chih-Jen Lin, “LIBSVM : a library for support vector machines,” ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [17] Lior Rokach and Oded Maimon, “DECISION TREES,” Department of Industrial Engineering, Tel-Aviv University, Chapter-9, pp.181.

First Author: Karan Bajaj, Btech(Computer Science & Technology) from Himachal Pradesh University and pursuing M.E(Computer Science & Engineering) from Chitkara University. He is presently working as Assistant Professor in Department of Computer Science & Engineering at Chitkara University. He has more than 3 years of teaching experience to his credit. He has attended various workshops and short-term courses in different domains.

Second Author: Amit Arora, ME(Computer Science & Technology) from IIT Madras. He is presently working as Assistant Professor in Department of Computer Science & Engineering at Chitkara University. His Research areas are Machine Learning, Artificial Intelligence and Data Structures and Algorithms.

A Novel Design to Increase Trust in Cloud IaaS model

Jitendra Kumar Seth¹, Satish Chandra²

¹Dept of Information Technology, Ajay Kumar Garg Engineering College,
Ghaziabad, India

²Dept of Computer Science and Engineering, Jaypee Institute of Information Technology,
Noida, India

Abstract

In IaaS services of cloud the customers demand for hardware resources as a service like memory, processor cycles, disk storage even software. IaaS services of cloud rely heavily on virtualization. The service is provided by means of virtual machine. It is not easy task for cloud users to store their valuable data over cloud because of matter of trust over cloud. Data integrity in cloud environment is ensured by the security of virtual machines. Recent survey shows the security is one of the primary concerns in adoption of cloud. In this paper a novel cloud design algorithm is proposed to ensure Virtual Machine integrity in which customer is also a part of the proposed security mechanism. The customer participation in cloud services increases customer's trust and adoption of cloud.

Keywords: Trust, cryptography, security, integrity, IaaS, hash code.

1. Introduction

Cloud computing is a Pay-per-Use-On-Demand model that can conveniently access shared IT resources through internet. Where the IT resources include network, server, storage, application, service and so on and they can be deployed with much quick and easy manner, and least management of and interactions with service providers [1]. Few examples of popular Cloud service providers are Microsoft Azure, Amazon EC2, and Google App Engine. Resource provisioning and flexibility are provided by means of service level agreement. During August 08/09 by IDC IT group [2], In the Cloud Computing Services Survey security, availability and performance issues still remain in the top 3 for both years the survey was done. Security is the main issue users are concerned with when considering Cloud computing solutions Cloud computing brings us the approximately infinite computing capability, good scalability, service on-demand etc. but also imposes the challenges like security, privacy, legal issues. Public

cloud [3, 4] offerings are very generic and offer multi-tenancy service which all organizations might not be comfortable with. Implementing an in-house cloud is more complex to implement and are burdensome on internal resources if the organization is not large enough. Cloud service providers are continuously evolving solutions to overcome the above mentioned hurdles. Some enterprises are seeing clear benefits in shifting to the cloud and are adopting it unconditionally while some enterprises are moving non-critical applications to test the waters. Some others want to wait and watch how the technology evolves before deciding.

Interaction of consumer and consumer devices online with cloud services imposes a series of security challenges like data leakage, Virtual Machine (VM) Security, data loss and protection, data authentication, intrusion detection and resolution etc. In this paper we are resolving the issue of VM Security and producing the mutual trust between cloud vendor and customer. Virtual Machine instances can be protected by applying security countermeasures to each guest virtual machines. VM (Virtual machines) interact with each other on hardware backplane or on internet a malicious VM can affect to another targeted VM. The network level security does not detect these threats. Another security concern is with migration of VM. An attack scenario may be the migration of a malicious VM in trusted zone. VM images are prone to attack or modifications on solution is to encrypt VM all time but this approach degrades the performance of service. When a virtual machine migrated to another server it should be ensured that not a single bit is left behind the disk so that it can be recovered by other user.

The recommendations by Cloud security alliances [5] are-

1. Implementers should associate self healing capability with VM.
2. Implementers should encrypt the VM images when not in use.

3. Implementers should divide the service into categorized zones.

Cloud infrastructure [6] is multi-tenants operates on shared resources provides resource utilization but there are VMs running on same server. There is no physical separation. Therefore a malware may spread throughout the cloud. Modern attacks like root kit attacks are very difficult to be detected by traditional antivirus products. These attacks infect the key components like hypervisor and drivers that causes malicious behavior.

As more and more information on individuals and companies are placed in the cloud, concerns are beginning to grow about just how safe an environment it is. Jianfeng Yang et. al. given top most concern and issues these are security, privacy reliability, legal issues, open standards, compliance, freedom, long term viability [2]. Many businesses are putting their data on cloud data centers to improve resources utilization, speed development and deployment and reduce cost; however these new platforms are having additional avenues for threats against data, systems, network and reputation. These threats are data-stealing malware, web threats, spam, phishing, Trojans, worms, viruses, spyware, bots, and more. For most of the part all these threat are presented in the same kind of attacks. In a recent survey conduct by trend micro [7] on cloud and virtualization used by industry and business worldwide almost 45 percent are using public cloud and 46 percent are using private cloud. Inter virtual machines communications are blind to the traditional security appliances. This is said to be blind spot problem. The solution is to install a virtual machine that continuously coordinates the communication between VMs.

2. Cloud Basics

In this section some cloud basics are discussed.

2.1 Cloud Architecture

Although no clear picture of cloud architecture is meeting in literature as there are no such international standards of implementing cloud and their interoperability, the architecture is varying as per the service provider's comfortability and the need of service level agreement and QoS. The essential components of cloud architecture are presented in figure 1. The cloud users are the clients who consume the services provided by cloud. Clients for the first time provide their personal and company detail and their email addresses, credit card details for billing to cloud provider to register for their cloud services. Provider then provides id and password to the consumer. Now the customer specifies their cloud service by login in their account. The SOAP request of the service is then forwarded to the cloud service provider.

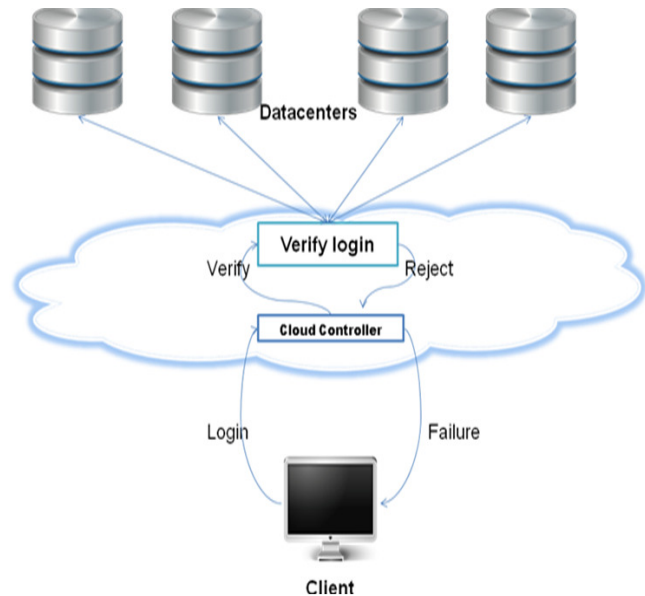


Figure 1: Basic architecture of cloud

The cloud controller accepts the request and authenticates the client for cloud service by verifying their id and password. The cloud controller can be the provider itself or some other third party. The cloud controller then checks for the host across data centers for availability of requested resource and initiate the virtual machine on host on suitable data center. Once client is connected with their VM on a host on data center the GUI of VM to client browser is loaded to interact with virtual machine. The billing of resource usage, provisioning of extra resource allocation, migration and load balancing of virtual machines and update of virtual machines all are the work of cloud controller.

