

IJCSIS Vol. 15 No. 3 Part II, March 2017
ISSN 1947-5500

**International Journal of
Computer Science
& Information Security**

© IJCSIS PUBLICATION 2017
Pennsylvania, USA

Indexed and technically co-sponsored by :



AUTHOR SERIES



IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS

International Journal of Computer Science and Information Security (IJCSIS) January-December 2017 Issues

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 Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.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.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site

Notification: see web site

Revision: see web site

Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
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
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>

arXiv.org

Google scholar

SCIRUS
search engine for science

ScientificCommons

Scribd

.docstoc
find and share professional documents

BASE
Bielefeld Academic Search Engine

CiteSeer^x beta

dblp.uni-trier.de
Computer Science
Bibliography

DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS



ProQuest

For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

Editorial Message from Editorial Board

It is our great pleasure to present the **March 2017 issue** (Volume 15 Number 3) of the **International Journal of Computer Science and Information Security (IJCSIS)**. High quality research, survey & review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and technology. It especially provides a platform for high-caliber academics, practitioners and PhD/Doctoral graduates to publish completed work and latest research outcomes. According to Google Scholar, up to now papers published in IJCSIS have been cited over 9800 times and this journal is experiencing steady and healthy growth. Google statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is indexed in major academic/scientific databases and important repositories, such as: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, Academia.edu and EBSCO among others.

A great journal cannot be made great without a dedicated editorial team of editors and reviewers. On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers for their outstanding efforts to review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status, making sure we deliver high-quality content to our readers in a timely fashion.

"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." We would like to thank you, the authors and readers, the content providers and consumers, who have made this journal the best possible.

For further questions or other suggestions please do not hesitate to contact us at ijcsiseditor@gmail.com.

A complete list of journals can be found at:
<http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 15, No. 3, March 2017 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):



Open Access This Journal is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source.



Bibliographic Information

ISSN: 1947-5500

Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:

IJCSIS Managing Editor

[\(ijcsiseditor@gmail.com\)](mailto:ijcsiseditor@gmail.com)

Pennsylvania, USA

Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

IJCSIS Editorial Board	IJCSIS Guest Editors / Associate Editors
Dr. Shimon K. Modi [Profile] Director of Research BSPA Labs, Purdue University, USA	Dr Riktesh Srivastava [Profile] Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE
Professor Ying Yang, PhD. [Profile] Computer Science Department, Yale University, USA	Dr. Jianguo Ding [Profile] Norwegian University of Science and Technology (NTNU), Norway
Professor Hamid Reza Naji, PhD. [Profile] Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran	Dr. Naseer Alquraishi [Profile] University of Wasit, Iraq
Professor Yong Li, PhD. [Profile] School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China	Dr. Kai Cong [Profile] Intel Corporation, & Computer Science Department, Portland State University, USA
Professor Mokhtar Beldjehem, PhD. [Profile] Sainte-Anne University, Halifax, NS, Canada	Dr. Omar A. Alzubi [Profile] Al-Balqa Applied University (BAU), Jordan
Professor Yousef Farhaoui, PhD. Department of Computer Science, Moulay Ismail University, Morocco	Dr. Jorge A. Ruiz-Vanoye [Profile] Universidad Autónoma del Estado de Morelos, Mexico
Dr. Alex Pappachen James [Profile] Queensland Micro-nanotechnology center, Griffith University, Australia	Prof. Ning Xu, Wuhan University of Technology, China
Professor Sanjay Jasola [Profile] Gautam Buddha University	Dr . Bilal Alatas [Profile] Department of Software Engineering, Firat University, Turkey
Dr. Siddhivinayak Kulkarni [Profile] University of Ballarat, Ballarat, Victoria, Australia	Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Reza Ebrahimi Atani [Profile] University of Guilan, Iran	Dr Venu Kuthadi [Profile] University of Johannesburg, Johannesburg, RSA
Dr. Dong Zhang [Profile] University of Central Florida, USA	Dr. Zhihan Iv [Profile] Chinese Academy of Science, China
Dr. Vahid Esmaeelzadeh [Profile] Iran University of Science and Technology	Prof. Ghulam Qasim [Profile] University of Engineering and Technology, Peshawar, Pakistan
Dr. Jiliang Zhang [Profile] Northeastern University, China	Prof. Dr. Maqbool Uddin Shaikh [Profile] Preston University, Islamabad, Pakistan
Dr. Jacek M. Czerniak [Profile] Casimir the Great University in Bydgoszcz, Poland	Dr. Musa Peker [Profile] Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Binh P. Nguyen [Profile] National University of Singapore	Dr. Wencan Luo [Profile] University of Pittsburgh, US
Professor Seifeidne Kadry [Profile] American University of the Middle East, Kuwait	Dr. Ijaz Ali Shoukat [Profile] King Saud University, Saudi Arabia
Dr. Riccardo Colella [Profile] University of Salento, Italy	Dr. Yilun Shang [Profile] Tongji University, Shanghai, China
Dr. Sedat Akleylek [Profile] Ondokuz Mayıs University, Turkey	Dr. Sachin Kumar [Profile] Indian Institute of Technology (IIT) Roorkee

Dr Basit Shahzad [Profile] King Saud University, Riyadh - Saudi Arabia	
Dr. Sherzod Turaev [Profile] International Islamic University Malaysia	

TABLE OF CONTENTS

1. PaperID 28021701: Load Flow Analysis of 11 kV Test Feeder with and without the Injection of DG (pp. 1-5)

*Umar Farooq, Shahryar Qureshi, Sanauallah Ahmad, IQRA National University (INU) Pakistan.
Dr. Fazal Wahab Karam, COMSATS Institute of Information technology (CIIT) Pakistan.*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

2. PaperID 28021704: Security Issues and Solutions in Wireless Sensor Networks (pp. 6-17)

*Chérif Diallo, Abdoulaye Sawaré, Maimouna Tedy Sow
Laboratoire Algèbre, Cryptographie, Codes et Applications (ACCA),
UFR Sciences appliquées et de Technologies (UFR SAT),
Université Gaston Berger BP 234 Saint-Louis, Sénégal*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

3. PaperID 28021705: A Detailed Approach to The Analysis and Design of Multi Agent System (MAS) using Multi Agent System Engineering (MaSE) Methodology (pp. 18-26)

*Nwagu, Chikezie Kenneth, Omankwu, Obinnaya Chinecherem and Inyama, Hycient
1 Mantrac Nigeria Limited, Lagos, Lagos, Nigeria,
2 Computer Science Department, Michael Okpara University of Agriculture, Umudike, Umuahia, Abia State,
Nigeria
3 Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

4. PaperID 28021706: Design of Robust Algorithm for Verification System Based on Fingerprint and Speech (pp. 27-38)

*Dr. Amit Verma (1), Seema Rani (2), Amandeep kaur (3)
(1) Professor and Head of Department, Computer Science & Engineering, Chandigarh Engineering College,
Landran, Punjab, India
(2) M. Tech. Research Scholar, Computer Science & Engineering, Chandigarh Engineering College, Landran,
Punjab,
(3) Associate Professor, Computer Science & Engineering, Chandigarh Engineering College, Landran, Punjab*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

5. PaperID 28021709: Development and Implementation of Arduino based Vehicle Monitoring & Tracking System (pp. 39-46)

*Fathima Jabeen A., Dr. B. Ramamurthy, Dr. N. Anju Latha
S K University, Anantapur, Andhra Pradesh*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

6. PaperID 28021710: Cloud Computing and Homomorphic Encryption (pp. 47-55)

Amit Verma (1), Ramandeep Brar (2), Amandeep Ummat (3)

(1) Professor and Head of Department, Computer Science & Engineering, Chandigarh Engineering College, Landran, Punjab, India

(2) M.Tech. Research Scholar, Computer Science & Engineering, Chandigarh Engineering College, Landran, Punjab,

(3) Associate Professor, Computer Science & Engineering, Chandigarh Engineering College, Landran, Punjab

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

7. PaperID 28021711: Comparison of SNR and SIR Based Fading Models in Determination of Normalized Channel Capacity of Cooperative Cognitive Radio Network (pp. 56-61)

Md Abul Kalam Azad, Associate Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342

Abu Sayed Md Mostafizur Rahaman, Associate Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342

Md. Imdadul Islam, Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342

Jugal K. Das, Professor, Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

8. PaperID 28021714: Hybrid Cloud: Multi-factor Authentication, a Solution to Hybrid Cloud in an Enterprise Environment (pp. 62-65)

Dieudonne Boupo Boupo, Computer Sciences and Technology, Beijing University of Technology, Beijing, China

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

9. PaperID 28021716: Design and Development of Robust Algorithm for Cryptography using Improved AES Technique (pp. 66-82)

Amit Verma (1), Simarpreet Kaur (2), Bharti Chhabra (3)*

(1) Professor and Head of Department, Computer Science & Engineering, Chandigarh Engineering College, Landran, Punjab, India.*

(2) M. Tech. Research Scholar, Computer Science & Engineering, Chandigarh Engineering College

(3) Assistant Professor, Computer Science & Engineering, Chandigarh Engineering College, Landran, Punjab, India.

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

10. PaperID 28021717: The Mutual Authentication Scheme to Detect Virtual Side Channel Attack in Cloud Computing (pp. 83-98)

Amit Verma 1, Megha Mittal (2), Bharti Chhabra (3),*

(1) Professor and Head of Department and Professor, Department of computer Science, CEC Landran, Mohali, Punjab.*

(2) Mtech Research Scholar, Department of computer Science, CEC Landran, Mohali, Punjab,

(3) Assistant Professor, Department of computer Science, CEC Landran, Mohali, Punjab

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

11. PaperID 28021722: Web Page Recommendations based on User Session Graph (pp. 99-104)

Sejal D, Vijay K M & Venugopal K R

Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore-560001

S S Iyengar, Florida International University, USA

L M Patnaik, INSA Senior Scientist, National Institute of Advanced Studies, IISc Campus, Bangalore

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

12. PaperID 28021723: Extraction of Exudates from Digital Retinal Images (pp. 105-120)

(1) M. R. N. Tagore, (2) Dr. E. V. Krishna Rao, (3) Dr. B. Prabhakar Rao

(1) Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, A.P, India;

(2,3) Lakireddy Bali Reddy College of Engineering, Mylavaram, A.P, India;

JNTU, Kakinada, A.P, India.

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

13. PaperID 28021725: Semantic Information Retrieval Based on Social Annotation (pp. 121-138)

E. ElSayed, A. ElKorany, A. Salah, H. Hafny

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

14. PaperID 28021734: TOAD-Based Reversible Gates for All-Optical Reed-Muller Techniques (pp. 139-153)

Sandip Mandal (1), Supriti Samanta (2), Goutam Kumar Maity (3) & Subhadipta Mukhopadhyay (4)*

(1) Raja N.L.K. Women's college, Paschim Medinipur, West Bengal, 721102, India

(2) Chaipat Girls' High School, Paschim Medinipur, West Bengal, 721148, India

*(*3) Netaji Subhash Engineering College, Techno City, Garia, Kolkata-700152, India*

(4) Department of Physics, Jadavpur University, Kolkata - 700032, India

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

15. PaperID 28021735: Detecting Honest Behavior for E-auction Bidders Using Intelligent Agent (pp. 154-167)

Laila Moharram Mohammed, Ali H. El.Bastawissy, Mahmoud M. Abd Elatif

Department of Information System, Faculty of Computers and Information, Mansoura University, Egypt

Faculty of Computer Science, MSA University Giza, Egypt

Department of Information System, Faculty of Computers and Information, Helwan University, Egypt

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

16. PaperID 28021738: Mobile Volunteered Geographic Information System for Real Time Reporting Egyptian Environmental Violations (pp. 168-180)

Gehad Abdallah (1), Mahmoud Abd Ellatif (2), Mohammed Nor El-Dien (3)
(1) Master student at Faculty of Computer Sciences and information systems, Information Systems Department, Mansoura University, Mansoura, Egypt.
(2) Professor at Department of Information System, Faculty of Computers and Information, Helwan University, Egypt.
(3) Associated Professor at Faculty of Computers and Information, Cairo University. Associated Professor at International College CIC.

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

17. PaperID 28021742: Design of Low Cost Underwater Network (pp. 181-190)

Md. Masud Hasan (1); Mohammad Badrul Alam Miah (2); Dr. Sajjad Waheed (2); Md. Anwar Hossain (3); Nargis Akter (4)
(1) Department of Information and Communication Technology, MBSTU and Assistant Engineer, National Electro-Medical Equipment Maintenance Workshop & Training Center, Ministry of Health & Family Welfare. Govt. of Bangladesh;
(2) Department of Information and Communication Technology, MBSTU,
(3) National Electro-Medical Equipment Maintenance Workshop & Training Center, Ministry of Health & Family Welfare. Govt. of Bangladesh;
(4) Department of Computer Science and Engineering, Asian University of Bangladesh.

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

18. PaperID 28021745: 3D Model Based Approach for Data Visualization (pp. 191-196)

K. Rajiv, Department of Information and Technology, Sridevi Women's Engineering College, Hyderabad, India
G. Ramesh Chandra, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engg. & Tech., Hyderabad, India
B. Basaveswara Rao, Computer Centre, Acharya Nagajuna University, Guntur-522510, A.P, India.

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

19. PaperID 28021746: Reliability Analysis of Distribution System using ETAP (pp. 197-201)

Sanaullah Ahmad, Department of Electrical Engineering, IQRA National University, Peshawar, Pakistan
Sana Sardar, Department of Electrical Engineering, IQRA National University, Peshawar, Pakistan
Azzam ul Asar, Department of Electrical Engineering, CECOS University, Peshawar, Pakistan
Fazal Wahab Karam, Department of Electrical Engineering, COMSATS Institute of Information technology (CIIT)

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

20. PaperID 28021748: Analysing the Research Ranking of Universities in Arabic Countries (pp. 202-211)

Radwa El Shawi, Princess Nourah Bint Abdulrahman University, KSA

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

21. PaperID 28021749: Pattern Taxonomy Deploying Model for Text Document Classification (pp. 212-218)

S. Brindha (1), Dr. S. Sukumaran (2)

(1) *Department of Computer Science Erode Arts and Science College, Erode, Tamilnadu, India.*
(2) *Department of Computer Science, Erode Arts and Science College, Erode, Tamilnadu, India.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

22. PaperID 28021753: A Comprehensive Survey on Techniques for Facial Emotion Recognition (pp. 219-224)

Renuka Deshmukh, M. E. Scholar, MIT, Pune
Vandana Jagtap, Professor, MIT, Pune

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

23. PaperID 28021755: An Authenticated Threshold Visual Cryptography Scheme with Perfect Reconstruction and Meaningful Shares (pp. 225-243)

R. Sathishkumar (1) & Gnanou Florence Sudha (2)
(1) *Dept of Electronics & Communication, Perunthalaivar Kamarajar Institute of Engg. & Tech., Puducherry, India*
(2) *Dept. of Electronics & Communication, Pondicherry Engineering College, Puducherry, India*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

24. PaperID 28021759: Detection of Cyberbullying on Social Media using Data Mining Techniques (pp. 244-250)

Hariani, Department of Informatics Engineering, Islamic University of Indonesia, Yogyakarta, Indonesia
Imam Riadi, Department of Information System, Ahmad Dahlan University, Yogyakarta, Indonesia

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

25. PaperID 28021760: New Approach for Text Mining of Arabic on the Web (pp. 251-257)

Eng. Nehad M. Abdel Rahman Ibrahim, Computers & Systems Dep., Faculty of Engineering AL AZHAR UNIVERSITY, EGYPT- CAIRO
Prof. Ass. Alaa Hammouda, Computers & Systems Dep., Faculty of Engineering, AL AZHAR UNIVERSITY, EGYPT- CAIRO
Prof. Sayed Abdel Hady Nouh, Computers & Systems Dep., Faculty of Engineering, AL AZHAR UNIVERSITY, EGYPT- CAIRO

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

26. PaperID 28021761: An Integrated Mobile-Based Complaints System (pp. 258-262)

Gamal H. El-Adl, Information Systems Department, Faculty of Computers and Information Sciences, Mansoura University, Egypt

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

27. PaperID 28021762: A Secure Document Image Encryption Using mixed Chaotic System (pp. 263-270)

Revanna C. R., Dept of ECE, Govt. Engg College, Ramanagara, Karnataka, India

Dr. C. Keshavamurthy, YD Institute of Technology, Kanakapura Road, Bangalore, India

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

28. PaperID 28021763: Parametric Optimization of Zigbee Parameter Using Soft Computing Techniques (pp. 271-279)

*Prativa P. Saraswala, V.T Patel Dept. of Electronics and Communication, Charusat, Changa, Gujarat, India
Dr. Jaymin Bhalani, Dept. of Electronics and Communication, Babaria Institute of Technology, Vadodara, Gujarat, India*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

29. PaperID 28021764: Forensic Analysis of Unmanned Aerial Vehicle to Obtain GPS Log Data as Digital Evidence (pp. 280-285)

*Septyan Eka Prastya, Departement of Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia
Imam Riadi, Department of Information System, Ahmad Dahlan University, Yogyakarta, Indonesia
Ahmad Luthfi, Departement of Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

30. PaperID 28021765: Business Process Modeling: Blueprinting (pp. 286-291)

Sabah Al-Fedaghi, Computer Engineering Department, Kuwait University, Kuwait

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

31. PaperID 28021769: ZigBee Nodes' EMI Immunity by Automatic Channel Selection Algorithm Paper (pp. 292-296)

*Ahmad Habboush, Computer Science Dept., Faculty of IT, Jerash University
Mohammad Alnabhan, Computer Science Dept., Faculty of IT, Mutah University
Mohammad Otoum, Computer Science Dept., Faculty of IT, Irbid National University*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

32. PaperID 28021771: Machine Learning based Systems for Organizational Support – The State of the Art (pp. 297-308)

*Gufran Ahmad Ansari, Dept. of Information Technology, Qassim University, Qassim, Saudi Arabia
Mohammad Tanvir Parvez, Dept. of Computer Engineering, Qassim University, Qassim, Saudi Arabia
Ali Al Khalifah, Dept. of Information Technology, Qassim University, Qassim, Saudi Arabia*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

33. PaperID 28021772: Social Trends: The Theory, Research and Sociality Discerned Through Twitter (pp. 309-317)

Jasmin D. Niguidula & Angelique D. Lacasandile

Technological Institute of the Philippines, Manila / National University - Manila

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

34. PaperID 28021775: Simulation of Artificial Life Modeling with Modified Biogenesis Model in Virtual Ecosystem (pp. 318-325)

*Md. Shahriare Satu (1), Mohammad Abu Yousuf (2)
(1,2) Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh*

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

35. PaperID 28021776: Employing Mean Opinion Score of Audio Lossy Compression Algorithms in VoIP Application (pp. 326-331)

*Ma. Emmie T. Delluza, Technological Institute of the Philippines-Manila/Catanduanes State University
Jasmin D. Niguidula, Technological Institute of the Philippines, Manila, Philippines*

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

36. PaperID 28021777: Performance Evolution of QoS in VoIP Using SIP for Ad hoc Routing Protocols (pp. 332-337)

*Ahmad Jahangir, Department of Computer Sciences, Pir Mehr Ali Shah, Arid Agriculture University, Rawalpindi Pakistan
Muhammad Usman Nasir, Department of Computer Sciences, Pir Mehr Ali Shah, Arid Agriculture University, Rawalpindi, Pakistan
Yawar Abbas Abid, Department of Computer Sciences, COMSATS institute of Information Technology, Sahiwal, Pakistan*

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

37. PaperID 28021780: Binary Search Optimization - Implementation and Amortized Analysis for Splitting the Binary Tree (pp. 338-341)

Dhaval Kadia, Department of Computer Science & Engineering, Faculty of Technology and Engineering, The Maharaja Sayajirao, University of Baroda, Vadodara, India

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

38. PaperID 28021781: Key Selective Secure Privacy Framework for User Management and to Improve Performance Parameters (pp. 342-352)

*Sandeep Kapur, Guru Kashi University, Talwandi Saboo, India
Dinesh Kumar, Associate Professor, Guru Kashi University, Talwandi Saboo, India*

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

39. PaperID 28021783: Secure Metadata based Search over Encrypted Cloud Data Supporting Similarity Ranking (pp. 353-361)

Mahreen Saleem, A.M.U, Department of Computer Engineering, Aligarh,UP, India
M. R. Warsi, A.M.U, Department of Computer Engineering, Aligarh,UP, India
Neelam Saleem Khan, NIT Srinagar, Department of IT, Srinagar,J&K, India

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

40. PaperID 28021786: Rola: An Equi-Matrik Chhand of Hindi Poems (pp. 362-364)

Kapil Kumar Kushwah, School of Electronics, Devi Ahilya Vishwa Vidhyalaya, Indore, India
Brijendra Kumar Joshi, Military College of Telecommunication Engineering, Mhow, India

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

41. PaperID 28021790: A Hybrid Approach for Compression of MPEG Videos (pp. 365-371)

Kanupriya, CSE, Chandigarh University, Mohali, Punjab, India.
Silky Narang, CSE, Chandigarh University, Mohali, Punjab, India.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

42. PaperID 28021792: Comparative Performance Analysis of Ad hoc Routing Protocol (pp. 372-378)

Muhammad Usman Nasir, Department of Computer Sciences, Pir Mehr Ali Shah, Arid Agriculture University, Rawalpindi, Pakistan.
Yawar Abbas Abid, Department of Computer Sciences, COMSATS Institute of Information Technology, Sahiwal, Pakistan.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

43. PaperID 28021793: A Comparative Evaluation Concept for Agent Oriented Methodologies (pp. 379-391)

Omankwu Obinnaya Chinecherem (1), Nwagu Chikezie Kenneth (2), Inyama Hycient (3)
(1) Computer Science Department, Michael Okpara University of Agriculture, Umudike, Umuahia, Abia State, Nigeria
(2) Nwagu Chikezie Kenneth, Computer Science Department, Nnamdi Azikiwe University, Awka, Anambra, Nigeria.
(3) Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

44. PaperID 280217100: Constructing a Predictive Model for an Intelligent Network Intrusion Detection (pp. 392-397)

Alebachew Chiche (1), Million Meshesha (2)
(1) Mizan Tepi University, School of Computing and Informatics, Tepi, Ethiopia
(2) Addis Ababa University, School of Information Science, Addis Ababa, Ethiopia

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

45. PaperID 280217101: Fine Tuning of TCP/IP for Ad Hoc Networks (pp. 398-403)

Sabahat Zahra, Department of Computer Sciences, Pir Mehr Ali Shah, Arid Agriculture University, Rawalpindi Pakistan.

Muhammad Usman Nasir, Department of Computer Sciences, Pir Mehr Ali Shah, Arid Agriculture University, Rawalpindi, Pakistan.

Yawar abbas abid, Department of Computer Sciences, COMSATS institute of Information Technology, Sahiwal, Pakistan.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

46. PaperID 31101645: Critical Issues and Challenges in Developing Mobile-Based Health Systems: Case of Tanzania (pp. 404-412)

Frederick Henri Chali, School of CoCSE, Nelson Mandela African Institution of Science and Technology, NM-AIST, Arusha, Tanzania

Eng. Dr. Zaipuna O. Yonah, School of CoCSE, Nelson Mandela African Institution of Science and Technology, NM-AIST, Arusha, Tanzania

Dr. Khamis Kalegele, Commission for Science and Technology, COSTECH, Dar es Salaam, Tanzania

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

47. PaperID 31121612: Perceptions of crime behavior and Relationships: Rough Set Based Approach (pp. 413-420)

Luai Al Shalabi, Arab Open University, Kuwait city, Kuwait

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

48. PaperID 30091531: VoIP Security: Common Attacks and their Countermeasures (pp. 421-428)

Abdullahi Mohammed Jingi, Department of Computer Science, Adamawa State University (ADSU), Mubi Mubi, Adamawa State, Nigeria.

Murtala Muhammad, Department of Computer Science, Adamawa State University (ADSU), Mubi Mubi, Adamawa State, Nigeria.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

49. PaperID 280217106: Improving Root Finding in Arabic Stemmers (pp. 429-433)

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

50. PaperID 310816197: Challenges That Restrict the Efficiencies of Security Frameworks in E-Commerce: A Review (pp. 434-439)

Kenneth Longo Mlelwa, PhD Cand., School of CoCSE, NM-AIST, Arusha, Tanzania

Eng. Dr. Zaipuna O. Yonah, Senior Lecturer, School of COCSE, NM-AIST, Arusha, Tanzania

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

51. PaperID 28021744: Pressure Oscillations in the Water Distribution Network before and behind the Solenoid Valve (pp. 440-452)

Saeid Jafari, Ali Ghaffari, Department of computer engineering, Tabriz branch, Islamic azad university, Tabriz, Iran

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

An Integrated Mobile-Based Complaints System

Gamal H. Eladl

Information Systems Department - Faculty of Computers and Information Sciences

Mansoura University, Mansoura, Egypt, 35516

gamalhelmyis@gmail.com, gamalhelmy@mans.edu.eg

Abstract -In the last decade; mobile devices and Smartphone have become a major part of our lives because of its countless advantages. These advantages are seen in Information and Communication Technologies (ICTs) such as data-sharing, uploading, downloading a lot of gigabytes, etc. On the other hand, what happened if we use one of these technologies; e.g. Mobile Technology (MT) with an integrated service such as Service Oriented Architecture (SOA) for developing any complaint system. Although, there are traditional means to handle customers 'complaints without moving toward the application of the recent technologies. Therefore, the major work of a complaint system is to find a solution to handle complaints. In this paper; a proposed mobile-based complaint system that handles complaints by using the advantages of SOA and also depending on a proposed Complaint Management Life Cycle (CMLC). The integrated system will have many advantages such as reducing the waiting time for uploading, receiving a complaint, guarantee a fast solution, make complaints easier to monitor, track until achieving customer satisfaction and preserve customer loyalty.

Keywords: ICT, MT, SOA, Complaints System, CMLC

1. INTRODUCTION

There is no company without complaints. Customers evaluate any institution according to the way of handling their complaints. Most of Middle East companies ignore the role of the complaints system inside its policies. There are many definitions of a complaints system. One of them is not only the system of communication of raising a problem for the organization owners or decision makers but also it is a discussion of unusual work or actions. In this context, a complaints system is defined as a set of procedures that is used to address complaints and resolve disputes or problems [1]. Companies should move towards applying the new development trend of complaint management such as the full automation system in order to preserve the customer's loyalty. Moreover, the innovative means of complaint handling will achieve one of the important company's goals which are customer's satisfaction. It appears that, ICT has a positive impact on improving life quality. Therefore, it required human to do any activities by optimizing the available up-to-date technologies that have a rapid development such as mobile technology. In this paper, the Mobile Technology will be used in handling complaints systems.

As we know that mobile application is becoming the most popular platform for doing business. These resulted from mobile popularity, flexibility. Furthermore; the Smartphone becomes equipped with a greatly reduced cost hardware and software technologies such as Bluetooth, digital camera, Wi-Fi network, GPS, etc. So, the wide use of mobile devices makes the Smartphone is the better choice for accessing point of a logical and conventional business application. A growing number of mobile employees want to make business decisions using their own cell phones. Moreover, a mobile-based system needs a platform in order to integrate different services and other client applications. This is will appear in the below section which focuses on the role of SOA that represent different system functionality and independent from underlying platforms [2, 3]. SOA will be used here as an interoperability framework in order to integrate heterogeneous systems.

The content of this paper is organized as follows: Section 2 reviews the related work on complaints system and section 2.1 summarizes the proposed CMLC phases. Section 3 introduces the importance of Mobile-based application usage in the complaints system. Section 4 presents the benefits of using SOA. The proposed complaints system and its major function are discussed in section 5. A discussion on the proposed system is presented in section 6. Finally, the conclusions are drawn in section 7.

2. RELATED WORK

There is no one hundred percent perfect body or company. Mistakes can occur from any person or service provider at any time. Complaints are a valuable source of information about how and where mistakes have occurred and can pinpoint weaknesses in the system. The matter that companies should have clear instruction responds to such incidents. It uses an effective response to fix the mistake in order to reduce the likelihood of incidents occurring again. There are many ways of complaints system representation such as customer call center, manual complaint box, or semi-automated complaints system which may be only the organization's email. There are different complaints systems. Its

methods focus on receiving complaints from complainants by any mean. The existing means of a complaints management handling are summarized as follow:

- 1- Online Complaints form/template
- 2- Telephone /Customer Call Center
- 3- E-Mail (e.g. info@company.com or any other mailbox).
- 4- The Customer Complaint form (paper-based form).
- 5- Face-to-face interaction (verbally).

Many organizations fail to integrate the essential components of an effective complaints management system, leading to lost opportunities for improvement and continued customer satisfaction.

2.1 A PROPOSED CMLC PHASES

After looking at many studies on complaints management system [1, 7], it basically works, focuses on one issue; the complaints system extracts communicative actions and their subjects and classifies a complaint as valid or invalid (submitted due to a customer misunderstanding, bad mood or other unrelated issues) in order to resolve it. A summarization of the most important phases are depicted on fig. 1 that represents the proposed life cycle of most complaint management steps that will be integrated by using SOA with a mobile-based application.

3. MOBILE-BASED APPLICATION USAGE

Basically, Mobile technology gives an opportunity to variety users in doing his activities without worrying about its physical connections or technical details. Generally, there is a general structure of the mobile technology; which is composed of three major components; hardware, software (mobile-based application) and communication. Figure 2 shows the three-tier architecture of the mobile-based system which is (mobile-client application, mobile-middleware and enterprise application) respectively. Whereas, mobile users access the enterprise applications (e.g. CRM, BI, ERP, Complaints System) via the connection of the network operators that allow mobile devices to establish and control the internet-connection by using the internet services providers (ISPs). The mobile network operators can provide services to mobile users such as accessing data that stored in databases and accessing different web services (e.g., web-server, enterprise application, and database servers). There are different kinds of mobile clients. The first; is a fat client application that is able to accept user input and store data in a local database until connectivity with the server is re-established and the data can be moved to the server. The second; is a thin client that haven't custom application code and completely rely on the server-tier for their functionality. The third; is a thick client that provides users with more features, graphics choices making the mobile-based application more customizable. The fourth; is a smart client that is internet-connected, allows the mobile user to interact with the server-based application through the use of web services (middle tier)[4-6].

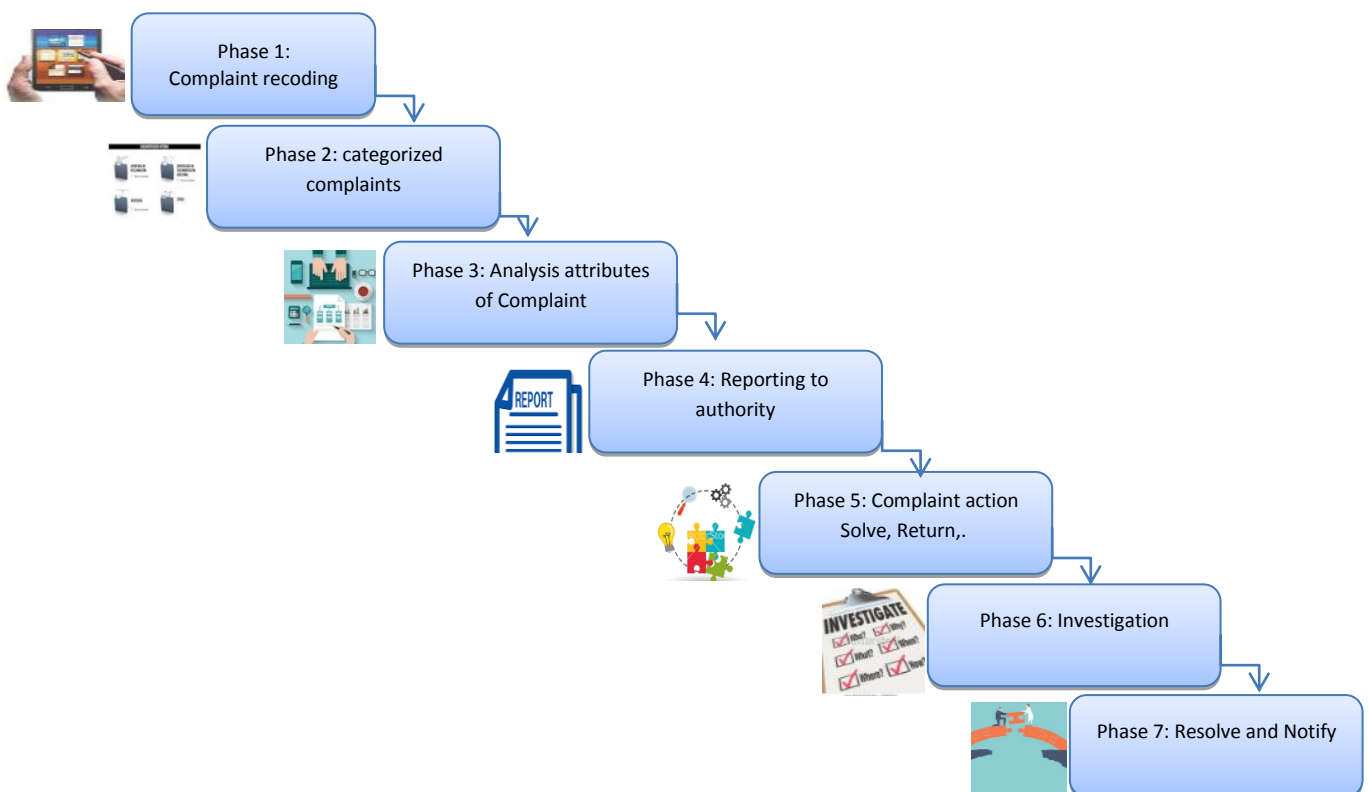


Figure 1. A proposed Complaint Management Life Cycle

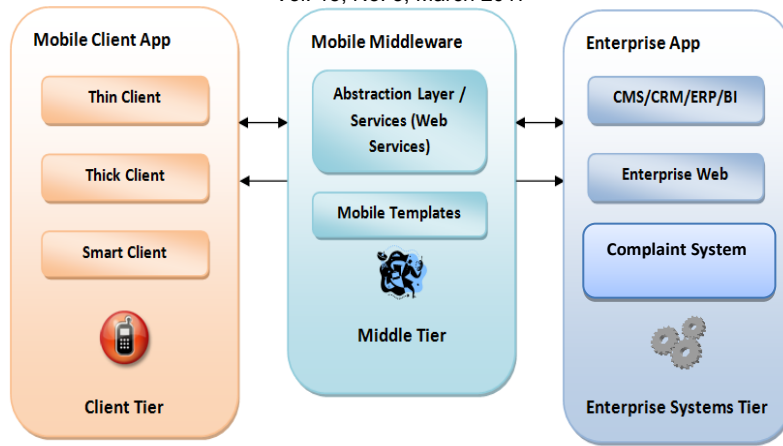


Figure 2. Three tier architecture of mobile-based application

Mobile-based application has different advantages such as providing anywhere, anytime availability, make easy to use and gives flexibility. It is frequently used in all our life activities. The complaints management system's activities will be performed in less amount of time, without bothering the time and contact limitations, with more efficient and effective manner and with the advantage of transparency, fast connectivity and flexibility of mobile-based application in order to find fast solutions.

4. WHY SOA? BENEFITS AND ADVANTAGES

SOA means an integration of different heterogeneous operating systems, software systems, and application infrastructures by using the SOA model. Figure 3 represents the SOA model which based on three roles (publish, find and bind) services. SOA's benefits are for example to establish and maintain federated data resources by using interoperability languages such as XML and RDF and SOAP services [5, 6].

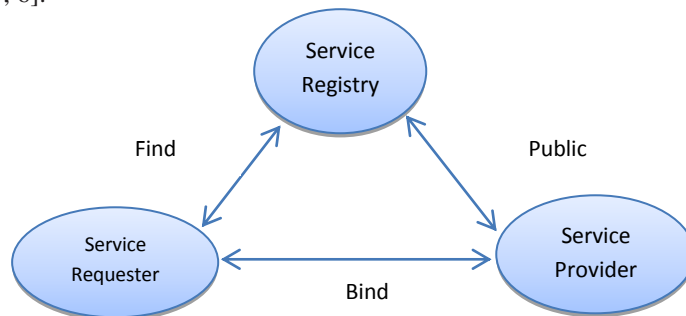


Figure 3. SOA Service Model

On the same context, SOA model for Complaints system is proposed in the figure (4) in order to integrate between complaint system's services and its mobile-based application .It represents the relationship between complainants as a service requester, complaint side /company as a service provider and complaint service as a service registry. More than two studies [7,8] proved that there is a robust technical relationship between SOA and Mobile-based applications.

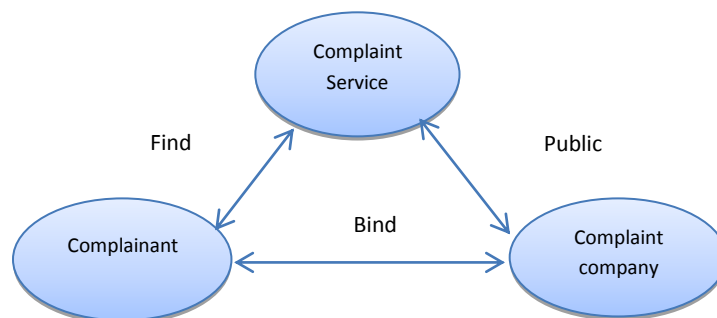


Figure 4. SOA model for Complaints system

We can summarize the benefits of SOA in these points as following:

- 1- Business-centric

- 2- High-degree of distribution
- 3- Loose-coupled of services
- 4- Utilize the service reusability
- 5- Rapid flexibility

5. SYSTEM ARCHITECTURE PERSPECTIVE CONSIDERATION

A workflow of the proposed system is showed in figure (5) that represent the pattern of business activity either processes or services. The complainant journey starts when he registers a complaint, the interaction is logged, and then a notification is sent to the appropriate/corresponding department for follow up and taking action. If action on the complaint is not documented, an SMS notification is generated to the complainant; a reminder is automatically generated and ensuring that the task is resolved. Therefore, the process of following-up complaint still monitors until reach a satisfactory solution to the complainant. At the end, the proposed system will be changed the status of the complaint by giving an alert for closing the complaint.

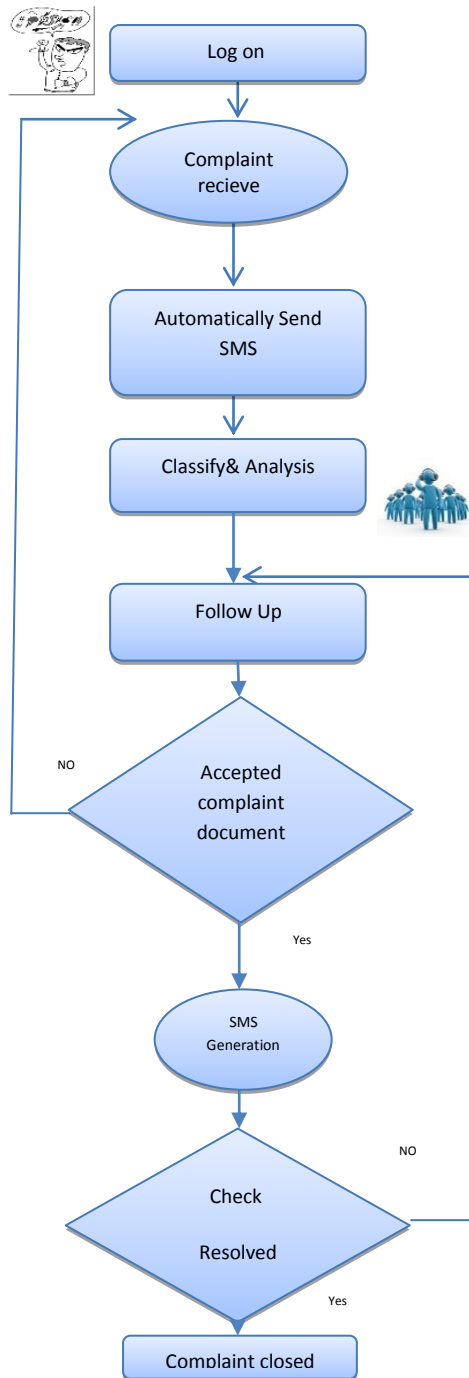


Figure 5. A flow chart for the proposed system

Furthermore, the GUI of the proposed complaints system is depicted on figure 6. It shows the process of registering a new complaint and checking the complaint's status as a demo version. Therefore, functions of the proposed complaints system are summarized as the following:

1. Allows the complainant to create/register a new complaint
2. Upload/send a complainant documents such as (text file, picture, video, etc.)
3. Determine the spatial location of the complaint by using GPS
4. Administration panel to follow up the complaint status
5. Send SMS alert for every action
6. Navigate and explore the complaint data anytime
7. Update and change the complaint anytime
8. Ensure the complaint privacy and confidential

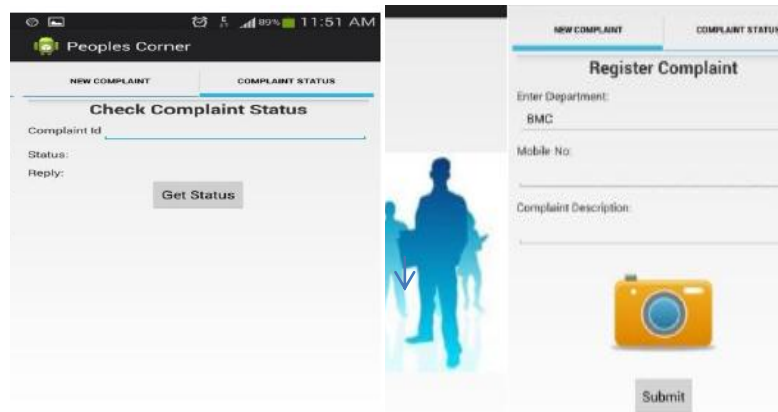


Figure 6. A demo GUI for the proposed complaints system

6. ADVANTAGES OF THE PROPOSED SYSTEM

After applying the proposed complaints system, it will be lead to achieve more than features. For example, complaints staff responsibilities are clearly outlined, the complaint's documents/data are never lost or forgotten and online available. Moreover, the proposed system will promising minimizing time, quick information assembling and gives a quick response. Moreover, companies will gain more than one benefit such as improved customer satisfaction and increased customer loyalty.

7. CONCLUSION

There is an urgent need to apply the proposed system for all companies that treat with customers. X-customers will return to the company after applying the proposed system. Any company will achieve its major objective to increase the productivity ratio. The integrated system will allows different users to access it anytime from anyplace in order to achieve one goal; "User satisfaction". A Complaint Management Life Cycle is presented in details to summarize the overall functions that will be mobile-based application. The integrated system will achieves important target such as reducing the waiting time for uploading, receiving a complaint, guarantee a fast solution. The advice now for all customer-dependent companies to apply this integrated system.