The cloud architecture is segregated into following seven different layers [8]:

User Layer: In this layer cloud user's profiles are entered and processed. User profile and login are processed and maintained at cloud data center. The cloud users are provided with cloud interaction interface to interact with their user account and cloud services by means of virtual machine interface.

Virtualization Layer: users demand services are provided by configured virtual machine as per user requirement.

Service oriented module: It provide the interface to reuse the services.

Cloud module: This module is what resizes the services and tracking the usage of services charges, billing are maintained in this layer.

Cloud offering layer: This layer offers users to use value added services offered by others cloud vendors without changing underlying basic architecture.

Cloud Information layer: data structure, business case history, business intelligence and way of extracting data from database is defined in this layer.

Cloud quality and monitoring layer: This module monitor for the services upto agreed QoS.

Cloud is a cost effective platform for the business that is of medium or of small scale or to those whose budget is low or just started and establishing. Basically, virtualization reduces complexity. In that way, a virtualized network is easier to manage, hence less cost to administrating a complex solution and enhances user's friendliness. Enhanced scalability, flexibility, and utilization provide additional cost savings. Automation is an additional functionality available through virtualization. All these features are transparent to user and give a clear picture of the cloud to the user [9]. Cloud should charge only when the resources are in use and not to charge for resource once assigned to their consumers and are idle. Automatic resource provisioning and scaling supports efficient functionality of the business.

3. Cloud Security Overview

In this section a brief overview of current security threats in cloud are discussed.

The analysis of management interface are done on Amazon EC2 and S3 services, the control interfaces could be compromised via the novel signature wrapping and advanced XSS techniques. Similarly, the Eucalyptus control interfaces were vulnerable to classical signature wrapping attacks, and had nearly no protection against XSS[10].The XML signature wrapping attack by using SOAP request occurred in cloud services. Using XML signature wrapping attack the intruder can bypass the user login verification and may prove them as legitimate user. Another threat to cloud management interface is cross site script (XSS) - The first script injection vulnerability discovered on the aws.amazon.com domain was caused by a download link used to retrieve X.509 certificates issued by Amazon.

Clouds have the following types of intrusion threats [11]:

- i. Insider attacks
- ii. Flooding attack
- iii. User to root attacks
- iv. Attacks on virtual machine (VM) or hypervisor
- v. Backdoor channel attacks etc.

According to Gurdeep Singh et. al. [12] critical information on cloud inspires the attacker to steal and threat them hence security is one of the prime concerns of

cloud. Author focuses on security of VM images which are foundation of cloud security. Today's cloud computing platforms are typically "opaque": Amazon EC2 users only receive virtual units of CPU and memory, and physical details of the platform are hidden. Such opacity prevents programs from online optimizations and deployment adjustment, and is penalizing the very applications cloud computing attempts to attract: high-performance software. On the other extreme, a completely transparent design of clouds would lead to severe security and reliability concerns. Yu David Liu et. al. had given an idea how we can customize our cloud features using given APIs. They are given two APIs, one to cloud side and another one to application side. Interface designs that can help cloud programs to fine-tune performance, and how it may impact on other important issues of cloud computing, such as security, scheduling, and pricing [13]. Concretely, they propose a new object-oriented programming language, iCloud, for cloud computing. iCloud is designed with the philosophy of Interaction-Based Programming. Here is cloud services can be abstracted as classes with connectors. Application side has a DB connector and Sch connector. This means that a MyApp object can participate in two and only two kinds of interactions with other objects at run time. Connectors serve as the complete specification of the cloud's exposure to the programmer. To avoid "too much" transparency, cloud service providers need to, and only need to, design connectors carefully. A connector-based design is a boon for security, so that access control policies are only needed on these well-defined interfaces. Each connector may have a number of imports and exports. Each export is a method the connector can provide to "the other party" (i.e. whoever is connected to this connector), and each import is a signature specifying what it expects the other party to provide. Each connector may also hold connection-specific data.

3.1 Security in cloud categories (SaaS, PaaS, IaaS)

Literature survey is being done on security in SaaS, PaaS and IaaS and is summarized as follows. Currently Cloud computing clients have to trust 3rd party cloud providers on many fronts, especially on the availability of cloud service as well as data security. Therefore the SLA forms an integral part of a client's first line of defense. The SLA thus becomes the solitary legal agreement between the service provider and client [14]. Figure 2 shows the different cloud delivery models and deployment models are matched up against the information security requirements with an "X" denoting mandatory requirements and an asterisk (*) denoting optional requirements. The only means that the cloud provider can gain the trust of clients is through the SLA; therefore the

SLA has to be standardized. The main aspects as a guideline, which the SLA contains, are:

1. Services to be delivered, performance,
2. Tracking and Reporting
3. Problem Management
4. Legal Compliance
5. Resolution of Disputes Customer Duties
6. Security responsibility
7. Confidential Information Termination.

		Cloud Delivery Models								
		Public cloud			Private Cloud			Hybrid Cloud		
Information security requirements	Identification and Authentication	X	X	*	X	X	*	*	X	*
	Authorization	X	X	X	*	X	*	*	X	*
	Confidentiality	*	X	*	*	X	X	*	X	*
	Integrity	X	X	*	*	X	X	X	X	X
	Non repudiation	*	X	*	*	X	*	*	*	*
	Availability	X	*	X	X	X	X	*	*	*
		IaaS	SaaS	PaaS	IaaS	SaaS	PaaS	IaaS	SaaS	PaaS
		Cloud Deployment Models								

Figure 2: Cloud security requirements

Potential Cloud organizations and vendors need to be aware that it may become easier for attackers to threaten clouds by moving towards a single cloud interface. The shift to Cloud computing moved much of a user’s normal activity to the Web browser. Web browsers generally store all of a user’s saved passwords, browsing history and other sensitive information in a single place. As such it is possible for malicious websites to exploit browser vulnerabilities in order to steal information associated with other existing or previous browsing sessions, such as a logged in email account or online banking session. It is for this reason that some security experts recommend that consumers use one web browser for general surfing, and another for more sensitive tasks, such as online banking. Often, usernames and passwords are transmitted to remote servers via unencrypted network connections. In cases where encryption is used, it is typically only used to transmit the initial login information, while all other subsequent data is sent in the clear. This data can easily be snooped on by hackers. This exposes users to significant risks when they connect to the services using public wireless networks to any Cloud Service.

SaaS and PaaS Security issues are as follows: SaaS (Netflix, MOG, Google Apps, Box.net, Dropbox and Apple’s new iCloud) typically focuses on managing access to applications, while PaaS (Google App Engine,

Microsoft Azure, Salesforce’s Force.com, the Salesforce-owned Heroku, and Engine Yard) focuses primarily on protecting data, and IaaS (Amazon, Microsoft, VMWare, Rackspace and Red Hat etc.) focuses on managing virtual machines. Since SaaS delivers applications from the cloud, the main risk is likely to stem from multiple passwords accessing applications [15]. “An organization can solve these issues by opting for a single sign-on option between on-premise systems and cloud. By leveraging a single sign-on option, users are able to access both their own desktops and any cloud services via a single password. This approach also reduces the incidences of dangling accounts – which are vulnerable to unauthorized usage – after users leave organizations.” PaaS can be inherently secure, but the risk is slow system performance. That’s because data encryption is recommended before data is sent to PaaS cloud providers. The risk is that encrypting every piece of data will also eat up consumer organizations’ CPU cycles and slow things down. Still, any solution implemented should broker the connection to the cloud service and automatically encrypt “confidential user data such as home addresses, social security numbers or even medical records.” Audit trails provide valuable information about how an organization’s employees are interacting with specific Cloud services, legitimately or otherwise.

Lombardi et. al. [16] shows the security of cloud by protecting the integrity of guest virtual machines and cloud components. Advance cloud protection system (ACPS) effectively monitors the integrity of guest virtual machines and infrastructure and remains transparent to VMs and cloud users. Accountability produces a record of action which can be examined when something goes wrong. All modules of ACPS are located on host side. ACPS makes use of Qemu to access the guest VMs. Suspicious guest activities (e.g. system call invocation) can be noticed by the Interceptor and recorded by the Warning Recorder into the Warning Pool, where the potential threat will be evaluated by the Evaluator component. Evaluation components are evaluator and hasher is always active. Although ACPS is a host based security mechanism of virtual machine it does not involves the client to ensure their virtual machine integrity; even the client does not know their virtual machine has been compromised. In this proposed security mechanism Client is also part of their security assurance hence produces mutual trust between cloud vendor and the client. Integrity one of the three major security aspects confidentiality, integrity and availability is resolved for VMs with client involvement. VMs are stored on vendor’s database as a unique file of customers. Our proposed idea detects the infections of the VM by checking integrity of theses stored files. As soon as the unauthorized modification to VM is detected at vendor side the appropriate security mechanism may run and recover from these unauthorized modifications and

restores the VM in user account. The detailed idea is described in section 4.

4. Proposed Security algorithm ensures integrity of VM

In cloud computing, to register with cloud services the customers provide their personal information including e-mail and credit card information to the cloud service provider. Customers are provided with unique user id and password via their registered e-mail from the cloud service provider. By using the provided user id and password customer login the cloud service and a pair of public and private key for the customer is generated and the public key is registered with the cloud host for secure transmission between both the parties. Now customer's encrypted request (with the private key) is served by cloud host [17]. Here clients do not play any role in their security assurance. Client is totally dependent on cloud service provider's security mechanism for his / her data security, running on cloud host.

4.1 Hash Code

Hash code is a cryptographic technique to check message integrity. A hash function accept a variable size message M as input and produces a fixed size output referred to as hash code $H(M)$. The hash code is also referred as message digest or hash value. The hash code is a function of all the bits of the message and provides error detection capability. A change to any bit or bits in the message results in a change to hash code. An important application of secure hashes is verification of message integrity. As of 2009, the two most commonly used cryptographic hash functions are MD5 and SHA-1. All well-known hash functions, including MD4, MD5, SHA-1 and SHA-2 are built from block-cipher-like components designed for the purpose, with feedback to ensure that the resulting function is not invertible.

4.2 The proposed security model

In our proposed model, after successful registration of client with the cloud provider a hash code checker module is installed at the client machine by cloud service provider, which reserves few KBs of memory in the disk for future use. This memory space is restricted and only used by hash code checker module. After successful installation of hash code checker module at client machine, cloud controller module which is a part of a provider's cloud service management software, sends a default valid hash value encrypted by cloud provider's public key to the hash code checker module at client which stores the encrypted hash

in client's machine restricted memory space. Whenever a client logs in to the cloud service the hash code checker appended the encrypted hash value with the login information and sent to the cloud provider. After verification of login information provider also decrypt the received hash value with his private key and compare the decrypted hash value with the stored virtual machine image hash value at provider's side if both the hash values are identical then client is successfully authenticated for cloud service. Once the virtual machine (VM) is allocated to the client and client workout on the allocated virtual machine after all the work done on VM client request for sign out, before logout confirmation to client and shutting down the virtual machine the hash value of saved up-to-date virtual machine image is computed at datacenter by using cryptographic techniques like MD5 or SHA-1 and encrypted with the cloud provider public key and transmitted encrypted hash value to the client which is received and stored in restricted memory area by the hash code checker module. Each time a client logs in to the cloud service the hash code checker module appends the last saved encrypted hash value of virtual machine with the login information.

Figures 3a & 3b shows the design of the proposed security mechanism discussed above.

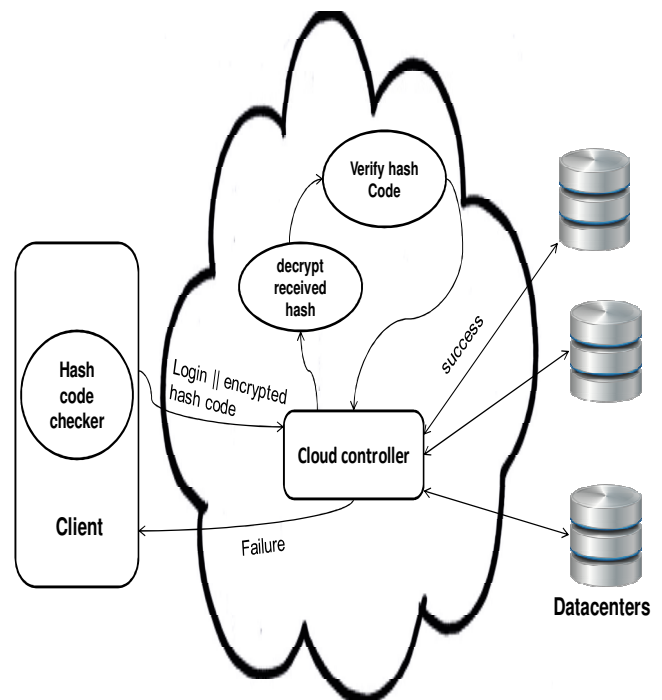


Figure 3a: Two way authentication of user by using login and hash code

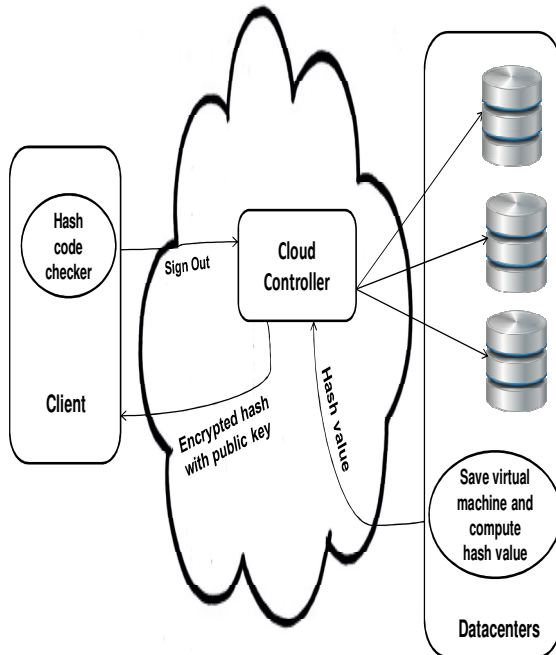


Figure 3b: Transfer of encrypted hash value to client machine after logout

4.3 Proposed Integrity check algorithm of virtual machine

The aforementioned approach can also be written in following steps:

1. Client registers with cloud service provider
2. After successful registration client is provided with user id and password.
3. Client logins and hash code checker is installed at client machine.
4. Hash code checker stores encrypted hash value to restricted memory area at client machine
5. Each time client login to cloud service the hash code checker module append encrypted hash code with login information
6. Cloud controller verify login information then decrypt received hash code by using provider's private key and compare with hash value of stored virtual machine image at data centers.
7. Once user is verified in both the ways successfully then only login is confirmed and last saved user's virtual machine image is loaded on data center and control is provided to user browser.

8. If user login fails in any of the ways then alternate mechanism addressed in section 6 points d and f executes.
9. After work out on virtual machine user signals logout
10. Cloud Provider save the virtual machine image and computes the hash value of VM image and encrypts it (by using Provider's Public Key).
11. Encrypted Hash value is sent to the Client by cloud controller.
12. Hash code checker module at client stores the received encrypted hash value and sends the acknowledgment to Cloud Provider, which then signals successful, sign-out to the Client.
13. Each time Client logins; the last saved encrypted hash value at Client is appended with login Information and all steps 6 to 12 repeated each time.

5. Experiment

In our experiment NetBeans 7.0.1 and jdk 1.6 were used with Libvirt API to interact with guest virtual machines. Qemu is used as a tool between Libvirt and Hypervisor. KVM hypervisor is used on Ubuntu 12.04. We have created Guest virtual machines from the .iso images of various Linux based operating systems like Ubuntu, Fedora, and DSL Linux etc. VM configuration is provided by using XML API of Libvirt. In our experiments we have given the configuration of VM creation as follows: one logical processor, 347 MB of RAM, and dsl linux .iso image. We have created, suspended, resume and saved virtual machines using Java. We have also computed the hash code of virtual machine using java security package [18] hash code algorithm SHA-1 and stored them in a file. Before the next boot of virtual machine the hash code of stored VM image file is again computed and compared with the stored hash value they are matched and VM is started.

The following depicts one of the output of our experiment:

```
Virtual Machine Name=mydslvm---Time of Creation:--Wed May 08 23:09:51 IST 2013
Virtual Machine Name=mydslvm---Time of Suspend:--23:11:39
Virtual Machine Name=mydslvm/nTime of Resume:--23:12:2
Virtual Machine Name=mydslvm---Time of Suspend:--23:12:24
Virtual Machine Name=mydslvm---Time of hashcode:--Wed May 08 23:12:45 IST 2013:12:45
Hash code is -
136763718569839279811069244876589845739944818196
```

For the next boot of VM

```
Virtual Machine Name=mydslvm---Time of hashcode:--Wed May 08 23:13:48 IST 2013:13:48
```

Hash code is -
136763718569839279811069244876589845739944818196

6. Benefits of proposed security mechanism

This approach is very much useful in following respects-

- a) It produces two way authentication of client to cloud host. One with something stored at cloud host (login information) second something stored (encrypted hash value of VM) with the client machine.
- b) This security approach prevents XML signature wrapping attack and script injection attack. If the signature (id, password and public or private key consumer side) is compromised [10] by such attacks even though the attacker still cannot prove their authenticity as client side hash code is encrypted by cloud providers public key and can only decrypted by providers private key.
- c) It enhances the mutual trust between the cloud provider and the cloud service consumer. Client machine is also playing a crucial role in their security mechanism.
- d) If the client VM is targeted and compromised by attacker and intruder, there are some unauthorized changes or modification to virtual machine at the host, hence there is a mismatch of hash values between received hash value and stored hash value on the host and client login is unsuccessful. It may also be the case with wrong password so first password is confirmed to client by using client e-mail id; if then after the login is not successful Then the strong intrusion detection mechanism by host to the targeted virtual machine is carried out and if found suspicious replaced by backup virtual machine confirm successful login to client.
- e) Even if there is a problem with restoration of up-to-date virtual machine then some previous version of virtual machine integrity is ensured by client side stored hash value and client data is restored upto some recent version.
- f) If client side security module, hash code checker is compromised then client login is again unsuccessful. Client is provided password by using their email id or other medium. If again unsuccessful login then a new copy of hash code checker module is installed at client machine and login password and encrypted last saved hash value at the host is provided to client security module.

7. Conclusion

From the proposed security approach it can be seen that this security scheme provides two way authentication and

also prevents xml signature wrapping attack and script injection attack. It also stores hash information at client machine so increasing level of mutual trust. Client trust is produced as client is assured any intruder cannot login the Client without knowing the hash value stored at Client machine even though login information is compromised. This scheme is also helpful in data recovery process and triggers intrusion detection mechanism as found suspicious. It provides two way security in which client is also playing an important role. If recovery is not upto date then some previous version of VM can be restored with the help of stored hash value at client machine and backup VMs on host. Hence this security approach reduces the threat level to cloud and also producing the trust to those peoples worrying to adopt cloud services for their data protection.

References

- [1] Junjie Peng et. al, "Comparison of Several Cloud Computing Platforms" in Second International Symposium on Information Science and Engineering, IEEE-2009.
- [2] Jianfeng Yang, Zhibin Chen, "Cloud Computing Research and Security Issues" in Computational Intelligence and Software Engineering (CiSE), IEEE-2010.
- [3] Kirchberg, Markus; Lee, Bu Sung," Efficient Migration of Virtual Machines between Public and Private Cloud", Cloud Computing Technology and Science (CloudCom), IEEE 2011.
- [4] Simmons, Bradley; Smit, Michael; Litoiu, Marin," An architecture for overlaying private clouds on public providers", Network and service management (cnsm), 2012.
- [5] Cloud security alliances, 2011.
- [6] Raghu Yeluri et. al, "Building Trust and Compliance in the Cloud for Services " in Annual SRII Global Conference, IEEE 2012.
- [7] Virtualization and Cloud Computing: Security Threats to Evolving Data Centers, Trend Micro, 2012
- [8] Amarnath Jasti et.al. ," Security in Multi-Tenancy Cloud" in Security Technology (ICCST), 2010 IEEE.
- [9] "Virtual Enterprise Model for Enabling Cloud Computing for SMMEs" in *ISWSA'11*, April 18–20, ACM 2011.
- [10] Juraj Somorovsky, Mario Heiderich et.al, "All Your Clouds are Belong to us – Security Analysis of Cloud Management Interfaces" in *CCSW'11 ACM-2011*.
- [11] Chirag Modi, "A survey of intrusion detection techniques in Cloud" in *Journal of Network and Computer Applications*, Elsevier 2012.