REFERENCES

- [1] Compliant system definition on line available wikipedia.org last accessed 1/12/2015
- [2] Albina Shapiro. "Service-oriented architecture for mobile applications", Proceedings of the 1st international workshop on Software architectures and mobility - SAM 08 SAM 08,2008
- [3] Yi Zheng. "Service-Oriented Mobile Applications for Ad-Hoc networks", 2008 IEEE International Conference on Services Computing, 07/2008
- [4] The Differences Between Thick, Thin & Smart Client By Vangie Beal online available http://www.webopedia.com/DidYouKnow/Hardware_Software/thin_client_applications.asp lastaccessed27-jan-2016.
- [5] V.lee and et.al, "Mobile-application Architecture Design and Development" online available www.informit.com last accessed 1-Feb-2016
- [6] José María. " SOA and its usage in mobile services" online available athttps://www.snet.tu-berlin.de/fileadmin/fg220/courses/SS12/osn-project/soa-in-mobile-services_toro.pdflast accessed 1-Jul.-2016
- [7] Farhan Shafiq et.al , " Solving Real-Life Problems: Future Mobile Technology " Mehran University Research Journal of Engineering & Technology, Volume 35, No. 3, July, 2016

A Secure Document Image Encryption Using mixed Chaotic System

Revanna C R
Assistant Professor
Dept of ECE
Govt. Engg College
Ramanagara, Karnataka, India
E-mail: revannacr2008@gmail.com

Dr. C Keshavamurthy
Former Principal
YD Institute of Technology
Kanakapura Road,
Bangalore, India
E-mail: Keshava_uvce@yahoo.com

Abstract

Providing security for the multimedia documents is of prime important in today's world, as there is rapid growth in internet and multimedia technology. This paper proposes a secure document image encryption scheme based on multi chaotic systems, which aims at fulfilling the purpose with the combination of the traditional bit stream ciphers technology and the spatial-domain encryption of digital image. This strengthens the security mechanism of the encryption process effectively instead of encrypting an image using a single chaotic signal directly. The proposed scheme combines two different chaotic sequences to provide higher security for multi-system. The first chaotic system used is the logistic map which is used to generate a chaotic sequence. Then using a threshold function this chaotic sequence is transformed into a binary stream. The second chaotic system used is the Henon map which is used to build a permutation matrix. First, using the stream of bits as a key stream, the pixel values of a plain image are modified randomly. Second, the altered image is encrypted again by using the permutation matrix. Detailed

Analysis of the encryption method depicts that the simulation results shown in the proposed scheme has a better efficiency than other existing methods and provides high security performance.

Keywords:

Image encryption; chaotic system; multi-chaotic sequence; Logistic map, Henon map

I. INTRODUCTION

Today advancing internet and multimedia technology is more evolved than before. Data containing image, video, audio and other information, is being transmitted over the internet. Images are being widely used in daily life. The application of image processing mainly includes documents,

military photographs and financial records etc., to interchange all kinds of data. Recently, the problem of image security has attracted increased attention.

Due to some intrinsic feature parameters of image, such as bulk information capacity and high correlation co efficient among the pixels, conventional encryption algorithms have however been found not appropriate in cases of image encryption. Further, because of this problem single chaotic system also fails to encrypt the image effectively. Currently, there are many image encryption algorithms such as affine transformation, Tan gram algorithm, magic cube transformation, Arnold map, and Baker's transformation [1-5] etc. In few algorithms effective separation of algorithm and secret-key is not possible. Because of this, the existing algorithms fail to satisfy the needs of modern cryptographic mechanism.

In order to overcome the above mentioned shortcomings, an image encryption technology based on multi chaotic systems has been proposed.

Chen Wei-bin et al. [6] presented a new image encryption algorithm based on the combination of Arnold cat map and Henon chaotic system in order to meet the requirements of secure image transfer. First, the positions of the pixels are shuffled using Arnold cat map method. Second, Henon's chaotic system is used to encrypt the shuffled-image based on pixel by pixel. Ali Soleymani et al. [7] proposed a novel encryption scheme using Arnold cat and Henon chaotic maps. The proposed scheme performs bit and pixel-level permutations on plain and secret images using Arnold cat map, while Henon map creates secret images and specific parameters for the permutations. The method yields a result with the entropy value equal to 7.997 and UACI 29.84. The multchaos-based image

encryption scheme proposed by Brindha Murugan et al. [8] Using Lorenz equation and Henon map with multiple levels of diffusion yields the entropy of 7.999 and the UACI of 33.44. Xingyuan Wang et al. [9] proposed image encryption algorithm based on DNA encoding rules and Chaos. Logistic Map and Piecewise Linear Chaotic Map (PWLCM) are applied to generate all parameters. Wang Yujing et al. [12] Presents Fast Fourier Transform (FFT) based mixed high dimensional chaotic system for image encryption scheme. Fadia TALEB [13] proposed a color image encryption algorithm by using the concept of confusion and diffusion. The author used One-dimensional chaotic map called logistic maps to generate the chaotic sequence. Hang, G., et al. [14] proposed a method in which the permutation box P-box is generated which is used to scramble the pixel positions totally to increase the encryption level. Teng, L, et al. [15] used a bit-level encryption of image using the algorithm which reduces execution time in the size of data during encryption and decryption. Wang, X., et al. [16] proposed a method where the confusion process is obtained by the Game of Life permutation using logistic map and diffusion is achieved using piecewise linear chaotic map (PWLCM) chaotic system. Wang, X., et al. [17] proposed an improved cat map method and dynamic random growth technique for fast image encryption. Both key and plain image contributes for pixel diffusion process. Chen, J.X et al in [18] presented, dynamic state variables which assigns a chaotic variables for pixel encryption dynamically. Boriga, R., et al. [19] proposed an image encryption consists of three 2D chaotic maps and the pixels are shuffled by using random permutation matrix chaotic map. Zheng, Y, et al. [20] presented an encryption scheme in which the pixel permutation is obtained by using Henon map and pixel shuffling is achieved using compound spatio-temporal chaos. Kumar, G., in [21] the image encryption is achieved using 3D chaotic dissipative based on Lorenz attractor. Lorenz attractor is used for image encryption by using the Confusion matrix.

Based on chaotic systems many digital image encryption algorithms have been presented. In this paper, a multi-image encryption scheme composed of two chaotic systems is proposed here. The scheme incorporates the traditional bit stream ciphers technology and the spatial-domain encryption of images. One of the chaotic systems used is the

logistic map that generates a chaotic sequence. This chaotic sequence is applied to threshold function in order to obtain its equivalent stream of ones and zeros. Another chaotic system Henon map is used to get the permutation matrix P. A random modification of the pixel levels of the input plain image is conducted by the application of the bits stream ciphers technology, then using the permutation matrix P, the obtained image is encrypted. By amalgamation of the above two methods, the security of the encryption system is further strengthened and the security is analyzed in detail at the final stage. The results depict a higher security quotient in comparison to the currently available methods.

II. THE CHAOTICS SYSTEMS CHARACTERISTICS

Chaotic systems are very suitable for message information encryption because of their many characteristics, for example, (a) chaotic sequence is not periodic and not convergent and their domain is limited. Along with time, the movement of the points traverses the entire domain, called as the periodicity of the chaotic orbit. (b) Through the limited domain, collapsing and flexing are carried continuously. The output of the chaotic systems like random noise is irregular in nature. (c) For initial conditions the chaotic systems are extremely sensitive. Hence, according to the exponent rule the movement of two closed points can be separated. These dynamic properties cause chaotic sequences to be pseudo-random, unmasked, hardly and wideband. Differences in initial conditions of values result in the generation of varied chaotic sequences. Hence, the space for encrypting is very wide. The possibility for encryption as well as decryption can be provided, because the chaotic sequence are very easy to generate and easy to control.

The basic one dimensional logistic map can be formulated as

$$X_{k+1} = a * X_k * (1 - X_k) \quad (1)$$

Where $X_k \in (0, 1)$. The value of a and the initial value X_0 are used as the key k_1 for the system (a, X_0) . The results obtained from the research indicates that the system is in chaotic condition when a ranges from $3.569 < a < 4.0$.

In discrete time dynamic systems Henon map exhibit good chaotic behavior thereby taking the point (X_k, Y_k) to a different point location in space.

Mathematically it can formulated as

$$X_{k+1} = Y_k - 1 + a * X_k^2 \quad (2)$$

$$Y_{k+1} = b * X_k \quad (3)$$

The initial value $X_0 \in (0, 1)$ and $Y_0 \in (0, 1)$ can be used as the key k_2 for the system (X_0, Y_0) . The Henon map mainly depends on two parameters a and b , the research results shows that the value for a is 1.4 and for b is 0.3 for which the Henon map exhibits chaotic nature.

III. ENCRYPTION SYSTEMS

Many of the existing algorithms for image encryption technologies work based on the image content such as Arnold transformation. Further in few algorithms, Key and algorithms cannot be separated effectively. These algorithms change only the position of the pixels in the image, not the pixel value. This will give more attention for the attackers and probability of successful attack may increase.

The multi-level encryption technique proposed in this system consists of two different systems which are used to produce two different chaotic sequences. One sequence generates a bit stream and other sequence used for permuting the matrix P. Using the bit stream the pixel values are modified in the plain image. Then the modified image is again encrypted using permutation matrix P.

A. Proposed Methodology

Let the plain image I, whose pixels can be accessed using i, j , can be represented as $I(i, j)$. And the keys for encryption can be represented as $k(k_1; k_2) = k(a, X_0; X_0, Y_0)$, where the key k_1 using logistic map and key k_2 using Henon map can be represented as $k_1 = (a, X_0)$, $k_2 = (X_0, Y_0)$. The scheme of multi chaotic encryption system consists of following steps.

1. The first chaotic sequence is obtained by applying k_1 to the logistic chaotic system. The values for k_1 being (a, X_0) .

2. Using a threshold function, transform the obtained chaotic sequence into a bits stream matrix.
3. From the obtained bits stream, modify the levels of the pixel in the input plain image I (i, j) and obtain the first round ciphered image I' (i, j) . Then transpose the bits stream matrix and obtain the second round diffused image I'' (i, j) . The operation performed is the bit-wise XOR.
4. The Permutation matrix P is constructed by making use of the Henon Chaotic system with the help of k_2 . The values for k_2 being (X_0, Y_0) .
5. Perform the second level encryption of the image I'' (i, j) by using the permutation matrix P and obtain the encrypted image I''' (i, j) .

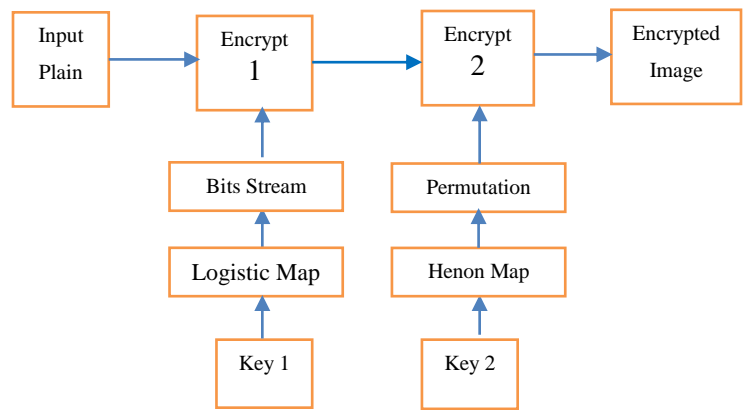


Figure1. The encryption scheme

B. Bit stream Generation

Using the equation (1) for the logistic map, the bit stream can be generated as follows:

1. Select the real value chaotic sequence $\{s_i; i=0, 1, 2, 3, 4, \dots\}$ generated by the chaotic system.
2. From the above chaotic sequence s_i , generate the bits stream using a threshold function T. The function T can be defined as below:

$$T(s) = \left\{ \begin{array}{ll} 00 \dots 00 & 0 \leq s < \frac{1}{2^K} \\ 00 \dots 01 & \frac{1}{2^K} \leq s < \frac{2}{2^K} \\ \vdots & \vdots \\ 11 \dots 10 & \frac{2^K-2}{2^K} \leq s < \frac{2^K-1}{2^K} \\ 11 \dots 11 & \frac{2^K-1}{2^K} \leq s < 1 \end{array} \right.$$

Where the value of K=8;

C. Permutation matrix generation

Using the equation (2) and (3) for the Henon map, the permutation matrix can be generated as follows:

1. The input image having size of A x B is divided into R squares of size U x V such that U x V x R=A x B. These squares are then placed one on top of the other to obtain matrix I having height T. The square matrix thus obtained by way of placing.
2. For the Henon map select the initial values X₀, Y₀ and the parameters, a, b and iterate (2) and (3) to obtain the consecutive T values.
3. At every iteration the values of X and Y are rounded to an appropriate decimal number.
4. Modulus is calculated between the obtained natural number and the height T, to obtain the matrix S having length 1 x T.
5. The T elements of Matrix S along the length array I are then filled up into another same sized matrix P (permutation matrix).
6. The Permutation matrix P is used to change the position of the T squares in the input image in round 1. In round 2, the pixel levels of the image in round 1 are again transposed with new chaotic sequence.

IV. SECURITY DISCUSSION

The improvised encryption system should enhance the resistance capacity against all types of attacks. The important attacks aimed at the chaotic encryption system contain statistical analysis, key space analysis, cipher text attack, plain text attack and so on. The proposed multi chaotic system

provides all types of security measures. The performed security analysis scheme as follows.

A. Analysis of the Key space

The maximum possible number of distinct keys used with the algorithm is termed as key space analysis. The keys used in the proposed multi-level encryption includes (a, X₀) of the logistic map and (X₀, Y₀) of the Henon map. Based on the floating point standard IEEE, In case of a double-precision number with 64-bit, the computation precision in this method is 10⁻¹⁵. The results in this algorithm provides key length larger than 2²⁵⁶ avoiding all types of Brute-force attack.

B. Key sensitivity test

The proposed method aims at finding the sensitivity of the encryption scheme in comparison to change-in values of initial conditions. The proposed multi-level encryption algorithm contains two keys key1 and key2 for both logistic and Henon map. It has been found that the proposed encryption scheme is very sensitive in terms of initial value and parameters. The key sensitivity test is experimented and it's shown in Figure 2 below. The result shows that the encryption system is very sensitive to initial conditions.

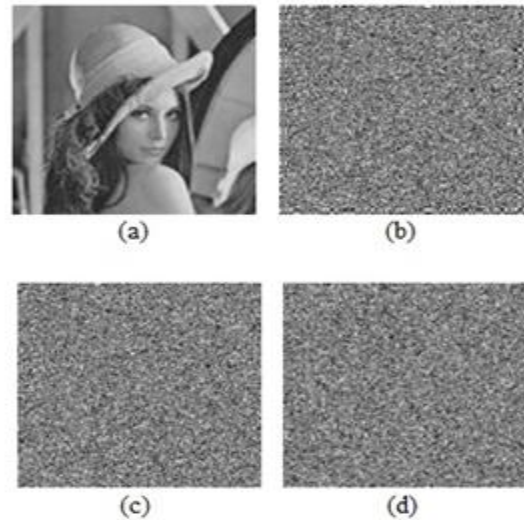


Figure2. Key sensitivity results

Picture (a) shows Lena's decrypted image by correct decrypting key, while picture (b), (c) and (d) show Lena's decrypted image by selecting the wrong decrypting key

C. Entropy analysis

To measure the randomness entropy is considered as a very imperative feature. It can be mathematically formulated as below:

$$ET(m) = -\sum_{i=0}^{L-1} p(m_i) \times \log_2(p(m_i)) \quad (4)$$

Here ET represents the entropy; L represents total grey scale values and $p(m_i)$ representing the probability of occurrence of pixel at each grey level m_i . Ideally the image to be considered random, the entropy value must be 8. The Table 3 shows the entropy variation for different schemes.

D. Statistical analysis

For any hacker the statistical analysis is the most important attack method. By way of calculating the correlations of adjacent pixels in different directions, the multi-level encryption statistical attack analysis can be arrived at. The Correlation coefficient can be calculated using equation (5)

$$C_{xy} = \frac{COVR(x,y)}{\sqrt{V(x)}\sqrt{V(y)}} \quad (5)$$

Where $COVR(x, y)$ is the Covariance between x and y can be formulated as

$$COVR(x, y) = \frac{1}{n} \sum_{i=1}^n E((x_i - \mu(x))(y_i - \mu(y))) \quad (6)$$

Where, y and x are two adjacent pixels values in the image, $V(x)$ is the variance of variable x,

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \mu(x))^2 \quad (7)$$

$\mu(x)$ Is the average of variable x.

$$\mu(x) = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

Position	Plain Image	Cipher image
Horizontal	0.9722	-0.0087
Vertical	0.9892	-0.0273
Diagonal	0.9578	-0.0049

Table 1. Correlation Co-efficient of the adjoining pixels along different directions.

The pixels that are distributed in the plain image as well as ciphered image is shows in Figure 3. Table 1 shows the correlation coefficients of the adjacent pixels along the horizontal, vertical and diagonal direction.

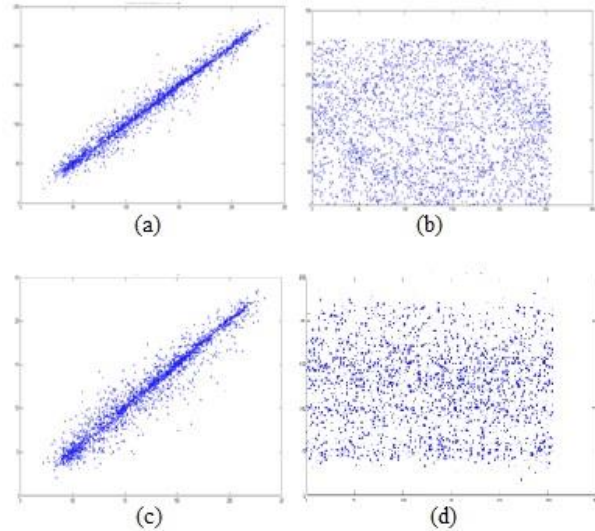
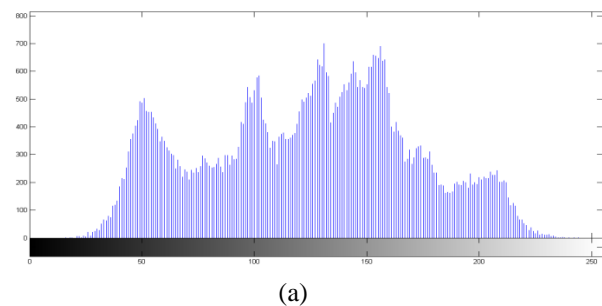


Figure 3. The pixel distribution of the Lena Image

- (a) The plain Lena image in horizontal
- (b) The ciphered Lena image in horizontal
- (c) The plain Lena image in vertical
- (d) The ciphered Lena image in vertical

E. Histogram

The histogram variations are calculated for both cipher images as well as for the plain image. Figure 4 shows the variations in the histogram for input plain Lena image and its output encrypted cipher image.



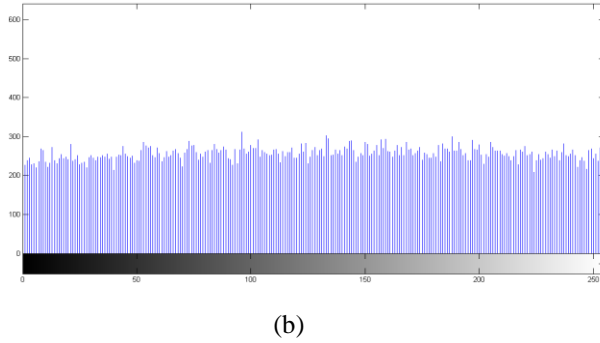


Figure 4. The histogram variations for Lena image
(a) Histogram variation for plain image
(b) Histogram variation for cipher image

F. Differential attack

- **Rate of Change of Pixel Number (NPCR):**

NPCR calculates the total number of pixels that have been changed in the ciphered images obtained for the plain image and its modified plain image at the same location. The Table 3 shows the NPCR variations among different encryption schemes.

$$NPCR = \frac{\sum_{i,j} D(i,j)}{(UXV)} \times 100 \quad (9)$$

Here U and V represent the height and width of the image.

$D(i,j)$ Can be defined as

$$D(i,j) = \begin{cases} 1 & Cip1(i,j) \neq Cip2(i,j) \\ 0 & Cip1(i,j) = Cip2(i,j) \end{cases}$$

$Cip1(i,j)$ Grey value of cipher image and $Cip2(i,j)$ Grey value of new cipher image.

- **Unified average changing intensity:**

UACI calculates the mean change in intensity between plain image and the encrypted images as a percentage with N. The maximum intensity pixel value. N for the grey image being 255. Table 3 depicts the UACI variations among various encryption schemes.

$$UACI = \frac{1}{(UXV)} \sum_{i,j} \left[\frac{abs(Cip1(i,j) - Cip2(i,j))}{(255)} \times 100 \right] \quad (10)$$

$Cip1(i,j)$ Represents ciphered image for the input plain image.

$Cip2(i,j)$ Represents encrypted image for the modified input image.

V. SIMULATION

The Simulated images are the color Lena image and different document images. These images are transformed to grey scale and then the encryption scheme is applied. The encryption results are shown in Figure 6 and Table 2 shows the variation of parameters for different document images. The initial key values used here for both logistic map and Henon map are

$$k(k_1; k_2) = k(a, X_0; X_0, Y_0) = (3.97, 0.3; 0.6315477, 0.18906343).$$



Figure 5. The Document plain images
(a) Doc1 (b) Doc 2 (c) Doc3

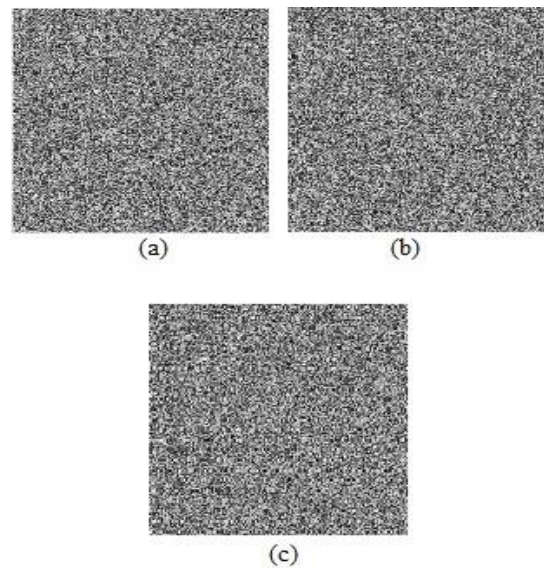


Figure 6. The encrypted Document images
(a) Doc 1 (b) Doc 2 (c) Doc

Images	Rounds	MSE	PSNR	UACI	NPCR	Entropy	UIQ	SSIM
DOC 1	Round 1	1.4948x10 ⁴	6.3848	39.4336	99.5605	7.9937	0.0018	0.0080
	Round 2	1.4965x10 ⁴	6.3799	39.4143	99.5636	7.9937	-1.558x10 ⁻⁵	0.0087
DOC 2	Round 1	1.8366x10 ⁴	5.4907	44.5601	99.3164	7.9899	0.0024	0.0112
	Round 2	1.8372x10 ⁴	5.4893	44.5507	99.3256	7.9899	0.0018	0.0099
DOC 3	Round 1	1.8074x10 ⁴	5.5602	44.0566	99.4705	7.9901	0.0031	0.0113
	Round 2	1.8081x10 ⁴	5.5585	44.0518	99.4339	7.9901	0.0023	0.0083
DOC 4	Round 1	1.3787x10 ⁴	6.7361	37.5908	99.5789	7.9926	2.123x10 ⁻⁴	0.0094
	Round 2	1.3817x10 ⁴	6.7268	37.6222	99.53	7.9926	-0.0037	0.0055
DOC 5	Round 1	1.5360x10 ⁴	6.2669	39.9917	99.5911	7.9945	-0.0013	0.0086
	Round 2	1.5371x10 ⁴	6.2639	40.0130	99.5529	7.9945	-0.0027	0.0056
Lena	Round 1	7.5846x10 ³	8.8050	28.28	99.627	7.997	-0.0016	0.0105
	Round 2	7.5617x10 ³	8.8180	28.1870	99.6582	7.997	0.0014	0.0105

Table 2. The variation of parameters (MSE, PSNR, UACI, NPCR, Entropy, UIQ and SSIM) for different document images with different rounds.

Scheme	NPCR	UACI	Entropy
Brindha Murugan et al[8]	99.6204	33.4612	7.994
Xingyuan Wang[9]	99.60	33.45	7.997
Yannick Abanda,et al[10]	99.57	35.08	7.996
Ritesh et al[11]	99.60	33.47	7.997
Proposed	99.627	28.28	7.997

Table 3. Comparison of NPCR, UACI and Entropy for different schemes using Lena Image.

VI. CONCLUSION

The image encryption scheme based on multi chaotic systems by applying both logistic and Henon map tested through simulation and security analysis reveal the efficacy of the proposed encryption method over the existing methods. The proposed encryption method by way of increased sensitivity to the key and its larger space reduces the instances of differential attack and statistical attack. From the above detailed analysis the proposed image encryption is an appropriate method for secured document image encryption.

REFERENCES

- [1]. Zhang Hao-Ran, Li Chang-Gang, and, Han Zheng-Zhi, "Image Encryption Techniques: A Survey", Journal of Computer Research and Development, Oct. 2002.
- [2]. Wei-qi Yan, Dong-xu Qi, and Wei Ding, "A Novel Digital Hiding Technology Based on Tangram and Conways Game", proceeding of 2000 International Conference on Image Processing, Sept. 2000.
- [3]. Zhao Xue-feng, "Digital Image Scrambling Based on the Baker's Transformation", Journal of Northwest Normal University, Feb. 2003.
- [4]. Ji Shi-ming, ShenJian-bin and Bao Guan-jun, "Magic Cube Transformation and Its Application in Digital Image Encryption", Computer Applications, Nov. 2002.
- [5]. Cao Changxiu, Hu Zhongyu, Zhu Guibin, et al., "An Image Scrambling and Encryption Algorithm Based on Affine Transformation", Journal of Computer-Aided Design & Computer Graphics, June. 2003.
- [6]. Zhang Xin, Chen Wei-bin "Image encryption algorithm based on Henon chaotic system". International Conference on Image Analysis and Signal Processing, 2009, IASP.
- [7]. Md Jan Nordin, Elankovan Sundararajan and Ali Soleymani, "A Chaotic Cryptosystem for Images Based on Henon and Arnold Cat Map" The Scientific World Journal, Volume 2014.
- [8]. Ammasai Gounden, Nanjappa Gounder and Brindha Murugan, "Image encryption scheme based on block based confusion and multiple levels of diffusion", the institution of engineering and Technology, computer vision 2016 Vol. 10, Issue. 6, pp. 593-602.
- [9]. Xingyuan Wang and Chuanming Liu, "A novel and effective image encryption algorithm based on chaos and DNA encoding", Multimedia Tools and Applications 2016, pp. 1-17.
- [10]. Yannick Abanda, Alain Tiedeu, "Image encryption by chaos mixing", the institution of engineering and Technology, image processing 2016, ISSN 1751-9659.
- [11]. Ritesh Bansal, Shailender Gupta and Gaurav Sharma, "An innovative image encryption scheme based on chaotic map and Vigenère scheme" Springer Science and Business Media. New York ,2016.
- [12]. Ritesh Bansal, Shailender Gupta and Gaurav Sharma, "An innovative image encryption scheme based on chaotic map and Vigenère scheme", Springer Science and Business Media. New York, 2016.
- [13]. Wang Yujing, Yang Fei, Kang Shouqiang, Jiang Yicheng,, Lan Chaofeng, "An image encryption scheme using mixed high dimensional chaotic system combined with fast fourier transform", IEEE 12th International Conference on Electronic Measurement & Instruments, 2015.
- [14]. Fadia TALEB, "A New Chaos Based Image Encryption Scheme Using Chaotic Logistic Maps" IEEE on Multimedia.
- [15]. Guoji Zhang., Qing Liu.: 'A novel image encryption method based on total shuffling scheme', Optics Communication, Volume 284, Issue 12, 1 June 2011, Pages 2775–2780
- [16]. Lin Teng., Xingyuan Wang.: 'A bit-level image encryption algorithm based on spatio temporal chaotic system and self-adaptive', Optics Communication., Volume 285, Issue 20, 15 September 2012, Pages 4048–4054.
- [17]. Xingyuan Wang, Canqi Jin.: 'Image encryption using game of life permutation and PWLCM chaotic system', Optics Communication., 2012, pp. 412–417.
- [18]. Xingyuan Wang.,Lintao Liu., Yingqian Zhang, 'A novel chaotic block image encryption algorithm based on dynamic random growth technique', Optics and Lasers in Engineering.,Volume 66, March 2015, Pages 10–18.
- [19]. Jun-xin Chen, Zhi-liang Zhu, Chong Fu, Hai Yu, Li-bo Zhang: 'A fast chaos-based image encryption scheme with a dynamic state variables selection mechanism', Communications in Nonlinear Science and Numerical Simulation. Volume 20, Issue 3, March 2015, Pages 846–860.
- [20]. Radu Eugen BORIG, Ana Cristina, Adrian Viorel: 'A new hyper chaotic map and its application in an image encryption scheme', Signal Processing. Image Communication. 2014, pp. 887–901.
- [21]. Yifeng Zheng, Jianxiu Jin, 'A novel image encryption scheme based on Hénon map and compound spatiotemporal chaoses, Multimedia Tools Applications., Volume 74 Issue 18, September 2015, pp. 1–18.
- [22]. Gelli M.B.S.S. Kumar, Chandrasekaran, V. 'A novel image encryption scheme using Lorenz attractor'. IEEE Conference. on Industrial Electronics and Applications (ICIEA), 2009, pp. 3662–3666.

Parametric optimization of Zigbee parameter using soft computing techniques

Prativa P. Saraswala

Research Scholar, V.T Patel Dept. of Electronics and
Communication, Charusat,
Changa, Gujarat, India

Dr. Jaymin Bhalani

Associate professor, Dept. of Electronics and
Communication,
Babaria Institute of Technology, Vadodara, Gujarat, India

Abstract— Zigbee protocol is based on IEEE 802.15.4 standard which mainly supports two types of routing protocols. The first is the well-known Ad-hoc on demand Distance Vector (AODV) routing protocol, which was considered for highly self-motivated application situations in wireless ad-hoc networks and second type is the Tree based routing scheme which is established at the time of network formation. True Time 2.0 is a MATLAB/Simulink based simulator used for real time control system. This paper uses the wireless block of True Time for the simulation of signal transmission range of Zigbee network using AODV routing protocol. Simulated results are also analysed using various soft computing techniques such as Fuzzy logic (FL), Artificial Neural Network (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS). On the basis of analysis done, comparison is done between the True Time simulated result and the soft computing based techniques for the signal transmission.

Keywords-com Zigbee and IEEE 802.15.4 standard, Zigbee protocol, AODV, ANFIS, ANN and Fuzzy logic.

I. INTRODUCTION

Zigbee based on IEEE 802.15.4, defines a specifications for low rate WPAN (LR-WPAN) which supports simple devices that consume less power and typically operate in the personal operating space (POS) of 10m [1]. The Zigbee Alliance [2] is an association of companies working together to develop standards (and products) for reliable, cost-effective, low-power wireless networking. Zigbee technology is used in a wide range of products and applications across consumer, commercial, industrial and government markets worldwide.[3] Zigbee is based on the IEEE 802.15.4 standard [4] which describes the physical and MAC layers which provides low cost, low rate personal area networks.

Zigbee technology has greater features when compared with other existing wireless methods like WIFI, Bluetooth, Ultra wide band, WiMaX and so on. Comparison between Zigbee, IEEE 802.11 based Wi-Fi, Bluetooth, Ultra Wide band and WiMaX can be obtained from [5]-[7]. Comparison shows that the data rate in Zigbee is least among all, but the power consumption is minimum. These two characteristics allow the Zigbee based devices to be in power saving mode when it is not used. Zigbee protocols are intended for use in embedded applications requiring low data rates and low power consumption. Though WPAN implies a reach of only a few meters, network has several layers, designed to enable the intrapersonal communication within the network, connection to

a network of higher level and ultimately an uplink to the Web [5].

This paper has focused on soft computing techniques which is basically a combination of artificial intellect methodologies used to do the modelling and solving the real world problems which are not easy to be solved mathematically [8]. Paper presents a review on different types of soft computing techniques and the application areas where they can be applied. Authors have presented the significance of power optimization in the field of Industrial WSN and explained the various method available commercially for power optimization and also dealt with the routing based methods for power optimization. In paper [9] authors have emphasized on prediction based shortest routing protocol for the optimization of energy in Zigbee standard. Paper [10] describes various routing protocols defined for Zigbee standard and proposes anew routing protocol which uses a modified AODV and Hierarchical Tree Routing. In [11], Authors have introduced an adaptable fuzzy-logic assisted configuration mechanism for Zigbee networks which is considered as most promising standards in the area of wireless sensor networks. Result shows that the use of fuzzy-logic principles can significantly improve the tree-based routing mechanisms if the set-up network time and the number of router nodes are reduced in the network, leading to the reduction of energy consumption by nodes by nodes. Paper [12] shows a reduction in the number of packets and overhead related to packets, communication delay and the improvement in route efficiency. Paper proposes the use of a novel fuzzy-logic-based metric to be used for the decision making process in AODV.

Paper is organized as follows: Section II will explain the 802.15.4 standard based Zigbee Protocol including the types of devices in Zigbee standard, Network topologies in Zigbee and the routing techniques in Zigbee. This section also includes the comparison between the traditional routing protocols and the modified protocols for further improvement in Zigbee performance metrics. Section III will discuss the necessary steps needed for the simulation process in MATLAB/ TRUETIME 2.0 including the AODV process and the features of MATLAB/ TRUETIME 2.0.

Section IV will explain the three soft computing techniques like Artificial Neuro Fuzzy Inference System, Artificial Neural Network (ANN) and Fuzzy Logic for the optimization of signal transmission range in Zigbee network. This section will explain the various soft computing techniques with involved mathematics. Section V will discuss the comparison between

different soft computing techniques and will provide the conclusion based on the comparison made.

II. IEEE 802.15.4 BASED ZIGBEE

- A. *Device types:* There are mainly three types of devices in Zigbee network. The Zigbee coordinator (one in the network) is a full functioned device which manages the whole network. An FFD can perform all the responsibilities defined by Zigbee standard based on IEEE 802.15.4 MAC layer. Third device is an end-device is a reduced functioned device which has limited tasks to perform. The whole comparison between the three types of devices are shown in table 1.
- B. *Topologies:* Zigbee is based on IEEE 802.15.4 standard for its physical layer and Medium access layer. IEEE 802.15.4 supports various topologies such as star, tree, cluster tree, and mesh; however, Zigbee provide support only star and peer to peer topologies. These two topologies are discussed in table 2 in brief.

- C. *Routing in Zigbee network :* In every system or network, the resources are fixed so it's essential to manage with the limited resources. Routing plays a very important role in *any* network in the direction of utilizing limited resources. An intelligent routing strategy can manage with the limited resources as well as it can be adaptable to changing network conditions like network size, traffic density and network partitioning [9]. In case of wired networks, link-state and distance-vector algorithms are used where each node maintains a list of its neighbours and periodically intimates its status to other nodes using flooding strategy. The traditional link-state and distance-vector algorithms do not work in large Mobile Ad Hoc networks (MANET's). This is because periodic or repeated route updates in a large networks may lead to a significant increase in consumption of available bandwidth, increase in channel collision and need every node to recharge itself regularly. In order to overcome these problems associated with the link-state and distance-vector algorithms, various routing protocols have been proposed for Mobile Ad Hoc networks.

Table 1 TYPES OF DEVICES IN ZIGBEE STANDARD [1], [13]

Zigbee Coordinator (ZC)	Zigbee Router (ZR)	Zigbee End Device (ZED)
<ul style="list-style-type: none"> Known as Coordinator in IEEE 802.15.4 standard. One and only one required per network Takes the initiation in network formation. Selects the channel to be used by the network. Does address assignment for nodes. Maintains a list of neighbors or routers to the coordinator. May act as router once network is formed Transfers application packets. 	<ul style="list-style-type: none"> Non-compulsory device in the network Generally contributes in multi hop routing. Used in tree and mesh topology to increase the network coverage. Router finds the best path between the sources to destination to transfer a message. A router has a similar characteristics as that of coordinator except the establishment of network. 	<ul style="list-style-type: none"> Non-compulsory device in the network. The end device can be connected to a coordinator or router. It works at very low duty cycle as consumes power only at the time of transmission. Leads to reduction in transmission time in Zigbee network. An RFD is use for very simple applications like switching a light or a fan or used in passive infrared sensor.

Table 2 NETWORK TOPOLOGIES

Star topology	Peer-to-Peer Topology
<ul style="list-style-type: none"> It consist of a coordinator and only the end devices. 	<ul style="list-style-type: none"> More complex network can be formed such as mesh topology.
<ul style="list-style-type: none"> All end devices (RFD) communicates with the PAN coordinator directly. 	<ul style="list-style-type: none"> Here a FFD can communicate with another FFD directly within the transmission range.
<ul style="list-style-type: none"> If there occur any malfunctioning in the coordinator then the communication between the two nodes get disturbed as the communication is via coordinator only. 	<ul style="list-style-type: none"> A peer-to-peer network can be of ad hoc, self-organizing, and self-healing [4].
<ul style="list-style-type: none"> Star topology Applications include home automation, personal computer (PC) peripherals, toys and games, and personal health care [4]. 	<ul style="list-style-type: none"> FFD's can transfer the data to other FFD still if there are out of range using intermediate FFD nodes forming a multi hop network.
<ul style="list-style-type: none"> The advantage of star topology is that it is simple and packets go through at most two hops to reach their destination [13]. 	<ul style="list-style-type: none"> Applications such as industrial control and monitoring, wireless sensor networks, asset and inventory tracking, intelligent agriculture, and security would benefit from such a network topology [4].

There are various performance metrics in Zigbee network such as packet delivery ratio, packet loss ratio, Hop delay, Average end to end delay, Network Lifetime, Optimal path length, Throughput and Media Access Delay and so on. Authors have done so much research for the improvement of these parameters based on two routing protocols in Zigbee network: Tree based routing and the AODV. Many researchers have modified these protocols for further improvement in performance metrics. Table 3 gives the brief of some of the modified algorithms based on AODV and tree based routing.

Table 3 COMPARISON BETWEEN TRADITIONAL ALGORITHMS WITH MODIFIED ALGORITHM

Traditional method	Modified algorithms
Tree based routing	<ul style="list-style-type: none"> Self-Learning Routing (SLR) protocol [14] is based on overhearing and a caching mechanism which leads to decrease in the end-to-end delay. Neighbor Tree Routing (NTR) [15], [16]: reduces the network costs and end to end delays, leading to the energy conservation and the real time of the network enhancement. Fuzzy- Logic based Tree Routing, FL-TR) [17] leads to reduction in energy consumed by nodes by reducing the network Set-up time and the number of router nodes in the network get reduce. In [18], leads to reduction in control overheads without influencing packet delivery ratio and path length of Zigbee network.
AODV	<ul style="list-style-type: none"> In [19], Author has proposed an energy-aware routing mechanism EA- AODV for the maximization of the use of limited energy and prolonging the lifetime of Zigbee network. In [20], Author has suggested a dynamic version of AODV called ARTO-AODV which outperforms with respect to packet delivery ratio, throughput, average end to end delay, and energy efficiency. Here in [21], proposed algorithm is based on Flooding and AODV. It has higher reliability and lower overhead than AODV. Proposed algorithm [22] reduces energy consumption, improve network energy efficiency, and avoid network segmentation. It also improves the reliability of network leading to extension of network lifetime. In [23], proposed method decreases both the routing message overhead and the average end to end delay by on an average 27.9%, 13.7% respectively less than the well-known AODV routing protocol which ultimately increases the throughput by 23.87% more than the traditional AODV routing protocol.

Table 4 DEFINITION OF AODV PARAMETERS USED FOR SIMULATION

Parameter	Values
Active Route Timeout	3000 milli seconds
Route time	2* Active Route Timeout
Hello Interval	1000 milli seconds
Delete Period	Allowed Hello loss*Hello Interval

III. IMPLEMENTATION OF AODV ROUTING PROTOCOL FOR ZIGBEE NETWORK

This section includes the explanation of AODV concept used here and how it is used in TRUETIME/MATLAB for Zigbee Network for the parametric optimization.

A. AODV

AODV [24], [25] is a reactive type of routing protocol where a path between the source and the destination is established only when there is a need. AODV defines three types of messages which are Route Requests (RREQs), Route Replies (RREPs) and Route Errors (RERRs). Figure 1 to 5 explains the steps for the data transmission in AODV routing protocol. Figure 6 shows the different notations used for the type of messages in AODV [26]. Definition of AODV parameters considered for the simulation process is shown in table 4.

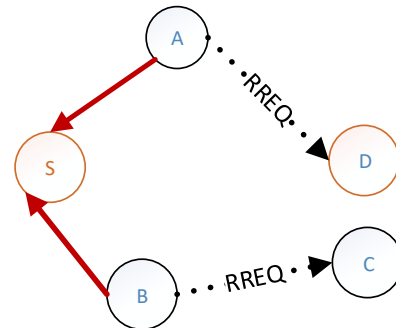


Figure 1: Node S wants to send a Data Packet to node D, S broadcasts an RREQ.

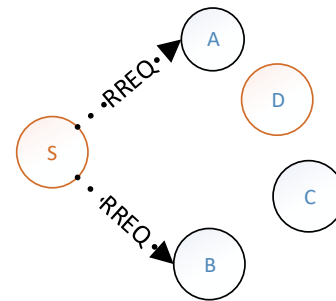


Figure 2: Reverse Route established between A & B and A & B rebroadcast RREQ to remaining nearby nodes.

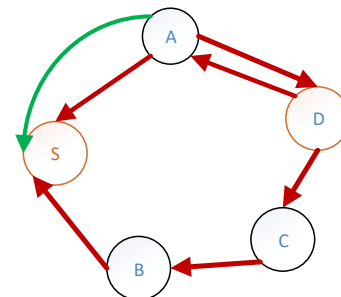


Figure 3: D establishes Reverse Route and A establishes Route and unicasts RREP

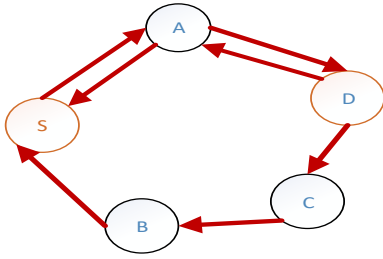


Figure 4: A path is established between Source S and Destination D

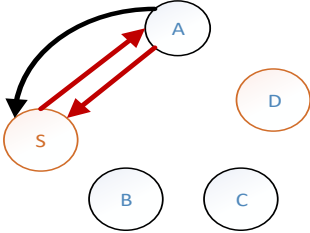


Figure 5: Link between node A and D is broken. A generates RERR

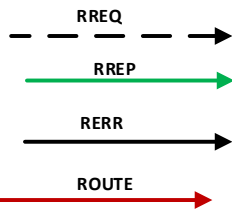


Figure 6: Symbols used for AODV messages

B. True Time 2.0/MATLAB

True Time is a Matlab/Simulink-based simulator, which provides co-simulation of controller task execution in real-time kernels, continuous plant dynamics and network transmissions [27], [28]. True Time library contains various blocks like True Time Kernel, True Time Network, True Time Send and Receive, True Time Wireless network etc. In our implementation, we have focussed on True Time Wireless Network, which can be used for data transmission in Ad-Hoc network. Mainly two network protocols are supported by True time wireless network block like IEEE 802.11b/g (WLAN) and IEEE 802.15.4 (Zigbee). Zigbee protocol mainly uses CSMA/CA MAC protocol for the data transmission. Flow chart for CSMA/CA [27] is shown in figure 7. Notations used here are

- NB: Number of backoffs.
- BE: Backoff exponent, macMinBE: The minimum value of the backoff exponent.
- MaxBE: The maximum value of the backoff exponent.
- macMaxCSMABackoffs: The maximum number of backoffs .

For simulating the AODV for Zigbee network, we need to provide the following parameters through MATLAB function:

- Transmitting power
- Receiver signal threshold
- Sending node name
- Receiving node name
- x and y coordinates of sending and receiving node

Current simulation time.Default Propagation model or the path loss function is used here is

$$P_{receiver} = \frac{1}{d^a} * P_{sender} \quad (1)$$

Where a is the pathloss function which can be chosen for modelling in different environment. When the sending node wishes to send a data packet. It checks whether the medium is free or not. For busy medium mac procedure is followed as explained in figure above. When the transmission starts, its relative distance d with other neighboring nodes in the network is calculated as:

$$distance(d) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

d is used for the calculation of power level as:

$$power = \frac{P_{transmit}}{(distance+1)^a} \quad (3)$$

Where a is taken as 3.5.

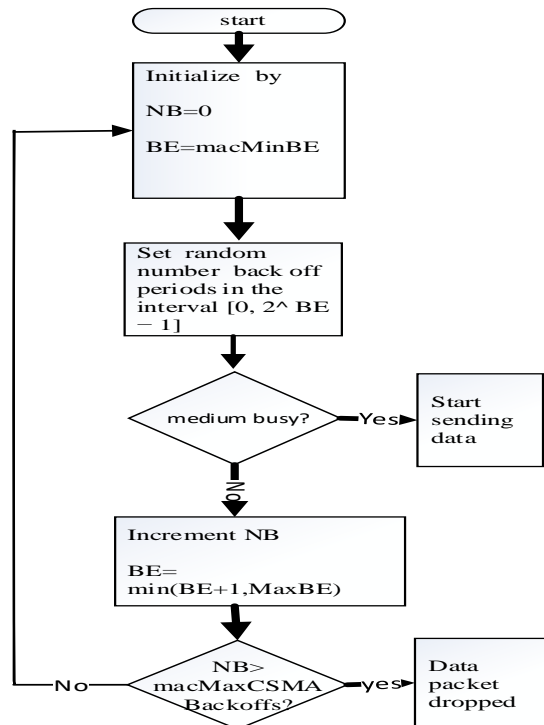


Figure 7: CSMA/CA algorithm in Zigbee Protocol

IV. SOFT COMPUTING TECHNIQUES

This section will explain the main features of three soft computing techniques, Fuzzy logic, ANN and ANFIS with their implementation in True Time 2.0.

A. Fuzzy logic

In a fuzzy system like a conventional system, input variables are continuously read and manipulated and thus a controller continuously generates values for the system output plus two more steps are included: fuzzification and defuzzification [29]- [31]. Fuzzification is the first step of fuzzy controller which reads the crisp input data and transforms the actual values of the variables to “fuzzy values”. Defuzzification is the last step of fuzzy logic system which again transforms the “fuzzy values” back to the actual values.

Figure 8 shows the fuzzy system scheme. Detail explanation of fuzzy logic implementation is given in [32] which includes the design of membership function and the rule base. The transmitting power and the threshold power are taken as the input to the fuzzy system and the signal transmission range is the output for the system.

B. Artificial neural network

Neuron is a fundamental building block for neural networks which contains three distinct functional operations: the scalar input let's say 'a' is multiplied with the scalar weight w. Secondly this product is added with the bias 'b' forming net input and thirdly the net input is given to the transfer function 'f' produced scalar output 'b' [33]. Process is shown in the figure 9.

There are various Transfer function (TF) used in neural network are log-sigmoid (logsig), tan-sigmoid (tansig) and linear transfer function purelin. Details can be obtained from [33]. For ANN training, various training algorithms are used such as Levenberg–Marquardt back propagation, gradient descent back propagation, quasi-Newton back propagation and Bayesian regularization back propagation etc. Table 5 enlists some of the training algorithms with brief description.

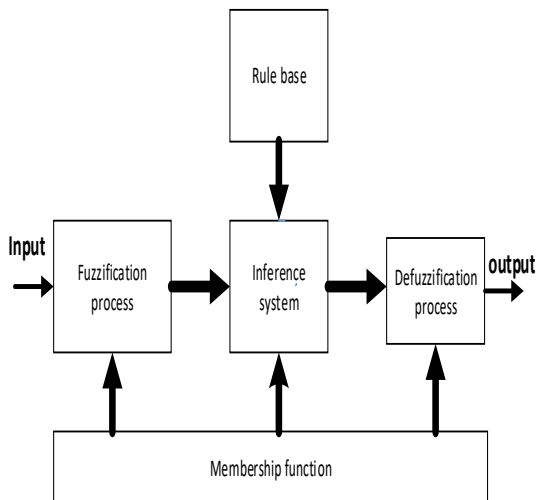


Figure 8 Fuzzy system scheme

Data set for ANN contains two inputs and one output obtained from the simulation of AODV routing protocol for Zigbee network in TRUETIME in MATLAB. Overall simulation process for ANN is shown in figure 10. Complete information regarding ANN implementation can be obtained from [37].

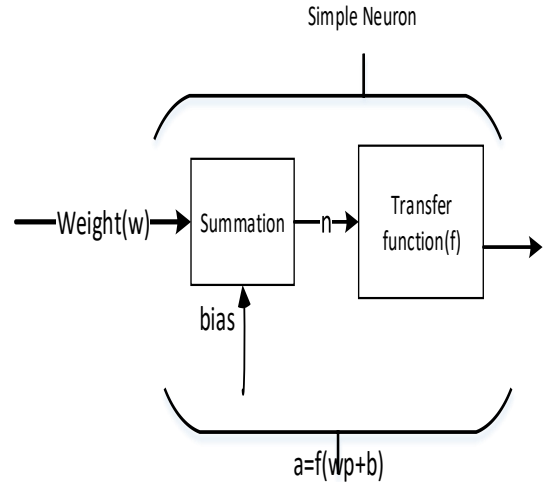


Figure 9 Basic Neural Network

Table 5 TRAINING ALGORITHMS IN ANN

Type of transfer function	Description
Levenberg – Marquardt back propagation (trainlm)	<ul style="list-style-type: none"> It is assumed to be fastest and default training function for feed forward net. It is widely adopted as the training algorithm due to its fast speed and high performance accuracy [33], [34].
BFGS quasi-Newton back propagation (trainbfg)	<ul style="list-style-type: none"> Less efficient for large networks having weights in thousands. [34]
Gradient descent back propagation (traingd)	<ul style="list-style-type: none"> Updates weight and bias values according to gradient descent [34].
Bayesian regularization back propagation (trainbr)	<ul style="list-style-type: none"> It updates the weight and bias according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights, for determining the correct combination to produce a network that can generalize better [34]. It mostly works best when the network inputs and targets are scaled to lie approximately in the range of [-1, 1]. [33]
Resilient back propagation (trainrp)	<ul style="list-style-type: none"> It has been used successfully in training multi-layer perceptron neural networks [35]. Memory requirements is relatively small. Much faster than gradient descent algorithms. This algorithm takes the advantage of the signs of the gradient in order to increase or decrease the learning rate. [35],[36]

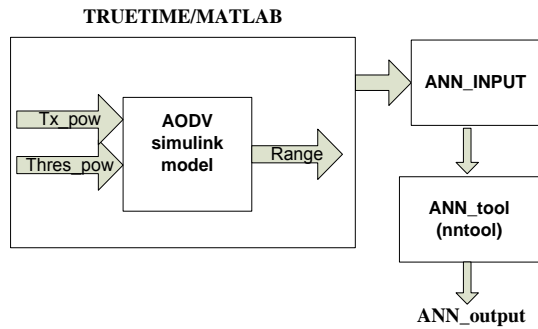


Figure 10 ANN simulation process [37]

C. ANFIS (Adaptive-Network-based Fuzzy Inference System)

One of the most important problems associated with the Fuzzy Inference System (FIS) is making appropriate decision about the parameters like membership functions, distributions of MFs and the fuzzy rules composition. Trial and error is one of the methods which can lead to better shape and location of ANFIS membership function for each fuzzy variable. Moreover, user's experience also affects FIS modeling. So because of these problem and the unsatisfactory result obtained from implementation of FIS leads us to combine both neural networks and fuzzy logic to minimize the error and achieve optimized method for maximum transmission range [38].

The ANFIS [39] is an adaptive network of nodes and directional links where all or some of the nodes have parameters which affect the output of the node. There is a learning relationship between inputs and outputs. ANFIS approach is to construct a FIS whose membership function parameters can be tuned with the help of only a back propagation algorithm or combining with a least squares type of method which allows FIS to learn from the input-output data set provided. Here, the Neuro-Fuzzy is applied to the network for optimizing Signal Transmission Range for Zigbee network using AODV routing protocol in TRUE TIME in MATLAB.

The ANFIS architecture for two inputs and one output [39] - [42] is shown in figure 11. The circular nodes represents the nodes that are fixed whereas the square nodes are nodes that have parameters to be learnt.

Layer_1: is called the Fuzzification layer where Neurons perform Fuzzification. In Jang's model, Fuzzification neurons have a bell activation function [43].

Membership function for bell function is given as

$$\mu_a = \frac{1}{1 + \left| \frac{x-c_i}{a_i} \right|^{2b_i}} \quad (4)$$

Where a_i, b_i, c_i are parameters to be learnt.

Layer_2: known as rule layer. Every neuron in this layer relates to a single Sugeno-type fuzzy rule. This layer collects the inputs from the respective fuzzification neurons and calculates the firing strength of the rule it signifies. Every node in this layer is fixed.

$$O_{2,i} = w_i = \mu_{A_i}(x) * \mu_{B_i}(y), i = 1,2 \quad (5)$$

Layer_3: is the normalization layer where every neuron in this layer receives the inputs from previous Rule layer and calculates the ratio of the firing strength of the rules, also called normalized firing strength of a given rule. Normalized firing strength is given as:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad (6)$$

Layer_4: It is called defuzzification layer where each neuron in this layer is connected to the respective normalization neuron. The nodes in this layer are adaptive and perform the consequent of the rules given as:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (7)$$

Where the parameters (p_i, q_i, r_i) are referred as consequent parameters which need to be determined.

Output_layer: This layer represents a single summation neuron which calculates the sum of outputs of all defuzzification neurons and produces the overall ANFIS output. The overall output obtained at this layer is given as:

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (8)$$

MATLAB is used to load the training pairs decided by the Sugeno Fuzzy Inference System (FIS) to train the neural network using ANFIS Editor [41]-[42]. Further information on ANFIS can be obtained from [44]. Model generated by training through ANFIS Editor Tool is shown in figure 12, where the inputs to ANFIS are transmitting power and the threshold power and the output is signal transmission range respectively.

Inputs to ANFIS is in form of matrix which contain inputs in form of column and the last column is of output. The input set taken for ANFIS is the data set created after the simulation of AODV model in TRUE TIME. The overall flow is shown in figure 13 and the surface viewer is shown in figure 14.

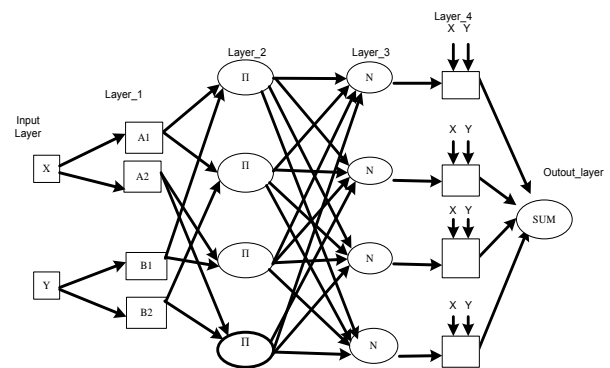


Figure 11 Overall ANFIS simulation process

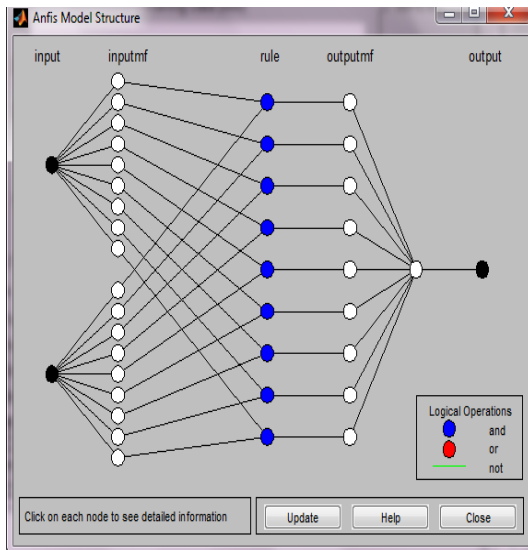


Figure 12 Generated ANFIS Model Structure

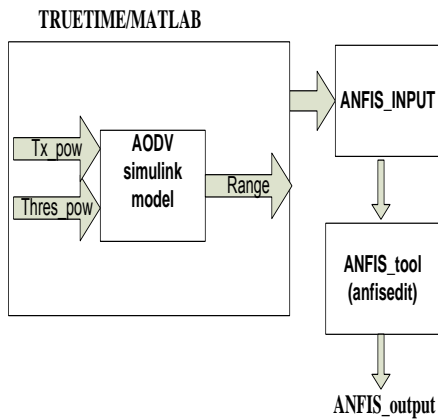


Figure 13 Overall ANFIS simulation process

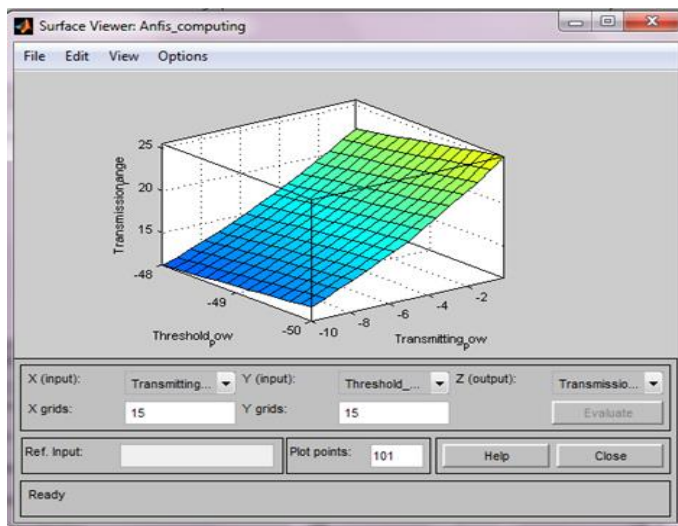


Figure 14: Surface view for generated model

V. DISCUSSIONS AND CONCLUSION

In previous sections, three soft computing techniques: ANFIS, fuzzy and ANN are discussed for a Zigbee network using AODV protocol in MATLAB, for data transmission between the source node and the destination node. On the basis of the discussion made in previous section, a comparison is done between these techniques for the optimization of transmission range in Zigbee network. Figure 15, 16 and 17 shows the comparison between the three soft computing techniques with the classical method at various threshold power: -48dbm, -49dbm and -50dbm respectively for the transmission range in Zigbee network using AODV protocol. It can be seen that ANFIS based and the ANN based method performs very well as compare to Fuzzy based method. With the change in threshold power of the signal, the transmission range is also getting affected which is governed by the expression (1). Fuzzy based transmission range varies abruptly with change in transmitting power. ANN and ANFIS based output is giving more promising output as compare to Fuzzy based method. Graph is plotted for three different values of threshold power: -48dbm,-49dbm and -50dbm respectively. It can be seen that, with the increase in the threshold value, the transmission range is also increasing. For threshold value of -48dbm, maximum transmission range is observed.

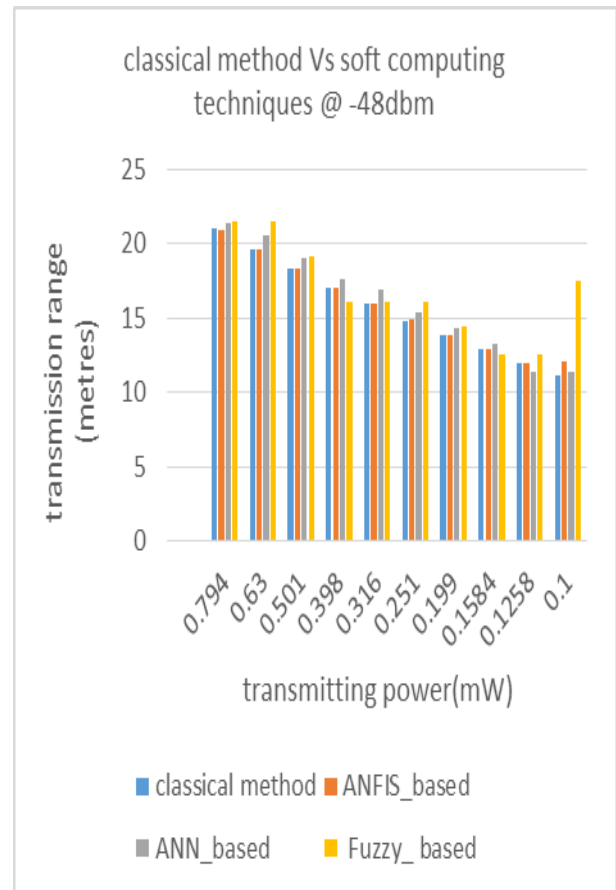


Figure 15 Comparison between various soft computing techniques with Classical Method at -48dbm threshold power

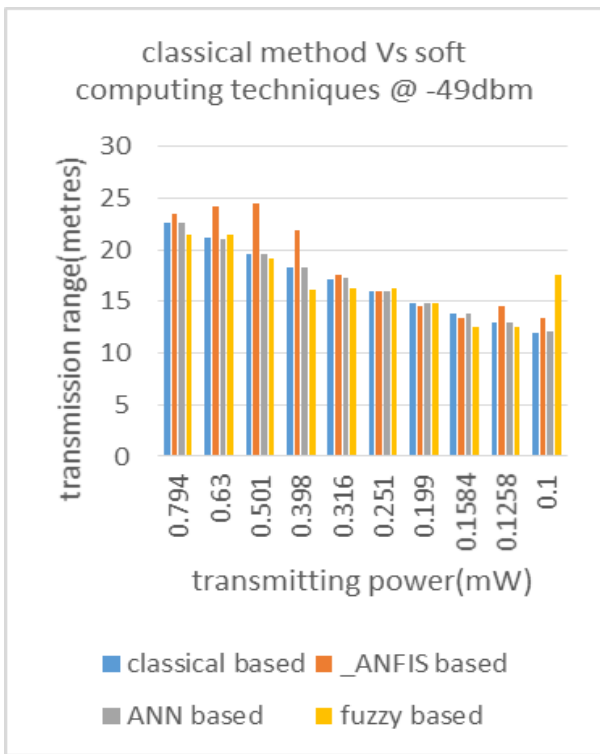


Figure 16 Comparison between various soft computing techniques with Classical Method at -49dbm threshold power

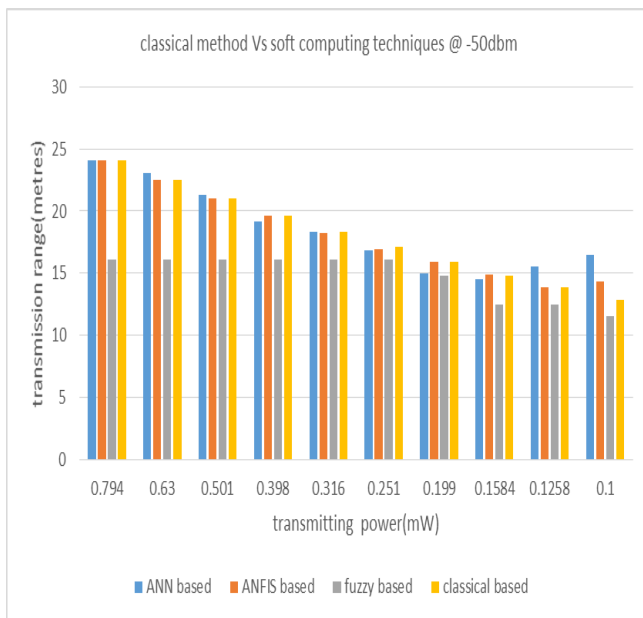


Figure 17 Comparison between various soft computing techniques with Classical Method at -50dbm threshold power

ANFIS and ANN soft computing seems to be more appropriate methods as compare to Fuzzy logic for the Zigbee parameter in TRUETIME/MATLAB.

ACKNOWLEDGMENT

I am thankful to my colleges, Assistant Professor Farah Pathan, Assistant Professor Sandhya Sharma, Assistant Professor Vishwas Raval, Research Scholar Varadha Tambe and Associate Professor Anuradha Deshpande. I am also thankful to Department of Electrical Engineering, Faculty of Tech and Engg, Maharaja Sayajirao University of Baroda, for providing me support.

REFERENCES

- [1] Javier Gracia Castro, Óscar Pérez Domínguez, A Presentation on “Zigbee: IEEE 802.15.4”, Tampere University of Technology, Institute of Production Engineering, 2007.
- [2] ZigBee Alliance, Zigbee Specifications, version 1.0, April 2005.
- [3] Paolo Baronti, Prashant Pillai, Vince W.C.Chook, Stefano Chess, Alberto Gotta and Y. Fun Hu, “Wireless sensor networks: A survey on the state of the art and the 802.15.4 and Zigbee standards”, published in Computer Communications, volume 30 in 2007.
- [4] IEEE 802 Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low- rate Wireless Personal Area Networks, IEEE Computer Society.
- [5] Bhavneet Sidhu, Hardeep Singh, and Amit Chhabra, “Emerging Wireless Standards - WiFi, ZigBee and WiMAX”, World Academy of Science, Engineering and Technology 25 2007.
- [6] Jin-Shyan Lee, Yu-Wei Su, and Chung-Chou Shen, “Comparative Study of Wireless Protocols: Bluetooth, UWB, Zigbee, and Wi-Fi”, published in the 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON) Nov. 5-8, 2007, Taipei, Taiwan.
- [7] ZigBee Technology Overview Available at: http://www.senanetworks.com/products/industrial_zigbee/ZigBee_Technology_Overview_for_SENA_Web_Page.V1.pdf.
- [8] Aderemi A. Atayero, and Matthew K. Luka “Applications of Soft Computing in Mobile and Wireless Communications” published in International Journal of Computer Applications (0975 – 8887), Volume 45– No.22, May 2012.
- [9] R.Nagarajan, Member, IEEE and R.Dhanasekaran, Senior Member, “Optimization of Power in Wireless Industrial Automation Networks” presented in IEEE ICCSP conference, 2015.
- [10] Dharendra Pratap Singh, Yash Agrawal and Surender Kumar Soni, “Energy Optimization in Zigbee using Prediction based Shortest Path Routing Algorithm”, 3rd National Conference on Emerging Trends and Applications in Computer Science, 2012.
- [11] Mohamed KASRAOUI, Adnane CABANI, Joseph MOUZNA, IEEE International Conference on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing, 2012.
- [12] Antonio M. Ortiz, Fernando Royo, Teresa Olivares and Luis Orozco-Barbosa, “Intelligent Route Discovery for Zigbee Mesh Networks”, published in IEEE, 978-1-4577-0351-5/11/ 2011.
- [13] Ata Elahi and Adam Gschwender, “Introduction to the ZigBee Wireless Sensor and Control Network” first edition, ISBN-13: 978-0137134854.
- [14] Chia-Hung Tsai, Meng-Shiuan Pan, Yi-Chen Lu, and Yu-Chee Tseng, “Self-Learning Routing for ZigBee Wireless Mesh networks”, National Chiao-Tung University, Hsin-Chu, 300, Taiwan.
- [15] Prativa Saraswala, “A Survey on Routing Protocols in Zigbee Network”, published in International Journal of Engineering Science and Innovative Technology (IJESIT), Volume 2, Issue 1, January 2013.
- [16] Taehong Kim ,Seong Hoon Kim, Jinyoung Yang , Seong-eun Yoo and Daeyoung Kim, “Neighbor Table Based Shortcut Tree Routing in ZigBee Wireless Networks”, Published in: IEEE Transactions on Parallel and Distributed Systems (Volume: 25, Issue: 3, March 2014).

- [17] Antonio M. Ortiz, Teresa Olivares and Luis Orozco-Barbosa, "Smart Routing Mechanism for Green ZigBee-based Wireless Sensor Networks", Albacete Research Institute of Informatics.
- [18] Jianpo Li, Xuning Zhu, Ning Tang, "Study on Zigbee Network Architecture and Routing Algorithm", published in 2010 2nd International Conference on Signal Processing Systems (ICSPS).
- [19] Yi Gong Peng, Ying Li Li, Zhong Cheng Lu, Jin Shou Yu, "Method for Saving Energy in Zigbee Network", published in IEEE, 978-1-4244-3693-4/09/2009.
- [20] Renu Kachhoria, "An Implementation in AODV based on Active Route Timeout between sensor nodes in Wireless Sensor Networks", published in International Journal on Computer Science and engineering (IJCS), Vol. 3 No. 3 Mar 2011.
- [21] Zheng Sun, Xiao-guang Zhang, Dianxu Ruan, "A Routing Protocol based on Flooding and AODV in ZigBee network", published in IEEE, 978-1-4244-3894-5/09/2009.
- [22] Jun Xiao and Xiaojun Liu, "The Research of E-AOMDVjr Routing Algorithm in ZigBee Network", published in IEEE, Chinese Control and Decision Conference (CCDC 2011), 978-1-4244-8738-7/11/26.00, pp. 2360-2365.
- [23] Maged Salah Eldin Soliman, Sherine Mohamed Abd El-kader, Hussein Sherif Eissa, Hoda Anis Baraka, "NEW ADAPTIVE ROUTING PROTOCOL FOR MANET", published in Ubiquitous Computing and Communication Journal Volume 2, Number 3, pp-16-22.
- [24] C. Perkins and E. Royer, "Ad Hoc On-demand Distance Vector (AODV) Routing", Internet Draft, MANET working group, draft-ietf-manetaodv-05.txt, March 2000.
- [25] C. E. Perkins, E. M. Royer, and S. R. Das, "Ad hoc on demand distance vector (AODV) routing," IETF Internet draft, draft-ietf-manetaodv-13.txt, Feb. 2003.
- [26] "Ad hoc On Demand Distance Vector (AODV) Routing Protocol", Chapter 6 from Book titled 'Ad Hoc Networking', published by Perkins, Addison Wesley, 2001.
- [27] M. Ohlin, D. Henriksson, and A. Cervin, "TrueTime 2.1 Reference Manual," Department of Automatic Control, Lund University, Sweden, 2007. [Online] Available: <http://www.control.lth.se/truetime>.
- [28] Anderson, Martin, Dan Henriksson, Anton Cervin and Karl-Erik arzen (2005). "Simulation of wireless networked control systems", Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference, Seville, Spain, 2005.
- [29] Tsukamoto Y. "An approach to fuzzy reasoning method", Gupta MM, Radade RK, Yager RR, eds Advances in fuzzy sets theory and applications, 1st edition New York, 1979. P. 137-149.
- [30] Lambert Torres G, Mukhedkar D. "On line fuzzy rule set generator". Advance Information Processing in Automatic Control Congress of IFAC, Nancy, France 1989.
- [31] Kwang Y. Lee, Mohamed A.EI- Sharkawi, "modern heuristic optimization techniques, theory and applications to power system".
- [32] Prativa P. Saraswala, Dr. Dharmistha D. Vishwakarma and Prof. Satish K. Shah, "Evaluation of Routing Protocol Performance for ZigBee Network Using Fuzzy Logic in MATLAB/TRUE TIME", published in International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 10, and October 2013.
- [33] Mark Hudson Beale, Martin T. Hagan and Howard B. Demuth, "User's Guide on Neural Network Toolbox™ 7".
- [34] Yizheng Xu, Graduate Student Member, IEEE, and Jovica V. Milanovic, Fellow, IEEE, "Artificial- intelligence-Based Methodology for Load Disaggregation at Bulk Supply Point", published in IEEE transactions on power systems, Vol. 30, MARCH 2015
- [35] Saduf, Mohd Arif Wani, "Comparative Study of Back Propagation Learning Algorithms for Neural Networks", pp. 1151-1156, Volume 3, Issue 12, December 2013.
- [36] M. Riedmiller, H. Braun, "A direct adaptive method for faster back-propagation learning: the RPROP algorithm," Proc. of the International Conference on Neural Networks, pp. 586-591, 1993.
- [37] Prativa P. Saraswala, Jaymin Bhalani and Sandhya Sharma, "Comparative Performance Analysis of AODV Parameter for Zigbee Network using Artificial Neural Network", published in International Journal of Computer Applications (0975 - 8887), Volume 140 - No.6, April 2016.
- [38] Pejman Tahmasebi, Ardeshir Hezarkhani, "Application of Adaptive Neuro-Fuzzy Inference System for Grade Estimation; Case Study, Sarcheshmeh Porphyry Copper Deposit, Kerman, Iran", published in Australian Journal of Basic and Applied Sciences, ISSN 1991-8178, pp. 408-420, 2010.
- [39] Yevgeniy, Gershteyn Larisa Perman, "Matlab: ANFIS Toolbox", presented on 04/22/2003.
- [40] Jyh-Shing Roger Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System", IEEE transactions on systems, man and cybernetics, vol. 23, may-june 1993.
- [41] Dr. Bob John, "ANFISnote" [Online]. Available www.scribd.com/doc/134927592/ANFISnote.
- [42] The Fuzzy Logic Toolbox for use with MATLAB, J.S.R. Jang and N. Gulley, Natick, MA: The Math Works Inc., 1995.
- [43] Expert System, Neural, Fuzzy system by Jamil Al Agha. Available online: <http://www.slideshare.net/JamiLALagha/expert-system-neural-fuzzy-system>.
- [44] Presentation on "Neuro- Fuzzy System", available at: <https://www.slideshare.net/sagahire/neurofuzzy-systems>, 2012.

AUTHORS PROFILE

Prativa P. Sraswala: She is a research scholar in V.T Patel Dept. of Electronics and Communication Charusat, Changa, Gujarat, India. She is working in the field of wireless communication. She is working as an Assistant Professor in the department of Electrical Engg., Faculty of Tech. and Engg., Maharaja Sayajirao University of Baroda, Gujarat, India.

Dr Jaymin Bhalani: Professor, EC Dept., Babaria Institute of Technology, Vadodara, Gujarat, India.

Forensic Analysis of Unmanned Aerial Vehicle to Obtain GPS Log Data as Digital Evidence

Septyan Eka Prastyana

Departement of Informatics
Universitas Islam Indonesia
Yogyakarta, Indonesia
septyan_ijyas@rocketmail.com

Imam Riadi

Department of Information System
Ahmad Dahlan University
Yogyakarta, Indonesia
imam.riadi@is.uad.ac.id

Ahmad Luthfi

Departement of Informatics
Universitas Islam Indonesia
Yogyakarta, Indonesia
Ahmad.luthfi@uii.ac.id

Abstract— In recent years, the use of drones by civilians is increasing rapidly by the presentation of total sales continued to increase rapidly every year. With the increasing possibility of Unmanned Aerial Vehicle (UAV) abuse, crime in the use of UAVs to be larger. Through forensic analysis of data using static forensic and live forensic to obtain data that allows it to be used as digital evidence. To dig up information that could be used as digital evidence in the UAV and controllers, as well as to know the characteristics of digital evidence on a UAV. The results showed that digital evidence on a UAV, the smartphone is used as a controller UAV has a very important role in the investigation. The findings in aircraft has a percentage of 50% and a camera memory card with 16.6%. DJI Phantom 3 Advanced GPS coordinates always store data in flight LOG; the data is always stored even when the flight mode is used does not use GPS signals to stability. Due to DJI Phantom 3 Advanced always use GPS on flights, file, image or video captured by the camera has the best GPS location coordinates to the metadata therein.

Keywords; UAV; Log; Forensic; GPS; Flight Data

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) or drones are also called, is a small aircraft without a pilot. This is now the most widely used in the military and those fans the hobby of photography/videography. In recent years, the use of drones by civilians increased rapidly, until it is mentioned by the UK House of Lords that 2014 is the "Year of the drones" [1].

Forbes magazine in 2015 wrote on its Web site sales distribution one trademark holder drones from its inception to the present. At the beginning of sales in 2009 to 2010, the presentation of the annual income more than 50% of overall sales are in North American. Moreover, in 2011 the total annual sales presentation to increase up to 280%, while sales in the North American presentation of only about 30% of total sales. Presentation of total sales continued to increase rapidly every year, the sale of drones in 2020 is expected to touch \$ 2.28 Billion [2].

Drone works with two parts; the first is the drone itself and a controller that functions to control the drones. In some types of drones, there is already no longer requires a controller for controlling the aircraft. This type of drone using GPS transmitters installed on the users, so when this is enabled drones, aircraft will automatically follow any direction from people who used the transmitter. Drone with works like this

has many sensors on the aircraft, which is useful to keep the drones remain safe from the surrounding environment when flying follow users.

Digital evidence that can be taken from the body of drones and the controller is the ID of the drone itself, the location where the drone ever flew, image or video is taken when drones were flown, log update the software used. While the controller can be found in the form of digital evidence storage of images or videos taken using drones, log the location of the use of drones, the software used to control the drone, the ID of the drones are connected.

Realm of this study is to gather information and conducts an analysis of digital evidence contained on drones along its controller by using static forensic and lives forensic with efforts to help complete the information on the forensic activity that uses GPS on the drone.

II. RELATED WORK

Several studies have been done on a UAV. Research about the small quadcopter demonstrates the utility of UAVs to safely and accurately mapped physical and biological characteristics of the unique habitat[3]. Then research explores the basics of the estimates and the flight controls for small winged UAV remains that covers common sensor and sensor configuration used small UAVs to be estimated[4]. Other studies have also been conducted, which is about the theory and practice of spoofing of Unmanned Aerial Vehicle (UAV) is captured and controlled by using a signal Global Positioning System (GPS)[5].

In 2014 conducted research to discover forensic methods in the search for artifacts that may be used for digital evidence on the device Garmin and Tom Tom satnav. The results obtained in the form of the acquisition method and the analysis and comparison of data obtained in the navigation system Garmin and Tom Tom[6]. Similar research is also done by exploration of digital evidence on Android Smartphone with through several stages. Results obtained in the form of technical image acquisition and analysis of digital evidence GPS on Android Smartphone, the application framework for the investigation, and provides several options framework that can be used[7].

Another study conducted on GPS data are based on a study reported in the relevant publication and focuses on variable directly to a GPS device such as clouds, weather, obstructions,

signal split and user preferences, and test the accuracy of three GPS devices[8].

III. BASIC THEORY

A. Log

Log files become a standard part of large and very important applications in the operating system, computer networks, and distributed systems. The log file is the only way how to identify and locate faults in the software because the log file analysis is not affected by the issues based on the time known as the probe effect. This contrasts with the analysis of the program when the analytical process can disrupt critical condition time or critical resources in the program being analyzed.

Log files are often very large and can have a very complex structure. Although the process of generating a log file is quite simple and straightforward, log file analysis can be a tremendous task requires very large computational resources, long and sophisticated procedures. This often leads to a common situation, when the log file is generated and continues to occupy precious space on a storage device, but no one uses them and utilizes the enclosed information [9]. The log has a large size. Therefore it is necessary steps to facilitate the process of storage and retrieval of information in databases [10].

Log (record keeping) is a file that records events in the computer program. Meanwhile, according to the definition of the log is a record of daily activities. Activities that are recorded directly called the transaction log. The log file can be used as a support in the process of cyber forensics to obtain digital evidence during the investigation stage [11].

While the GPS log is a collection of GPS points, each GPS point containing latitude, longitude, and timestamp [12].

B. GPS Forensics

The GPS-enabled device uses satellite readings to determine the geocentric receiver. Coordinates associated with the center of the earth, and the information to be read by some satellites, which optimally at least four satellites. Gps is defined as a group of satellites in earth orbit that sends the right signal, to enable GPS receivers to calculate and display accurate location, time and speed information to the user [8].

GPS evidence is digital evidence capable of determining a particular geographic location with incredible accuracy. It shows directly to a user's location, so it is easy to find, as well as with a particular user is sought in criminal cases. Evidence in the form of data GPS latitude and longitude [7].

C. Unmanned Aerial Vehicles (UAVs)

Unmanned Aerial Vehicles (UAV) is an unmanned flying vehicle. As for civilian purposes, UAVs can be used for mapping isolated areas, volcano monitoring, monitoring of congestion or shooting area after the tsunami disaster. UAV system consists of the air vehicle (aircraft), payload and control station [13].

Hartmann and Steup [14] describes the flow of information between system components UAV with its ground station that shows in Figure 1.

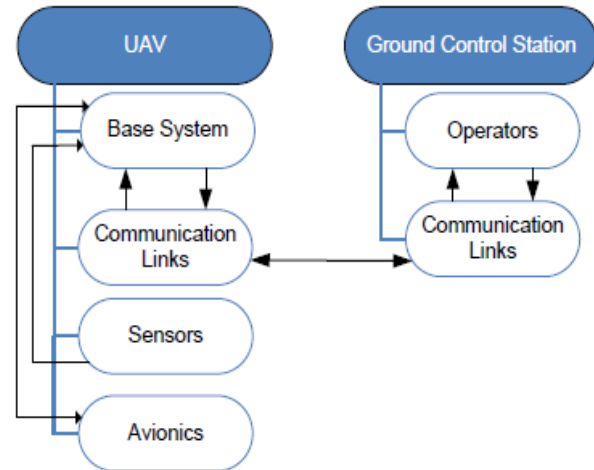


Figure 1 Flow of Information between UAV with Ground Station

"UAV base system" is the basis of the UAV, linking together the components UAV. It is necessary to allow inter-component communication and sensor control, navigation, avionics and communication systems. This can be considered as a UAV "operating system" [14].

D. Flight Data

Jiang and Huang [15] explains that in a conventional aircraft identification system, the various testing technology required for flight data. In general, these technologies can be divided into two methods that measure the parameters of the external and internal parameter measurements. Instantaneous position, trajectory, velocity, and acceleration, etc. can be measured with an external parameter measurement. This data can then be compared with data measured by the air system to test the accuracy of the air system. Measurement parameters of external equipment include photography, radar measurements, laser measurement and others. Measurement parameters include the internal apparatus of global positioning system (GPS) receiver, angular velocity gyroscope, accelerometer, an angular accelerometer, altimeter, airspeed meters and beyond.

IV. RESEARCH METHODS

To support experiment on this research, hardware, and software that necessary used is listed below:

- DJI Phantom 3 Advanced and controller.
- Android smartphone and PC.
- DJI GO for Android version 2.8.4(415)
- FTK Imager version 3.4.2
- DatCon version 2.3.0
- PhotoMe version 0.79R17

To simulate the scenario created from the use of drones, while the UAV usage scenarios in this study will be described as a Figure 2.

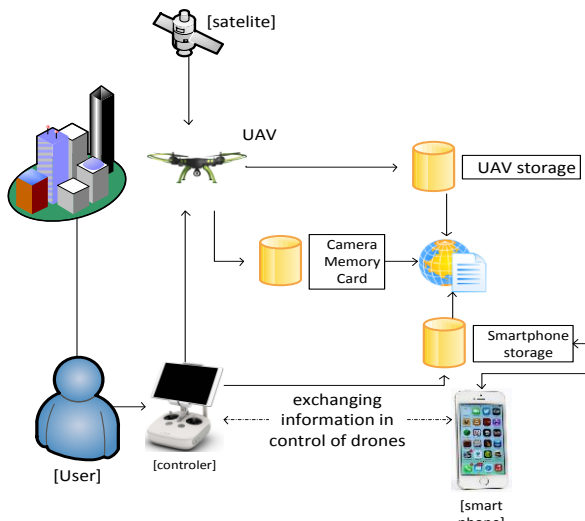


Figure 2 usage scenario of UAV

Scenarios designed in this study are operated a drone to do some flying with different flying mode and taking pictures and video on the site. In each of these locations, when flown navigation sensors inside drone receive location data from GPS and GLONASS satellites which then stored into the database on drones. Controller and smartphone as a ground station are used as a controller and video signal receiver of aircraft. All data received from the UAV ground station then stored in a database on the smartphone as a signal receiver

V. RESULT

A. Scenario

Scenarios used in this study through the multiple activities carried out using three different modes during the flight. The first flight is done by using the P-mode (Positioning), which in this mode using GPS and Vision Position System works together. In this mode, there are three circumstances that are automatically selected by DJI Phantom 3 Advanced based on the signal strength of GPS and Vision Positioning Sensor. As for the form of three circumstances:

- P-GPS: GPS Positioning and Vision sensors are available in this mode UAV using GPS for the position.
- P-OPTI: Vision Positioning available but GPS signal strength is not sufficient, in this mode using only UAV Vision Positioning System for the position.
- P-ATTI: GPS signal and Vision Positioning is not available in this mode using only UAV barometer for the position, so only the height that can be stabilized.

B. Acquisition

The process of acquisition of the UAV performed in three parts, the first of the aircraft used during flight. The second camera storage media used in flight. Moreover, the third is on the controller or ground station that is here is to use a smartphone Lenovo P70.

The acquisition process is in aircraft storage and the memory card found on the aircraft done by way of physical

(sectors per sector or bit-stream copy) so that the imaging results will be the same as physical evidence. Imaging files saved with the extension .dd.

C. Analysis

1) GPS evidence extraction

In this study, it was found that the logs that contain GPS information data have the file extension DAT on aircraft storage and file with a .txt extension on the smartphone. Log data on storage UAV with DAT extension that contains GPS information stored in the directory /root/FLY019.DAT. Whereas the log results on a smartphone with a .txt extension stored in the directory /root/DJI/dji.pilot/FlightRecord/DJIFlightRecord_2016-08-29_[16-25-49].txt.

a) P-mode (Position)

In the log found on the aircraft analyzed by reading the file FLY019.log.txt outcome use of the DatCon application on file FLY019.DAT known flight mode, the location of the home point recorded, and the duration of the flight. For more detail can be seen in Figure 3.