[12] Gurdeep Singh Bindra et. al, "Cloud Security: Analysis and Risk Management of VM Images", in International Conference on Information and Automation (ICIA), IEEE 2012.

[13] Yu David Liu and Kartik Gopalan. " Interaction-Based Programming Towards Translucent Clouds (Position Paper)" in APLWACA ACM-2010.

[14] Ramgovind S et. al, "The Management of Security in Cloud Computing" in Information Security for South Africa (ISSA), IEEE-2010.

[15] Joe McKendrick "SaaS, PaaS and IaaS: three cloud models; three very different risks" on ZDNet April 2012

[16] Flavio Lombardi et. al, "Secure virtualization for cloud computing" in Journal of Network and Computer Applications 2010.

[17] Sven Bugiel et. al, "AmazonIA: When Elasticity Snaps Back" in CCS'11, October 17–21, 2011, Chicago, Illinois, USA, ACM 2011.

[18] <http://www.roseindia.net/java/java-security.shtml>

Jitendra Kumar Seth is an Assistant professor in Information Technology Department at Ajay Kumar Garg Engineering College, Ghaziabad, India. He did his B.Tech in CSE from UPTU Lucknow and M.Tech in CSE from Shobhit University Meerut. Currently he is pursuing his PhD in CSE from Jaypee Institute of Information Technology, Noida, India. His area of interest is cloud computing.

Satish Chandra is an Assistant professor in the Department of Computer Science & Engineering at Jaypee Institute of Information Technology, Noida, India. He did his B.E. and M.Tech. from Birla Institute of Technology, Mesra, Ranchi and PhD from JUIT, Solan, India. His area of interest includes Cloud Computing, Machine Learning, Artificial Intelligence, Biologically Inspired computing and Medical Image Analysis.

Continuous Improvement of Production system in Algerian Industry

Aouag-Hichem ¹, Mechenene Athmane ²

¹ Productic and Automation Laboratory, Batna University, Algeria
Batna, 05000/Algeria

² Industrial engineering department Batna University, Algeria
Batna, 05000/Algeria

Abstract

This paper presents the methodology for process improvement of Algerian Company, which implements a new process by the following of business process reengineering approach. For this, we conducted a comprehensive study of the company activity through a study of the current manufacturing process followed by an evaluation of various performance indicators, this has allowed us to propose a new model production process and identify weaknesses of the activities and services that interact with the company machining process

Keywords: Improvement continuous, Kaizen, performance, Réingeneering, Six sigma.

1. Introduction

Today, companies are obliged to adapt to a disruption: market developments, changing technology and customer requirements, which orients companies to make a change and it is more difficult to carry and control. And despite all, the company remains a requirement for continuous improvement which consists of implement actions to achieve the best levels of quality and productivity. In this work we will propose a new production process model based on the principles and tools of continuous improvement.

2. What is a process

The process is a sequence of activities from one or more inputs produces a result (output) representing a value for a customer.

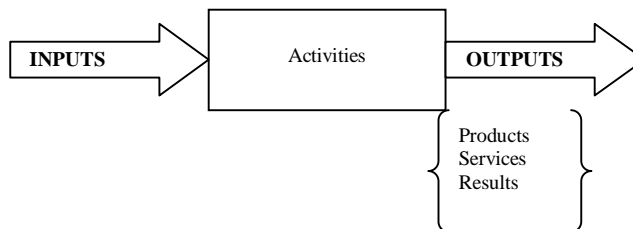


Fig. 1. Process Définition

3. What is a Performance

A performance in the company is all that and only that which helps to improve the pair value- cost, ie to improve the net creation of value

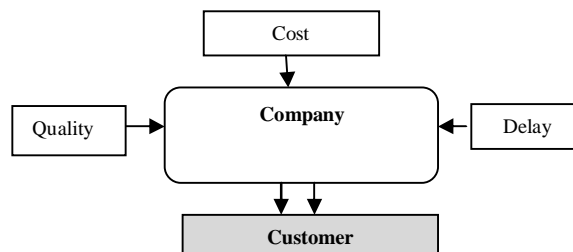


Fig. 2. Performance by the trilogy: quality- cost-delay

4. Type of process

There is no single type of business processes, however, we can facilitate the identification with two selections [3]:

4.1 According to the importance and value of a process: As shown in the table below:

Table 1. Type of process

Process	Type
active	One that provides a distinct value to the company in terms of capacity, reputation, competitive differentiation, cost, efficiency.
passive	One that turns the company capital without providing benefit.
identity	Central object of the company.
priority	Important element of the company (logistics, quality)
Background	That the company should do, but without spending too much effort or resources
obligatory	Laws and regulations.

4.2 According to according priority to three main families: The following figure shows this typology

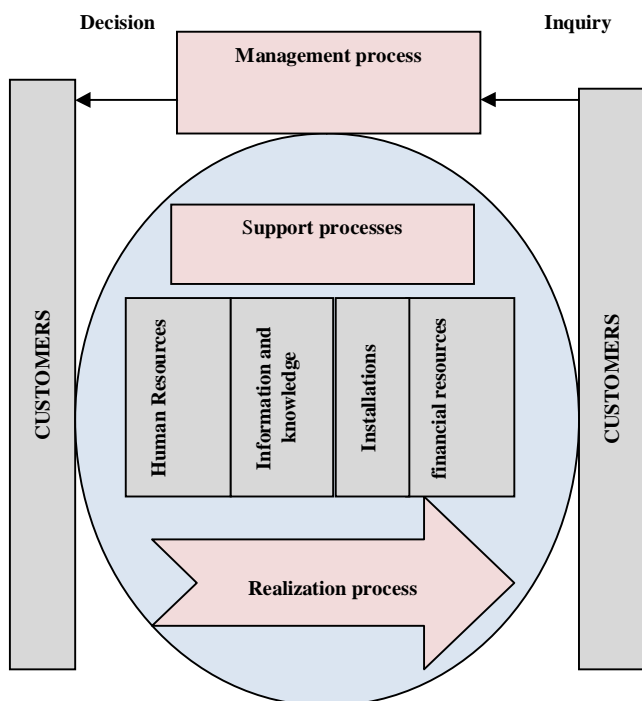


Fig. 3. Process Typologie [12]

5. Principles for improving industrial systems

It's possible to distinguish four main types of approaches to business. Those are [12]:

5.1 Approach oriented to eliminating waste

These are the steps of Lean Production and Kaizen seeking to eliminate waste. We can see these two approaches more akin to a system management. The objectives are to reduce and eliminate waste in the industrial world .

5.2 Approach oriented to fault rectification

The two approaches presented here are fairly considered as working methodologies directly applicable only as most general principles of management.

5.3 Approach oriented resource efficiency

In contrast to previous approaches that are process-oriented (reduction of defects and waste), we will now discuss an approach that focuses on the performance of the resources used by these processes, and these resources can be equipment or Human resources

5.4 Approach focused on change, improvement and development

The Business Process Reengineering (BPR) comes from the United States. It has become the leading approach in the 1990. Addition to his intrinsic qualities, the success of BPR is also because it was breaking with the existing, come mainly from Japan (JIT, Kaizen, 5S, TPM, Kanban, Poka-yoke) and allowed to run a new dynamic among business consultants

6. BPR approach

Most companies which had succeeded in changing radically had applied the same series of rules and techniques. In contrast, the failures were usually due to one or more identical reasons companies that sought more than a modest improvement and successful does not pose the same questions: "How does better, faster and less expensive things?" Why we do what do we make?"