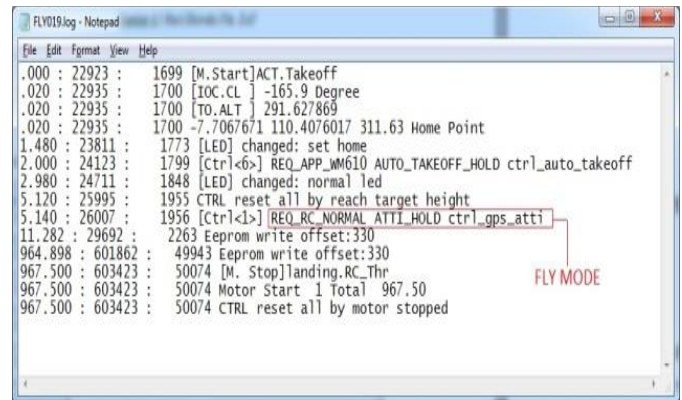


Figure 3 Results Event Log on File FLY019.DAT

In the logs found on the UAV's storage media can be seen the collection of coordinates and flight path conducted by uploading a log file FLY019.DAT to web applications based on https://www.mapsmadeeasy.com/log_viewer to get a list of coordinates to a .csv file. The examples of GPS log data can be seen in Table 1.

TABLE I SAMPLE COORDINATES GPS LOG RESULTS ON FILE FLY19.DAT

Longitude	Latitude	Altitude (m)
110.407602	-7.70676706	292.17404
110.407602	-7.70676696	291.9357
110.407603	-7.70676250	378.17572
110.407603	-7.70676783	400.34015
110.407602	-7.70676827	408.8675
110.407606	-7.70676712	387.056

b) A-mode (Attitude)

In the log found on the aircraft analyzed by reading the file FLY021.log.txt outcome use of the DatCon application on file FLY021.DAT known flight mode, the location of the home point recorded, and the duration of the flight. For more detail can be seen in Figure 4.

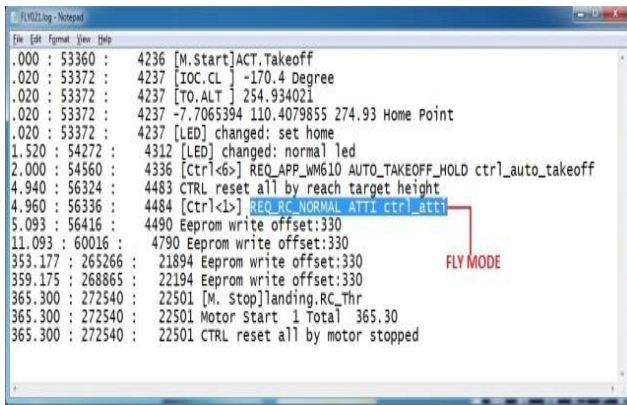


Figure 4 Results Event Log on File FLY021.DAT

For the flight path with this mode can still be found using the same process as before. This is by uploading a log file FLY021.DAT to web applications based on https://www.mapsmadeeasy.com/log_viewer address to get a list of coordinates to a .csv file. The examples of GPS log data can be seen in Table 2.

TABLE II SAMPLE COORDINATES GPS LOG RESULTS ON FILE FLY021.DAT

Longitude	Latitude	Altitude (m)
110.407974	-7.70653814	256.1733
110.407877	-7.70667488	280.66656
110.407814	-7.70676682	279.22858
110.407825	-7.70669853	281.3734
110.407841	-7.70661924	282.16977
110.407827	-7.70658926	282.1325

c) F-mode (Function)

For a flight path with this mode can still be found using the same process as before. That is by uploading a log file FLY022.DAT to web applications based on https://www.mapsmadeeasy.com/log_viewer address to get a list of coordinates to a .csv file. The examples of GPS log data can be seen in Table 3.

TABLE III SAMPLE COORDINATES GPS LOG RESULTS ON FILE FLY022.DAT

Longitude	Latitude	Altitude (m)
110.407070	-7.70664311	271.31238
110.407070	-7.70664396	277.22327
110.407069	-7.70664478	277.67554
110.407073	-7.70664458	286.21893
110.407111	-7.70661816	286.2421
110.407074	-7.70650167	283.9515

2) GPS evidence conversion

Further analysis done is to convert log files other than that contained in aircraft storage, camera memory cards, and smartphones. This method is done by reading the file containing metadata or GPS location information; the file may include images, video, and others. In this process, after the captured image files in the storage UAV, and smartphones are found. The file exported using FTK Imager application to read metadata therein by using the application PhotoMe. In detail, the information on file .dd GPS coordinates can be seen in Figure 5.

Field	Content	Tag-ID	Tag Name	Data Format
GPS tag version	Version 3.2	0000	GPSVersionID	BYTE(4)
North or South Latitude	South latitude	0001	GPSLatitudeRef	ASCII(2)
Latitude	7° 42' 21.647"	0002	GPSLatitude	RATIONAL(3)
East or West Longitude	East longitude	0003	GPSLongitudeRef	ASCII(2)
Longitude	110° 24' 27.4888"	0004	GPSLongitude	RATIONAL(3)
Altitude reference	Sea level	0005	GPSAltitudeRef	BYTE
Altitude	328.963 m	0006	GPSAltitude	RATIONAL

Figure 5 GPS Information on File Image

From the result of the conversion files from storage media both the drone and controller Investigators can strengthen the evidence obtained from the GPS information found in the case of crime in the use of drones.

D. Analysis Result

After going through the process of extraction and conversion of digital evidence the GPS, the next stage performed is present in the form of presentation. Digital forensic presentation in a series of activities carried out by forensic experts in demonstrating his findings in court to explain a case in assisting judges in making decisions. Presentation of GPS digital evidence for a different shape GPS presented visually using Google maps or applications that are relevant in presenting the coordinates of the location.

1) Data log on smartphone

This research presentation made by using the website address <https://healthydrones.com/> to display flight information stored in log files DJIFlightRecord_2016-08-31_[17-00-34].txt contained in smartphones. For more details, flight information can be seen in Figure 6.

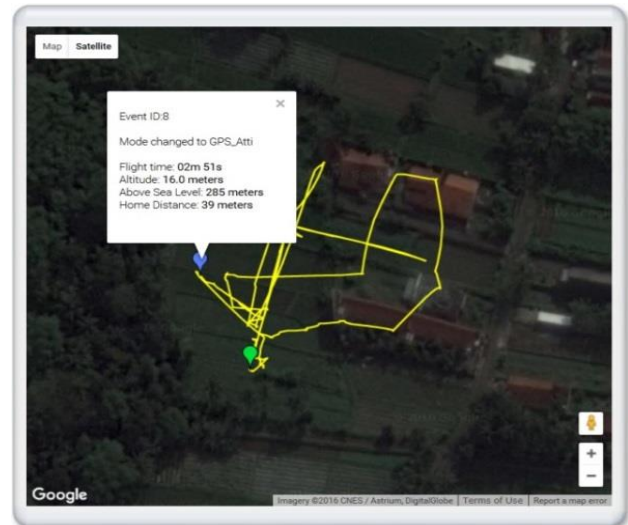


Figure 6 Presentation Log File DJIFlightRecord_2016-08-31_[17-00-34].txt

Presentation of the results can be known UAV flight path. In a green dot "F-mode" is executed and the function Follow me on UAV work, and on the blue dots altitude of 16 meters

and as much distance as 39 meters from the home point, the F-mode switched off and changed using P-mode.

2) Data log on UAV storage

To log file with the extension DAT contained in the storage UAV can use https://www.mapsmadeeasy.com/log_viewer site address in the presenting location, flight path, speed, altitude, and a wide range of useful information as evidence UAV. For more details log FLY021.DAT data presentation can be seen in Figure 7.



Figure 7 Presentation Log File FLY021.DAT

In the log data with the extension DAT, the information displayed are still not rich when compared with the information obtained from the log of the controller (smartphone) which is used to control the flight.

3) The result of image conversion on UAV

For GPS information, the presentation of the results of conversion image files taken by UAVs camera, use the Google Maps app. In Figure 8 can be seen results location of the coordinates contained in the metadata file `org_a8ccc30f7ce0c44f_1472617871000.jpg`



Figure 8 presentations of Information GPS Data File
`org_a8ccc30f7ce0c44f_1472617871000.jpg`

Results of the presentations can be known location coordinates of $-7.706890, 110.408255$. The location of the pictures taken is in Sardonharjo, Nganglik, Sleman Regency, and Special Region of Yogyakarta.

E. Characteristics of Digital Evidence

The characteristics of GPS digital evidence from the UAV is known in detail as follows:

Evidence of digital GPS is very susceptible to changes in migration; a little movement can influence the change point of latitude and longitude.

In the DJI Phantom 3 Advance, GPS coordinates and flight data information is always written into the log in aircraft storage and smartphone. Data will continue to be written and saved even used flight mode does not use GPS signals to aircraft stability.

On aircraft storage, flight log information is stored in the file with the extension DAT, while the flight log information on the controller is stored in the application folder DJI with a .txt extension.

At the camera's memory card found on the UAVs, GPS digital evidence that must be obtained through a process conversion of metadata from the pictures or video files contained therein. Digital evidence in storage is only in the form of latitude and longitude coordinates of the location at the time of an image or video was taken.

Digital evidence information obtained from a smartphone as a controller has a wealth of information that is more than the information obtained from the aircraft storage or a memory card in the camera UAVs.

As the result of analysis, a detailed comparison of the information obtained from the UAV storage, camera memory cards, and storage media within DJI applications on smartphones can be seen in Table 4.

TABLE IV COMPARISON OF DIGITAL EVIDENCE IN UAV

Digital Evidence Informations	Storage		
	UAV	Memory Card	Smartphone
Acquisition Method	Live	Static	Live / Static
Type of Image	Physical	Physical	Logical
Image Format	.dd	.dd	.ad1
Acquisition Tool	FTK Imager	FTK Imager	FTK Imager
GPS Location	√	√	√
Log coordinate flight path	√	-	√
UAV configuration information	√	-	√
Pictures/ Videos	-	√	√
Flight Mode Information	√	-	√
UAV user information	-	-	√
UAV flight data information	√	-	√
Directions shooting	-	-	√
UAV signal strength information	-	-	√
Information UAV sensor condition	-	-	√
UAV power condition information	√	-	√
Information condition UAV controller	-	-	√

Information included in the storage controller (smartphone) has more potential as digital evidence. To obtain digital evidence UAV flight path can be found in the storage UAV and smartphones.

VI. CONCLUSION

Forensics on the UAV can be done by using a static method on a forensic acquisition UAV storage media devices. Except for aircraft storage can not use static methods forensics because the system is turned on in order to be accessible.

GPS data which could potentially be used as digital evidence is always stored in the system log UAV contained on aircraft storage, memory cards and smartphones. GPS data is always stored information even if the system uses UAV flight mode without using GPS. From this research, it has been known to the potential of the whole digital evidence contained information on the device UAV. Percentage of digital evidence found in the storage UAVs has 50% of the overall findings and on the memory cards has 16.6% of the overall findings.

While most digital evidence found to exist on the storage of smartphones used as UAV controllers. In the smartphone storage media found almost the whole information that can be obtained from the UAV.

REFERENCES

- [1] G. Horsman, "Unmanned aerial vehicles: A preliminary analysis of forensic challenges," *Digit. Investig.*, vol. 16, pp. 1–11, 2016.
- [2] H. Shao, "Drone Overlord Frank Wang On DJI's Milestones, Miscarried GoPro Partnership & Corporate Espionage," *Forbes Asia*, 2015.
- [3] A. Nishar, S. Richards, D. Breen, J. Robertson, and B. Breen, "Thermal infrared imaging of geothermal environments and by an unmanned aerial vehicle (UAV): A case study of the Wairakei – Tauhara geothermal field, Taupo, New Zealand," *Renew. Energy*, vol. 86, pp. 1256–1264, 2016.
- [4] J. D. Barton, "Fundamentals of Small Unmanned Aircraft Flight," *Johns Hopkins Apl Tech. Dig.*, vol. 31, no. 2, pp. 132–149, 2012.
- [5] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys, "Unmanned Aircraft Capture and Control Via GPS Spoofing," *J. F. Robot.*, vol. 31, no. 4, pp. 617–636, 2014.
- [6] A. Arbelet, "Garmin satnavs forensic methods and artifacts: An exploratory study School of Computing," no. August, 2014.
- [7] Sukriadi and Y. Prayudi, "Analysis of Digital Evidence of Global Positioning System (GPS) On Android Smartphone," *Kns&I Stikom*, no. 11, 2014.
- [8] D. Huang, "Evidential Problems with GPS Accuracy: Device Testing," 2013.
- [9] J. Valdman, "Log File Analysis," 2001.
- [10] A. Iswardani and I. Riadi, "Denial Of Service Log Analysis Using Density K-Means Method," vol. 83, no. 2, pp. 299–302, 2016.
- [11] M. Zulfadhilah, "Cyber Profiling using Log Analysis and K-Means Clustering A Case Study Higher Education in Indonesia," vol. 7, no. 7, pp. 430–435, 2016.
- [12] Y. Lou and W. Wang, "Map-Matching for Low-Sampling-Rate GPS Trajectories," no. c, 2009.
- [13] M. A. Lukmana and H. Nurhadi, "Design of Unmanned Aerial Vehicle (UAV)," pp. 1–5.
- [14] K. Hartmann and C. Steup, "The vulnerability of UAVs to cyber attacks - An approach to the risk assessment," *Cyber Confl. (CyCon)*, 2013 5th Int. Conf., pp. 1–23, 2013.
- [15] T. Jiang, J. Li, and K. Huang, "Longitudinal parameter identification of a small unmanned aerial vehicle based on modified particle swarm optimization," *Chinese J. Aeronaut.*, vol. 28, no. 3, pp. 865–873, 2015.

Business Process Modeling: Blueprinting

Sabah Al-Fedaghi

Computer Engineering Department
Kuwait University
Kuwait
sabah.alfedaghi@ku.edu.kw

Abstract—This paper presents a flow-based methodology for capturing processes specified in business process modeling. The proposed methodology is demonstrated through re-modeling of an IBM Blueworks case study. While the Blueworks approach offers a well-proven tool in the field, this should not discourage workers from exploring other ways of thinking about effectively capturing processes. The diagrammatic representation presented here demonstrates a viable methodology in this context. It is hoped this explicit analysis of diverse fundamental approaches will benefit the research in the field and also advance current practices.

Keywords—Software system models; System modeling languages; capturing processes; conceptual model; diagrammatic representation

I. INTRODUCTION

Current tendencies in development of information and communication systems are leaning toward higher levels of abstraction in process modeling. *Process modeling* aims at capturing the reality of processes [1] through diagrams that symbolically depict the capture, manipulation, and handling of data and information between a system and its environment and among system components. Business process modeling is an example of such a process. The term “business” is interchangeable with “organization,” a term applicable to all sorts of organizations such as government agencies and departments, charities, mutuals and cooperatives, etc. [2]. The term Business Process Model refers to a diagram that describes flows and activities in a particular business or organizational unit [2].

Process modeling takes diverse approaches found in the literature as well as in practice; e.g., software engineering, enterprise modeling, knowledge modeling, simulation and quantitative analyses, and workflow systems. Without loss of generality, this paper focuses on process modeling as adopted in *IBM Blueworks* [3-4], which captures business process blueprints and creates BPMN diagrams. Here the methodology is to start the “blueprinting process” [3] by generating *milestones*, followed by identifying *activities* in “discovery sessions” with the process owner and subject matter experts. Identifying milestones reflects a thinking paradigm, a certain way of generating elements, and directs attention to identifying interactions in a manner that affects how we *link* concepts.

The most noteworthy aspect of the new Blueworks Live offering extends it beyond being a platform for process discovery and modelling – and into a platform for process automation and execution. Blueworks ... focused on helping tackle the many dozens or hundreds of lightweight processes that are not properly with IT today. [6]

According to Peisl [6],

More and more customers are using IBM Blueworks Live for business process discovery and modeling, as well as for simplified analysis and ... people LOVE to model their business processes with this tool, which is hosted in the cloud / managed by IBM and does not require any IT support ...

This paper investigates the underlying theoretical principle at the base of the “blueprinting process” and proposes a flow-based methodology of thinking to apply to capturing processes in business process modeling. The proposed methodology is demonstrated through re-modeling of a Blueworks case study. This explicit analysis of diverse basic approaches should benefit the research in this field and could also advance current proven practices such as Blueworks.

The next section introduces a flow-based diagrammatic language to be used both as a thinking style and as a vehicle for depicting the outcome of captured processes. The rest of the paper applies this language to a case study taken from the IBM Blueworks literature.

II. FLOWTHINGS MACHINE (FM)

The FM Model was inspired by the many types of flows running through diverse fields, including information flows, signal flows, and data flows in communication models [7-10]. This model is a diagrammatic schema that uses *flowthings* to represent a range of items, for example, electrical, mechanical, chemical, and thermal signals, circulating blood, prepared and ingested food, shared concepts, pieces of data, activities, and so on. *Flowthings* (hereafter *things*) are defined as what can be *created*, *released*, *transferred*, *processed*, and *received* as stages of a flow machine (hereafter *a machine*), as shown in Figure 1, a generalization of the typical input-process-output model used in many scientific fields (Figure 2).

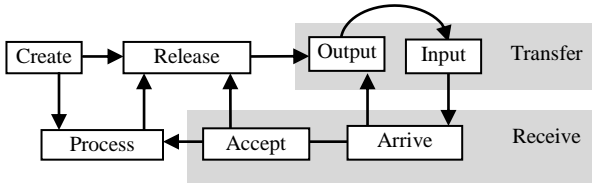


Figure 1. Flow machine

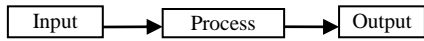


Figure 2. Input-process-output model

The stages of the flow machine can be briefly explained as follows:

Arrive: A thing reaches a new machine, e.g., data flow to a process

Accepted: A thing is approved/not approved to enter a machine, e.g., a datum is not of the right type. If arriving things are always accepted, *Arrive* and *Accept* can be combined as a **Received** stage.

Processed (changed): A flowthing goes through some kind of transformation that changes it without creating a new thing, e.g., different number representation.

Released: A flowthing is marked as ready to be transferred outside the machine.

Transferred: A flowthing is transported somewhere from/to outside the machine.

Created: A new thing is born (created) in a machine.

In general, a flow machine is conceived as an abstract machine that receives, processes, creates, releases, and transfers things. The stages in this machine are mutually exclusive (i.e., a thing in the Process stage cannot be in the Create stage or the Release stage at the same time) with respect to atomic flowthings (things that are not constructed from other things). An additional stage of *Storage* can also be added to any machine to represent the storage of things; however, storage is not an exclusive stage because there can be *stored processed* flowthings, *stored created* flowthings, etc.

Flow machines also use the notions of *spheres and subspheres*. These are the network environments and relationships of machines and submachines, e.g., a stomach is a machine in the digestive system *sphere*. Multiple machines can exist in a sphere if needed. A sphere can be a person, an entity (e.g., a company, a customer), a location (a laboratory or waiting room), a communication medium (a channel, a wire). A flow machine is a subsphere that embodies the flow; it itself has no subspheres.

Example: Bækgaard and Andersen [11] define an *entity* as a chunk of space-time. An *event* is an entity with defined temporal boundaries, e.g., the Second World War. A *process* is an action composed of identifiable sub-events. The authors [11] propose a basic set of *action* types called interaction primitives: “An interaction primitive can be viewed as a pattern that

defines a dynamic relation between two elements. One of the elements performs an action” [11].

The two diagrams shown in Figure 3 represent this type of modeling style. Figure 4 shows the corresponding FM diagrams. In the upper diagram, the book on the shelf flows to the librarian, then to the borrower. The shelf is represented in a barrel shape as way to indicate a storage entity. For simplicity and because only the book flow is shown, the flow machines in *Librarian* and *Borrower* are not enclosed in boxes.

To make the Speaker-Word-Listener diagram (machine) more interesting, we have added *Concepts* and *Understanding* in the lower diagram. The creation of a concept triggers the creation of a word that flows to the listener to trigger the creation of understanding.

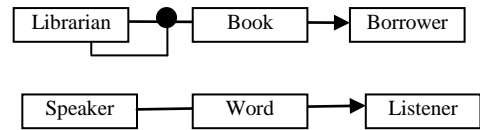


Figure 3. Some modeling cases (redrawn from [11])

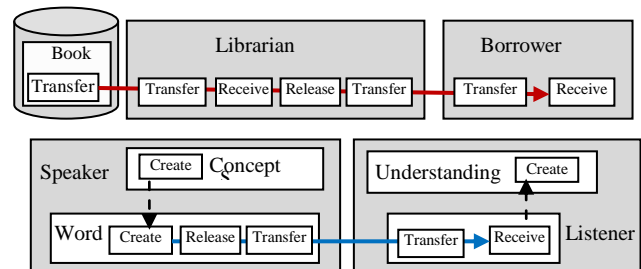


Figure 4. FM representation, with the lower diagram extended to include flow of concepts and understanding

III. CASE STUDY: CALL CENTER COMPANY C

IBM Blueworks (Live) is a flexible, graphical design and visualization tool that enables users to capture business process outlines and create BPMN diagrams.

IBM Blueworks Live does not enforce any process classification framework. How you intend to use the tool for modeling mainly depends on the purpose of your process models. [3]

King et al. [3] give a scenario to demonstrate how to analyze processes and define “solutions” in Blueworks. Their example uses a fictitious company, *Call Center Company C*, that provides call center services to its offices in the US, India, and China. When a large deal is made, Company C cannot serve the huge call volume with its current staff and must hire a significant number of call center representatives (onboarding). The process for hiring new call center employees is manual and not documented.

The onboarding process of call center representatives involves the following players and their functions and requirements:

Recruiter: identifies potential job candidates and manages communications with them.

Job candidate: participates in a job interview, performs a test, and, if he or she accepts a job offer, changes role from job candidate to new hire.

New hire: has successfully passed the hiring process but must attend call center training starting work as a call center representative.

Hiring manager: performs interviews with job candidates and negotiates contracts, and manages communication with the recruiters. If a new hire's probationary work period is unsuccessful, the hiring manager creates a performance plan.

Call center manager: plans the work schedule and runs a probationary review with new employees after 7–10 days of work.

Human resources (HR) administrator: helps new employees get started with their first days on the job and also enters employee information into the employee database.

This process for onboarding of new call center employees was performed manually. A team of analysts was established to redefine it by using first IBM Blueworks and then IBM Business Process Manager for implementation of the process to increase process efficiency.

IBM Blueworks is used as the tool to capture process models. The best practice is to first capture the current state process through discovery, then analyze the process information, and, finally, design a future process for IBM BPM implementation" [3].

King et al. [3] start "the blueprinting process" by generating a default process with two *milestones* and one *activity* in the "discovery map view" which involves discovery sessions with the process owner and subject matter experts. This leads to identifying the start and end points of the process, with no focus on exceptional or error conditions. The facilitator listens to the stakeholders and adds *milestones* and *activities*. A milestone is defined as a logical grouping of activities, including,

- a specific *event* occurring or a *condition met*
- a group of activities that represent a *phase* to deliver an output.

In the case study, on the first pass the requirements are separated into the following *logical groups of activities* (milestones):

- Selection of candidates
- Conduct interviews
- Offer negotiation
- Orientation
- Training
- Probationary work review
- Performance plan initiation

After discussions with stakeholders, milestones are converted to activities, and vice versa, e.g., Selection of candidates: *conduct interviews* and *offer negotiation*. Activity is a generic word and can be atomic (task) or compound (process, subprocess, and so on). Then, the seven milestones might be converted to sub-processes.

"Now that you have a sense of the flow of the process, it is time to start adding details to each of the processes and activities" [3]. Next, the following information is gathered: Who supplies the inputs, What are the inputs, What are the processes or activities, What are the outputs, and Who the outputs are for.

Several diagrams are produced in this methodology, as shown in Figures 5, 6, and 7. For example, Figure 7 represents the background check process, involving a criminal background check, a credit check, and a social media background check. The *HR Admin* conducts the background tests and sends the results to the HR Manager for review.

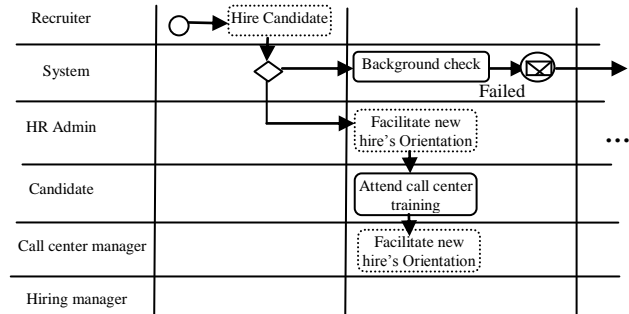


Figure 5. Call center to-be (redrawn, partial from [3])

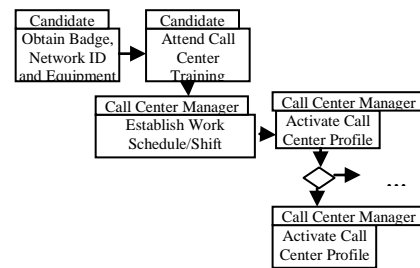


Figure 6. Playback process flow split (redrawn, partial from [3])

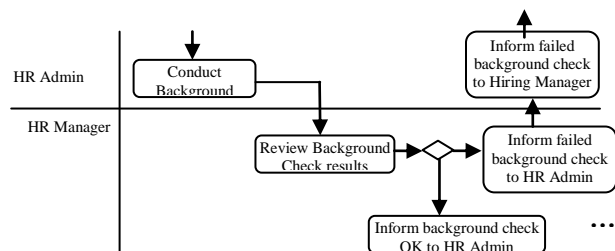


Figure 7. Background check (redrawn, partial from [3])

Clearly, it is not possible to present a fair description of the example, given the limitations of a conference paper; however, this description of the case study along with the diagrams provides a general idea of the methodology adopted to capture business process outlines and create a BPMN diagrams.

In the material that follows, we offer an FM-based alternative methodology that could benefit this type of application. To understand the diagrammatic description of Call Center Company C, we have to juxtapose diagrams resulting from the original seven *milestones*. Because juxtaposing several different diagrams is difficult, our understanding of the details of the study case may be imperfect or incomplete; however, these inaccuracies do not detract from our purpose of contrasting two styles of modeling. Any inaccuracies can be redrawn using the FM model notions.

While the Blueworks approach offers a well-proven tool in the field, this utility should not discourage exploration of other lines of thinking about capturing processes. This would benefit the research and also further current proven methodologies such as Blueworks.

IV. ALTERNATIVE APPROACH

An organization can be described in abstract form as an abstract machine (FM diagram) and realized in reality through its existence over time. *Events*, in FM, are slices of time when sub-machines become “active” to form processes and subprocesses. Figure 8 shows the machine description of *Create job offers* (in black in the online version of the paper). The machine is “activated” by four events: selecting candidates, creating offers, setting deadlines for response, and sending offers. Note that the FM “process” (a stage in the diagram) of an event means that the event *takes its course* after creation. Accordingly, a *process* (in general) is defined as a diagram or subdiagram that incorporates events. Because of its event/time association, a process has a start, a time duration, and an end, and these phases are made obvious in the event descriptions shown in Figure 8.

The Blueworks way of identifying milestones reflects a thinking paradigm that adopts a certain way of generating elements and directs attention to interactions with the world in a way that affects how we link concepts. The Blueworks way of analysis starts as follows:

- Identify milestones (e.g., in this example, select candidates, conduct interviews, negotiate, ...) as disconnected pieces of reality. Then:
- Find the connections among these pieces.

It is a process-less (see the definition of a process given previously) approach that aims at capturing processes in reality.

In this paper, the aim is not to scrutinize the method; rather, the purpose is to explain different philosophies of analysis, as in the case of contrasting the Blueworks approach with the proposed approach in this paper. Use of a certain approach influences the analyst’s thinking by “automatically” constraining meanings.

Our proposed approach in the paper aims at modeling reality through processes that *exhibit themselves through flows* (solid and dashed arrows in FM). For example, in Figure 8, we start with Event 1: selecting candidates, and the triggering *leads* us to Event 2 of creating offer. This event *leads* to

creating a deadline and starting the flow of an offer to the candidate.

To further contrast Blueworks and our proposed approach, consider the following example from [12].

Example: To clarify the milestones-collection methodology of capturing processes in *Call Center Company C*, consider the analogy of a scientist who wants to model the *phosphorus cycle*. In the Blueworks approach, the modeling process is “jump-started” by identifying milestones, as shown in the left diagram of Figure 9. The connections (shown to the right in Figure 9) are found by identifying activities within each milestone.

In contrast, the FM approach begins with the subprocesses of flow of inorganic material in rocks to water, then to ocean, then to sediments, etc., as shown in the steps in Figure 10.

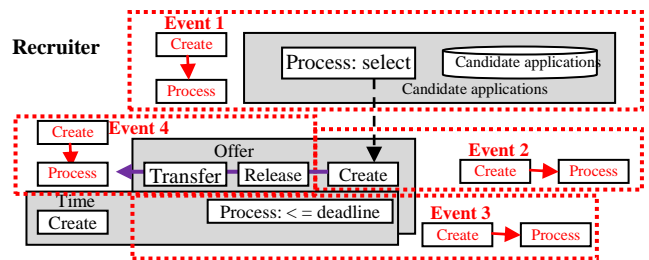


Figure 8. Illustration of a process formed from four events

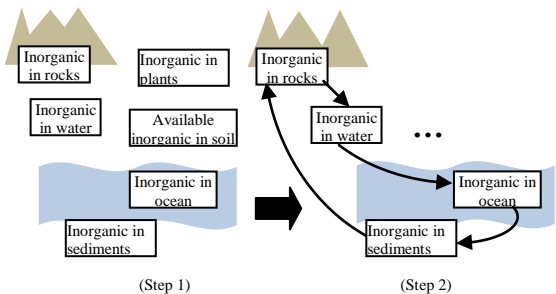


Figure 9. Blueworks approach begins with identifying milestones

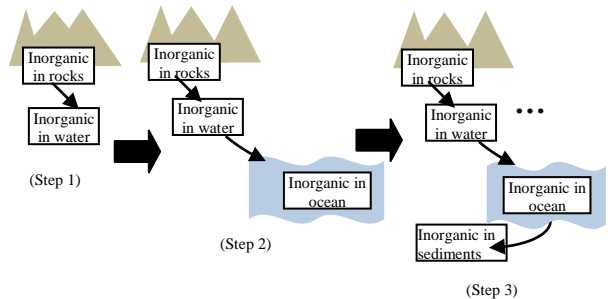


Figure 10. FM approach to capturing processes

V. APPLYING FM TO CALL CENTER COMPANY C

Applying FM to the call center case results in not only a different diagram but also a different developmental method, as explained in the previous section.

Figure 11 shows different phases of building the FM representation. These phases are shown to emphasize the methodology of development. We will focus only on the *final diagram* at the bottom of the figure since it is the culminating representation.

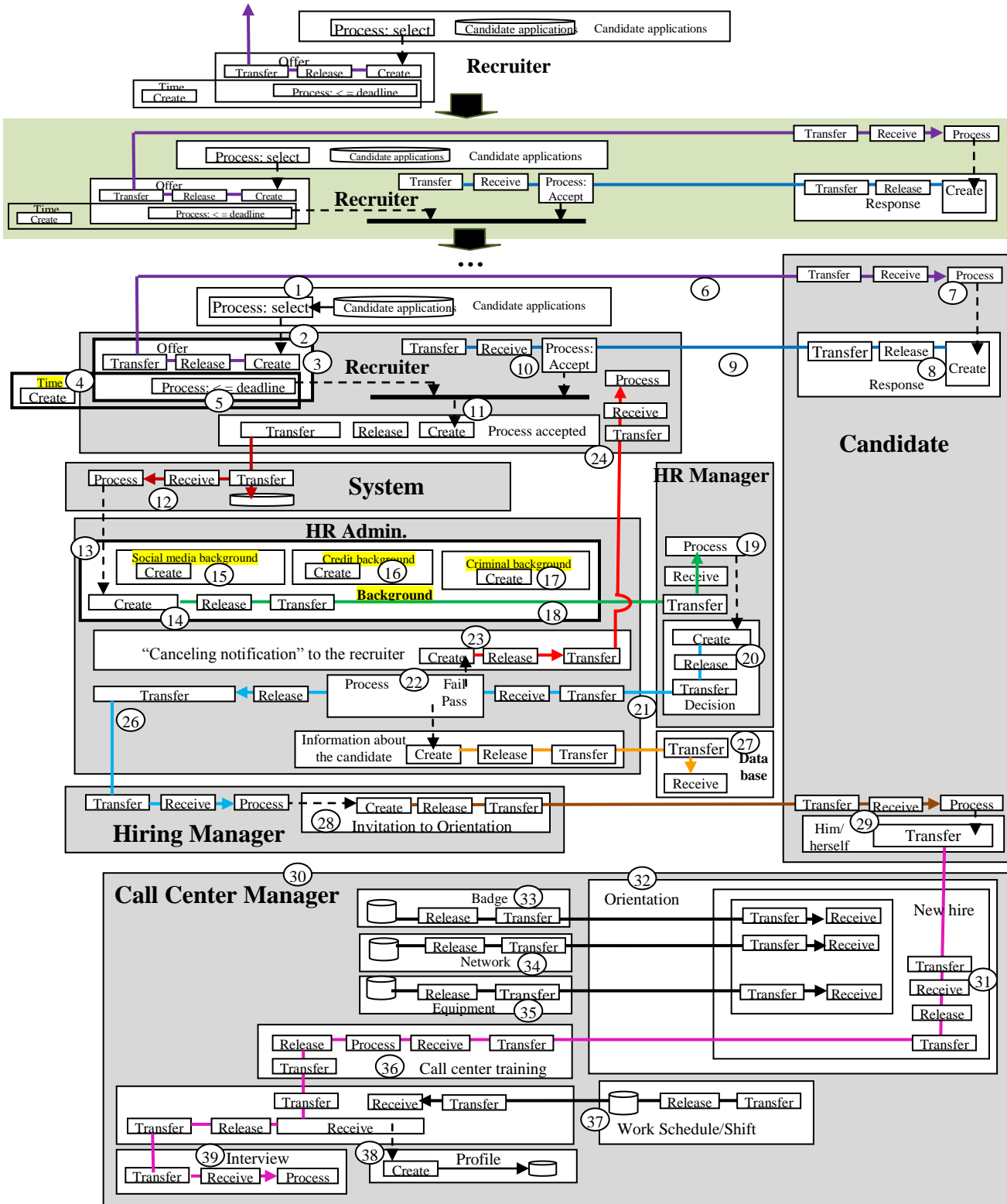


Figure 11. FM representation of Call Center Company C

In Figure 11, the process starts with selection of a candidate (circle 1 in the figure) which triggers (2) the creation of an offer. The offer includes the element of time (4), a thing created universally outside the domain of the recruiter and the processing of time (5) performed by the recruiter. The offer flows to the candidate (6) who processes it (7) to trigger the creation of a response (8) that flows to the recruiter (9). The recruiter processes the response (10); if the candidate has accepted the offer and the deadline period has not expired, then these two conditions trigger (11) the decision to start *the hiring process*. The thick horizontal line indicates that both conditions are realized. The scenario could be modeled in FM as the merging of two flows (of triggering); however, for simplicity, we have used the familiar computer science notation.

This decision to start the hiring process is transferred to the system (12; as described in [3]) and this, in turn, triggers (13) the *HR Admin* to create (14) a background check. The created background check has a structure that is formed from:

- Social media background (15)
- Credit background (16), and
- Criminal background (17).

The background report flows to the HR manager (18), who processes it (19), triggering creation of a hiring decision (20) that flows to the HR manager (21). If the decision is negative (22), a Cancellation notification is sent to the recruiter (23 and 24). If the decision is positive, it is sent to the Hiring manager (26), who registers the candidate's information in the database (27). The Hiring manager also invites (28) the person (29) to an orientation with the Call Center manager (30).

Upon arrival at the call center (31), the person enters an orientation session (32), where he/she is given a badge (33), network account (34), and equipment (35). Then he/she moves to training (36), is given a Work Schedule/Shift (37), creates a profile (38) and finally goes for the interview (39).

VI. CONCLUSION

This paper has presented a flow-based methodology for thinking about how to capture processes in business process modeling. The proposed methodology is demonstrated through remodeling of a Blueworks case study. The resulting flow-based diagrams reveal a way of thinking in system analysis that is based on flows of things passing through flow machines. Accordingly, an analyst's "thought machine" forms a train of thought that is different from other ways of thinking such as the one reflected in the Blueworks approach. The paper emphasizes this thinking style as a unifying method with potentially diverse applications. The remodeling of the Blueworks case sample seems to point to merits that deserve further development.

REFERENCES

[1] O. Svatoš, Conceptual process modeling language: Regulative approach. In Proceedings of the 9th Undergraduate and Graduate Students eConf. and 14th Business & Government Executive Meeting on Innovative Cross-border eRegion, Univ. of Maribor, 2007

[2] A. Chapman, Business process modelling, 2016. <http://www.businessballs.com/business-process-modelling.htm>

[3] J. King, N. Chidambaram, P. Lee, P., Schume, and D. Teran, Process Discovery Best Practices Using IBM Blueworks Live, IBM Redpaper REDP-5111-00, International Technical Support Organization, Sept., 2014.

[4] IBM, Process modeling at enterprise scale using BlueworksLive and IBM BPM, WebSphere User Group, IBM Corporation, 2014. http://www.google.com.kw/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwitm-OXmp_RAhUXN1AKHQEXDroQFggIMAA&url=http%3A%2F%2Fwww.websphereusergroup.co.uk%2Fwug%2Ffiles%2Fpresentations%2F38%2FWUG2014%2FBlueworks%2FLive%2FEnterprise%2FScale.pdf&usq=AFQjCNEqmhhuqyhb7ITzONwEJAIOKhPQ_A

[5] N. Ward-Dutton, Vendor Insight: IBM breaks new ground with Blueworks Live, Premium Advisory Report, Nov. 2010. http://www.google.com.kw/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjT-IPOnJ_RAhVWVeFAKHSb1CukQFggeMAE&url=https%3A%2F%2Fwww.blueworkslive.com%2Fdownload%2Fbpm_vi_IBM_Blueworks_Live_1110.pdf%3FpostId%3Da93c1a523f%26fileItemId%3Da93c1a3981&usq=AFQjCNGGXSERpJmBmb28kwHIU95VWQ6U9Q&bvm=bv.142059868.d.ZWM

[6] R. Peisl, Have fun with collaborative business process modeling in the Cloud with IBM Blueworks Live (BPM), Learnquest Conference, IBM Cloud Technical University, 25–28 October 2016, Madrid, Spain.

[7] S. Al-Fedaghi, "Conceptual modeling in simulation: a representation that assimilates events," *Int. J. Adv. Comput. Sci. Appl.*, Vol. 7, No. 10, 2016, pp. 281-289.

[8] S. Al-Fedaghi, "Toward a philosophy of data for database systems design," *Int. J. Database Theory Appl.*, Vol. 9, No. 10, 2016.

[9] S. Al-Fedaghi, "Diagrammatic modeling language for conceptual design of technical systems: a way to achieve creativity," *Int. Rev. Autom. Control*, Vol. 9, No. 4, 2016.

[10] S. Al-Fedaghi, 2016. "Flowcharting the meaning of logic formulas," *Int. J. Adv. Res. Artif. Intell. (IJARAI)*, Vol. 5, No. 10, Oct, 2016.

[11] L. Bækgaard, and P. B. Andersen, Using Interaction Scenarios to Model Information Systems, Working Paper I-2008-04, Department of Information and Media Studies, University of Aarhus, 2008. http://www.google.com.kw/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0ahUKEwj-kfrFop3RAhULIsAKHW5WDusQFggMAM&url=http%3A%2F%2Fpu.re.au.dk%2Fportal%2Ffiles%2F2570%2FI_2008_04.PDF&usq=AFQjC NFqJFyjBumdKklU5ZR5YHiCEN90Eg

[12] Science Learning Hub, The phosphorus cycle, University of Waikato, 2007, <http://sciencelearn.org.nz/Contexts/Soil-Farming-and-Science/Sci-Media/Images/The-phosphorus-cycle>

AUTHORS PROFILE

Dr. Sabah Al-Fedaghi holds an MS and a PhD in computer science from the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois, and a BS in Engineering Science from Arizona State University, Tempe. He has published two books and more than 260 papers in journals and conferences on software engineering, database systems, information systems, computer/information privacy, security and assurance, information warfare, and conceptual modeling. He is an associate professor in the Computer Engineering Department, Kuwait University. He previously worked as a programmer at the Kuwait Oil Company and headed the Electrical and Computer Engineering Department (1991–1994) and the Computer Engineering Department (2000–2007).

ZigBee Nodes' EMI Immunity by Automatic Channel Selection Algorithm Paper

Ahmad Habboush
Computer Science Dept.
Faculty of IT
Jerash University
ahmad_ram2001@yahoo.com

Mohammad Alnabhan
Computer Science Dept.
Faculty of IT
Mutah University
m.alnabhan@yahoo.com

Mohammad Otoum
Computer Science Dept.
Faculty of IT
Irbid National University
moh_atoom1979@yahoo.com

Abstract— Electromagnetic inference (EMI) is a major issue for all electronic devices, and it represents a major problem in wireless communication schemes. ZigBee is a wireless communication technology based on IEEE 802.15.4 standard, mainly using 2.4 GHz band in most ISM and electronic applications, but the performance of ZigBee nodes is seriously impaired by the interference generated in this channel frequency. This paper investigates the interference immunity of ZigBee devices and presents a new adaptive algorithm for selecting a new channel with low interference levels during network transmissions. The proposed algorithm achieves high packet delivery rates and low delay ratio within different levels of channel interference.

Keywords-component; Wireless Communications, Zigbee, WiFi, Interference.

I. INTRODUCTION (HEADING 1)

Wireless communication is used in various systems and applications, with wireless sensors networks (WSN) receiving particularly high attention in the most advanced research [1]. ZigBee technology is a special wireless communication module whose specifications were instituted for optimum compatibility with WSN [2, 3].

ZigBee is considered a low-cost, low-power wireless network stack up operating in ISM 2.4GHz frequency band. ZigBee nowadays is being embedded in costumer electronics, home automation and medical applications etc. ZigBee standard has four basic layers: Physical (PHY), Media Access Control (MAC), Network (NWK) and Application (API). Each layer adds its header to the transmitted frame [4]. Physical layer activates and deactivates the transceiver to transmit data in radio frequency with (2.4-2.48) GHz band divided into sixteen channels, with 5MHz band guards between each channel. The transmitted signal power is about 3.5 dbm and QPSK modulation is used. The physical layer uses a low data rate of 250Kbps, and tests the energy in each channel to determine noise. In this concern, PHY uses Clear Channel Assessment (CCA), adopting Carrier Sense Multiple Access with Collision Avoidance (CSMA-CA).

MAC layer handles all access to the physical radio channel and generates network beacons if the device is a coordinator, supporting Personal Area Network Identifier (PANID) for each network, and this layer also supports multi-level security. The encryption and decryption topology is implemented in MAC layer. The Application layer uses all previous layers for specific applications like home automation and wireless sensor network etc., and implements the API interface with the network layer for data and management services. The basic structure of ZigBee is based on a single channel for the network operation, determined once, and changing it requires modifying network structure. The fixed frequency channel of ZigBee increases the problem of interference. Hence, ZigBee is suitable for low data rate transmissions, and optimized for wireless sensor networks [5].

The cause of electromagnetic interference could be internally by another ZigBee device or external devices that use a similar 2.4GHz protocol. The existence of too many ZigBee devices within a controlled sequence allows all devices to send and or receive data at the same time, increasing the receiver detectable power, which interferes with the received signal. This interference causes many problems, including increasing the time required for retransmission, increasing the power consumption that occurs because of repeated re-transmissions (as each retransmission consumes a lot of power that equals to the original transmission power), and increasing the possibility of data losses [6, 7]. In fact, the single channel utilized by ZigBee devices results in advanced interference levels. This means that, if the network devices are working on channel 1 of the ZigBee spectrum (2.40GHz to 2.48GHz), and interference happens on that channel, there is no way to avoid this interference other than repeated retransmission and error detection schemes. If this is the case, changing the interfered channel will temporarily solve the problem by avoiding the channel interference, thus saving time and energy [7].

This work presents a new scheme for automatic channel selection used by ZigBee devices, increasing its immunity to interference effects caused by neighbouring devices operating within the same frequency band and scope. This paper is structured as follows: the next section describes the literature and related work; section 3 presents the proposed channel selection algorithm; section 4 explains experimental work

conducted to evaluate the proposed model, and illustrates and discusses achieved results; and finally, section 5 concludes this work.

Literature Survey

The ISM band frequency used for domestic and commercial user applications is 2.4GHz. This band is commonly used for different applications and technologies. Wi-Fi, Bluetooth, and ZigBee are the most common 2.4 GHz standards that all fall within IEEE 802 specifications. This makes the interference problem more common between applications utilizing these types of technologies. Bluetooth is more immune to interference than Wi-Fi and ZigBee because of its ability to change the channel and due to its frequency hopping topology, however Bluetooth is a peer-to-one communication protocol [7, 8].

A study [9] of the impact and interference between Wi-Fi and ZigBee devices operating within the same home area network, analysing the spectrum of different working channels of ZigBee and the existence of different channel power over Wi-Fi signals, indicated that the use of ZigBee devices in home applications within the area of Wi-Fi networking causes high loss rate in ZigBee data transfer and higher packet errors. In addition, it was found that the interference is not uniform, and depends on the distribution of each ZigBee channel and the interference level between the two devices working within the same spectrum. The interference issue between Wi-Fi and ZigBee devices was also investigated by [10], finding that ZigBee is considered more sensitive than Wi-Fi and is affected by its signal spectrum because it uses lower power level than the Wi-Fi. The experiments conducted indicated that channel interference increased the ratio of packet loss in ZigBee and affected the ratio of retransmissions, with an increased packet delivery delay. The constraints of using ZigBee in industrial applications were investigated in [11, 12], finding that interference in the communication channel between ZigBee and other wireless modules in the same ISM spectrum band was considered the main issue requiring further investigation.

A methodology for monitoring the impact of Wi-Fi networks on Zigbee networks is described in [14]. The methodology was used to conduct intensive experiment in a real-time environment consisting of 20 Wi-Fi access points and several mobile nodes. Experiment result reveals that Zigbee transmissions are severely suffering from vast and regular WiFi interference. In addition, the integration between WiFi and Zigbee in packet level was observed to mitigate the interference within MAC layer. In [15] a performance analysis of IEEE802.11 and IEEE 802.15.4 ZigBee model was conducted using NS2. The performance metrics were Packet Delivery Ratio, data loss, end-to-end delay and Throughput.

Moreover, [16] analyses the effect of foliage on 2.4GHz frequency band. The study revealed that ZigBee signal degrades in more than 83% in a range of scenarios with different levels of obstacles, interference, and foliage. A new performance metric described as channel idle indicator (CII) was presented in [17], which is used to quantify channel quality. CII and logistic regression are utilized to build a channel idle state prediction model for ZigBee devices.

Experimental results indicate that prediction model ensures an acceptable level of ZigBee performance.

In the same concern, this work investigates the interference immunity of ZigBee devices operating in the same communication channel with Wi-Fi devices. The effects of interference on ZigBee network performance are analyzed and discussed, and a solution to reduce interference effects based on optimal channel selection and hopping is presented.

II. PROPOSED CHANNEL SELECTION ALGORITHM

There are 16 possible channels in the ZigBee spectrum. Channel bandwidth of 3MHz and 5MHz is used for pitch spacing between centers of two adjacent channels. Figure 1 illustrates the spectrum of ZigBee and channels distribution [6].

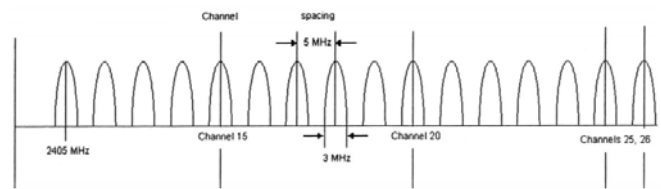


Figure 1: ZigBee spectrum and channels distribution

The most common interference source over the ZigBee spectrum is the Wi-Fi spectrum; Wi-Fi is a ubiquitous technology that uses the same spectrum band, and highly affects ZigBee due to its wider channel band and use of a higher signal power. Figure 2 illustrates the spectrum and channel distribution of the Wi-Fi, which falls under specifications of IEEE 802.11.

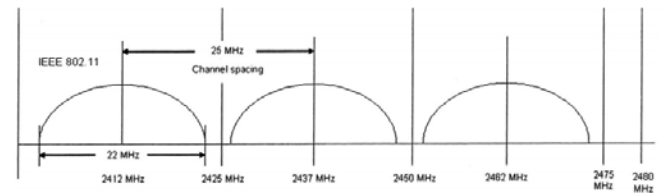


Figure 2: Wi-Fi spectrum and channels distribution

The idea of fixed channel ZigBee network implies that all network members are working on the same frequency channel, thus every node can leave and join the network without any request (i.e. parent / child connectivity). The use of single channel makes this functionality easy and the joining nodes are already working on the correct frequency. However, if the channel frequency changes, then nodes are unable to reconnect to the network because they are using different frequency from the network.

Wireless sensor networks could be either hierarchical or ad-hoc. In hierarchical networks, the head of cluster holds the management functionality over all cluster nodes, and has the

authority to lead to cluster settings, while the base service station does the same with respect to heads of clusters, and leads the functionality to control the settings of all cluster heads. Conversely, ad-hoc networks do not have cluster heads, and all network members are planners, thus the network nodes can talk together without the need for another medium node. In this case, the network settings are stored and held by a specific node called the “coordinator”, which is not intended to interact with the environment in order to enable physical functionality or send/receive data. Accordingly, the base station and head of cluster in hierarchical clustering networks, and the network coordinator in ad-hoc networks, can carry-out the change of channel process.

In this work the problem of channel interference in ZigBee network is solved by presenting a new scheme responsible for handling channel changing event. The new scheme is based on measuring the received power on a specified channel. If the received power is high and no data is detected, this means that another data structure is interfering in the working channel spectrum. If this condition is continued for a specific amount of time, the channel change decision is made, with selection carried out by the base station or cluster head acting as the functionality of the network coordinator node in ad-hoc networks.

Figure 3 illustrates the proposed channel handling scheme. After detecting interference, the coordinator changes the channel by one hop to the next frequency channel and then tests the received power; if this is low, this indicates that the channel has no or limited interference and the network can utilize it. Before leaving the current channel, the coordinator broadcasts a request message to network nodes indicating that the channel should be changed, and notifying them of the new channel configuration. Nodes receiving the broadcast message reply to the coordinator affirmatively and immediately switch to the new channel. Afterwards, the coordinator starts working with responded nodes on the new clean or low interference channel. The coordinator saves nodes’ responses. If all nodes responded to the broadcast call, then there will be no remaining nodes in the old channel. However, if one or more nodes did not respond by confirmation, then the coordinator will enter a recheck back loop. This loop requires the coordinator to go back again to the old channel and re-send the broadcast with new channel configurations to remaining nodes, and wait for confirmation messages from them.

The recheck back loop is repeated at specific intervals until all working nodes within the current channel change their frequency to the new spectrum channel. In this situation, old records of confirmation messages are deleted, previous channel settings and information is removed, and settings go to the new channel.

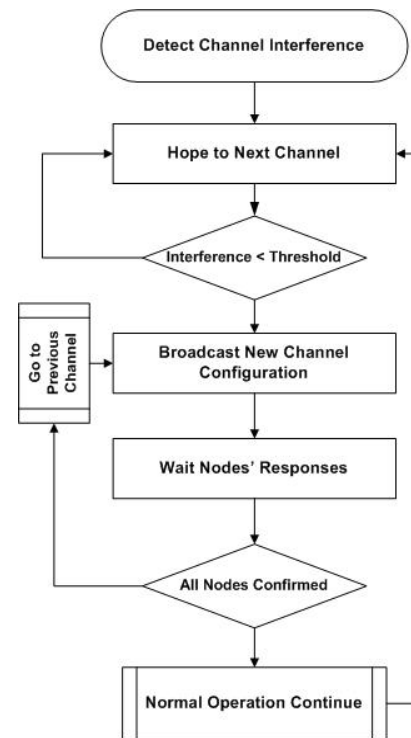


Figure 3: Channel selection algorithm flow chart

An additional situation that takes place when trying to switch to the new channel, is the interference level in the new channel being hopped to. If the interference is above a predefined threshold the next corresponding channel should be chosen, ignoring the current one. This process is repeated until a clean channel is found, and then the process of broadcast call and confirmations starts.

III. RESULTS ANALYSIS AND DISCUSSION

The presented channel selection algorithm was implemented in a network consisting of 100 ZigBee nodes. The network was designed to be an ad-hoc network structure. A low data rate was used to send and receive information between nodes. The frame structure in application layer was simple, merely representing a data packet and data transmission. Within the scope of the target ZigBee network, 10 Wi-Fi access points were used during the experiment representing different Wi-Fi interference sources. Figure 4 describes the relation between the packet delivery ratio and the number of access points during the experiment. It can be seen that the increase in Wi-Fi access points results in higher interference levels, increasing packet loss rate and packet delivery delay. Thus, the ratio of delivered packets is reduced dramatically. The relationship is directly proportional.

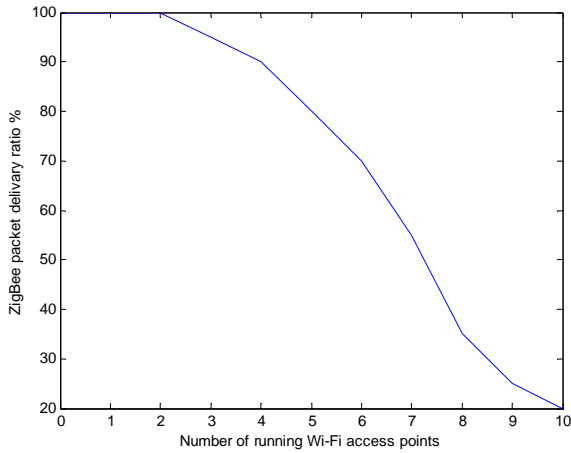


Figure 4: ZigBee delivery ratio over different levels of interference

Figure 5 illustrates the packet delivery time (delay) in case of different interference sources running on a working ZigBee channel. It is obvious that the delay increases while the number access points increases, due to the high level of interference and large number of re-transmissions.

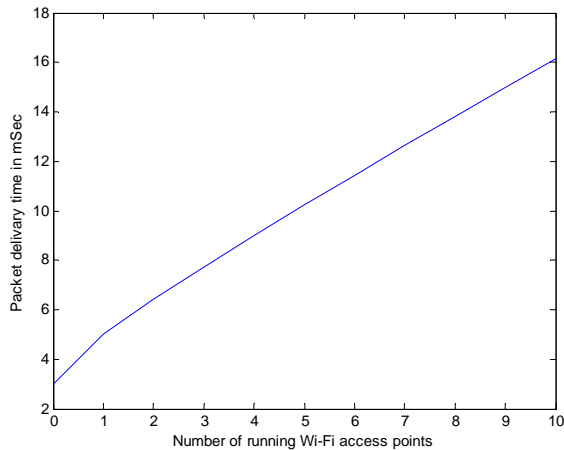


Figure 5: Packet delivery time in milliseconds over different levels of interference

When applying the proposed channel selection algorithm, the noisy channel will be continuously overridden and a new clean channel is used to ensure clean and fast transmission of data. Figure 6 shows that the packet delivery ratio attains 100% with the new clean channel, because of the absence of the interference.

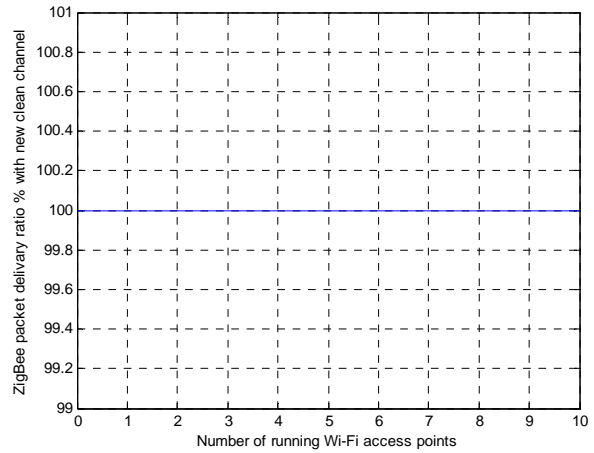


Figure 6: ZigBee delivery ratio when jumping to the new clean channel

IV. Conclusions

This work has focused on the limited interference immunity of ZigBee devices. A new channel selection algorithm used to increase the noise and interference immunity of ZigBee network nodes was described in this work. After detecting the interference on the current working channel, the algorithm starts the process of selecting the best suitable channel with low interference. The algorithm automatically and continuously changes the working channel by hopping to the next clean channel. The proposed channel selection algorithm was designed to ensure engaging all network nodes in the channel hopping process.

REFERENCES

- [1] James F. Kurose and Keith W. Ross "A Top-down Approach Featuring the Internet", Third Edition, 2005. New York: Pearson Higher Education.
- [2] Jin Zhu, "Development of an Application-Oriented Wireless Communication", Technology Interface Journal, Volume 10, Number 3, 2010.
- [3] Makoto Miyake, "A Survey on Progress and Standardization Trends in Wireless Communications", Journal of Communications, 2009, Volume 4, Number 7, 2009.
- [4] David J. Nagel, "Wireless Sensor Systems and Networks: Technologies, Applications, Implications and impacts", Centre of Strategic and International studies, 2005, ISBN 0892064625.
- [5] Nisha Ashok Somani, Yask Patel, "ZigBee: A Low Power Wireless Technology for Industrial Applications", International Journal of Control Theory and Computer Modelling (IJCTCM) Volume 2, Number 3, 2012.
- [6] P.Rohitha, P. Ranjeet Kumar, N. Adinarayana, T. Venkat Narayana Rao, "Wireless Networking Through ZigBee Technology", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 7, 2012 .

- [7] Dharmistha D. Vishwakarma, “*IEEE 802.15.4 and ZigBee: A Conceptual Study*”, International Journal of Advanced Research in Computer and Communication Engineering, Volume 1, Issue 7, 2012.
- [8] Soyoung Hwang, Donghui Yu, “*Remote Monitoring and Controlling System Based on ZigBee Networks*”, International Journal of Software Engineering and Its Applications, Volume 6, Number 3, 2012.
- [9] Tulin Mangir, Lelass Sarakbi, and Harvy Younan, “*Analysing the Impact of Wi-Fi Interference on ZigBee Network Based on Real Time Experiments*”, International Journal of Distributed and Parallel Systems (IJDPS) Volume 2, Number 4, 2011.
- [10] Wail Mardini, Yaser Khamayseh, Reem Jaradat, and Rana Hijjawi, “*Interference Problem between ZigBee and WiFi*”, Proceeding of International Scientific Association of Computer Science and Information Technology. Volume 30, Singapore, Hong Kong, September 2012.
- [11] Tomas Lennvall, Stefan Svensson , Fredrik Hekland, “*A Comparison of Wireless HART and ZigBee for Industrial Applications*”, IEEE International Workshop on Factory Communication Systems, 2008. WFCS 2008.
- [12] Emanuele Toscano and Lucia Lo Bello, “*Comparative Assessments of IEEE 802.15.4/ZigBee and 6LoWPAN for low-power industrial WSNs in realistic scenarios*”, 9th IEEE International Workshop on Factory Communication Systems (WFCS), 2012.
- [13] Sofie Pollin, Mustafa Ergen, Sinem Coleri Ergen, Bruno Bougard, Liesbet Van Der Perre
- [14] Ingrid Moerman, Ahmad Bahai, Pravin Varaiya, Francky Catthoor. “*Performance Analysis of Slotted Carrier Sense IEEE 802.15.4 Medium Access Layer*”, IEEE Transactions on Wireless Communications. Volume 7, Number. 9, pp. 3359-3371, 2008.
- [15] Yue Tao, Xiang-Yang Li, and rd Cheng Bo. 2013. *Performance of Coexisted WiFi and ZigBee Networks*. IEEE 33rd International Conference on Distributed Computing Systems Workshops (ICDCSW), 8th-11th of July, Philadelphia, PA, USA
- [16] Elmustafa S.Ali Ahmed , Bakhita El.Sheikh Ali, Eman Omer Osman and Tagwa Amin M. Ahmed. *Performance Evaluation and Comparison of IEEE 802.11 and IEEE 802.15.4 ZigBee MAC Protocols Based on Different Mobility Models*. International Journal of Future Generation Communication and Networking Vol. 9, No. 2 (2016), pp. 9-18.
- [17] Kuruvilla Mathew and Mujahid Tabassum. *Analysis of Bluetooth and Zigbee Signal Penetration and Interference in Foliage*. Proceedings of the International MultiConference of Engineers and Computer Scientists, IMECS, March 16 - 18, 2016, Hong Kong
- [18] Junyu Hu, Zhenquan Qin, Yingxiao Sun, Lei Shu, Bingxian Lu, and Lei Wang. *A Light-Weight Mechanism for ZigBee Performance Assurance under WiFi Interference*. Proceeding of 25th International Conference on Computer Communication and Networks (ICCCN). August 1 – 4, 2016, Hawaii, USA.

Machine Learning based Systems for Organizational Support – The State of the Art

Gufran Ahmad Ansari
Dept. of Information Technology
Qassim University, Qassim
Saudi Arabia
gufranansari@qu.edu.sa;
ansariga1972@gmail.com

Mohammad Tanvir Parvez
Dept. of. Computer Engineering
Qassim University, Qassim
Saudi Arabia
m.parvez@qu.edu.sa

Ali Al Khalifah
Dept. of Information Technology
Qassim University, Qassim
Saudi Arabia
a.alkhalifah@qu.edu.sa

Abstract—This work examines and evaluates the state of the art research works on Machine Learning (ML) based support systems in two key areas: Business and Education. The purpose of this paper is to review the most recent works in ML based support systems (MLSS), since the research works in MLSS are evolving quite rapidly. We review the major reported works published during the period 2010 to 2016. We have divided the research works reviewed in this paper into three broad categories. Two categories are reserved for the works in MLSS for product development and business control. The third category of papers covers the Educational Data Mining (EDM) sector. For each of the categories, we identify the key issues addressed by the researchers, followed by the approaches adopted in those works. We then critically appraise the reported works in terms of techniques, data used and application domains. Finally, some suggestions on future research works are given for each category of the reported works

Keywords— Machine learning, Organization Support Systems, Decision Making Systems

I. INTRODUCTION

Machine Learning (ML) based systems have radically altered the possibilities and ways of analysis and data processing. Although many researchers and practitioners have written about the benefits of machine learning based systems, few research works have been reported on how machine learning can support the business decisions in the extremely competitive business environment. ML can facilitate the implementation of ideas or the making of a business decision in the competitive market, as poor decision making can push down the reputation and trust of an organization.

For these reasons, in this work we review the state of the art research works on Machine Learning (ML) based support systems, called *MLSS*, in two prime areas: Business and Education. These two areas are selected due to the reason that education and business sectors represent the knowledge (the base) and the outcome (product of that knowledge) of a society. We have reviewed the major reported works published during the period 2010 to 2016. We have also referred some previous review papers in this area, in case readers are interested in research works prior to 2010. We have divided the research works reviewed in this paper into three broad categories. Two categories are reserved for works in MLSS for product development and business control. The third category of papers cover the Educational Data Mining

This work was supported by the Deanship of Scientific Research, project no. 3176, Qassim University, Saudi Arabia.

(EDM) sector. For each of the categories, we identify the key issues addressed by the researchers, followed by the approaches adopted in those works. We then critically appraise the reported works in terms of techniques, data used and application domains. Finally, some suggestions on future research works are given for each category of the reported works.

Before we go into the details of the various methods, a brief discussion of the structure of such ML based support systems is given here. For this purpose, consider Fig.1 which is adapted from [Delen2013] for the illustration purpose. Prior to the availability of MLSS, a decision maker could consult other domain experts before making a crucial decision. With evolve of data science, now decision makers can ‘consult’ automated systems that can infer knowledge from raw data. Thus, various business processes can provide data and these data can in turn be used to infer ‘opportunities’ that further improve those business processes. ML based systems thus works on the business data to provide decision makers with suggestions or answers to queries which are otherwise difficult to achieve. With the rapid development of techniques in Big Data Analytics, the relevance of ML based support systems for decision makers is now more than ever.

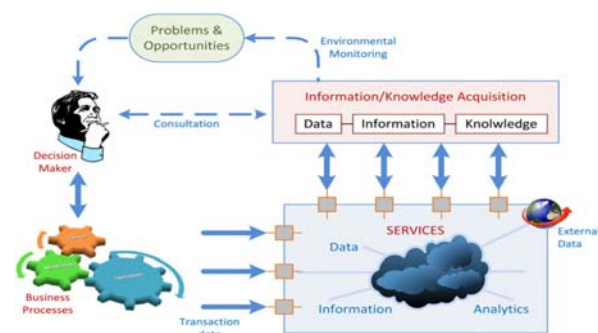


Fig.1. A framework for service oriented decision support systems using data science, adapted from [Delen2013].

The contributions of the current work can be summarized as follows:

- A systematic review of most recent works in machine learning based support systems.
- Identification of key issues with future research directions to help researchers advance the state of the art.

The rest of the paper is organized as follows. In Section 2, we review research works related to the product development sector. Section 3 deals with works that target business environments. Then, in Section 4, we discuss research works in Educational Data Mining (EDM). Finally, Section 5 is the conclusion.

II. MACHINE LEARNING FOR PRODUCT DEVELOPMENT (ML-PD)

A number of works have been reported where machine learning is used for improving product development process [Babiceanu2016, Cato2015, Dutta2015, Hamid 2015]. As outlined in Section 1, we first discuss the issues addressed by the researchers in this area, followed by a discussion on the techniques, open challenges and future work. Table 1 summarizes the reported works in ML-PD. A more detailed discussion on the techniques are given below.

2.1 Issues in ML-PD

Businesses that aim to improve their products (either hardware or software products) can utilize ML at a multitude of levels.

Machine learning is widely used in prediction problems, so it is no surprise that researchers have used ML for deciding products' fate. Hamid and Peng addressed the issue of measuring products' future changes and quantified uncertainties (external or internal) in product design process [Hamid2015]. Edwards et al. used seven different machine learning algorithms for predicting future hourly electrical consumptions in residential and commercial buildings in US [Edwards2012]. Another work tried to assess sustainability integration in company decision systems for product development levels [Hallstedt2010].

A number of works have used Big Data Analytics for product development process [Babiceanu2016, Cato2015, Dutta 2015]. Dutta and Bose [Dutta2015] used big data models to build a performance measurement system using several metrics. Their system was capable of analyzing performance at real time across multiple levels such as districts, regions, zones, dealers, etc. Another work [Babiceanu2016] used big data analytics, virtualization and cloud-based services for planning and control of manufacturing operations (see Fig.2. for the structure of such a system).

Some researchers focused on green product development and market analysis [Lee2012, Lin2013]. A framework for sustainable marketing for green market analysis, green product development, sustainable operations management and customer acquisition was discussed in [Lee2012]. [Lin2013] examines how market demand affects green product innovation. This work addressed questions like how does market demand influence a firm's green product innovation or how can green product innovation affect firm's performance.

Several works are reported on improving software products. Suma et al. presented a software defect prediction model based on Random Forest (RF) ensemble classifier for

large-scale software system [Suma2014]. Another work for cross-company software defect prediction system was reported in [Ying2012] using algorithms based on Naive Bayes called Transfer Naive Bayes.

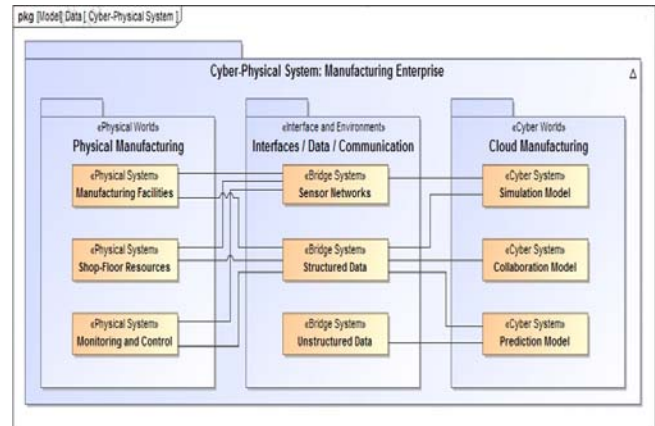


Fig.2. Big data analytics in manufacturing process [Babiceanu2016].

Interface issues of logistics and supply chain management for bio-energy production was discussed in [Gold2011]. The authors addressed a number of issues and challenges regarding the design and operation of biomass chains. The goal of the work was to secure stable and competitively-priced feedstock supply for bio-energy plants. For that purpose, the authors classified the issues in a bio-energy plant into a number of operations, namely harvesting and collection, storage, transport, and pre-treatment techniques, in addition to overall design of supply system.

The study in [Yao2015] identified the task of manufacturing into two types: smart and additive manufacturing. The authors aimed to utilize both forms manufacturing as a paradigm shift in industry. They discussed the methodologies to enable human capabilities in generating ideas for product development utilizing information and communication technologies. The goals of such generated ideas were multifold: (i) products should be economical in small or large scale, (ii) efficient enough for mass production, (iii) flexible enough to be customized to satisfy the needs of individuals/customers. The authors also discussed from the organizational semiotic perspective various social aspects, SCPS-based (socio-cyber-physical system) manufacturing.

2.2 Approaches Used in ML-PD

Edwards et al. [Edwards2012] used seven ML methods, namely Linear Regression, Feed Forward Neural Network, Support Vector Regression, Least Squares Support Vector Machine, Hierarchical Mixture of Experts, Fuzzy C-Means with Feed Forward Neural Networks, Temporal dependencies for predicting energy consumption.

Ying et al. [Ying2012] used TNB algorithm for building a system to guide optimal resource allocation strategies with two goals in mind: (i) reduce software testing cost and (ii) increase effectiveness of software testing process.

Their algorithm was based on the weighted Naive Bayes model that worked in two steps. In the first step, the system calculated each attribute information of the target data. Then in the second step, the similarity score for each sample in the source data was computed by comparing the features extracted in the first step.

The approach in [Suma2014] used Random Forest Classifier for the efficient prediction of defects in a project. Again, the goal of such a system was to enhance the decision making capabilities of project managers/developers prior to start of the project.

In [Dutta2015], the authors argued how enterprise manufacturing and business decision support could be improved using cloud business solutions and cloud manufacturing. They used state-of-the-art cloud manufacturing solutions and Big Data Analytics to propose a framework for the development of predictive *cyber-physical* manufacturing systems. Such a system incorporates a number of capabilities: capability to attach to Internet of Things (IoT), capability for complex event processing and Big Data algorithmic analytics.

The general interoperability model for SCPS-based manufacturing proposed in [Yao2015] can also be applied to other integrated manufacturing systems. Their proposed framework for SCPS-based manufacturing incorporated Internet of People (IoP), Internet of Knowledge (IoK), Internet of Things (IoT) and Internet of Services (IoS).

In [Afshari2015], two methods were proposed to model and quantify uncertainty in the product life cycle. Changes of user preferences are considered as the external uncertainty. An innovative application of Big Data Analytics (BDA) was proposed to evaluate the external and internal uncertainties in product design.

Based on a case study, challenges and difficulties faced by the OEM were identified in [Lee2012] and the potential solution and performance measurement of reverse logistics were suggested.

2.3 Appraisal of Methods in ML-PD

In this section, we critically appraise the reported works in ML-PD, pointing the possible weakness in methods or results.

Some works has used a single case study for reaching conclusions. [Dutta2015] drew inferences based on one single case study which could be fraught with its own bias. In addition, the case study used a limited amount of data. Similar comments can be made for the approach in [Suma2014]. Likewise, the work in [Cato2015] suffer from small data set. Other works have used only a few of the features of data. For example, [Ying2012] used simple information of the target data, i.e. the max and min values of each attribute. Similar limitations exist for the work in [Gold2011].

In a number of other works, comparative evaluations were not provided. In [Yao2015], previous comparative data of 'industrial revolution' were not provided. The study in [Afshari2015] lacks the access to other sources of Big Data in order to compare with the used sources. Similarly, the main focus of the work in [Lin2013] is the motorcycle

manufacturing industry in Vietnam, with no comparative study with other markets are provided.

The work in [Edwards2012] used various sensors in homes for collection of data. Since the total number of possible sensor combinations could be exponential, the authors used approximate solutions for their purpose. This approach is rather expected though, as a large number of sensors means that it would be expensive and time consuming to explore all possible sensor combinations. However, the authors argued that even these approximate solutions could help reduce the cost and overhead of energy usage in homes, making sensor-based energy modeling more viable.

2.4 Research Directions in ML-PD

In this section, we provide some future research directions for different methods discussed earlier. Obviously, the discussion here cannot be comprehensive and new research directions can always be found.

Some reported works can be extended by implementing and testing them on a broader scale. For example, the framework presented in [Dutta2015] was a theoretical and holistic one, thus it remains a challenge how the framework could be implemented for different production companies. Likewise, the assessment approach presented in [Hallstedt2010] can be tested in other companies to verify its generic applicability and to find potentials improvements.

[Ying2012] used Naive Bayes classifier for software defect prediction. It will be interesting to explore other transfer algorithms for cross-company software defect prediction on more software defect data sets.

The work in [Babiceanu2016] can be extended by adapting Big Data algorithms for the requirements in manufacturing processes. For example, the task of scheduling a set of jobs across a set of machines using algorithms for Big Data platforms should consider distributed file systems like HDFS. In addition, these algorithms should analyze the system performance at very high load.

The authors in [Lee2012] proposed some future work in terms of confirmatory factor analysis to analyze the measurement properties of the four constructs to determine how well the items represent the latent factors. Also their methods can be employed for empirical investigation.

Other works are constrained by the utilization of specific factors. Thus, an obvious future work can be the utilization other factors! For example, in [Afshari2015], uncertainty in the product design phase was modeled. This work can be extended by evaluating the effects of the quantified uncertainty on different design objectives (e.g. product cost, development time, environmental impacts of a product, etc.). Here, to goal is to evaluate and optimize design parameters under uncertainty when multiple design objectives are demanded. Likewise, Future studies of the work in [Cato2015] include wider variables to examine the effects of market demand, green product innovation and firm performance in more details.

Table 1: Summary of Results in ML-PD.

Ref.	ML Techniques Used	Issues Addressed	Limitations/Future Works
Suma, V2014 [35]	Random forest classifier.	Software defect prediction for large-scale software system.	Single Case Study.
Edwards, 2012[13]	Linear Regression, Feed Forward Neural Network, Support Vector Regression, Least Squares Support Vector Machine, Hierarchical Mixture of Experts, Fuzzy C-Means with Feed Forward Neural Networks, Temporal dependencies.	Predicting future hourly Residential and commercial buildings electrical consumption in US.	Predicting residential electrical consumption This is a pivotal problem that must be solved, because it is not practical to install and support 140 sensors for new homes.
Ma, Ying, et al.2012 [38]	TNB (Transfer Naive Bayes) algorithm applies the weighted Naive Bayes model.	Cross-company software defect prediction.	Only use simple information of the target data, i.e. the max and min values of each attribute.
Babiceanu 2016 [4]	Cloud business solutions and Big Data algorithmic analytics.	A review of the current status of virtualization and cloud-based services for manufacturing systems and of the use of Big Data Analytics for planning and control of manufacturing operations	Missing case study and comparative data.
Cato, Patrick et al. 2015 [7]	Big Data Analytics.	Analytical approach they are used.	Small data set used only few publications have specifically addresses reliability may have some biases because data does not capture incidentally Could not include the ACM digital library and Springer Science direct should include in future
Afshari, Hamid 2015 [1]	Big Data Analytics (BDA) is proposed to evaluate the external and internal uncertainties in product design.	Quantify external and internal uncertainties in product design process. The research addresses the measure of product future changes.	Lack of access to other sources of Big Data to compare with the used sources. Should evaluate effects of the quantified uncertainty on different design objectives (e.g. product cost, development time, environmental impacts of a product, etc.). Should evaluate and optimize design parameters under uncertainty when multiple design objectives are demanded.
Lin,2013 [25]	Empirical study used for findings.	Examines how market demand affects green product innovation, and firm performance in the context of Vietnamese motorcycle industry.	The main focus is only on the motorcycle manufacturing industry in Vietnam. Why it's not generalized study. No comparison study with other methods.

III. MACHINE LEARNING FOR BUSINESS CONTROL (ML-BC)

In this section, we review some recent works that address issues related to business operations. These methods differ from the methods discussed in the previous section in their focus area. Here, we deal with issues that affect business operations, factors that affect the product development process, but does not improve the product itself directly. Such issues include predicting customer behavior, supplier selection problem, evolution of business relationships, etc. Again, in this section, we follow the same structure of discussions as followed in the previous section. Table 2 summarizes the reported works in ML-BC. A more detailed discussion on the techniques are given below.

3.1 Issues in ML-BC

A lot of work have been reported for the case of supplier selection problem. The work in [Chai2013] reviewed Data Mining (DM) techniques for supplier selection. They considered four aspects: decision problems, decision makers, decision environments, and decision approaches. By examining the research trends on uncertain supplier selection, 26 DM techniques were identified from three perspectives: (1) Multi-criteria decision making (MCDM) techniques, (2) Mathematical programming (MP) techniques, and (3) Artificial intelligence (AI) techniques. Another work that reviews approaches used for supplier selection is [Govindan2013]. This work mainly deals with the issue of green supplier evaluation that considers environmental aspects. In the following, we discuss some recent works on supplier selection problems for a completeness of the review.

[Lin2015] used Analytic Network Process (ANP) to supplier selection issues at a Taiwanese Electronics Company. The selection process was based on a technique called Triple Bottom Line (TBL). A framework was developed to enable general application of ANP. A case study involving ten executives from an anonymous Taiwanese company was used to illustrate the framework and to identify the key factors in supplier selection.

An integrated methodology for product line design and supplier selection problem was proposed in [Deng2014]. They formulated a multi-objective optimization model to determine the specifications to select suppliers that ensured maximization of the profit, quality and performance as well as minimization of the cost of the product line. The work in [Rezaei2014] deals with supplier selection in the airline retail industry. The authors argued that supplier selection in supply chain management was a highly complex process, due to the involvement of many, sometimes conflicting, qualitative and quantitative criteria.

In [Kar2014], the author utilized heuristic models to enable supports for group decision making. The authors integrated AHP, fuzzy set theory and mathematical programming theories for predictive decision support system for the supplier selection problem. Other works that deal with

supplier selection issues include: selection of suppliers for a semi-conductor industry [Deswal2015], supply chains (SC) activities that requires the need for coordination between supply chains partners to maximize the efficiency (to solve supply chain coordination problem) [Igoulalene2015].

The work in [Fu2012] utilizes GSDP (Green Supplier Development Programs) categorizations to decide on multi-functional managerial inputs within a telecommunication systems provider to evaluate the GSDPs.

Another slightly different problem was dealt in [Ofoghi2013]. This work describes the implementation of machine learning techniques that assist cycling experts in the crucial decision-making processes for athlete selection and prediction athletic performances Olympics games.

Other than supplier selection problem, several other works focus on other issues related to business control. An integrated model predicting whether a customer buys or does not buy a specific product for target marketing strategy was proposed in [Bae2010]. Ali and Khan developed a software outsourcing partnership model (SOPM) to identify and analyze factors that were important for vendors in conversion of their existing outsourcing relationship to partnership [Ali2016].

[Akin2015] considered Seasonal Auto Regressive Integrated Moving Average, n -Support Vector Regression, and multi-layer perceptron type Neural Network models to evaluate performances on monthly tourist arrival data to Turkey from different countries. Based on these results, this study proposed a model for selection of tourism time series. Another related work on supporting tourist information retrieval and decision making is [Gretzel2011].

3.2 Techniques and Approaches in ML-BC

In this sub-section, we briefly present the ML techniques used for business control operations.

A number of ML methods have been used in works that deal with supplied selection. Multi-criteria decision making (MCDM) techniques, Mathematical Programming (MP) techniques and Artificial intelligence (AI) techniques were used in [Chai2013]. [Lin2015] adopted fuzzy theory, in addition to ANP model, to collect, analyze and express experts' judgments. A methodology for integrated product line design and supplier selection based on a multi-objective optimization was proposed in [Deng2014].

Another work that utilize Fuzzy Logic for supplier selection problem is [Kar2014]. This work integrated two techniques: fuzzy Analytic Hierarchy Process (AHP) and fuzzy goal programming. The goal of the first techniques was to enable group decision making, while the second one for discriminant analysis. The authors utilized evaluation criteria used by the decision makers and developed some performance vectors consisting of the standardized prequalifying supplier scores. Similarly, fuzzy analytic hierarchy process (AHP) was also used in [Akin2015] to evaluate suppliers based on a number of criteria.

Igoulalene et al. [Igoulalene2015] used two approaches for the problem of multi-stakeholder multi-criteria (MSMC) decision making. The first technique was the hybrid approach that combined fuzzy consensus based possibility measure and fuzzy TOPSIS method. The second hybrid approach combined the fuzzy consensus-based neat OWA and goal programming model that explicitly included the participation of stakeholders in the decision-making process. Similarly, multi-criteria decision making (MCDM) technique for supplier selection in a semi-conductor industry was used in [Deswal2015].

As for green supplier problem, [Govindan2013] used fuzzy based single model approaches. For a similar problem a grey-based DEMATEL (Decision-Making Trial and Evaluation Laboratory) approach was used in [Fu2012].

As for the case of athlete selection, Omnium (a multiple race event in track cycling) competition data from the World Championships (since 2007) were analyzed by statistical, machine learning based and probabilistic approaches [Ofoghi2013]

Works that deal with consumer behavior utilized various ML methods. 4 tree-based models of CHAID, CART, QUEST, and C5.0 was used in [Bae2010] for predicting customer buying habits. [Ali2016] developed a model, called SOPM (Software Outsourcing Partnership Model), with the intent to assist SDO vendor organizations in measuring their capabilities for successful conversion of their contractual outsourcing relationship to outsourcing partnership.

3.3 Appraisal of Methods in ML-BC

In this section, we identify the limitations of the various works described in the previous two sections. Again, we first discuss the methods related to supplier selection.

The study in [Lin2015] focused on TBL only. To evaluate the green performance of suppliers, it examined three scopes of sustainable development as critical factors, while there are other factors that may have similar implications. The empirical study in that work considered data from only one firm in the electronics industry in Taiwan. Similarly, [Rezaei2014] considered data from just only some specific airlines. In addition, no comparative results can be found in [Lin2015]. Chai et al. [Chai2013] mainly focused on the application of DM techniques for SS and important aspects such as criteria analysis and evaluation in SS processes were not considered. Karconducted multiple cross-validations to provide comparable performances for the proposed approach [Kar2014]. However, the outcome of that work would always be dependent on the context specific data inputs and distribution of data used during the prediction stages.

As for the works dealing with green supplier problem, the method in [Govindan2013] needs experimental designs, in addition to application validation. The authors could have addressed multiple criteria decision analysis concerns, which might be useful in socially sensitive decisions like green supplier selection. Additionally, [Fu2012] discussed their limitations in terms of utilization of some assumed grey-

scale values for a linguistic variable. The work in [Ofoghi2013] (athlete selection) didn't use any data, the theoretical results had no data-related supports.

We also discussed some works related to consumer data. [Akin2015] used monthly tourists coming to Turkey only, without considering the case for other countries. As for the work predicting consumer behavior [Bae2010], the authors predicted the purchase behavior of customers for some specific products. However, the authors only considered a limited number of products. In addition, ML based techniques can be used in an integrated approach, as opposed to using single individual models.

3.4 Research Directions in ML-BC

In this section, we provide some future research directions for the general are of ML-BC. Again, we start our discussion with the works related to supplier selection.

In addition to supplier analysis, the scopes of the model in [Lin2015] might be extended to include other issues, like strategic planning (i.e. decision of merge and acquisition) and strategic alliance (i.e. deciding on business partners essential for sustainable development). In addition, the model in [Lin2015] could be further extended by incorporating environmental aspects like environmental conservation or environmental education.

The method in [Igoulalene2015] (for decision making) can be enhanced in a number of ways. First, the technique can use a hybridization of two consensus methods to obtain a new consensus with an interactive aspect (consensus-based neat OWA). Second, the method can be modified to keep as much as possible the stakeholders' preferences (consensus-based possibility measure).

In the development of the proposed methodology in [Deng2014], it was assumed that the market demand is static. In reality, market demand can be quite dynamic. Thus, a future work could consider the dynamic effects in the integrated problem.

As for the work in [Kar2014], one can compare/integrate the proposed approach with other techniques (i.e. techniques like TOPSIS, VIKOR, ELECTRE) for group decision making in supplier selection problem. In addition, the performance of the integrated approach can be compared and extended on group decision making in the supplier selection domain. Furthermore, other theories (like multi-attribute theories, analytic network process, etc.) could be explored for possible application for supporting group decision making. Similarly, in [Deswal2015], TOPSIS can be used with fuzzy set theory to include linguistic terms of decision making.

In [Rezaei2014] the authors recommended examining the general applicability of the supplier selection methodology in the airline retail industry. They proposed to integrate more supply chain activities such as procurement, purchasing and protocols between the stakeholders to better coordinate. Additional issues to deals with include how the coordination

between the stakeholders' decisions can help to mitigate the risks such as delays in deliveries.