7. Reengineering

Reengineering is a fundamental questioning and a radical redefinition of business processes to achieve spectacular gains in the critical performance which constitute today: the cost, quality, service and rapidity

7.1 The Process Reengineering

The reengineering does not see the company as a set of functions but as a set of processes to satisfy the customer, these processes are often cut across the company. Each process is defined by a class of customer (internal or external to the organization) and must offer added value to the client.

7.2 The elements of reengineering

Reengineering approach is based on four essential elements:

7.2.1 The critical process of the company

- The strategic process: medium and long term
- business processes: daily
- activities supporting the strategic and operational processes (process supports)

7.2.2 The process towards the client (business processes)

These are the processes for product realization. They include the activities of the identification of customer requirements to customer satisfaction. These processes can be found in the activities related to production, design, procurement, commercial approach

7.2.3 Focus on the company's business

Consultation on the objectives set by the company.

7.2.4 Use of modern information technology

Use other forms of management.

7.3 The principles

The reengineering is based on seven principles of reorganization:

- Have a person responsible for the process
- Facilitate access to information
- Make available the data processing for the process actors
- Centralize up similar activities
- To compare the activities
- Develop the autonomy of the process actors
- Eliminate the redundancies of information

7.4 The stages of Business Process Reengineering

BPR can be described as a sequence of six steps.

7.4.1 To adopt goals

From an inventory management to set goals and get the message to all employees. The message should be structured in two points: the need to change, and the status of processes and organization of the company after the change effected. Communication is a key element in the application of the method.

7.4.2 Identify the process to reconfigure

The BPR, even if it is "radical" should remain modest: we can revolutionize the entire company at one time it is necessary to locate the process changed.

7.4.3 Assess the factors favoring the reconfiguration

We need to know if the Human Resources, Information Technology, organization, corporate culture helps to change. A good evaluation of these factors have a significant impact on the success of the proposed reconfiguration

7.4.4 Understand the current process

The current process must be understood and diagnosed (advantages, disadvantages, outputs, performance), especially if one is a technological improvement of the process and not an outright change.

7.4.5 Establish the presentation of the new process

This step is the most creative We start from scratch and it suspends all rules, procedures and methods. It uses only the basic principles of the method, and perhaps the feedback from other cases of BPR.

7.4.6 Install the new process

In this step, leaders hold a key role, since they must make an effort to communicate to everyone involved and feel involved in the project implementation process. Moreover, we must ensure that the BPR project has achieved its objectives, comparing with the original objectives achievement.

The reorganization is based in the processes and the reengineering and structure based on certain number of principles, we will mention later

8. Six Sigma approach:

Six sigma is a disciplined, data-driven methodology for eliminating defects in any process. To achieve six sigma quality, a process must produce no more than 3.4 defects per million opportunities. Six sigma's basic value proposition is that principles for process improvement, statistical methods, a customer focus, attention to processes, and a management system focusing on high-return

improvement projects result in continuous improvement and significant financial gains.

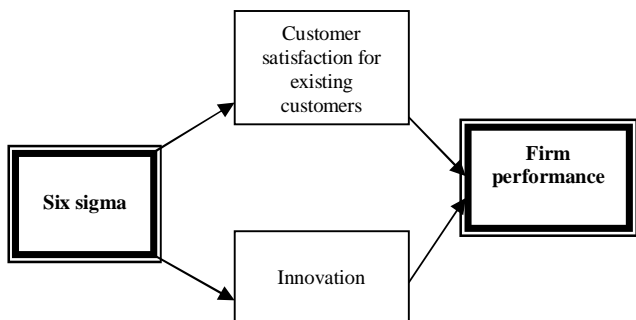


Fig.4 The effect of six sigma projects on firm performance [9]

9. Presentation of company CABAM (Algeria)

Is a national economic belongs to the industrial complex holding wood, with a capital of 500.000.000,00 Algerian Dinar (5 000 000 euro), it was established in March 1998 from the national unity of the joinery and construction precast, and it was created after the restructuring of the institution of cork and wood SNLB in 1988, which obtained ISO9001 certification

10. The product CABAM

It consists of manufacturing prefabricated cabins with different types or it can be changed in terms of length, width, nature of the cabin, and the characteristics of interior finishing and all this according to the desire of the customer.



Fig. 5. The product CABAM

11. Implementation of the reengineering approach

10.1 The objective

Our goal is to try to make reengineering in a process of realization chosen after a follow-up steps in our approach to obtain a new process model

10.2 Identify the process to reconfigure

To manage the process, it is necessary to have a global view of business operations. It is therefore important to develop a schematic representation, systematic process validated by all stakeholders: mapping process in the following figure, this representation is used to align the organization on customer needs. It must also define the functional relationships and interfaces of these processes by providing a real system of process

10.3 Implementation of the balanced approach

We take a score of 1 to 5 from the selected criteria, and then the value 1 will say that there is little impact of a thing of our process on the test, and the contrary for the other.

The criteria chosen for the new process

1.The impact on customers:

We take some criteria:

- PT : production time
- NQ: no quality
- QS: quality standards
- CP: Production cost

Table 2. The impact on customers criteria

diagnosis		PT	NQ	QS	CP	Total score
processus						
Machining	1					14
	2					
	3		x	x		
	4	x			x	
	5					
Primary assembly	1					13
	2		x			
	3	x				
	4			x	x	
	5					
Assembly of the cabin	1		x			11
	2	x				
	3					
	4			x	x	
	5					
Finishing	1					15
	2					
	3		x			
	4	x		x	x	
	5					

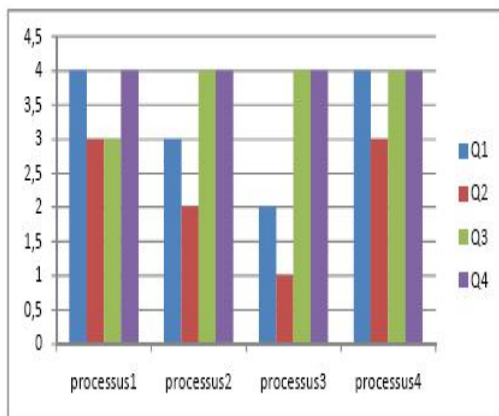


Fig .6 Impact on customer graphic

Interpretation: we can note that the finition process has the most impact on the customer (total score = 15) because works of this process are extremely clear for the customer. On the second level we have the note 14 which relates to the machining process due to the importance of production time and the relative costs

2.Possibility of process improving

We take some criteria:

AW: Accidents at work

DPA: differences between planned and actual production quantities

DPT: differences between planned and actual time

M: the maintenance

Table 3. Possibility of improving criteria

diagnosis		A W	D P A	D P T	M	Total score
Machining	1					17
	2					
	3					
	4	x	x	x		
	5				x	
Primary assembly	1					9
	2	x	x	x		
	3				x	
	4					
	5					
Assembly of the cabin	1					9
	2	x	x	x		
	3				x	
	4					
	5					
Finishing	1	x			x	6
	2		x	x		
	3					
	4					
	5					