Works dealing with green suppliers can be further extended. The authors in [Govindan2013] pose a number of additional issues that need additional research. For example, longitudinally are the results in their work associated with the greenest suppliers that are selected, confirmed months or years after the selection? Are the decisions that are made good ones? In [Fu2012], the authors suggested future research into a comparative analysis across multiple companies to determine the generalizability of GSDP relationships.

The data driven methodology in [Ofoghi2013] accepts data from a variety of sources of evidence (similar to

the rankings of the participants in the six events of the omnium racing competition). These data can be used in various domains where epistemological prediction of certain outcomes are needed, like winning a medal, for a given dependent event (e.g., finishing the omnium in certain overall rankings). An example of such application domain is Resource Allocation problem in Financial Planning.

As for the work on consumer behavior, the results from the study [Bae2010] should be generalized as the purchase behavior of customers were predicted for some specific products.

Table 2: Summary of Results in ML-BC.

Ref.	ML Techniques Used	Issues Addressed	Limitations/Future Works
Akın, Melda 2015 [3]	Decision Trees, n-Support Vector Regression, Multi-layer Perceptron.	Monthly tourist arrival data to Turkey from different countries.	Data from one country only.
Ofoghi, Bahadorreza 2013 [28]	Utilization of statistical, machine learning-based, and probabilistic approaches.	Implementation of ML techniques that assist cycling experts in the crucial decision-making processes for athlete selection and strategic planning in the track cycling omnium.	No data used, theoretical result no data support
Chai, Junyi 2013 [8]	Multi-criteria decision making (MCDM) techniques, Mathematical programming (MP) techniques and Artificial intelligence (AI) techniques	Selection of suppliers using DM techniques.	A survey paper of limited coverage of recent works.
Govindan, Kannan 2013[16]	Applied techniques are mostly fuzzy based single model approaches.	Supplier evaluation and selection, with green supplier evaluation that considers environmental impacts.	Experimental designs, not just application validation, are needed for future research. Additional investigations in experimental settings are needed.
Gretzel, Ulrike 2011 [17]	Theoretical discussions.	Support tourist information search and decision making as well as work processes. The development of intelligent systems, which will likely increasingly permeate consumption and work processes related to tourism.	Considers the tourism domain only.
Deng, S., et al 2014 [9]	Multi-objective optimization techniques.	Integrated product line design and supplier selection. Joint-spacing mapping is used to help estimate market share of products and indicate positions of product variants.	Static market data is used.
Rezaei, Jafar 2014[30]	Fuzzy analytic hierarchy process (AHP).	Supplier selection in supply chain management. Select the most suitable supplier(s) that meet a company's specific needs.	Supplier selection in the airline retail industry only.
Igoulalene 2015 [20]	Fuzzy consensus based possibility measure and fuzzy TOPSIS method.	Addressed supply chain coordination and strategic supplier selection problems.	Explore hybridization of the two consensus methods to obtain a new consensus and keep the stakeholders' preferences as much as possible.
Deswal M 2015 [10]	Multi-criteria decision making (MCDM). TOPSIS are used.	Supplier selection in a semi-conductor industry	Only theoretical discussion.

IV. MACHINE LEARNING FOR EDUCATIONAL DATA MINING (ML-EDM)

In this section, we present some research works that use machine learning for educational data mining. Again, we follow the same structure of discussion as done in the previous two sections.

The primary goals of an EDM system and its various components are shown in Fig.3. (adapted from [Romero2007] for illustration purpose). As shown in Fig.3, educators and students have different goals from an EDM system, which basically draws its inferences upon the data provided by the traditional class rooms and educations processes. As compared with the process outlined in Fig.1 the educational system can be compared with ‘business processes’ and the educators/students take the role of the decision makers. Table 3 summarizes the reported works in ML-EDM. A more detailed discussion on the techniques are given below.

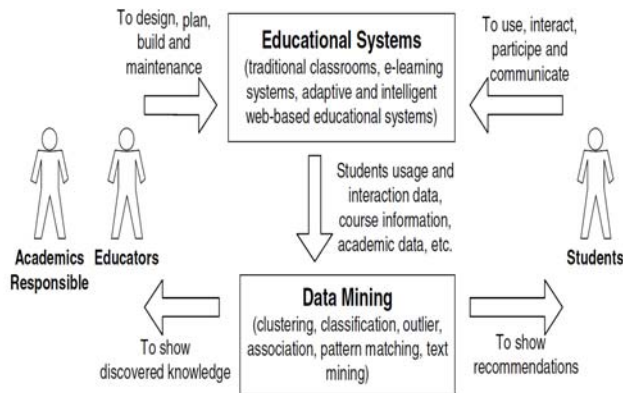


Fig.3. The cycle of applying data mining in educational systems [Romero2007].

4.1 Issues in ML-EDM

Siemens discussed the relationships between data and education, dividing the research in this respect into two research communities: Educational Data Mining (EDM) and Learning Analytics and Knowledge (LAK) [Siemens2012]. A good review paper in EDM is [Li2015] where the authors examined the research literature published in the last five years and constructed a systematic review of learning analytics. Another survey in EDM area is [Romero2010] where the authors classified each study in EDM not only by the type of data and DM techniques used, but also and more importantly, by the type of educational tasks they resolve.

As outline in [Romero2010], the objective in EDM “is to be able to make recommendations directly to the students with respect to their personalized activities, links to visits, the next task or problem to be done, etc., and also to be able to adapt learning contents, interfaces, and sequences to each particular student.”

Berland et al. investigated the relevance EDM techniques to help provide a basis for quantitative research on constructionist learning (a student-centered, discovery based learning) which does not abandon the richness seen as essential by many researchers in that paradigm [Berland2014]. Siemens built up a substantial base of techniques for analyzing discourse, social networks, sentiments, predictive models, and semantic content (they call as “intelligent” curriculum) [Siemens2012b].

There are a number of tasks that various reported techniques in EDM tried to handle. One these tasks is predicting students’ performance. Xing et al. built a student performance prediction model with the goal of making the model both practical and understandable for users [Xing2014]. They used data from a collaborative geometry problem solving environment called Virtual Math Teams with Geogebra (VMTwG). The authors synthesized three approaches: learning analytics approaches, educational data mining (EDM) and HCI theory with the goal to develop more usable prediction models.

In another related work [Huang2013], the authors predicted students’ academic performance. They developed and compared four types of mathematical models to predict student academic performance in ‘Engineering Dynamics’ – a high-enrollment, high impact, and core course that many engineering undergraduates were required to take.

As students’ performance is strongly related to the knowledge they attain, a number of works in EDM have attempted to measure such knowledge levels. Levy and Wilensky[Levy2011] used educational data mining methods for detecting students’ knowledge-in-action and the broader question of how conceptual and mathematical forms of knowing interact in exploring complex chemical systems. The authors determined their expectations regarding the model’s mathematical behavior. It was found that conceptual knowledge of the fundamental causes underlying the system’s behavior impacts educational strategies.

Other works related to EDM include a deep study and interactive graphical representation for e-commerce in educational system [Muruganathan2016].

4.2 Techniques and Approaches in ML-EDM

Various data mining methods have been used in ML-EDM. In this section, we briefly mention some of these techniques for the reported works discussed earlier.

Four types of mathematical models including multiple linear regression (MLR), multilayer perception network (MLP), radial basis function network (RBF), and the support vector machines (SVM) were used in [Huang2013] for predicting students’ academic performance.

Xing et al. demonstrated that connecting the structure of VMTwG trace data to a theoretical framework and processing that data using the GP (Genetic Programming) algorithms could outperform traditional models in prediction rate and interpretability [Xing2014]. They described a methodology that connected perspectives from learning

analytics, EDM theory and application to solve the problem of predicting students' performance in a CSCL learning environment with small datasets.

In [Muruganathan2016], the data model proposed was an agent based intelligent system that constructed the data model in the form of 3D cubes for every classification types. The cubical data models were given as the inbound inputs to the processing tools like OLAP, OPAC, etc.

4.3 Appraisal of Methods in ML-EDM

In this section, we critically evaluate some of the works discussed previously. This discussion forms the basis for future work in EDM.

In [Xing2014], the authors did not consider the quality of ultimate artifacts or objects that may be generated at the end of a course. Researchers less familiar with the VMT (Virtual Math Teams) environment and without the experience for analyzing interactions in the environment may difficulties replicating the reported results in a different context.

[Siemens2012b] only discussed theoretical approaches and no data were used. In addition, the work lacked comparisons with other approaches. The proposed models in [Muruganathan2016] have been demonstrated as the result for the selective datasets and have not tested with any universal data mining DB's.

The limitation of the method in [Levy2011] lies in the way data mining methods were used. The authors used expert knowledge to presume the indicators of knowledge, thereby limiting the possibilities of other factors. The authors should have used fewer assumptions, which in turn may enable the discovery of wider information. A stronger approach could be to utilize a larger set of factors related to students' activities. Then, clustering can be used to form the labels on students' behavior.

The authors in [Huang2013] reported high average prediction accuracy (APA) of 81%–91%. The grades that a student earned in pre-requisite courses might not truly reflect the student's knowledge of those topics. In addition, this study made no differentiation between norm-referenced and criterion-referenced scores in the data collected.

4.4 Research Directions in ML-EDM

Here, we provide some guidelines regarding future research approaches based on the discussions in the previous sections. We first describe the general guidelines in EDM that various researchers have suggested in the reported works. Then some specific research guidelines are provided.

The authors in [Romero2010] suggested that, instead of the present reported individual proposals and methods, researchers should develop more unified and collaborative studies. They believe that in this way the full integration of DM techniques in the educational sector can become a reality.

As a result, fully operative implementations of EDM tools could be made available for researchers, developers and for external users. Similarly, in [Berland2014], the authors argued that interdisciplinary subfields could make assessments more feasible in large scale. This approach may enable the creation of smarter technologies for real-time feedback, streamlining and optimizing the process of giving feedbacks to students. In turn, this approach can offer researchers a deeper insight into the learning processes in constructionist learning environments.

Some comments are now made regarding the reported works and their possible improvements. The study in [Xing2014] considered the quantitative aspect of the data. A possible future work could be to incorporate qualitative aspect into the models. For example, using natural language processing to process the chat logs of students, and then adding more factors to the community dimensions to see whether it can improve the prediction rate. Also, more experiments could be done in other learning environments using this methodology and test its transferability.

The 3D cubic data model proposed in [Muruganathan2016] can be applied as data visualization techniques in Big Data Analytics or in traditional DMM. In [Levy2011], Levy and Wilensky suggested that future research work should address the differences among the different models where the models differ in the type of prior knowledge used. The authors commented that students could employ various strategies in exploring the models, this in turn could result in different conceptual and mathematical challenges.

The authors in [Huang2013] utilized eight cognitive factors to develop the predictive models. The accuracy of the predictive models could possibly be increased by considering more factors. Such factors include some psychological factors such as learning styles, self-efficacy, achievement goals, motivation and interest, teaching and learning environment, etc.

V. CONCLUSION

In this work, we have reviewed state of the art research works in machine learning based support systems. Systems in two key areas, namely business and educational sectors, are covered, as these two sectors represent the base and the outcomes of a society. We have attempted to critically appraise the reported methods, instead of just listing them. The current paper can benefit researchers in ML based support systems in identifying key issues to work on.

Acknowledgment

This work was supported by the Deanship of Scientific Research, project no. 3176, Qassim University, Saudi Arabia.

Table 3: Summary of Results in ML-EDM.

Ref.	ML Techniques Used	Issues Addressed	Limitations/Future Works
Siemens, 2012 [33]	Theoretical discussions.	Educational Data Mining (EDM) and Learning Analytics and Knowledge (LAK).	No data was used. Only theoretical discussions.
Li, Kam Cheong 2015 [24]	Analysis and learning analytic methods.	Examine the research literature published in the last five years and construct a systematic review of learning analytics.	No comparisons made with other techniques.
Romero 2010 [31]	Data Mining technique used	A survey of collecting and consulting all the published bibliography in EDM area.	No general tools or reusing tools that can be applied to any educational system.
Berland, Matthew 2014 [6]	Quantitative research.	Quantitative research on constructionist learning.	Authors integration of constructionist pedagogical approaches with EDM only, it should be made general.
Siemens, George 2012 [33]	Learning analytics approaches.	Proposed an integrated and holistic vision for advancing learning analytics as a research discipline and a domain of practices.	Just theoretical approach, no data used. No comparison is made with other approaches.
Muruganathan,2016 [27]	Construct the data model in form of 3D cubes, with inbound inputs to the processing tools like OLAP, OPAC.	A study and interactive graphical representation for e-commerce in educational system.	Proposed models were demonstrated for the selective datasets and were not tested with any universal data mining DB's.
Levy, Sharona T 2011[23]	Data mining methods for detecting students' knowledge-in-action.	Detecting students' knowledge-in-action and the broader question of how conceptual and mathematical forms of knowing interact in exploring complex chemical systems.	Should have used fewer assumptions, which in turn may enable the discovery of wider information.
Huang, Shaobo, 2013 [19]	Predictive mathematical models: multiple linear regression, multilayer perception, radial basis function network and support vector machine.	Predicting student academic performance in core courses that many engineering undergraduates are required to take.	The predictive models only take into account eight cognitive factors. The study made no differentiation between norm-referenced and criterion-referenced scores in the data collected.

REFERENCES

- [1] Afshari, Hamid, and Qingjin Peng. "Modeling and quantifying uncertainty in the product design phase for effects of user preference changes." *Industrial Management & Data Systems* 115.9: pp. 1637-1665, 2015.
- [2] Ali, Sikandar, and SiffatUllah Khan. "Software outsourcing partnership model: An evaluation framework for vendor organizations." *Journal of Systems and Software* 117: pp.402-425, 2016.
- [3] Akin, Melda. "A novel approach to model selection in tourism demand modeling." *Tourism Management* 48 *Tourism demand modeling*: pp. 64-72, 2015.
- [4] Babiceanu, Radu F., and RemziSeher. "Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook." *Computers in Industry* 81: pp. 128-137, 2016.
- [5] Bae, Jae Kwon, and Jinhwa Kim. "Integration of heterogeneous models to predict consumer behavior." *Expert Systems with Applications* 37.3: pp. 1821-1826, 2010.
- [6] Berland, Matthew, Ryan S. Baker, and Paulo Blikstein. "Educational data mining and learning analytics: Applications to constructionist research." *Technology, Knowledge and Learning* 19.1-2: pp. 205-220, 2014.
- [7] Cato, Patrick, Philipp Gölzer, and Walter Demmelhuber. "An investigation into the implementation factors affecting the success of big data systems." *Innovations in Information Technology (IIT)*, 2015 11th International Conference on. IEEE, 2015.
- [8] Chai, Junyi, James NK Liu, and Eric WTNgai. "Application of decision-making techniques in supplier selection: A systematic review of literature." *Expert Systems with Applications* 40.10:pp. 3872-3885.

- [9] Deng, S., et al. "Integrated product line design and supplier selection: A multi-objective optimization paradigm." *Computers & Industrial Engineering* 70: pp. 150-158, 2014.
- [10] Deswal M. and S. K. Garg. "Application of Topsis: A Multiple Criteria Decision Making Approach in Supplier Selection." 2nd International Conference on recent Innovation in Science, Engineering and Management, JNU Convention Centre, Jawaharlal Nehru University, New Delhi, 22nd Nov. India, 2015.
- [11] Dutta, Debprotim, and Indranil Bose. "Managing a big data project: The case of Ramco Cements limited." *International Journal of Production Economics*, 2015.
- [12] [Delen2013] Delen, Dursun, and Demirkan, Haluk. "Data, information and analytics as services", *Decision Support Systems* 55: pp. 359-363, 2013.
- [13] Edwards, Richard E., Joshua New, and Lynne E. Parker. "Predicting future hourly residential electrical consumption: A machine learning case study." *Energy and Buildings* 49: 591-603, 2012.
- [14] Fu, Xiaoyong, Qinghua Zhu, and Joseph Sarkis. "Evaluating green supplier development programs at a telecommunications systems provider." *International Journal of Production Economics* 140.1: pp. 357-367, 2012.
- [15] Gold, Stefan, and Stefan Seuring. "Supply chain and logistics issues of bio-energy production." *Journal of Cleaner Production* 19.1: pp. 32-42, 2011.
- [16] Govindan, Kannan, et al. "Multi criteria decision making approaches for green supplier evaluation and selection: a literature review." *Journal of Cleaner Production*, 2013.
- [17] Gretzel, Ulrike. "Intelligent systems in tourism: A social science perspective." *Annals of Tourism Research* 38.3: pp. 757-779, 2011.
- [18] Hallstedt, Sophie, et al. "An approach to assessing sustainability integration in strategic decision systems for product development." *Journal of Cleaner Production* 18.8: pp. 703-712, 2010.
- [19] Huang, Shaobo, and Ning Fang. "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models." *Computers & Education* 61: pp. 133-145, 2013.
- [20] Igoulalene, Idris, LyesBenyoucef, and Manoj Kumar Tiwari. "Novel fuzzy hybrid multi-criteria group decision making approaches for the strategic supplier selection problem." *Expert Systems with Applications* 42.7: pp. 3342-3356, 2015.
- [21] Kar, Arpan Kumar. "Revisiting the supplier selection problem: An integrated approach for group decision support." *Expert systems with applications* 41.6: pp. 2762-2771, 2014.
- [22] Lee, Carmen Ka Man, and Jasmine Siu Lee Lam. "Managing reverse logistics to enhance sustainability of industrial marketing." *Industrial Marketing Management* 41.4: pp. 589-598, 2012.
- [23] Levy, Sharona T., and Uri Wilensky. "Mining students' inquiry actions for understanding of complex systems." *Computers & Education* 56.3: pp. 556-573, 2011.
- [24] Li, Kam Cheong, Hoi Kuan Lam, and Sylvia SY Lam. "A Review of Learning Analytics in Educational Research." *Technology in Education. Technology-Mediated Proactive Learning*. Springer Berlin Heidelberg: pp. 173-184, 2015.
- [25] Lin, Ru-Jen, Kim-Hua Tan, and Yong Geng. "Market demand, green product innovation, and firm performance: evidence from Vietnam motorcycle industry." *Journal of Cleaner Production* 40: pp. 101-107, 2013.
- [26] Lin, Chinho, et al. "Developing an assessment framework for managing sustainability programs: A Analytic Network Process approach." *Expert Systems with Applications* 42.5: pp. 2488-2501, 2015.
- [27] Muruganathan, V., and B. L. ShivaKumar. "An adaptive educational data mining technique for mining educational data models in elearning systems." *Indian Journal of Science and Technology* 9.3, 2016.
- [28] Ofoghi, Bahadorreza, et al. "Supporting athlete selection and strategic planning in track cycling omnium: A statistical and machine learning approach." *Information Sciences* 233 (2013): pp. 200-213, 2013.
- [29] Piety, Philip J., Daniel T. Hickey, and M. J. Bishop. "Educational data sciences: framing emergent practices for analytics of learning, organizations, and systems." *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*. ACM, 2014.
- [30] Rezaei, Jafar, Patrick BM Fahim, and Lori Tavasszy. "Supplier selection in the airline retail industry using a funnel methodology: Conjunctive screening method and fuzzy AHP." *Expert Systems with Applications* 41.18: pp. 8165-8179, 2014.
- [31] Romero, Cristobal, and Sebastian Ventura. "Educational data mining: A survey from 1995 to 2005." *Expert systems with applications* 33.1 : pp. 135-146, 2007.
- [32] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40.6: pp. 601-618, 2014.
- [33] Siemens, George, and Ryan SJ d Baker. "Learning analytics and educational data mining: towards communication and collaboration." *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM, 2012.
- [34] Siemens, George. "Learning analytics: envisioning a research discipline and a domain of practice." *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. ACM, 2012.
- [35] Suma, V., T. P. Pushphavathi, and V. Ramaswamy. "An Approach to Predict Software Project Success Based on Random Forest Classifier." *ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II*. Springer International Publishing, 2014.
- [36] Xing, Wanli, et al. "Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory." *Computers in Human Behavior* 47: pp. 168-181, 2014.
- [37] Yao, Xifan, and Yingzi Lin. "Emerging manufacturing paradigm shifts for the incoming industrial revolution." *The International Journal of Advanced Manufacturing Technology*: pp. 1-12, 2015.
- [38] Ma, Ying, et al. "Transfer learning for cross-company software defect prediction." *Information and Software Technology* 54.3: pp. 248-256, 2012.

AUTHORS PROFILES



Dr. Gufran Ahamd Ansari received his Bachelor degree (B.Sc. Computer Science) from Shia P.G. College, Lucknow in 1997, MCA from DR. B.R. Ambedkar University Agra in 2002 and Ph.D. (Computer Science) from Babasaheb Bhimrao Ambedkar (A central) University, Lucknow, U.P., India in 2009. He is currently working as an Assistant Professor in the department of Information Technology, College of Computer, Qassim University, Saudi Arabia. He has over 14 years of experience in teaching undergraduate as well as postgraduate students of Computer Science, Information Technology, and Computer Applications.

He has authored more than 30 research papers on Software Engineering in leading international journals and conferences. His current research interests are Software Engineering, UML, Machine Learning, Modelling, Testing, Artificial Intelligence, Data Mining, Software Quality, and Software Security.



Mohammad Tanvir Parvez obtained his B.Sc. and M.Sc. Engineering in Computer Science and Engineering (CSE) from Bangladesh University of Engineering and Technology (BUET), Dhaka in 2004 and 2006 respectively and a Ph.D. in CSE from King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia in 2010. His research interests include Pattern Recognition, Image Processing and Machine Learning with special interest in handwriting recognition using structural approach. He has received several awards including Best Graduate Students' Award (KFUPM, 2010), First Prize in Graduate Seminar Day (KFUPM 2010), BUET Academic

Performance Scholarship (1999–2003), etc. Dr. Parvez is an Assistant Professor in Computer Engineering Department at Qassim University, Saudi Arabia since 2011.



Ali Alkhalifah received a BS in computer science from Qassim University in 2007, and master (honor) in IT from the University of Newcastle in 2010, and PhD in information systems from the University of New South Wales, Australia in 2013. He is an Assistant Professor in Computer College at Qassim University. Until recently he was the head of IT department. He has been involved in several program committees and is being a Reviewer in different international conferences and journals. Dr. Ali has a number of research interests including e-business, identity management systems, evaluation of the World Wide Web, and the semantic Web.

SOCIAL TRENDS: THE THEORY, RESEARCH AND SOCIALITY DISCERNED THROUGH TWITTER

Jasmin D. Niguidula
Technological Institute of the Philippines-Manila
jasmin.niguidula@tip.edu.ph

Angelique D. Lacasandile
Technological Institute of the Philippines-
Manila / National University - Manila
adlacasandile@national-u.edu.ph

Abstract— Communication is essential in every human activity it enables a person to express one's thought. There are different ways on how to convey the message to another person or a group, still interpretation of this messages are misinterpreted and worst the receiver of the message cannot make sense of the message at all. In the advent of technology, different tools and platform started to emerge to make communication better and faster. This paper discusses the theory, research and sociality discerned through Twitter. The data set used were retrieved from twitter between the midnight of Monday, February 27, 2017 and the morning of February 28, 2017. The extracted tweet were analyzed, identifying the top trending topic of the day #Oscars, analyzed the characters within and discovered tweet and tweet entities utilizing Frequency Analysis. The lexical diversity of the tweet was computed and histogram was used to visualize the frequency of data. Future directions for social trends includes the following: testing on how twitter conversation and topics affects a specific community or decision on controversial issues; implement an additional approach; and utilizing the sociality discerned through Twitter in predicting behavior of a person or a group of people and in other field that resolve human activities and interaction.

the numerous users of a commerce-oriented technological development, social network sites, such as Facebook, Instagram, LinkedIn, Pinterest, Flickr Tumblr, Twitter and many more.

People reconstruct their network of friends in the virtual world and dominates the web usage. It attracts attention, despite of concepts and the initiative of different researchers to create research in this phenomenon, little percentage focus behind the computation and theory which made possible to connect people in the cyberspace. There are many aspects that makes this technology successful and fond to use.

In this paper, the researcher focused on taking advantage of the available information in the web and review the technology behind how social networking sites discern trending topics. To take advantage and to know the people needs and convert this tool into a powerful mechanism that bridge their business to engage and understand their customers through social networking sites.

1. INTRODUCTION

“Communication is a basic human activity” and “enables us to connect with each other” [5] Michael Langley, (2016). There is no doubt that communication plays a vital role in our everyday lives. It is the foundation of all human relationships. As the years go by, communication evolves as society moves forward into the digital age. People are increasingly being dependent on digital information and this encouraged the emergence of a new platform to communicate. Social networking sites seems to dominate the digital space. Social networking spaces allows people to create their own personal profile that can be accessible by other people via internet. The significant development connected to social media is

This paper presents the theory, research and sociality discerned through social networking sites. Analyzing one of the most commonly used social networking sites in the Philippines, Twitter who make “trending” and “hashtags” popular in the digital space, the researcher explores an exciting field of study that reveals how information behaves using frequency analysis technique determines the most common topics being talked about in the cyberworld.

2. RELATED WORKS

From a social networking sites standpoint, the wealth of data that it stores about the individuals, groups and product gives a great opportunity to synthesize everything and turned into a valuable

insights as stated by Russell (2013).[11] The researcher acclimated to the journey towards discerning the sociality in the trends. However, it is very common to misinterpret the purpose or usage of a social networking site. It seems appropriate to discuss first how we define Twitter.

According to Fried, (2011) the twitter chairman / CEO Jack Dorsey cannot define what his own creation is. The researcher quote Jack Dorsey “Twitter, he said, is different things to different people. “Twitter is the world”. This definition as perceived could built a new dimension or a world in the computer system that supports the message of the founder [3]. In addition, in the biography written by Biography.com Editors, (2015) it was stated in the article that the creation of twitter which was formerly called Obvious, Dorsey approached Silicon Valley called Odea and pitched the concept and we quote. “He came to us with this idea: ‘What if you could share your status with all your friends really easily, so they know what you’re doing?’”. Within two weeks after presented the idea, Dorsey had built a simple site where he execute his mission and vision where users could instantly post short messages up to 140 characters or less, known today in a Twitter parlance as “tweets” [1].

On the other hand, Facebook perhaps is the heart of the social web and somewhat an all-in-one wonder, Facebook is unquestionably the largest social network in the world, according to the latest statistics and facts, Popper & Erlick (2017) stated that the company has shown steady growth when it comes to user and its revenue, with reported more than 1.23 billion people use the social network every day, and 1.15 billion log in from mobile devices each day. It was projected at the current growth rate it will hit 2 billion monthly active users by the middle of this year [9]. Facebook defined by TechTerms (2008) was just created to provide an easy way for friends to keep in touch and for individuals to have a presence on the web without needing to build a website. From the vision of the CEO and co-founder Mark Zuckerberg, Facebook was noticed and has connected friends and families across the globe [13]. On Swisher & Wagner (2017) emphasized that “Today we are close to taking our next step. Our greatest opportunities are now global-like spreading prosperity and freedom, promoting peace and understanding, lifting people out of poverty, and accelerating science.”[12] In addition Zuckerberg wrote. “Our greatest challenges also need global responses-like ending terrorism, fighting climate change, and preventing pandemics. Progress now requires humanity coming together not just as cities

or nations, but also as global community”, Quartz (2017). [10]

Every social networking sites have distinct features and each platform has different focus on dealing with their business and operations. Still there are a lot of similarities on their features and one of this is they are capable to show users which topics, photos, videos and posts are trending or as of the moment talked of the town. Virtually all analysis boils down to the simple exercise of counting things and showing people the most numbered views, likes and shares. It seems simple when you think how can they determine it but how to translate it in the computer system is a different thing. Russell (2013) shares and discussed the incredible feature behind this amazing technology in both Twitter and Facebook [11]. The researcher was able to understand that even if it works the same on the perspective of the user when they use a computer, it is not the same at the back of it when you try to look on the perspective of a developer. Although Facebook is at the top of all social networking site, as of 2017 statistics, the researcher focused on the network that according to Patterson (2014), the network that brought us the hashtag is Twitter and it is the most popular site to use it on. In their feature users can easily find the trending topics on the left hand side of the Twitter interface [7].

In the study of Manning et. al (2009) which discussed information retrieval and states that, “A query in which the terms of the query are typed freeform into the search interface, without any connecting serach opeartors. This query style is extremely popular on the web, views the query as simply a set of words. A scoring mechanism is to compute a score that is the sum, over the query terms, of the match scores between each query term and the document.”[6] Furthermore, the researcher assign to each term in a document a weight for that term, that depends on the number of occurrences of term in the document. Then it will be simply compute a score between a query term t and a document d , based on the weight of t in d . The simplest approach to assign the weight to be equal to the number of occurrences of term t in document d . This weighting scheme is referred to as term frequency and denoted as:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

FIGURE I. TERM FREQUENCY FORMULA

Fig. 1 shows the term frequency formula that is famously known in the literature as the bag of words model.

The application of this formula in Twitter when it comes to identifying top topics, photos, videos and post had different terms depends on the platform and behaves differently when it comes to execution of the programmers logic and algorithm. Manning, Raghavan, & Schutze (2009) emphasizes on their paper that the collection of frequency (cf) and document frequency (df) behave differently, as in this example from the Reuters collection.

TABLE I. SAMPLE COLLECTION FREQUENCY AND DOCUMENT FREQUENCY

Word	cf	df
try	10422	8760
Insurance	10440	3997

3. METHODOLOGY

Communication can be in different form, from written, non-verbal, visual and communications utilizing a social networking sites. Given the rise in the popularity of this sites, it seems only natural to ask what is the trending topics today? Twitter is the social networking site considered in identifying trending topic for the day. The platform that provides social information has been driven by their ability to help users manage, organize and share online resources because humans want to be heard, wanted to satisfy curiosity, wanted it easy and wanted it now.

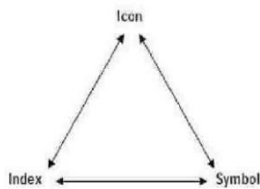


FIGURE II. PIERCE'S TRICHOTOMY OF SIGNS

Fig. 2 shows Peircian model Ghode (2017) emphasizes that for a sign to exist, it must consist of all three part (the object, representamen and the interpretant) and the interaction between them is a process Peirce termed semiosis [4]. He described this process as an action, an influence, which is, or involves, a cooperation of three subjects, such as a sign, its object and its interpretant, this tri-relative influence is not being in any way resolvable into actions between pairs, Peirce (1932).[8]

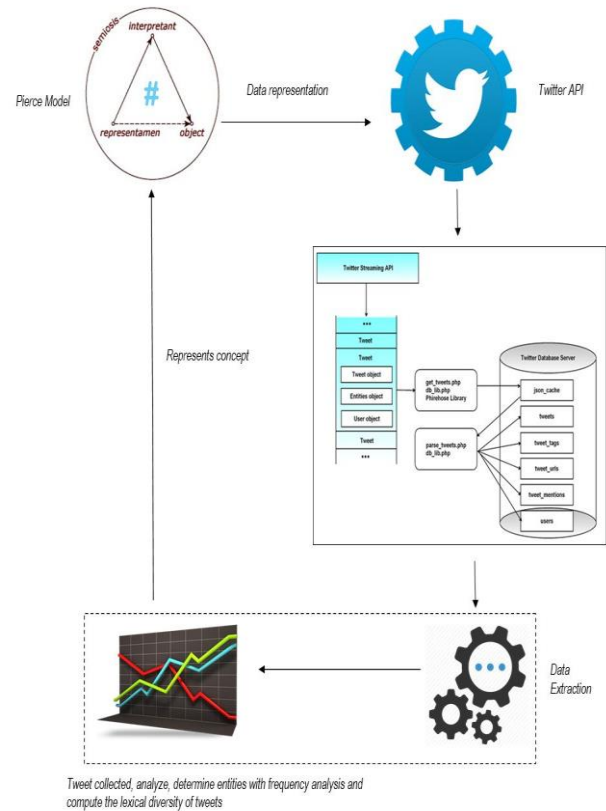


FIGURE III. FRAMEWORK ADOPTING PIERCE'S TRICHOTOMY OF SIGNS

Fig 3 shows the representation of the process to discern the theory and sociality in Twitter. With an authorized API connection in place, trending topics was explored to uncover sociality in social networking sites. Among the things that the researcher analyzed are trending topics in the Philippines, search for tweets, extract a single tweet from the search results and stored in a variable. The researcher distill entities and the text of the tweets into a convenient structure for further examination and then extracts the text, screen names, and hashtags from the tweets that are collected, analyze tweets and tweet entities with frequency analysis and computed the lexical diversity of tweets. Finally, the researcher discussed the value of this collaboratively generated data and used Pierce's trichotomy of signs to represent a concept.

3.1 RETRIEVING DATA IN TWITTER

Our analysis was performed on the data retrieved from Twitter, which the researcher retrieved. Data was gathered between the midnight of Monday, February 27, 2017 and the morning of Monday February 28, 2017.

We begin by retrieving trends in the Philippines by searching first the (WOE) ID, which can be search in the Where On Earth ID system, Elliot (2017), it is an API unto itself that aims to provide a way to map a unique identifier of any place on Earth. The following text is used to retrieve raw result of the trending topics [2].

```
<?php
require_once("TwitterAPIExchange.php");

/** Set access tokens here - see: https://dev.twitter.com/apps/**

$settings = array(
    'oauth_access_token' => "2832580312-
CoNuwXnNscEEJAmAhqKfItpxbXqd3PfrkDi5Uno",
    'oauth_access_token_secret' =>
"HIAAuFozVBuF5DQ2gDHseMiuOglV5iz0DX1nnlZKkV6BZ",
    'consumer_key' => "sJ0sjnc8e3Gs2mXfWJtvKwVwu",
    'consumer_secret' =>
"K3oO4AkYRPJR0JtD37oA3W3XDpx86VhllQ6k6vscJ6ZsjetQbS"
);

//**** TRENDING TWEET

$url = "https://api.twitter.com/1.1/trends/place.json";
$requestMethod = "GET";
$getfield = '?id=23424934';

$twitter = new TwitterAPIExchange($settings);

$json_string = $twitter->setGetfield($getfield)
->buildOauth($url,$requestMethod)
->performRequest();

echo $json_string;
```

FIGURE IV. RETRIEVING TRENDS

Fig. 4 shows a semireadable response that is the retrieved collection of the trending topics in the Philippines in Java Script Object Notation (JSON) formatter and validator (<https://jsonformatter.curiousconcept.com/>), before proceeding further, the researcher format the response to be more easily readable using the results retrieved contains a URL for a trend represented as a search query that corresponds to the hashtag (#). The trending topics retrieved on the day that the data was collected are shown in the figure 5 below:

```
#BadtripAkoKasi -
#PBBAngKatotohanan - 84250
#ALDUBxDTBYTadhana - 44302
#TeamHangingOutS1Finale -
AlwaysJaDine USATour2017 - 54549
#FLIGHTLOGARRIVAL - 219726
Pastillas Girl -
Senate - 60345
Moonlight - 1850608
Pacquiao - 14762
German - 65730
Duterte Youth -
Nokia 3310 - 119479
Viola Davis - 599450
Pilipinas - 35399
Warren Beatty - 239859
BB KISSES to 2366 - 24420
LuscaAzas -
Best Actress - 157286
Kim Chiu Hosts Showtime -
#MTBMakuhakaSaTingin - 29686
#ALTLNextMove - 22167
#iheartGGV - 80416
#MondayMotivation - 110621
#Oscars - 8230427
#EncantadiaKapahamakan - 43180
#FPJAPBangaan - 17212
#MDHMangkalaban - 37135
#PBBNoniEvent - 27939
#MyExAndWhysHello250M - 278227
#ExordiumPCD -
#HKSLCharDawnTheBeginning -
#iheartGGV -
#KathNielASAPFEBorito -
#PKLSLLabenSantana - 16235
#TransportStrike -
#DestinedToBeYoursBukasNa - 17375
#WildflowerBunyang -
#EverydayMusicVideo - 57822
#LEdwardHelpsGrace - 18073
#JasmineTWBA -
#TBHBagongBuhay -
#TGLAwakeningMoment -
#NoToJeepneyPhaseOut -
#IvoLiveKYRU -
```

FIGURE V. TRENDING TOPICS IN THE PHILIPPINES

The idea behind was harnessed by the researcher and described how it works behind its users, JSON formatter and validator shows the name, url, promoted content, the query and the tweet volume a snippet of the code is shown below.

```
{
  "name": "#Oscars",
  "url": "http://twitter.com/search?q=%23Oscars",
  "promoted_content": null,
  "query": "%23Oscars",
  "tweet_volume": 8230427
},
```

FIGURE VI. NAME, URL, PROMOTED CONTENT, QUERY AND TWEET VOLUME OF A TREND

Fig. 6 shows that a particular trend can be identified by its name and redirects its user in a particular content and it counts the number of times people visited the url provided in a particular hashtag. Moreover, as described by Russell (2013), JSON provides a way to arbitrarily store maps, lists, primitives such as numbers and strings, and combinations thereof. In a brief analyzation, the researcher can theoretically model just about anything with JSON whatever you desire to do so [11].

3.2 SEARCHING FOR TWEETS

One of the common items between the sets of trending topics as shown in Fig. 4 turns out to be the hashtag (#) symbol. The researcher used #Oscars as the basis of search query to fetch some tweets related to Oscars that will be used for further analysis and exploration of social trends. The researcher extracted from the list a single tweet and stored in variable. The figures below shows the data retrieved.

status text

RT @FOXsoccer: And you thought the mixup at the #Oscars was bad...
 RT @weirdornotharry: Name a more iconic duo. I'll wait #Oscars
 If the #Oscars mix-up had happened in Naja, this is how we'd have handled it
 RT @badestoutfits: Priyanka Chopra's #Oscars
 If the #Oscars mix-up had happened in Naja, this is how we'd have handled it
 RT @IniciativaV: Nuestra #CaptainMarvel. @BrieLarson no aplaudirá a #CaseyAffleck en los #Oscars por su apoyo contra el acoso hacia las mujeres.
 RT @13StevenGraves: Ughhh I need to sleep, but these #Oscars follow-up reactions got me fired up! I truly feel #Moonlight was robbed of a gift.
 RT @TheRealStanLee: No "Best Cameo" award #Oscars #OscarsMistake
 RT @LamSteveHarvey: Call me Warren Benty. I can help you get through this! #Oscars
 RT @funder: Time to talk Trump impeachment #impeachment #trump #resist #theresistance #nodapl
 RT @ACLU: "I'm wearing an ACLU ribbon because they're fighting incredible fights right now for American ideals."
 RT @jaser182: Se la la hand fosse il film della mia vita ci sarebbero 17 nomination e 0 vittorie e la locandina diventerebbe così!
 RT @Gerdalomc: Cuando te dicen que ya no hay racismo en EEUU pero ponen a dos de los Avengers a vigilante durante los #Oscars
 RT @Fon_Reloaded: Los #oscar resumidos en 19 segundos
 RT @OSWALDRIOSM: Los #Oscars en una imagen...

FIGURE VII. STATUS TEXT

Fig. VII shows only fifteen (15) current status text on the time the researcher is writing this paper which is posted by different people relating to the topic about #Oscars.

Screen names

Irrviiiiinnn
 ziamondsdancing
 engraliyu80
 FCBaouz
 HiveMediaNG
 QueenChopped
 adboa95
 emelia_drysdale
 paolardz99
 LAUrbanBohemian
 BeanieButler
 RUGGESH4WN
 XaviOutes
 GHMartista
 fjs955

FIGURE VIII. SCREEN NAMES

Fig. VIII shows the fifteen (15) screen names used by the people who posted the retrieved status text as shown in Fig VI.

Table 2. Word Count

Word	Count	Word	Count
RT	13	Https://	1
#Oscars	12	@Fon_Reloaded	1
a	5	una	1
en	4	imagen	1
the	4	https://t.co/pKon	1
los	3	@OSWALDRIOS	1
I	3	resumidos	1
this	2	19	1
ls	2	racismo	1
Naija	2	hay	1
had	2	Nomination	1
in	2	0	1
Happened	2	vittorie	1
wat	2	17	1
it	2	mia	1
e	2	vita	1
la	2	ci	1
to	2	impeachment	1
me	2	#impeachment	1
no	2	diventerebbe	1
handled	2	dicen	1
mix-up	2	que	1
have	2	ya	1
how	2	te	1
you	2	Cuando	1
Los	2	@gerdalomc	1
was	2	della	1
#oscar	2	@lamSteveHarvey	1
if	2	@Iniciativa	1
#trump	1	Nuestra	1
@jaser182	1	#CaptainMarvel	1
@lin-manuel	1	Priyanka	1
#Oscars	1	@BrieLarson	1
#theresistance	1	contra	1
La	1	el	1
#russiagate	1	acoso	1
fosse	1	apoyo	1
ideals	1	su	1
#resist	1	#CaseyAffleck	1
Se	1	por	1
for	1	@badestoutfits	1
I'm	1	bad	1
film	1	mixup	1
@ACLU	1	@FoxSoccer	1
wearing	1	And	1
an	1	thought	1
ribbon	1	@weirdornotharry	1
ACLU	1	Name	1
https://	1	wait	1
#nodapl	1	I'll	1
fights	1	duo	1
rights	1	more	1
now	1	iconic	1
Incredible	1	hacia	1
fighting	1	las	1
because	1	#OscarsMistake	1
they're	1	Call	1
American	1	Warren	1
locandina	1	award	1
Avengers	1	Cameo	1
vigilante	1	@TheRealStanLee	1
durante	1	No	1
de	1	Best	1
dos	1	Beatty	1
EEUU	1	can	1
Pero	1	@funder	1
ponen	1	Time	1

Table 2 shows that the entities with a frequency greater than two are interesting, the result reveals that “RT” outnumber other words at table 2 showing that it was a very common token implying that there were a significant number of retweets. The #Oscars was able to gain a numerous token and also at the top together with “RT”. Finally, as expected, the extracted tweet #Oscars and a couple of case-sensitive variations dominated the single extracted tweet that was being experimented, a data processing is essential to normalize each word, screen name, and hashtag to lowercase when tabulating frequencies since there will inevitably be variation in different tweets.