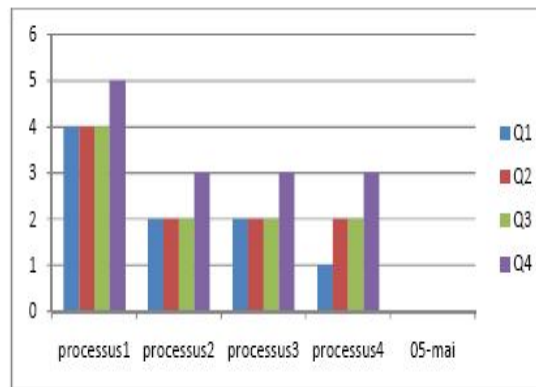


Fig.7 process improving graphic

Interpretation: It is clear that the first process is very important because the note is almost stationed into 4. All accident work saw in machining process because it's the only sector which contain machines

3. Adaptability Process

We take some criteria:

TA: The technical assistant in case of use problems

UCT: Using computer technology

EAT: Estimated average time to order processing

SE: social environment

Table 3. Possibility of improving criteria

diagnosis		T A	U C T	E A T	S E	Total score
proces						
Machining	1					19
	2					
	3					
	4				x	
	5	x	x	x		
Primary assembly	1					16
	2					
	3			x		
	4	x			x	
	5		x			
Assembly of the cabin	1					16
	2					
	3			x		
	4	x			x	
	5		x			
Finishing	2					14
	3	x				
	4			x		
	5				x	
	1		x			

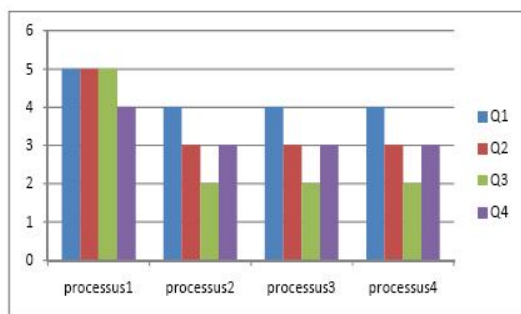


Fig.8 process Adaptability graphic

Interpretation: The impact on the adaptability for the first three diagnoses is maximum for the machining process. It is very important to think here of the action in order to face the new changes.

4. The Business impact:

We take some criteria:

L : Labor

CM: The cost of maintenance

QD: Quantities deterred

AE: The aging of equipment

Table 5. The business impact criteria

Diagnostic		L	CM	QD	AE	Total Score
processus						
Machining	1					19
	2					
	3					
	4				x	
	5	x	x	x		
Primary assembly	1					12
	2			x		
	3		x		x	
	4	x				
	5					
Assembly of the cabin	1					12
	2			x		
	3		x		x	
	4	x				
	5					
Finishing	1					12
	2			x		
	3		x		x	
	4	x				
	5					

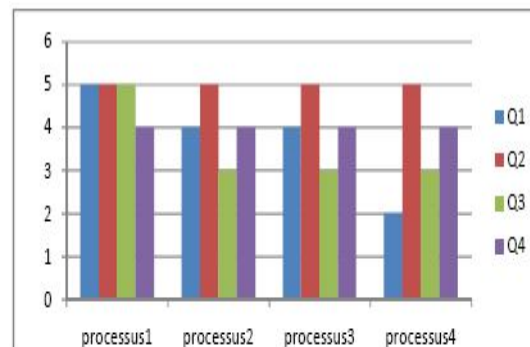


Fig.9 Business impact graphic

Interpretation: it is almost perfect that this process is characterized by a great effect on the investment of change of the machines, the three processes which remains have the same effect but the machining process will be always the choice.

General interpretation

We can see that the machining process has a great importance in the workshop and it will be selected for analysis in order to be improved.

12. Diagnosis of the state of processes

Competitiveness and the company growth in a very competitive sector depends largely on the technology option, the mechanization and automation that are essential to the company.

Our analysis of the weighted approach was chosen as the machining axis change, then we will analyze the situation of that to obtain improvements in the performance of workshop production.

13. Presentation of the new model for the machining process

Our model is a new process that is based on changes in the tools and technologies in the process, and no doubt it brings gains to the different senses, each idea of these changes is a proposal must have a special study committee by special order to be applied, however, we will simply present these ideas:

- Automation in a functional analysis
- Integration of industrial data for the workshops supervision
- Evolution of the process Vision
- Change the mode of transition pieces
- Integration of Robotics

14. Diagram of new machining process

The diagram (fig.10) presents the new process proposed by our approach:

- The implementation of a computer platform using the production management computer and maintenance management computer software
- Automatic transmissions using conveyors such as gravity rollers, powered roller
- Identification of each type of part by kanban (labels)
- Installation of a LAN to ensure excellent communication
- Proposal for Turntable for the movement of parts