4. RESULTS AND DISCUSSION

As the paper further investigate social trends it was able to discern the advanced measurement that involves calculating simple frequencies and was applied to unstructured text in this metric is called lexical diversity. The concept can be discerned mathematically, this is an expression of the number of unique tokens in the text divided by the total number of tokens in text. This definition can be correlated to the weighing scheme also know as term frequency, denoted and shown at Figure 1. Term Frequency Formula. Lexical diversity as stated by Russell (2013) is an interesting concept in the area of interpersonal communications because it provides quantitative measure for the diversity of an individual’s or group’s vocabulary [11]. This can be worth considering primitive statistics for measuring how broad or narrow the subject matter that an individual or group discusses. For exploration and research purposes it would be interesting to measure if there is significant difference between the lexical diversity of the two controversial movies, La La Land and Moonlight comparing the conversation of people on what happened during that moment and will compute #Oscars lexical diversity as well.

```
words 0.82
screen_name 0.9333333333333333
hashtags 0.40909090909091
status_text 16.666666666667
```

FIGURE VIII. #OSCAR'S LEXICAL DIVERSITY RESULT

Fig. VIII. Shows the result of the computed lexical diversity of #Oscars. In the figure, the following observation is worth considering in the results:

- The lexical diversity of the words in the text of the tweets is around 0.82. It can be perceived that each status update has 82% unique information. Given that the average number of words in each tweet as indicated in the status text is around 17. It can be discerned that there is about 14 unique words per tweet. The #Oscars shows that there are a lot of people would want to convey message on the said hashtag.
- The lexical diversity of screen names, however, is even higher, with a value of 0.93, which means that most people are not providing the same responses for the #Oscars.
- The lexical diversity of the hashtags is extremely low at a value of around 0.41, indicates that very few term can be coined in #Oscars.

```
words 0.51476793248945
screen_name 0.9333333333333333
hashtags 0.39130434782609
status_text 15.8
```

FIGURE IX. #MOONLIGHT LEXICAL DIVERSITY RESULT

Fig. 9 Shows the result of the computed lexical diversity of #Moonlight. In the figure, the following observation is worth considering in the results:

- The lexical diversity of the words in the text of the tweets is around 0.51. It can be said that each status update has 51% unique information. Given that the average number of words in each tweet as indicated in the status text is around 16. It can be discerned that there is about 8 unique words per tweet. Based on the data #Moonlight is a trending hashtag that solicit a response has a fewer words compared to #Oscars and #Lalaland. Still they hold the title but not in the computation of its lexical diversity.
- The lexical diversity of screen names, however, is even higher, with a value of 0.93, which means that most people are not providing the same responses for the Moonlight.
- The lexical diversity of the hashtags are extremely low at a value of around 0.39, indicates that very few term can be coined in #Moonlight.

```
words 0.65277777777778
screen_name 0.933333333333333
hashtags 0.58974358974359
status_text 14.4
```

FIGURE X. #LALALAND LEXICAL DIVERSITY RESULT

Fig. 10 Shows the result of the computed lexical diversity of #Lalaland. In the figure, the following observation is worth considering in the results:

- a. The lexical diversity of the words in the text of the tweets is around 0.65. It can be perceived that each status update has 65% unique information. Given that the average number of words in each tweet as indicated in the status text is around 14. It can be discerned that there is about 9 unique words per tweet. The researcher sense that based on the data #Lalaland is a trending hashtag that solicit a response that has more vocabulary and that there is many words that a particular person wants to convey and compared to #Moonlight in the world of lexical diversity #Laland has more unique words and it shows that it is talked of the town compare to #Moonlight in this experiment.
- b. The lexical diversity of screen names, however, is even higher, with a value of 0.93, which means that most people are not providing the same responses for the #Lalaland.
- c. The lexical diversity of the hashtags is extremely low at a value of around 0.59, indicates that very few term can be coined in #Lalaland.

4.1 VISUALIZING FREQUENCY DATA WITH HISTOGRAMS

This paper reveals insight into the underlying sociality discerned in Twitter, frequency distribution also known as frequency formula shown in Figure 1 is the key and the approach taken to identify the most popular tweet in Twitter. To visualize frequency data histograms was used to plot data taken. This is an easiest way to represent tuple where each tuple is a (word, frequency) pair. The x-axis value correspond to the index of the tuple that was retrieved by the researcher and the y-axis correspond to the frequency for the word in that tuple.

Table 2. Word Count shows where the maximum and the minimum value on the table to compute for the range by simply subtracting the two value. After such decision on how many intervals to set by counting the records in the table and get the square root of the number of observations, it is ideal to use a round number. The interval width should be approximately equal to the range divided by the number of intervals and the following table was analyzed using excel and was able to generate the histogram by using the data analysis feature. Results are shown in the figure below.

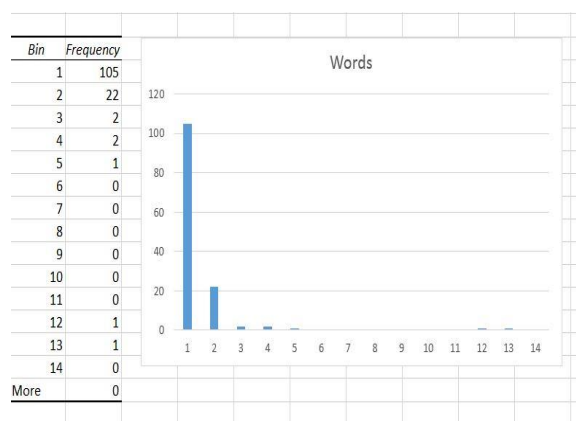


FIGURE XI. WORD RANK

Fig. 11 shows the plot displaying the frequency of words computed and analyzed by simply identifying the maximum and the minimum value on table 2 and utilized excel to generate the histogram. It can be gleaned that there are 105 words that are used once in the data retrieved by the researcher about #Oscars. It can be stated that based on the result #Oscars has a large volume of people that uniquely brought up idea about the topic that shows high side for lexical diversity of ordinary human communication. Given the nature of data, it shows that it is difficult to discern people's sentiment or common interest in this kind of result.

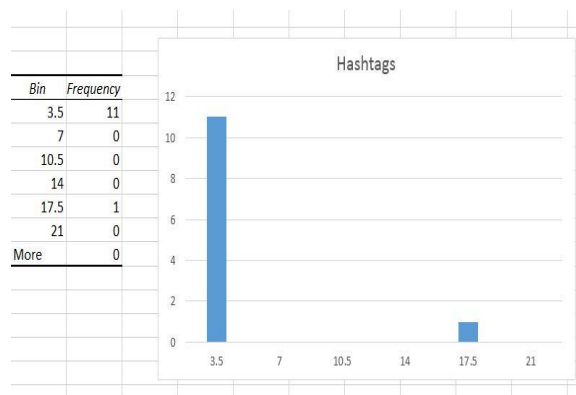


FIGURE XII. HASHTAGS

Fig. 12 shows the number of times a particular hashtag appears. In this case, it can be seen that Oscars apparently capture the attention of the users because of the result in this figure.

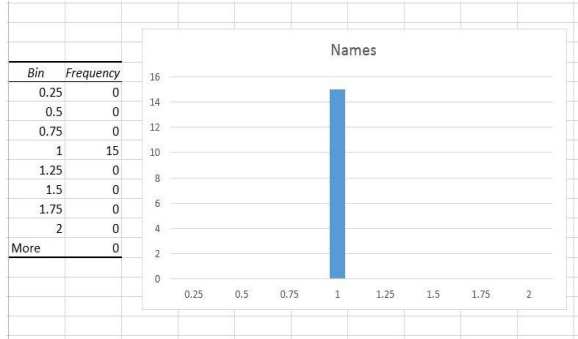


FIGURE XIII. SCREEN NAMES

Fig. 13 shows the number of times a particular screen name appears. It can be interpreted that no similar screen names appeared on our data tested and all the screen names are identical and unique users.

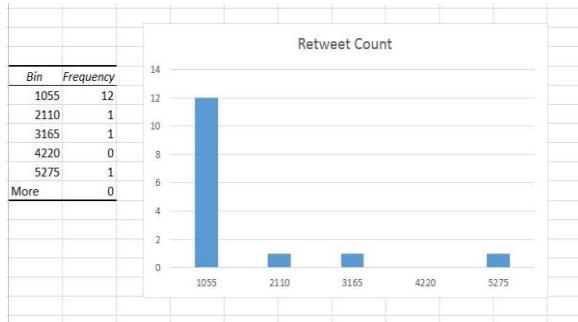


FIGURE XIV. RETWEET

Fig. 14 shows a histogram of retweet frequencies. The figure reveals a high retweet count that shows meeting of the minds between users and people apparently share the same sentiment and common interest. This is one of the primary data structure that supports human interaction, and needs further analysis.

5. CONCLUSIONS

The researcher discerned from an analytical standpoint is that counting is generally the first step in all complicated mathematical calculations. This is the key to a meaningful quantitative analysis on this study. Although basic frequency analysis is easy and simple it is a powerful tool that leads to unimaginable technology of today and many advance research and technology depends on it. These modest technique leads to the following conclusions:

- (a) Term frequency captures values and generate a list of comprehension. The generate word, hashtags, screen name and retweet count to display popular tweets are usually coined as trending topics.
- (b) Identifying trending topic is a valuable one because it allows users to create a concept that leads to discover a group of other users that apparently share the same sentiment or interest. This can be a good platform to understand each other and create a symbology that represent a concept.

6. RECOMMENDATIONS

The data that was used to discover the research, theory and the sociality discerned in Twitter led to the following recommendation which the researcher foresee in the next generation of technology:

- (a) The ability of twitter to capture the number of times a word was able to post or mentioned. This can be a good tool to track the behavior of the user and can be a precise crime-fighting tool in the future.
- (b) Trending topics can help to identify important conversations within a community. It can be a powerful tool to decide on a controversial issue and as more and more are getting connected to each other collaboration can be a new direction on this platform.

REFERENCES

- [1] Biography.com Editors. (2015, October 15). *Jack Dorsey Biography*. Retrieved from bio: <http://www.biography.com/people/jack-dorsey-578280#billionaire-businessman>
- [2] Elliot, R. (n.d.). *Where on earth identifiers*. Retrieved from woeid lookup: <http://woeid.rosselliot.co.nz/>
- [3] Fried, I. (2011, June 1). *Twitter Chairman Jack Dorsey Still Can't Define What Twitter Is*. Retrieved from All Things D: <http://allthingsd.com/20110601/twitter-chairman-jack-dorsey-still-cant-define-what-twitter-is/>
- [4] Ghode, D. R. (n.d.). *Understanding Facebook Popularity through its Semiotic Analysis*. Pune. Retrieved from Academia.
- [5] Langley, M. (2016). Retrieved from <http://michael.langley.id.au/blog/posts/37>.
- [6] Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Retrieved from Stanford: <http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>
- [7] Patterson, M. (2014, August 15). *How to Use Hashtags on Every Social Media Network*. Retrieved from sprout social: <http://sproutsocial.com/insights/how-to-use-hashtags/>
- [8] Peirce, C. S. (1932). *Collected Papers of Charles Sanders Peirce*. United States of America: Vol. II of Elements of Logic at Harvard University Press, 1932.
- [9] Popper, B., & Erlick, N. (2017, February 1). *Facebook is closing in on 2 billion monthly users*. Retrieved from The Verge: <http://www.theverge.com/2017/2/1/14474534/facebook-earnings-q4-fourth-quarter-2016>
- [10] Quartz. (2017, February 16). *Mark Zuckerberg doesn't need to be US president; he can be president of the global world society*. Retrieved from Quartz: <https://qz.com/913176/mark-zuckerbergs-manifesto-reads-like-a-state-of-the-union-address/>
- [11] Russell, M. A. (2013). *Mining the Social Web*. United States of America: O'Reilly Media, Inc., 1005 Gravenstein Highway North. Sebastopol, CA95472.
- [12] Swisher, K., & Wagner, K. (2017, February 16). *Facebook's Mark Zuckerberg wrote 6,000-word letter addressing fake news and saving world*. Retrieved from recode: <http://www.recode.net/2017/2/16/14632726/mark-zuckerberg-facebook-manifesto-fake-news-terrorism>
- [13] TechTerms. (2008, January 14). *Facebook*. Retrieved from TechTerms: <https://techterms.com/definition/facebook>

AUTHORS PROFILE



FULL NAME: Jasmin De Castro Nigidula
TITLE: Chair Computer Science / Information Systems at Technological Institute of the Philippines
Citations: 7 Since 2012
h-index: 2



FULL NAME: Angelique Dulay Lacasandile
University / Affiliation: Technological Institute of the Philippines – Manila (Graduate Studiers)
TITLE: Academe Linkage Coordinator / Faculty at National University – Manila

GOOGLE SCHOLAR

<input type="checkbox"/>	Title	+ Add	More	1-3	Cited by	Year
<input type="checkbox"/>	BeWarned: An Android-Based Application for Public Safety and Disaster Preparedness within the adopted Barangay of National University					2015
	AC Angelique Lacasandile, Aerial Tablante, Jayvee Baluyot, Emily Yabut... IEEE Humanitarian Technology Conference					
<input type="checkbox"/>	I-Health: A Nutrition Fact Reader with Workout Plan Assistant Utilizing QR Code Technology					2015
	A Lacasandile, J Caballero 3rd International Conference in Technology 2015					
<input type="checkbox"/>	E-viaje: An Android-Based Mobile Route Planner on Philippine Road Maps Utilizing Haversine Formula					2014
	AD Lacasandile http://www.icodel.org/home/ICODEL%202014%20CONCURRENT%20SESSIONS.pdf Page 47					

Simulation of Artificial Life Modeling with Modified Biogenesis Model in Virtual Ecosystem

Md. Shahriare Satu ¹, Mohammad Abu Yousuf ²

^{1,2} *Institute of Information Technology*

Jahangirnagar University

Dhaka, Bangladesh

Email: ¹shahriar.setu@gmail.com, ²yousuf@juniv.edu

Abstract—Artificial life is a field where researchers evaluate natural life related system, its processes, evolution to use its different field of applied science. Simulated model of artificial life that shows living behaviors of organisms virtually that is considered the addressing questions of this research. A virtual ecosystem called Biogenesis which simulates several processes that involves in the evolution of unicellular organism of nature. With no enough definition of life, this model is represented artificial creatures whose tried to achieve living characteristics from its simulated ecosystem. In this work, proposed biogenesis model is implemented in biogenesis x mode model by combining some feature of the biogenesis color mode model that simulates artificial creatures in the virtual ecosystem. Then we analyze possible outcome of proposed biogenesis model and compares the result with traditional biogenesis model, biogenesis x mode model and biogenesis color mode model. Then, it shows that this proposed biogenesis model is more interactive to represent life-like attributes than traditional biogenesis model, biogenesis x mode model and biogenesis color mode model.

Keywords—Artificial Life, Organism, Population, Segment, Gene, Genetic Code, Atmosphere

I. INTRODUCTION

Life is a characteristic that distinguishes physical entities by containing self-sustaining and biological processes. Various forms of life exist in the environment such as plants, animals, fungi, protists, archaea, and bacteria. Simulation of living organisms is a great attempt to synthesise and understand life-like behaviors. Chris Langton is the founder of the research discipline known as artificial life [1]. He defines the subject as the study of man-made systems that exhibit behaviors, characteristics of natural living systems. There are defined two types of artificial life which are called weak artificial life and strong artificial life. Proponents of weak artificial life represent life by developing synthetic models associated with living organisms on computer and other artificial media. But it does not demand that any parts of their models are actually living themselves. On the other hand, proponents of strong artificial life claim that the end product will be deserved the term 'living' as biological organisms by instantiating such processes in artificial media. It must be embedded within a physical, symbolic environment that is provided by a computer program might be enough for this purpose.

There are remaining many philosophical and practical obstacles to create artificial life. A major problem is to consider the enormity of life in terms of the numbers of molecules and billions of years over evolution which has been proceeding in the biological world. Another argument concerns the nature of the major evolutionary innovations in biology, which provides organic evolution with an unlimited supply of phenotypic novelty.

A computer model that can represent the essential materials of the real world. If there is simulated a model of artificial life with different properties, it might be expected complex forms to emerge in a complex environment. The potential benefits of this work is contributing to our scientific knowledge of biological evolution. Even if we are ultimately unsuccessful to reach the grand goal of synthesising artificial life, but the nature of the ways will be instructive. Most of the models are used as systems in evolutionary biology and ecology. When the system is designed a model, those systems are tracked the changes in macroscopic measures. So, it is hard to build model incomplete mixing and spatial inhomogeneity. This type of research can enable to evolve an unlimited variety of complex adaptations is obvious significance in areas such as artificial intelligence, machine learning, evolutionary computation, computer games.

In this paper, we represent our work in a different section. Section I represent the introductory overview of this work. Section II shows previous related works in this field. Then, Section III describes our proposed biogenesis model which is integrated, modified from different version of biogenesis. Section V shows a possible outcome of proposed biogenesis model and compare this with the outcome of existing biogenesis models. Section VI summarized the contents and future plan of this work.

II. RELATED WORKS

The concept of ecosystem first appeared by ecologist Tansley in an article which assess the correlation between the organism and its environment [2]. The behavior of artificial creatures impacts significantly by the environment. The movement of the organism plays an important role and most active and diverse research in the biological field [3].

Craig Reynolds explained how to simulate and animate the behavior of flocks and herds that are based on the actor model of distributed computation in 1987 [4]. Millar created a model of the mass-spring system in snacks and worms [5]. Zelter and McKenna developed a Hexapod model using biologically-based control schema [6]. Drougoul showed a multi-agent simulation of ant colonies that is based on reactive agents [7]. A lot of different work about evolving behaviors of artificial creatures in three-dimensional, physical environment are based on Karl Sims [8]. He also presented a system which is created virtual agents by genetic language [9]. Some of work was developed to study the evolution of different behaviors like light following behaviors [10] [11] [12] [13], predator-prey co-evolution [13] [14], foraging behaviors [15] [16] [17] in virtual ecosystem. Chaumout provided one of the first examples of foraging supported evolution for three legged animals that are physically simulated in a complex environments. Some creature has bizarre morphology and different from real animals. Cliff has simulated co-evolution between a predator and prey. Palmer has developed an evolutionary track between two spider species during interacting with them [18]. Xiaoyuan Tu showed artificial fishes where each animal was represented as autonomous agent [19]. The agent control was divided into perception, intention generator, behavior and motor control. Besides, his original work was extended including learning and better perception simulation. Dickerson proposed fuzzy cognitive maps to simulate a virtual environment [20]. Kandrop presented the simulation of growth and forms of sponges and corals [21]. Blumberg described a system of user interaction with animated virtual dog and other creature [22]. Technosphere is an another interesting work which allows people to create virtual creatures and monitor their evolution. Recently, Ito represented a model on the 3D simulated environment that interacts between evolutionary dynamics and population using an artificial life approach [23].

III. PROPOSED BIOGENESIS MODEL

Biogenesis is an open source platform that simulates the evolution of unicellular organisms [24]. The idea of biogenesis is taken from primordial life. Proposed biogenesis model is implemented in biogenesis x mode model with some features of biogenesis color mode model [25] [26]. Both of them modified version of traditional biogenesis model [24]. The goal of this work to improve the performance of showing life-like behaviors like self-replicating reproduction, interacting with other organisms, regulating genes of organisms rather than other Biogenesis models (traditional biogenesis model, biogenesis x mode model, biogenesis color mode model). So, to fulfill our mission, we analyze and modify existing biogenesis models by increasing number of color segments and change some configuration parameters of biogenesis and improve its performance compare to the different version of biogenesis model. Figure 1 shows the architecture of proposed biogenesis model. There are implemented different

types of elements which mimic the real world in the virtual ecosystem that describes briefly as follows:

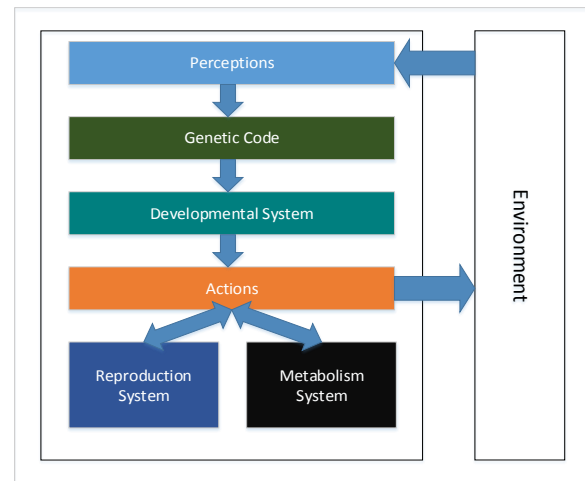


Fig. 1: Proposed Biogenesis Model Architecture

A. The World

The world is a discrete two-dimensional spatial environment. It consists of organisms, atmosphere and different actions (behaviors) in the world [24] [26] [25]. The spatial space is considered as the fluid. So, organisms can move around the world using their cilia that contains cyan segments. At the start of each unit of time, a number of instructions are started to create new organisms, energy, and analysis the simulated world. The organism is tried to represent several interactions like track, feed, weaken, kill, copy, reproduce and rejuvenate. However, some instructions are necessary to maintain organisms movements, collisions, time in the environment.

B. Atmosphere

An atmosphere is a layer of gasses that held by the gravity and surrounds the world or other material bodies [24] [26] [25]. It provides sufficient energy to the organisms. It contains oxygen and carbon dioxide gasses. Oxygen is a secondary product of the photosynthesis that occupies green segments and all organisms use it in the respiration process. On the other hand, carbon dioxide is the primary source of carbon for the organisms and the major factor in every metabolism. It is produced as a result of respiration and is consumed in the photosynthesis process.

C. Organism

Organism is the phenotypical representation of living being by genotype(genetic code) [24] [26] [25]. Since the world has begun, a number of generations that have exist by counting its descendants. It is achieved living behaviors and analysis

interactive performance of the world.

Organism has contained energy for usual activities that achieved from different resources. Each segment has consumed energy to maintain different kinds of activities like organism's longevity, reproduction of descendants. By touching some specific segments of others, energy can be lost of an organism. If organism is growing, then update the energy contained in the body using following equation 1 & 2.

$$E_{com} = (1 + BB_{EF}) * \frac{10 - G_R}{10} * E_R \quad (1)$$

$$E_{min} = BB_{EF} * \frac{10 - G_R}{10} * E_R \quad (2)$$

Where E_{com} and E_{min} are comfort and minimum energy of organism. Besides, BB_{EF}, G_R, E_R are denoted as body bound energy fraction, growth ratio, energy of genetic code for reproduction. Growth of the organism is considered how to increase size of the organism at a time. In Biogenesis, mass and inertia of the organism and accelerate growth functionality using following equation 3 & 4.

$$m = 2 * \sqrt{\frac{m}{M}} \quad (3)$$

$$i = \frac{\theta}{2} \sqrt{\frac{i}{I}} \quad (4)$$

Where m, i are indicated as mass, inertia at initial state. But M, I are updated mass and inertia after executing the instructions related to growth of organism. Besides, when an organism exceeds life expectancy, it is dying and corpse of this organism is dispersed slowly on the environment.

Now, we consider our proposed biogenesis model into three different sections called genetic code, reproduction, metabolism which are described as follows:

1) *Genetic Code*: Genetic code is the set of instructions that maintains existing organisms in the virtual ecosystem [24] [26] [25]. Segments can be drawn several times to the organisms with other characteristics that are called genes. Each segment of organism's body generate a particular instructions to the genetic code. Genes are the key factor of creating organism's genetic code. Organism's position depends on the neighbors and structures of genes. Some attributes of Biogenesis Color Mode Model also integrates in biogenesis x mode model like generation battle, siblings battle, altruist, familiar, social, passive, clockwise, mimical, modified spink, modified scream, modified lilac, colonyplus. So, it is increased the diversity of existing organisms in the virtual ecosystem.

There are considered two different things called segments and gene structure which is needed to create genetic code of

Algorithm 1 Randomised color segment generation

Input: Color segments and their Probabilities

Output: Create randomized color segments sequence for new gene.

```

1: Initialisation :
2: Find Maximum Probability  $P_r^m = \sum_{s=1}^{28} P_s$ .
3: Select the Probability  $p_r$  of color segment randomly within Maximum Probability  $P_r^m$  .
4:  $i = 1$ 
5: while  $i < 28$  do
6:    $apr = apr + p_r$ 
7:   Set Color Segment Probability to  $apr$ 
8:   if  $p_r < apr$  then
9:     Set the Pigment Color to  $p_i$ 
10:    return  $p_i$ 
11:  end if
12:   $i = i + 1$ 
13: end while
14: Set the Pigment Color as yellow to  $p_i$ 

```

the organism that are described as follows:

- **Segments:** There are integrated 21 color segments of biogenesis color mode model in 7 color segments of biogenesis x mode model in the organism's genetic code. Algorithm 1 shows how to generate randomised color segments for creating new gene of the organism. In this proposed biogenesis model, segments are considered into two major category called primary and secondary segments.

- **Primary Segments:** Primary segments are essential for fundamental operations of organisms. Each color has different functionality in the environment. Green segments obtain energy to consume carbon dioxide and release oxygen. This amount of oxygen is similar to the carbon dioxide consumption. Red segments get energy from other organisms and organic materials and effective against any other segment except blue segments. A part of consumed energy is freed as carbon dioxide. Blue segments are secured from the effects of red, white and gray segments. The organism is shown in blue momentarily when a blue segment becomes activated. Cyan segments allow to move the organism through the surroundings. They get activated randomly and impel the organism in the directed segments points. White segments causes infection to the organisms when they communicate its in the organism. Infected organism have affected a particular species using its minimum energy instead of its own. But an organism can be infected once in the environment. This segments are effective against any other segment except blue. Gray

segments are harmful segments in the environments. When they touch one to another, it becomes dying. Yellow segments are increased this fertility. Each segment produce more descendant of maximum symmetric property.

- **Secondary Segments:** Secondary segments are used to add extra features in the organism. There are existing 21 segments. Dark segments are mimic other segments. Fire segments are worked as omnivore between red and orange. If the other segment is blue, it acts as a shield. Forest segment enhance photosynthesis in a colony. Lilac segments are weaken organisms. Magenta segments are heal all sick segments. Maroon segments are tried to get energy from plants and feed on ochre, sky, dark olive and cracked light blue segments. Mint segments are removed an infection, corrupt all white segments. Orange segments are tried to get energy from the other organisms. If the other segments are blue, it acts as a shield. Pink segments are tried to get energy from corpses, weak organisms, viruses and parasites. Silver segments are infected to all other organisms, if it has more infections, duels with other absorbing segments. Sky segments are freed to another organism by disabling its photosynthetic and movement segments, protects itself with a frost shield. Violet segments are poisoned to the another segments and make it useless.

- **Gene Structure:** A gene is the basic physical and functional unit of heredity. They are represented as an organism using a particular forms or shapes. This forms are divided into following two segments. Symmetry indicates that segment pattern repeats into certain periods of times. Besides, mirror specifies if the organism has symmetry that can cross the center or not with respect to a straight unit of genes.

2) *Reproduction:* Reproduction is the process where new individual organisms are produced from their parents [24] [26] [25]. Sufficient energy is needed for reproduction of an organism. To consider unicellular organisms, there are considered asexual reproduction called mutation in this work. It consists of some functionality like add or remove segments, swap segments, flip gene sequences, split or merge segments, different random mutations, change angle, color, gene length mutations, shrink and enlarge organism slightly. There are considered some extra features of mutation like major mutation rate, minor mutation rate, spontaneous mutation rate, stress resistance. There are used following equations for reproduction.

$$E_C = E * \frac{(1 - Y_I^R)}{N_C} + \frac{Y_I^R}{N_C * (L_Y^{av} - G_{min})} \quad (5)$$

$$D_G = G_{max} - G_{min} \quad (6)$$

$$E_C^{av} = \begin{cases} \frac{1}{2} * \frac{E_C}{D_G} & N_C > 1 \\ \frac{1}{2} * E & N_C = 1 \end{cases} \quad (7)$$

Where E , E_C and E_C^{av} are indicated as net energy, energy of child organism and average energy of child organism. Besides, Y_I^R , N_C , L_Y^{av} , G_{max} , G_{min} , D_G are denoted as yellow reproduction influence, number of children, average length of yellow segments, maximum and minimum length of gene, distance between gene length.

3) *Metabolism:* Metabolism is the set of actions of an organism that acquires energy from convenient materials in the environment [24] [26] [25]. Energy can be acquired by accomplishing photosynthesis, respiration and organic matter processing. Carbon dioxide is transformed into energy and oxygen is freed to the environment in the photosynthesis process. Green segments are responsible to maintain and execute photosynthesis of an organism. The length of the photosynthetic segments is proportional to the amount of energy. Besides, energy is acquired from processing organic matter and transferred from the prey to the predator. Red segments can consume energy from other organic matters that is proportional to their lengths. Blue segments are indestructible to the red segment's effects. They obtain and decrease energy from organic matter that synthesized by other organisms. This is called Chemoautotrophic Metabolism. On the other hand, organisms are accomplished in the respiration process by transforming stored energy to the effective energy. In this case, carbon dioxide is freed and oxygen is consumed in the environment. To keep longevity, this process has executed continuously through respiration environment.

IV. RESULT AND DISCUSSION

Biogenesis is a project of unicellular organisms simulation that implemented in Java SE (Standard Edition). In this experiment, We work with biogenesis 0.8 version of traditional and modified biogenesis project. Proposed biogenesis model is implemented in biogenesis x mode model by integrating some functionality of biogenesis color mode model. Our proposed biogenesis model is used 28 color segments to generate genetic code and each segment has unique feature to generate new organisms of those features. Runtime view of proposed biogenesis model is shown in Fig 2.

In Fig 2, virtual ecosystem is contained black surface that is conceptually considered area of liquid. The organism of proposed biogenesis model is initiated with 100 organisms in virtual ecosystem. We execute traditional biogenesis model, biogenesis x mode model, biogenesis color mode model and proposed biogenesis model within 1.30 hours and compare the outcomes of traditional biogenesis model, biogenesis x mode model, biogenesis color mode model and proposed biogenesis

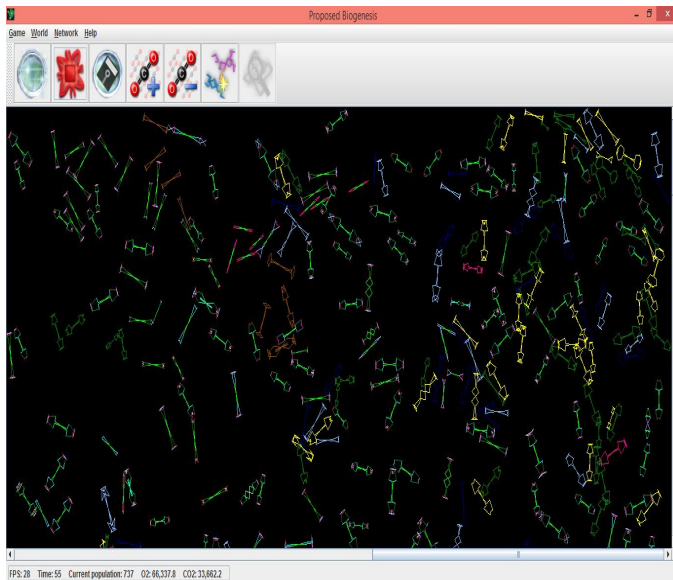


Fig. 2: Screenshot of Simulated Proposed Model

model. Performance of expressing life like behaviors of proposed biogenesis model is showed in Table I based on population that comparing with traditional biogenesis model, biogenesis x mode model and biogenesis color mode model in the terms of average population(AP), average birth rate(ABR), average moral rate (AMR), average infection rate (AIR), maximum population (MXP), minimum population (MIP), mass extinction (ME). Besides, performance analysis of energy consumption of proposed biogenesis model is also showed in Table II comparing with traditional biogenesis model, biogenesis x mode model and biogenesis color mode model in the terms of maximum carbon dioxide consumption (MXC), minimum carbon dioxide consumption (MIC), maximum oxygen consumption (MXO) and minimum carbon dioxide consumption (MIO). Then, analyze the experimental outcomes which are found from this simulation.

In this experiment, we show that our proposed biogenesis model can perform much better to represent life like behaviors rather than traditional biogenesis model, biogenesis x mode model and biogenesis color mode model which is depicted graphically in Fig 3.

proposed biogenesis model is showed highest generated organisms rather than traditional biogenesis model, biogenesis x mode model and biogenesis color mode model to maintain average population(AP), average birth rate(ABR), average morality rate(AMR) and average infection rate(AIR). Now we discuss comparative study of proposed biogenesis model, traditional biogenesis model, biogenesis x mode model and biogenesis color mode model and try to find the causes of achieving highest generated organisms rather then others.

Traditional biogenesis model has less capability to show

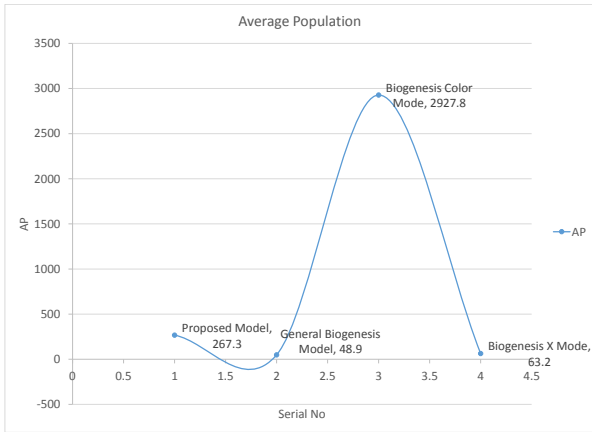
TABLE I: performance analysis of models based on population

	General Biogenesis	Biogenesis Color Mode	Biogenesis X Mode	Proposed Biogenesis
Average Population	48.9	2927.8	63.2	267.3
Average Birth Rate	3.5	495.5	5.9	415.7
Average Morality Rate	3.5	482.8	5.9	415.2
Average Infections Rate	2	26647.7	1.4	629.4
Generated Organisms	1651	28764	27259	458195
Maximum Population	86	3378	172	1276
At Time	195	47	50	24
Minimum Population	15	2311	19	44
At Time	0	7	2304	1035
Mass Extinctions	11	1	65	7

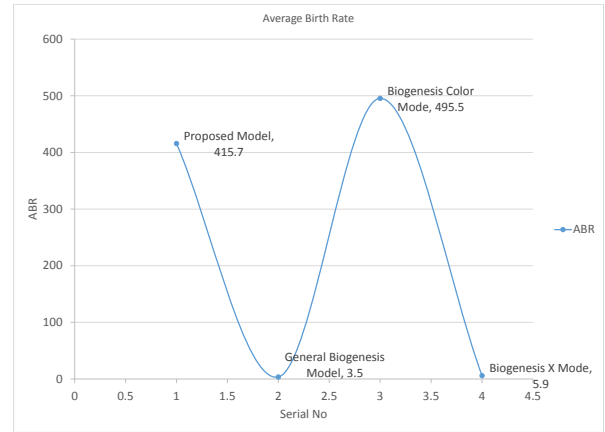
better outcomes in terms of average population(AP), average birth rate(ABR), average morality rate(AMR) and average infection rate(AIR), generated organisms(GO), maximum population(MXP) and minimum population(MIP) rather then existing modified models of Biogenesis because they use only 7 color segments for generating genetic code of new organisms. Besides, it is not used extra functionalities to increase life like behaviors.

Movement of metabolism and reproduction activities of biogenesis x mode model is faster and more interactive rather than traditional biogenesis model and biogenesis color mode model. But, average population, average birth rate(ABR), average morality rate(AMR) and average infection rate(AIR) of biogenesis x mode model are less than proposed biogenesis model because they also use only 7 color segments for generating genetic code which creates new organisms which quantity is so much less than proposed biogenesis model in the virtual ecosystem. Besides, our proposed biogenesis model is modified version of biogenesis x mode model. So, approximately all of the characteristics of biogenesis x mode model is remaining in our proposed biogenesis model with additional functionalities from traditional biogenesis model and biogenesis color mode model. There are existing another capabilities like energy consumption, growth, reproduction of organisms are less than proposed biogenesis model.

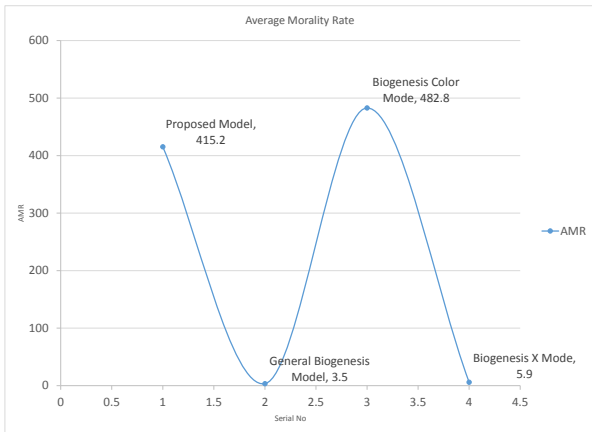
The biogenesis color mode model initially started with 2500 organisms which is more huge amount than our proposed biogenesis model. average population, average birth rate(ABR) and average morality rate(AMR) of biogenesis color mode model is higher rather than our proposed biogenesis model but average infection rate(AMR) also higher than proposed biogenesis model. So, reproduction of generated organisms is decreased in biogenesis color mode model raheer then proposed biogenesis model. Besides, reproduction and metabolism activities of our proposed



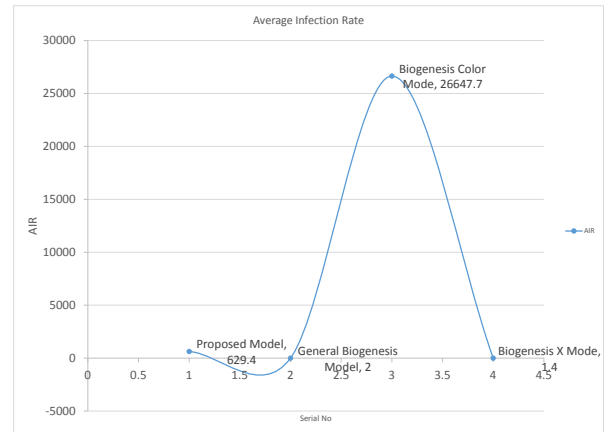
(a) Average Population



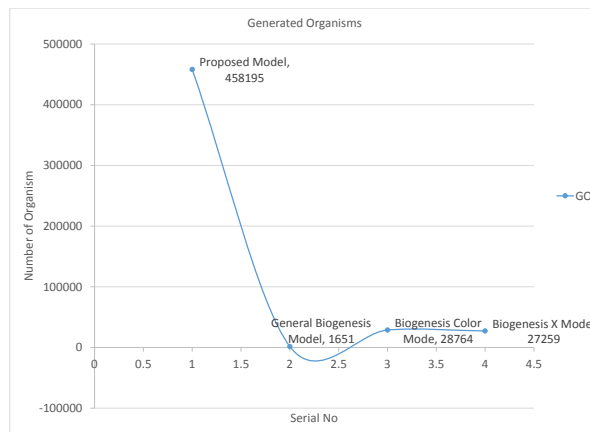
(b) Average Birth Rate



(c) Average Mortality Rate



(d) Average Infection Rate



(e) Generated Organisms

Fig. 3: Performance analysis of Biogenesis models based on population

TABLE II: performance analysis of models based on atmosphere

	General Biogenesis	Biogenesis Color Mode	Biogenesis X Mode	Proposed Biogenesis
Maximum Carbon dioxide (CO ₂)	5000	100000	7000	100000
At Time	0	0	0	0
Minimum Carbon dioxide (CO ₂)	568.2	1319.4	656.2	2175.2
At Time	415	6	1474	21
Maximum Oxygen (O ₂)	4431.8	98428.9	6343.8	97824.8
At Time	415	6	1474	21
Minimum Oxygen (O ₂)	0	0	0	0
At Time	0	0	0	0

biogenesis model is more faster and more interactive than biogenesis color mode model because proposed biogenesis model is modified version of biogenesis x mode model.

In Fig 4, we represent a pictorial comparative study of traditional biogenesis model, biogenesis color mode model, biogenesis x mode model and proposed biogenesis model based on generating organisms. In this case, we can show that our proposed biogenesis model can generate more organism than traditional biogenesis model, biogenesis color mode model, biogenesis x mode model.

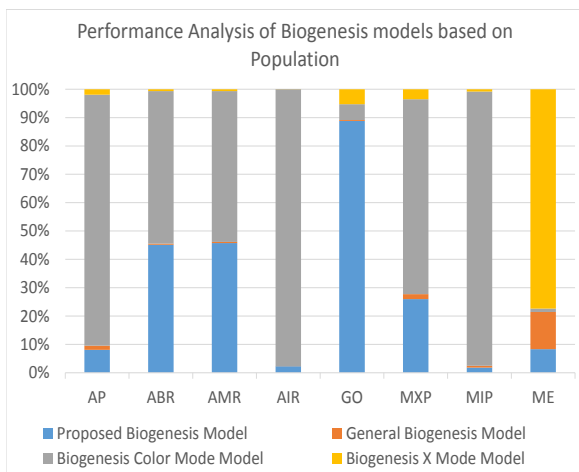


Fig. 4: Comparison to the performance of proposed biogenesis model with existing biogenesis model based on population

Energy consumption of biogenesis models are measured by photosynthesis and respiration process in the virtual ecosystem. So, we consider some parameters of maximum and minimum consumption of carbon dioxide and oxygen at a time. Initially, we set amount of oxygen 0 and amount of

carbon dioxide 100000 in proposed biogenesis model which is similar to biogenesis color mode model. Some Traditional Models like traditional biogenesis model, biogenesis x mode model are initialized less energy because they are used 7 color segments to produce new organisms and manipulate as usual tasks of environment. In Table II, we observe that energy (oxygen and carbon dioxide) consumption rate of proposed is greater than existing Biogenesis models because our proposed biogenesis model is used 28 color segments to create genetic code and can generate a lot of new organisms. In this case, it is needed more energy to produce new organisms and maintain some tasks of organisms like metabolism, reproduction in the environment. In Fig 5, we represent a pictorial comparative study of traditional biogenesis model, biogenesis color mode model, biogenesis x mode model and proposed biogenesis model based on energy consumption of organisms in different models.

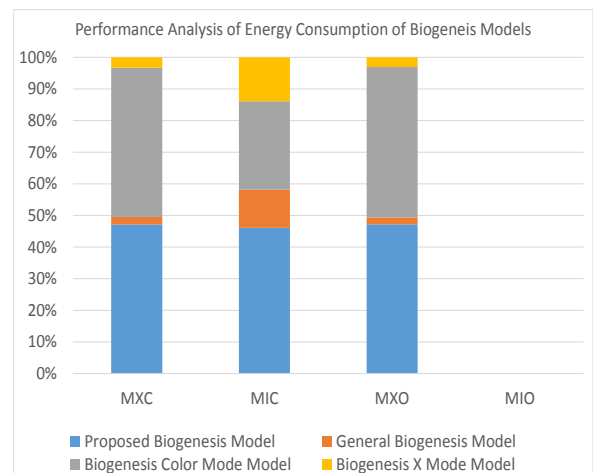


Fig. 5: Comparison to the performance of proposed biogenesis model with existing biogenesis model based on energy consumption

V. CONCLUSION AND FUTURE WORK

This research work is tried to simulate unicellular organisms in a open source project called biogenesis and improve its performance than traditional and modified version of biogenesis. But this work contain some limitations that organisms have a hardwired structure and that their interactions with other organisms are not sufficient. However, the lack of a precise and satisfactory definition of life led us to concentrate on more specific issues. This application is not well organised than some other artificial life application (e.g. Avida). So, in future, it will try to remove those kinds of weakness of this work and make more feasible system of artificial life modeling which will tread to us toward true artificial intelligence.