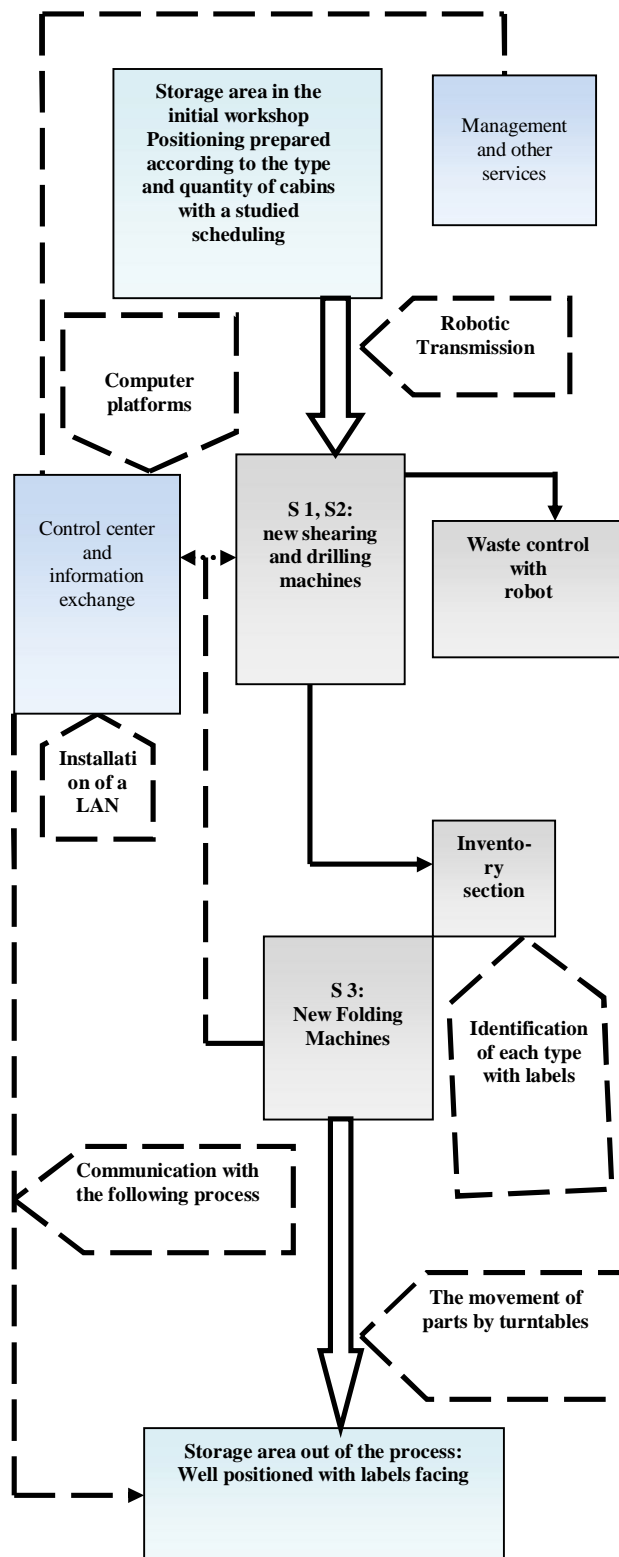


Fig..10. The new process proposed by BPR approach

15. Future research

The paper addressed the effect of reengineering process on improvement and firm performance using ideas from process working

It is believed that the new model is needed to further validate the propositions.

It is recommended that the type of industry, the labor and the customer base taken into account. In addition, the size of the organization should be considered as a complexity criteria..

One of the challenges in conducting research in business process reengineering is to clearly integrated with Six Sigma approach.

With this integration and the exploitation of the Six sigma principles, organizations achieves the goal which is always the improvement of performance and competitiveness

16. CONCLUSIONS

This work has enabled us to both discover the company through one of its strategic functions, and discover one of the important aspects of continuous improvement that is contributing to the improved performance of the company CABAM a new process proposed by our approach.

But training remains the point of success of the BPR approach by the commitment of the company. Therefore, all staff must be integrated in the process decision to ensure continuous improvement of the company

References

[1] C.Hohmann, *Techniques de productivité Comment Gagner Des Points De Performance*, Organisation Editions, 2009.

[2] C.Marxt, F.Hacklin, *Design, product development, innovation: all the same in the end? A short discussion on terminology*, Journal of Engineering Design

[3] H.Brandenburg, J.Wojtyna, *l'approche processus*, Organisation Editions, 2003.

[4] I.B. Hipkin, C.D.Cock, TQM and BPR: lessons for maintenance management, Omega International Journal of Management science, V 28 (2000) 277-292

[5] I.D.Prajogo, A.S.Sohal, The relationship between organization strategy, total quality management (TQM),

and organization performance, European Journal of Operational Research 168 (2006) 35–50

[6] K.J.Fryer, J.Antonny, A.Douglas, Critical success factor of improvement continuous, The TQM magazine, V 19 Issue 5, 497-517

[7] K. V. Scyoc, Process safety improvement: Quality and target zero, Journal of Hazardous Materials 159 (2008) 42–48

[8] M.Hammer, J.Champy, *Le Reengineering, Réinventer L'entreprise Pour Une Amélioration Spectaculaire De Ses Performances*, Dunod Editions, 1993.

[9] M.Parast, The effect of Six Sigma projects on innovation and firm performance, International Journal of Project Management 29 (2011) 45–55

[10] P.Bradley, J. Browne, S. Jackson, H. Jagdev, Business Proces Re-engineering (BPR)--A Study of the Software Tools Currently Available, Computers in Industry, vol. 25

[11] T.Goldsby, R. Martichenko, J. Ross, *lean six sigma logistics : Strategic Development to Operational Success*, Publishing, 2005

[12] Y.Mougin, *Processus : Les Outils d'optimisation de la Performance*, Organisation Editions, 2004.

Aouag Hichem is a Professor assistant at the Industrial engineering department ,Batna University (Algeria), He received the Magister degree in 2006 from the industrial engineering department (Batna university, Algeria), His research interests include Quality Management, Improvement continuous and system management

Mechenene Athmane is a Professor of management at the Industrial engineering department, Batna University, Algeria. His research interests include Quality Management, Production Management and Project management

IJCSI CALL FOR PAPERS JANUARY 2014 ISSUE

Volume 11, Issue 1

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas. See authors guide for manuscript preparation and submission guidelines.

Accepted papers will be published online and indexed by Google Scholar, Cornell's University Library, DBLP, ScientificCommons, CiteSeerX, Bielefeld Academic Search Engine (BASE), SCIRUS, EBSCO, ProQuest and more.

Deadline: 10th December 2013

Online Publication: 31st January 2014

- Evolutionary computation
- Industrial systems
- Evolutionary computation
- Autonomic and autonomous systems
- Bio-technologies
- Knowledge data systems
- Mobile and distance education
- Intelligent techniques, logics, and systems
- Knowledge processing
- Information technologies
- Internet and web technologies
- Digital information processing
- Cognitive science and knowledge agent-based systems
- Mobility and multimedia systems
- Systems performance
- Networking and telecommunications
- Software development and deployment
- Knowledge virtualization
- Systems and networks on the chip
- Context-aware systems
- Networking technologies
- Security in network, systems, and applications
- Knowledge for global defense
- Information Systems [IS]
- IPv6 Today - Technology and deployment
- Modeling
- Optimization
- Complexity
- Natural Language Processing
- Speech Synthesis
- Data Mining

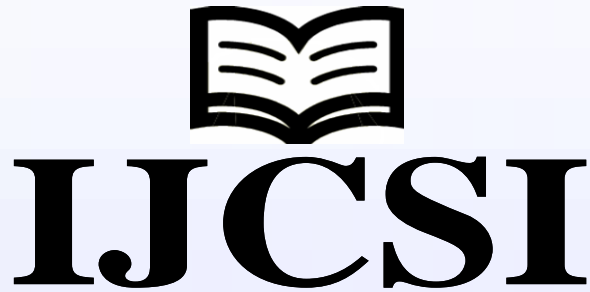
For more topics, please see <http://www.ijcsi.org/call-for-papers.php>

All submitted papers will be judged based on their quality by the technical committee and reviewers. Papers that describe on-going research and experimentation are encouraged. All paper submissions will be handled electronically and detailed instructions on submission procedure are available on IJCSI website (www.IJCSI.org).

For more information, please visit the journal website (www.IJCSI.org)

© IJCSI PUBLICATION 2013

www.IJCSI.org



The International Journal of Computer Science Issues (IJCSI) is a well-established and notable venue for publishing high quality research papers as recognized by various universities and international professional bodies. IJCSI is a refereed open access international journal for publishing scientific papers in all areas of computer science research. The purpose of establishing IJCSI is to provide assistance in the development of science, fast operative publication and storage of materials and results of scientific researches and representation of the scientific conception of the society.

It also provides a venue for researchers, students and professionals to submit ongoing research and developments in these areas. Authors are encouraged to contribute to the journal by submitting articles that illustrate new research results, projects, surveying works and industrial experiences that describe significant advances in field of computer science.

Indexing of IJCSI

1. Google Scholar
2. Bielefeld Academic Search Engine (BASE)
3. CiteSeerX
4. SCIRUS
5. Docstoc
6. Scribd
7. Cornell's University Library
8. SciRate
9. ScientificCommons
10. DBLP
11. EBSCO
12. ProQuest