ACKNOWLEDGMENT

We are grateful to Ministry of National Science and Technology, Bangladesh for financing our work. We also thankful to all of teachers, staff, students of IIT, JU for collaborating this work.

REFERENCES

- [1] Zhongzhi Shi. *Advanced artificial intelligence*, volume 1. World Scientific, 2011.
- [2] Arthur G Tansley. The use and abuse of vegetational concepts and terms. *Ecology*, 16(3):284–307, 1935.
- [3] Marcel Holyoak, Renato Casagrandi, Ran Nathan, Eloy Revilla, and Orr Spiegel. Trends and missing parts in the study of movement ecology. *Proceedings of the National Academy of Sciences*, 105(49):19060–19065, 2008.
- [4] Craig W Reynolds. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH computer graphics*, 21(4):25–34, 1987.
- [5] Gavin SP Miller. The motion dynamics of snakes and worms. *ACM Siggraph Computer Graphics*, 22(4):169–173, 1988.
- [6] Michael McKenna, Steve Pieper, and David Zeltzer. Control of a virtual actor: The roach. *ACM SIGGRAPH Computer Graphics*, 24(2):165–174, 1990.
- [7] Alexis Drogoul and Jacques Ferber. Multi-agent simulation as a tool for modeling societies: Application to social differentiation in ant colonies. In *European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, pages 2–23. Springer, 1992.
- [8] Karl Sims. Evolving 3d morphology and behavior by competition. *Artificial life*, 1(4):353–372, 1994.
- [9] Karl Sims. Evolving virtual creatures. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 15–22. ACM, 1994.
- [10] Yoon-Sik Shim and Chang-Hun Kim. Generating flying creatures using body-brain co-evolution. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 276–285. Eurographics Association, 2003.
- [11] Keita Nakamura, Ikuo Suzuki, Masahito Yamamoto, and Masashi Furukawa. Acquisition of swimming behavior on artificial creature in virtual water environment. In *European Conference on Artificial Life*, pages 99–106. Springer, 2009.
- [12] Kenji Iwadate, Ikuo Suzuki, Masahito Yamamoto, and Masashi Furukawa. Evolving amphibian behavior in complex environment. In *European Conference on Artificial Life*, pages 107–114. Springer, 2009.
- [13] Dave Cliff and Geoffrey F Miller. Co-evolution of pursuit and evasion ii: Simulation methods and results. 1995.
- [14] Tiago Francisco and Gustavo Miguel Jorge dos Reis. Evolving predator and prey behaviours with co-evolution using genetic programming and decision trees. In *Proceedings of the 10th annual conference companion on Genetic and evolutionary computation*, pages 1893–1900. ACM, 2008.
- [15] Michal Joachimczak and Borys Wróbel. Evolving gene regulatory networks for real time control of foraging behaviours. In *ALIFE*, pages 348–358, 2010.
- [16] Nesrine Ouannes, Noureddine Djedi, Yves Duthen, and Hervé Luga. Toward the construction of a virtual ecosystem by evolving virtual creature’s behaviours. In *Multimedia Computing and Systems (ICMCS), 2012 International Conference on*, pages 350–355. IEEE, 2012.
- [17] Nesrine Ouannes, Noureddine Djedi, Yves Duthen, and Hervé Luga. Following food sources by artificial creature in a virtual ecosystem. *VIRTUAL WORLDS-Artificial Ecosystems and Digital Art Exploration*, pages 99–116, 2012.
- [18] Michael E Palmer and Andrew K Chou. An artificial visual cortex drives behavioral evolution in co-evolved predator and prey robots. In *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*, pages 361–364. ACM, 2012.
- [19] Xiaoyuan Tu and Demetri Terzopoulos. Artificial fishes: Physics, locomotion, perception, behavior. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 43–50. ACM, 1994.
- [20] Julie A Dickerson and Bart Kosko. Virtual worlds as fuzzy cognitive maps. *Presence: Teleoperators & Virtual Environments*, 3(2):173–189, 1994.
- [21] Jaap A Kaandorp. Simulation of growth and form of sponges and corals and the influence of the physical environment. In *Modeling Complex Data for Creating Information*, pages 199–208. Springer, 1996.
- [22] Bruce M Blumberg and Tinsley A Galyean. Multi-level direction of autonomous creatures for real-time virtual environments. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 47–54. ACM, 1995.
- [23] Takashi Ito, Marcin L Pilat, Reiji Suzuki, and Takaya Arita. Population and evolutionary dynamics based on predator–prey relationships in a 3d physical simulation. *Artificial life*, 2016.
- [24] Biogenesis. <http://biogenesis.sourceforge.net/>, 2006. Online; accessed 18-February-2016.
- [25] Biogenesis x mod. <https://sourceforge.net/projects/biogenesisxmod/>, 2015. Online; accessed 13-June-2016.
- [26] Biogenesis color mod. <https://sourceforge.net/projects/biogenesiscolor/>, 2014. Online; accessed 25-July-2016.

Employing Mean Opinion Score of Audio Lossy Compression Algorithms in VoIP Application

Ma. Emmie T. Delluza

Technological Institute of the Philippines-Manila
Catanduanes State University
emsdell@gmail.com

Jasmin D. Niguidula

Technological Institute of the Philippines
Manila, Philippines
jasmin.niguidula@tip.edu.ph

Abstract— Communication is vital. It is constructed from one individual to a group of people or vice versa. Nowadays, people always want to be connected anytime and anywhere while using communication devices. Thus, there are great demands of VoIP applications that offer quality audio and low resource utilization. In this study, some audio compression algorithms also known as codecs are considered and integrated into VoIP application. Computation of jitter, packet loss, and bandwidth is examined and analyzed. Also, the G.107 Mean Opinion Score is applied to measure the scalar quality rating value of voice using different codecs in VOIP Application.

Keywords—jitter, codecs, packet

I. INTRODUCTION

Global communication applications are familiar among all groups of people, young and old alike. Every day many communication systems offer and build effective ways of communication that are both enjoyable and beneficial to users. Originally, the Internet was not designed to convey audio or voice for real-time communications. That is why there had been several technical challenges in implementing telephone services while utilizing the Internet in relation to jitter, bandwidth, packet loss that have direct impact on VoIP audio quality. [1]

Over the years, the way people communicate has changed markedly; people want to interconnect from anywhere at any time using their own mobile devices. With this, there have been several demands for an application with the finest audio quality, low resource requirements, and fast Internet rate. Therefore, audio data compression algorithm is considered as a determinant of audio quality and resources requirements.

Several types of audio compression algorithms have been developed and standardized purposely to condense or adjust digital audio data in order to reduce bits requirements. Compression algorithms are usually employed in software. [2] Audio compression process is known as encoding; decompression, as decoding. Algorithm is popularly named as codecs, short for coder and decoder.

Generally, there are two (2) categories of data compression - the lossless compression algorithms, which do not remove

original data permanently but rather reproduce their exact duplicates by decoding a losslessly compressed file; and the lossy compression algorithms, which remove or modify audio data permanently [3]. The latter, lossy compression algorithm being used in various VoIP applications is the focus of the present study.

Likewise, the study evaluated compression in terms of resources utilization, and Mean Opinion Score of codec algorithm utilized by VoIP application.

II. RELATED WORKS AND LITERATURE

Codecs, short for compression and decompression, started in 1830 through the formation of Morse code in which letters are transmitted using short and long signals known as “dits” and “dahs” across telegraph wires [4]. Years later, the telegraph was replaced by a communication system with frequency modulation, bandwidth, and vocoder developed in 1928 by Homer Dudley. The vocoder analyzes the sound formed by a person and reforms the sound to another end [5]. In 1940s, Claude Shannon’s theory of information and the Mathematical theory of communication stand as a turning point for the birth of modern communication application such as Instant messaging, VoIP, among others [6]. Another method, Huffman Coding, finds optimal prefix codes [7][8]. From then on, other algorithms have been developed.

A. A-law Algorithm

A-law Algorithm is a standard companding algorithm used in European digital communications systems to enhance the dynamic range of an analog signal (voice) for digitizing a 16-bit LPCM data reduced to 8 bits of logarithmic data. [9] This algorithm is another type of PCM also known as G.711 standard which was industrialized by ITU-T. [10] The term companding used in this algorithm is coined from the words compressing and expanding. Companding is a process wherein a range of signal is reduced and is returned to its similar form. [11] A-law algorithm is good for low-amplitude signals and is commonly used in WAVE Audio File Format [9].

B. Warped Linear Predictive Code (WLPC)

WLPC was proposed by Hans Werner Strube in 1980. It is an improved LPC in which spectral representation of the system is modified. In designing the coding algorithm in WLPC, Strube utilized the characteristics of a human hearing since in WLPC system the spectral resolution can be adjusted to closely estimate the frequency resolution of a human being. [12][13]

C. Code-excited Linear Prediction (CELP)

CELP algorithm was originally proposed by Manfred R. Schroeder, a German physicist, and Bishnu S. Atal, an electrical communication engineer in 1985 [14]. This algorithm offers a better quality than that of the usual residual-excited linear prediction and linear predictive coding (PLC) vocoders even though CELP technique is based from LPC. The technique is utilized to model the vocal tract, codebook (adaptive and fixed) entries as input (excitation) of the linear prediction performed in closed-loop in a “perceptually weighted domain”, and to implement vector quantization [15][16]. The CELP functions in a low bit rate speech compression employed in MPEG-4 audio speech coding. Audio codec such as G.718, G.728, and G.729 are based from CELP [17][18][19].

G.729 is shaped from the coding model known as CS-ACELP. This coding format refers to an audio data compression that compacts digital voice signal into packets of 10 milliseconds duration. This G.729 works in a low bandwidth requirement of 8 kbps bit rate and is typically used in Voice over Internet Protocol applications such as Viber and Skype [20].

D. Linear Predictive Coding (LPC)

LPC is a method used for signal source modeling commonly utilized in audio signal processing and speech processing. LPC transports a prevailing speech analysis technique, delivers a quality speech at a low bit-rate, and offers accurate estimates for speech parameters [21]. One of the audio codec compressions which are implemented based from LPC is SILK. This audio codec format was developed by Skype Limited. It works at a bit rate of 6 to 40 kbps with a frequency of 8, 12, 16, and 24 kHz. SILK employed in a low algorithmic delays of 25ms and extended to the Internet standard is identified as Opus codec [20][22].

E. Modulated Lapped Transform

MLT is utilized in both audio and video data compression. It was developed by Malvar [23] as an efficient tool for local frequency decomposition of signals to transform signal processing. MLT is used in a most modern audio coding system such as Dolby Av-3, MPEG-2 Layer III, and Siren14 [24]. Siren14 algorithm is finalized with G.722.1 on the same transform coding system. This audio codec was established and certified by PictureTel Corporation, and was later developed by Polycom. Inc. [25]. Siren14 supports stereo and audio with a sampling frequency of 32kHz, 48 kbps for mono, 48, 64, 96 48 kbps for stereo, and offers 40 ms algorithmic delay utilizing

20-millisecond frame lengths [26]. One of the implementations of Siren14 is ooVoo application.

III. METHODOLOGY

In this research, the tools, procedure, settings, configurations, and equations used include the following:

A. Files

Table I presents the audio files used to transmit. Each file has the same audio content but with different audio file format and size.

TABLE I. LIST OF AUDIO FILES

File Format	Percentage	Size (kb)	File Format	Percentage	Size (kb)
.ogg	14%	96	.aac	5%	33
.aiff	11%	74	.voc	5%	32
.au	11%	74	.wma	5%	32
.flac	10%	67	.ra	4%	28
.mka	8%	55	.mid	3%	18
.ac3	8%	54	.amr	1%	4
.wav	5%	37	.mid	3%	18
.mp3	5%	34	.ra	4%	28

B. System Configuration

Client computer system configuration is shown in Table II. No other application was used except for the application needed for the test.

TABLE II. CLIENT COMPUTER SYSTEM SPECIFICATION

System Component	Description
Memory	8 GB
CPU	Intel Core i7-4510U CPU @2.0GHz 2.6 GHz
Storage	914 GB
Operating System	Window 8.1
Network Adapters	Broadcom 80211n Network Adapter, Microsoft Hosted Network Virtual Adapter, Realtek PCIe GBE Family Controller

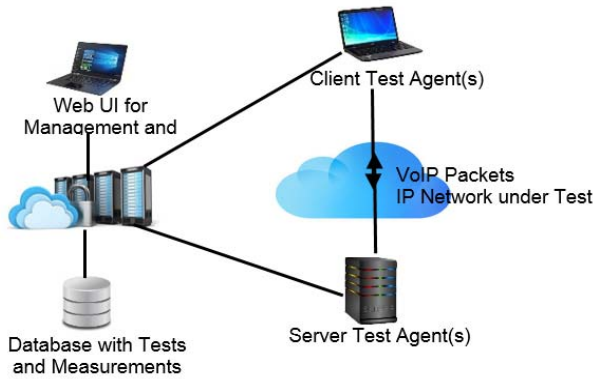
C. Test Detail Configuration

Table III reveals the test detail configuration. Figure 1 shows the network setup of the VOIP test.

TABLE III. TEST DETAIL SPECIFICATION

Test Configuration	Description
G.729	Audio Codec
1 minute	Interval measurement
100 ms	Jitter buffer size
1	Number of channels
14 minutes	Application measurement duration
startrinityServer1	Server
7am-9am, 1pm-3pm, 7pm-9pm, 1am-3am,	Test Schedule

FIGURE I. TEST DETAIL SPECIFICATION



D. VOIP Application

The VOIP application used three different protocols with two different codecs in each application [20][25], as shown in Table IV.

TABLE IV. VOIP APPLICATION

Application	Skype	Viber	ooVoo
Authentication	Login	Phone number	Phone number
Architecture	P2P	P2P	P2P
Chat	Yes	Yes	Yes
Encryption Algorithm	Public key distribution Data: AES-256 Key: RSA-1024 Certificate: RSA-2048	No encryption	No encryption
Audio codecs	SILK (8-24 kHz) G.729 (8 kHz) G.711 Opus	G.729 Opus	Polycom Siren14/ G.222.1
Minimum Internet speed required for communication	Voice: 60 kbit/s Video: 200 kbit/s	500 kbps	Video: 230kps

E. Packet Loss Computation

A packet is a set of data employed by digital devices to communicate with each other. Each packet is composed of (a) control information that determines the flow of data on where and how the data will be sent, and (b) user data that contain the actual information [27]. When packets of data are transmitted over a computer network but are unable to reach the specific destination caused by a network congestion, this results to a packet loss. The packet loss is used in this study as a measurement method to MOS. This packet loss is measured by the percentage of packets and the total number of packets sent [28][29].

$$N_n = \frac{\sum f(x)}{N} \times 100\%$$

Where: N_n - packet loss rate
 N - total number of packets
 $\sum f(x)$ - summation of packet received

F. Bandwidth Computation

Bandwidth is measured as a bit-rate of available or spent information capacity conveyed typically in metric multiples of bits per second. The maximum bandwidth is calculated with the formula [30]:

$$\text{Throughput} \leq \frac{RWIN}{RTT}$$

Where: RWIN - TCIP Receive Window
 RTT - Round-trip for the path

G. Jitter Computation

Jitter is the result of assuming periodic signal, regularly relative to a reference clock signal. A random jitter, also known as Gaussian jitter, is the changeable electronic timing noise. The deterministic jitter is the non-Gaussian probability distribution. To get the total jitter is based on the formula calculated below [31]:

$$T = D_{\text{peak-to-peak}} + 2 \times n \times R_{\text{rms}}$$

Where: T - total jitter
 D - deterministic jitter
 R - random jitter
 N - based on the bit error rate (BER)

Maximum RFC3550 Jitter is used in this study to get the maximum value of RTP stream jitter per call, calculated according to RFC3550 standard similar to Wireshark. It specifies changing delays in media flow. Large values (above 50ms = 0.05s) that show overloads in IP network or media server's CPU. The RFC3550 jitter value has straight effect with the quality of audio. Maximum instant jitter is utilized same with the RFC3550 jitter but not smoothed by a low-pass filter and is measured to the instant deviation of packet time [33].

H. MOS

Mean Opinion Score is applied in the VOIP test to measure the perceived quality of received voice after being transmitted and compressed using different codecs. The measurement is the result of underlying network attributes in different time schedules and a good VoIP test tool in determining issues that affect VoIP quality.

The G.107 e-model of Mean Opinion Score (MOS) applied to measure the scalar quality rating value called as "Transmission Rating Factor -R" using the specified jitter buffer settings of 100 milliseconds is used in this study. R is calculated based on the relation between the different impairment factors, as shown in the given formula [32]:

$$R = R_o - I_s - I_d - I_{e.\text{eff}} + A$$

Where: R_o - the basic signal to noise ratio
 I_s - impairments that occur more or less simultaneously with the signal
 I_d - the total of all impairments due to delay and echo effects
 $I_{e.\text{eff}}$ - "effective equipment impairment factor" that represents impairments caused by low bit-rate codecs and packet-losses of random distribution
 A - "advantage factor" that allow "advantage of access"

The G.107 MOS rating scales are expressed based on the estimated listener's satisfaction level, as shown in Table V:

TABLE V. MOS LISTENER'S SATISFACTION LEVEL

Rating	Level
4.3 – 5.0	Very Satisfied
4.0 – 4.3	Satisfied
3.6 – 4.0	Some Users Satisfied
3.1 – 3.6	Many User Dissatisfied
2.6 – 3.1	Nearly All Users Dissatisfied
1.0 – 2.6	Not Recommended

I. VOIP System Resource Utilization

The efficiency of the transmission speed of the different VoIP algorithms implemented may be affected by the CPU, Memory, and Network utilization. The Microsoft Resource Monitor is employed to monitor CPU, Memory and Network with similar application analysis and energy efficiency analysis.

IV. RESULTS AND DISCUSSION

A. VOIP Application Resource Utilization

Table VI shows the CPU, Memory, Disk, and Network used by the VoIP applications during transmission based on the test schedule. It is observed that ooVoo has the lowest CPU utilization, Viber for the memory utilization, and ooVoo with the lowest network utilization.

On the other hand, Viber uses a large amount of CPU, and Skype utilizes the maximum memory and network.

TABLE VI. VOIP APPLICATION RESOURCE UTILIZATION

Test Schedule	ooVoo			Skype			Viber		
	CPU	RAM	Network	CPU	RAM	Network	CPU	RAM	Network
Test1 (7-9am)	5%	26%	0	7%	28%	691	27%	5%	25
Test2 (1-3pm)	1%	28%	0	14%	28%	25	27%	1%	27
Test3 (7-9pm)	1%	32%	1	3%	35%	57	3%	1%	35
Test4 (1-3am)	7%	31%	26	2%	33%	47	32%	7%	25
	4%	29%	6.75	7%	31%	205	22%	4%	28

B. Mean Opinion Score

The first VoIP test conducted from 7:00 AM to 9:00 AM shows that the current speed of the Internet for download is 1.94 mbps and for the upload is 1.97 mbps. It is noteworthy that in the download stream test results, Skype application has the highest maximum instant jitter, maximum RFC3550 jitter, maximum packet loss, lowest bandwidth, lowest packets per seconds, and worst G.107 Mean Opinion Score. On the other hand, Viber application has the highest results for the instant jitter, RFC3550 jitter, packet loss, and lowest result in terms of bandwidth, packets/seconds and G.107 MOS in the upload stream, as shown in Table VII.

TABLE VII. VOIP TEST RESULT AT 7AM TO 9AM

Test 1	Download			Upload		
	Viber	Skype	ooVoo	Viber	Skype	ooVoo
Max Instant Jitter	53.973	76.069	45.754	244.5	167.24	42.31
Max RFC3550 Jitter	6.644	15.306	6.797	23.33	17.18	15.66
Max Packet Loss	0.033	0.467	0.033	4.97	0	0
Bandwidth (bps)	12799.7	12787	12799.7	12754.7	12800.2	12800.6
Packets/sec	49.9665	49.92	49.96	49.77	49.95	49.95
Worst G.107 MOS	4.099	4.03	4.099	3.05	4.1	4.104

Table VIII shows the result of the second VoIP test conducted between 1:00 PM to 3:00 PM; the current speed of the Internet is 1.84 mbps for the download and 2.0 mbps for the upload. In the download stream test results, ooVoo application has the highest maximum instant jitter while Viber gets the maximum RFC3550 jitter. Skype has the maximum packet loss, lowest bandwidth, lowest packets per seconds, and worst G.107 Mean Opinion Score. In the upload stream, Skype application has the highest results in relation to instant jitter, RFC3550 jitter, and lowest result in bandwidth. However, the three applications have the same results in packet loss, packets/sec, and G.107 MOS.

TABLE VIII. VOIP TEST RESULT AT 1PM TO 3PM

Test 2	Download			Upload		
	Viber	Skype	ooVoo	Viber	Skype	ooVoo
Max Instant Jitter	70.785	84.497	86.781	42.201	166.66	57.768
Max RFC3550 Jitter	15.938	11.855	13.282	12.715	18.253	16.399
Max Packet Loss	0.067	0.0801	0.3	0	0	0
Bandwidth (bps)	12799.1	12778.6	12790	12800.1	12800	12800
Packets/sec	49.96	49.88	49.92	49.95	49.95	49.95
Worst G.107MOS	4.094	3.974	4.057	4.104	4.104	4.104

The third VoIP test conducted from 7:00 PM to 9:00 PM shows that the current speed of the Internet for download is 1.97 mbps and for the upload is 1.98 mbps. It is noticed that in the download stream test results, Viber application gets the maximum instant jitter, maximum RFC3550 jitter, and lowest bandwidth. The three applications get the same results in maximum packet loss, packets/sec, and G.107 MOS. On the other hand in the upload stream, ooVoo application gets the highest results – instant jitter, RFC3550 jitter, packet loss, and lowest result in bandwidth, packets/seconds and G.107 MOS, as shown in Table IX.

TABLE IX. VOIP TEST RESULT AT 7PM TO 9PM

Test 3	Download			Upload		
	Viber	Skype	ooVoo	Viber	Skype	ooVoo
Max Instant Jitter	72.22	23.867	69.346	41.528	42.337	42.351
Max RFC3550 Jitter	10.562	4.798	10.417	14.259	14.636	15.65
Max Packet Loss	0	0	0	0.033	0.033	0.067
Bandwidth (bps)	12799.73	12799.74	12799.8	12799.1	12799.3	12798.5
Packets/sec	49.97	49.97	49.97	49.95	49.95	49.94
Worst G.107 MOS	4.104	4.104	4.104	4.099	4.099	4.094

Table X shows the result of the last VoIP test conducted between 1:00 AM to 3:00 AM with 1.95 mbps current speed of the Internet for the download and 1.98 mbps for the upload. It

is observed that in the download stream test results, ooVoo application gets the highest results in terms of instant jitter, RFC3550 jitter, packet loss, and the lowest results in bandwidth, packets per second and G.107 MOS. At the same time Skype also gets the highest result in packet loss and worst result in G.107 MOS. In the upload stream, Skype application has the highest results in relation to instant jitter; ooVoo application in RFC3550 jitter; and Viber has the lowest result in bandwidth and packet/second. However, the three applications have the same results in packet loss and G.107 MOS.

TABLE X. VOIP TEST RESULT AT 1 AM TO 3 AM

Test 4	Download			Upload		
	Viber	Skype	ooVoo	Viber	Skype	ooVoo
Max Instant Jitter	75.284	63.797	186.46	27.389	46.446	42.781
Max RFC3550 Jitter	11.656	8.683	19.211	12.49	15.452	16.034
Max Packet Loss	0.033	0.067	0.067	0	0	0
Bandwidth (bps)	12798.5	12798.6	12797.7	12800.2	12800.7	12800.3
Packets/sec	49.962	49.963	49.959	49.952	49.954	49.953
Worst G.107MOS	4.099	4.094	4.094	4.104	4.104	4.104

The summary of all test results is presented in Table XI. It can be observed that the applications that get the highest instant jitter and RFC3550 jitter is ooVoo in the download stream and Skype in the upload stream. It is notable that both results in the download and upload streams in terms of low bandwidth, low packet per second and high packet loss have direct effect in the G.107 MOS results.

TABLE XI. SUMMARY OF VOIP TEST RESULT

Summary of Average Test Results	Download			Upload		
	Viber	Skype	ooVoo	Viber	Skype	ooVoo
Max Instant Jitter	68.066	62.058	97.085	88.905	105.671	46.303
Max RFC3550 Jitter	11.2	10.161	12.427	15.699	16.380	15.936
Max Packet Loss	0.0333	0.1535	0.1000	1.2508	0.0083	0.0168
Bandwidth (bps)	12799.3	12791	12796.8	12788.5	12800.0	12799.8
Packets/sec	49.965	49.933	49.952	49.906	49.951	49.948
Worst G.107MOS	4.099	4.0505	4.0885	3.83925	4.1016	4.1015

Summary of Average Test Results	Average (Download/Upload)		
	Viber	Skype	Viber
Max Instant Jitter	78.485	83.86413	71.69388
Max RFC3550 Jitter	13.44925	13.27038	14.18125
Max Packet Loss	0.642	0.080888	0.058375
Bandwidth (bps)	12793.9	12795.51	12798.33
Packets/sec	49.93506	49.94213	49.95025
Worst G.107 MOS	3.969125	4.076125	4.095
Rank	3	2	1

V. CONCLUSION

There are several audio codecs that are utilized by VoIP application. In addition, there are vital characteristics in determining the quality of audio to be received at the other end of the communication network, which are the jitter, packet loss, bandwidth, packets per seconds, and resources. Mean Opinion Score provides a new standpoint in the assessment of VoIP test.

It was identified that the best codec is Siren14/G.222.1 utilized in both download and upload streams with its algorithm getting the highest score in G.107 Mean Opinion Score that measures the quality of sound. Also, it was found out that the higher packet loss means decreases the performance of the VoIP application than the jitter results.

REFERENCES

- [1] Mahbub Hassan, Alfandika Nayandoro, Mohammed Atiquzzaman, (2000), "Internet Telephony: Services, Technical Challenges, and Products", in Proceedings of IEEE Communications Magazine
- [2] Mahdi, O.A.; Mohammed, M.A.; Mohamed, A.J. (November 2012). "Implementing a Novel Approach an Convert Audio Compression to Text Coding via Hybrid Technique" (PDF). International Journal of Computer Science Issues. 9 (6, No. 3): 53–59.
- [3] Alexander Benjamin (2010), "Music Compression Algorithms and Why You Should Care"
- [4] History.com Staff (2009), "Morse Code and the Telegraph", A+E Networks, Retrieved from <http://www.history.com/topics/inventions/telegraph>, Access Date: March 6, 2017
- [5] Robert Pepper (2014), "The History of VoIP and Internet Telephones", Retrieved from <https://getvoip.com/blog/2014/01/27/history-of-voip-and-internet-telephones/>
- [6] Shannon, Claude Elwood, "A mathematical theory of communication." ACM SIGMOBILE Mobile Computing and Communications Review 5.1 (2001): 3-55.
- [7] Shannon, Claude Elwood, "A mathematical theory of communication.", Reprinted with corrections from The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October 1948.
- [8] D.A. Huffman+, Associate IRE, "A Method for the Construction of Minimum-Redundancy Codes", Retrieved from http://compression.ru/download/articles/huff/huffman_1952_minimum-redundancy-codes.pdf
- [9] digitalpreservation.gov (2014), "A-Law Compressed Sound Format", Retrieved from <http://www.digitalpreservation.gov/formats/fdd/fdd000038.shtml>
- [10] Wikipedia, "a-law algorithm", Retrieved from https://en.wikipedia.org/wiki/A-law_algorithm
- [11] Cisco.com (2006), "Waveform Coding Techniques", Retrieved from: <http://www.cisco.com/c/en/us/support/docs/voice/h323/8123-waveform-coding.html>
- [12] Wikipedia, "Warped linear predictive coding", Retrieved from https://en.wikipedia.org/wiki/Warped_linear_predictive_coding
- [13] Aki Härmä and Unto K. Laine, "A comparison of warped and conventional linear predictive coding", Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.1455&rep=rep1&type=pdf>
- [14] Atal, Bishnu S(2006), "The history of linear prediction." IEEE Signal Processing Magazine 23.2, 154-161
- [15] Speex Codec Manual, <https://speex.org/docs/manual/speex-manual/node9.html>
- [16] Schroeder, M., and B. S. Atal. "Code-excited linear prediction (CELP): High-quality speech at very low bit rates." Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85. Vol. 10. IEEE, 1985.
- [17] International Telecommunications Union, Standardization Sector (ITU-T), Study Group 15 (1993-1996), Recommendation G.729, March 1996.
- [18] ITU-T G.728, Retrieved from: <http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-G.728>
- [19] ITU-T G.718 Recommendation, Retrieved from: <http://www.itu.int/rec/T-REC-G.718/en>

- [20] MTT Company, "Comparative analysis of popular Internet instant messengers and the technology used", Retrieved from: <http://webcache.googleusercontent.com/search?q=cache:1UY0E1p0IpMJ:www.mttgroup.ch/upload/Comparison%2520Stealthphone,%2520Viber,%2520Skype,%2520WhatsApp.doc+&cd=1&hl=en&ct=clnk&gl=ph>
- [21] Robert M. Gray, IEEE Signal Processing Society, Distinguished Lecturer Program, Retrieved from: <http://www-ee.stanford.edu/~gray/dl.html>
- [22] SILK, Retrieved from: https://en.wikipedia.org/wiki/SILK#cite_note-IETF79-2
- [23] H. S. Malvar, "Lapped transforms for efficient transform/subband coding," IEEE Trans. Acoust., Speech, Signal Processing, ASSP-38, pp. 969-978, June 1990.
- [24] H.S. Malvar, "A modulated Complex Lapped Transform and its Application to Audio processing", Retrieved from: http://www.mirlab.org/conference_papers/International_Conference/ICASSP%201999/PDF/SCAN/IC992457.PDF
- [25] Jeff Rodman (2009), "Polycom® Siren14 - Information for Prospective Licensees", Polycom, Inc, Retrieved from: http://docs.polycom.com/global/documents/company/about_us/technology/siren14_g7221c/info_for_prospective_licensees.pdf
- [26] Business Wire (2001-03-26). "PictureTel Announces New Siren Wideband Audio Technology Licensing Program". thefreelibrary.com. Retrieved from <http://www.thefreelibrary.com/PictureTel+Announces+New+Siren+Wideband+Audio+Technology+Licensing...-a072261882>
- [27] Wikipedia.org, "Packet (Computing)", Retrieved from: [https://simple.wikipedia.org/wiki/Packet_\(computing\)](https://simple.wikipedia.org/wiki/Packet_(computing))
- [28] Wikipedia.org, "Packet Loss", Retrieved from: https://en.wikipedia.org/wiki/Packet_loss
- [29] Bolot, Jean-Chrysotome. "End-to-end packet delay and loss behavior in the Internet." ACM SIGCOMM Computer Communication Review. Vol. 23. No. 4. ACM, 1993.
- [30] Measuring network throughput, Retrieved from: https://en.wikipedia.org/wiki/Measuring_network_throughput
- [31] Jitter, Retrieved from: <https://en.wikipedia.org/wiki/Jitter>
- [32] ITU-T Rec. G.107 (March 2005), "The E-model, a computational model for use in transmission planning", Retrieved from: <https://www.itu.int/rec/T-REC-G.107>
- [33] Startrinity.com (2014), "SIP Tester Tutorial", Retrieved from: <http://startrinity.com/VoIP/SipTester/SipTesterTutorial.aspx>

AUTHORS' PROFILE



FULL NAME: Jasmin De Castro Niguidula
TITLE: Chair Computer Science /
Information Systems at Technological
Institute of the Philippines
Citations: 7 Since 2012
h-index: 2



FULL NAME: Ma. Emmie T. Delluza
University / Affiliation: Technological
Institute of the Philippines – Manila
(Graduate Studies)
TITLE: Assistant Professor at Catanduanes State
University

Performance Evolution of QoS in VoIP Using SIP for Ad hoc Routing Protocols

Ahmad Jahangir

Department of Computer Sciences,
Pir Mehr Ali Shah, Arid Agriculture University
Rawalpindi Pakistan
csahmadjahangir@gmail.com

Yawar abbas abid

Department of Computer Sciences,
COMSATS institute of Information Technology
Sahiwal, Pakistan
yawar.abid@ciitsahiwal.edu.pk

Muhammad Usman Nasir

Department of Computer Sciences,
COMSATS institute of Information Technology
Sahiwal, Pakistan
usmannasir@ciitsahiwal.edu.pk

ABSTRACT-To discuss voice quality of service of VoIP call sessions in Mobile Ad Hoc Networks (MANETs) we have to consider network bandwidth and appropriate routing protocol with all its pros and cons to get acceptable delays for call registration, initiation, retransmission and successful termination. The pros and cons of routing protocols are affected by physical distance and distribution of nodes in a network. There is a lot of research work done for proactive, reactive, or hybrid routing protocols for VoIP in MANETs. In this research work we did a comparative analysis of three on demand routing protocols named DSR, AODV and TORA over IPv4 addresses. VoIP quality of service and SIP signaling have been examined for number of active calls and End-to-End Delays for transmissions. The simulation concludes that Adhoc on demand distance vector routing protocol exhibits the highest throughput gain for Static, Uniform, Random and Random All mobility models. After that DSR shows better for Static and Uniform mobility models and TORA shows worst in all mobility models. This work helps most appropriate reactive routing protocol to be used for VoIP sessions to achieve QoS over different types of mobility models.

Keywords-VoIP; SIP; TORA; AODV; DSR.

I. INTRODUCTION

Mobile Ad-hoc networks (MANETs) is a need in today's era for 4G networks and many recent projects such as EU IST-DAIDALOS project because of their self configurable nature in dynamic distributions and faster recovery options under high mobility [3]. MANETs is mobile adhoc networks and as the name suggests, several independent, highly moving nodes, without presenting a particular infrastructure communicate and process information through wireless technologies. MANETs are connected to 4G network using internet via some gateway nodes that may exhibit multihop paths for information processing. MANETs use different routing protocols such as Ad hoc On Demand Distance Vector (AODV), Dynamic Source Routing (DSR), and Temporally Ordered Routing Algorithm (TORA). Voice over IP (VoIP) is a multimedia session to get implemented in MANETs, we have to consider

several factors such as different design mobility types, how far two parties are kept, counting hops, sizes of nodes, and communication time setup to provide quality of service (QoS). For VoIP multimedia sessions there is a constant need of signal passing. While on the other hand, in typical MANETs, limited bandwidth and delayed transmissions turn down the performance of VoIP. To overcome these issues a defined instant messaging protocol named, the Session Initiation Protocol (SIP) needs to be introduced [1]. Two main types of signaling that define VoIP QoS are: TCP based, SIP signaling and UDP based, voice Signaling. Because in MANETs nodes are independent there comes many limitations while implementing proxies/registrars based SIP services, in the network by MANETs. All SIP messages must have to pass through proxy/registrar. In Vehicular area networks VoIP QoS is described in terms of average jitter, latency between two consecutive hops and amount of data loss [5] where moving vehicle nodes make it difficult to maintain connectivity.

II. RELATED WORK

In [9], authors have made a comparative study of reactive, proactive and hybrid mobile adhoc networks routing protocols. Based upon different number of nodes OLSR, TORA and Geographic Routing Protocol (GRP) using FTP traffic is evaluated. In [10], an evaluation matrix (data loss, transmission delay, no of retransmissions) is observed on OLSR and TORA. OLSR gives good performance while the only issue was retransmissions attempts. In [2] authors propose, that standard approach of SIP is not good enough to provide good service for a user and certain alternatives must be applied on SIP architecture. And In [4] authors have studied VoIP with GSM voice codecs and SIP signaling and then compared DSR, TORA and AODV routing protocols throughput, in OPNET modeler, for different number of nodes and different mobility models such as static, uniform, random & random all. Where AODV gives best results when number of nodes gets larger and mobility model is random or random

all, where DSR produces best results for uniform mobility model and TORA is best for static model. In [6], authors have designed an ITU-T E-Model for VANETs and measure QoS when mobile vehicles move from one network to another. A solution to this problem is described in [7] by using a reserved SIP proxy modules that exhibit MPLS and DiffServ, QoS mechanisms, to decrease processing time in handover from one network to another. In [8] authors have introduced an authenticated VoIP QoS solution in VANETs. First of all, to ensure security and fast processing, first of all, a registration procedure is performed by moving vehicle to choose a road side unit (RSU) for next handover. This handover is secured by exchanging key sessions between the mobile vehicle and connected servers in the network. In [11], with different number of nodes, increasing size of nodes and high CBR traffic flow, different routing protocols are studied. It is observed that AODV out performs TORA because of no temporary loops formation and runtime routing decisions. In [12], DSR, OLSR and ZRP are compared in NS2 using CBR (UDP) traffic under different mobility models resulting in high throughput of OLSR. In [13], the authors have discussed the throughput of multimedia streaming that is a real time session for a reactive OLSR routing protocol in MANETs .

III. VOIP IN AD-HOC NETWORKS

A. MANET Architecture

As described above, MANETs are highly independent, moving nodes that configure, connect and communicate among different wireless networks themselves [3]. So three types of MANETs routing protocols are described as:

Reactive routing protocols: In this protocol, nodes discover routes based upon the request of some demand submitted. Some examples are Ad hoc On Demand Distance Vector (AODV), Dynamic Source Routing (DSR), and Temporally Ordered Routing Algorithm (TORA) and it is a need of the era to compare among routing protocols to evaluate with different metrics under different conditions.

Proactive routing protocols: In this case the nodes search and share their paths (routing tables) after some defined periodic time and as a result produce high overhead which makes them slow to process as node capacity increases or the mobility model become Random or RandomAll. Examples are optimized link state routing protocol (OLSR) and destination sequenced distance vector (DSDV).

Hybrid routing protocols: These protocols try to combine the functions of proactive and reactive routing protocols to attain maximum throughput and least delays in all mobility models. Examples are Zone Routing Protocol (ZRP).

B. SIP Architecture

A standard signaling protocol defined by Internet Engineering Task Force (IETF) in Request For Comments (RFC 2543) is generally used for VoIP sessions [14]. It is an application layer protocol that can control one or more coded media sessions at a time. Two main components are User side and server side. Where SIP server side is further divided into 1) registrar server: to keep track of mobile nodes' location. 2) location server: is a data storage used by region server to locate nodes 3) proxy server: does data transfer to the required location 4) redirect server: initiates and maintains direct connection between sender and receiver by exchanging locations. So SIP based VoIP calls must do initial registration and after that, call initiation and call termination is introduced by the above mentioned SIP servers. Because SIP is a TCP protocol all TCP delays for ensured in-order delivery of transmissions adversely affects performance of VoIP.

Let's understand the mechanism of VoIP calls in SIP. In REGISTRATION caller send messages to register itself and is stored into Registration server with some pre-requisites. After that a location IP and port number are assigned to this caller and saved on Location server (Registration Phase). After that the caller sends a session initiation message (INVITE) to Proxy server that forwards the session towards the callee. This call is redirected and maintained by the Redirect Server. After successful establishment of call the callee sends an acknowledgment message (ACK) to the caller. In case of unsuccessful session the caller can send CANCEL messages for termination. BYE messages should be used to end a successfully established call session. This call transfer procedure is supported by some voice codecs that are categorized as Low Quality Speech (LQS), Pulse Code Modulation (PCM), GSM FR Quality Speech, and IP Telephony [15]. Different routing protocols, number of nodes, nodes movement, noise, mobility models network bandwidth, congestion and calling devices voice compression can affect the characteristics of voice codecs. There are a number of SIP versions available such as proxy-based SIP, distributed SIP, service discovery SIP and peer to peer SIP with Distributed Hash Tables (DHT). Different VoIP QoS parameters that are taken under consideration are: Amount of traffic sent, traffic received, end to end delay and jitter under different node capacity, mobility models and routing protocols in MANETs.

Comparative analysis of DSR, AODV, TORA on VoIP:

The analysis of reactive routing protocols takes IEEE 802.11n with data rate of 13 Mbps as the wireless network standard to study different mobility models in MANETs. The implemented simulation tool OPNET modeler is tested with Static, Uniform, Random, and Random All, mobility models using different parameters. SIP architecture for mobility models is shown in Fig1.

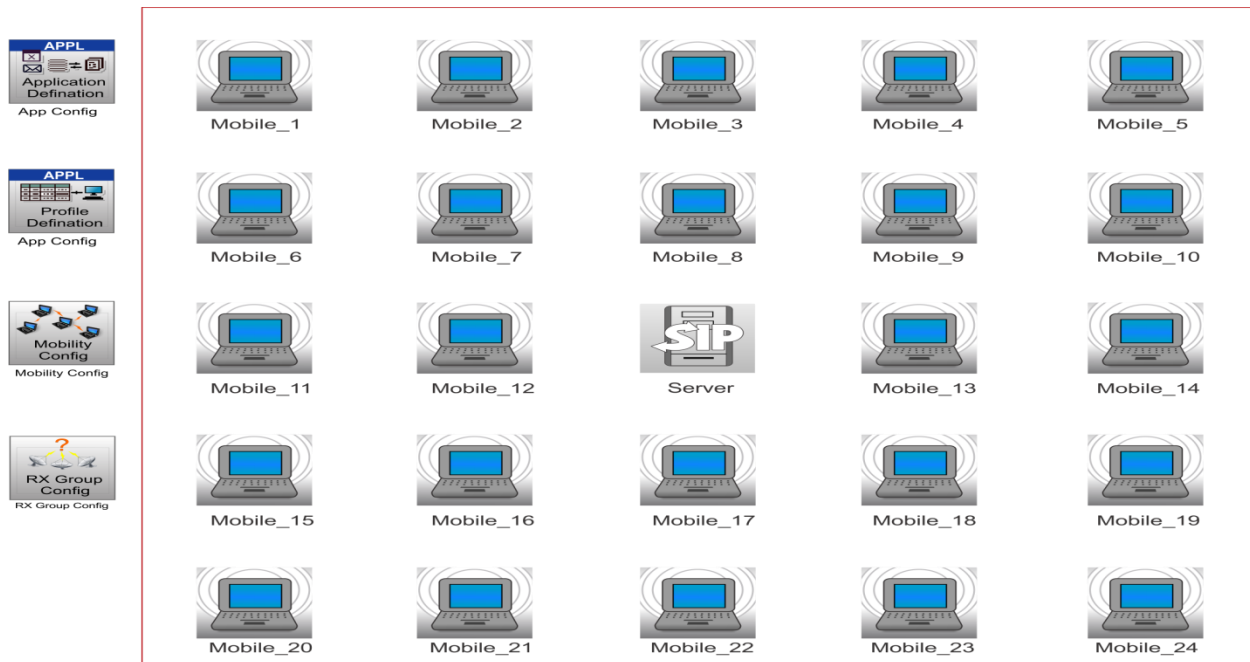


Fig.1: SIP Architecture

In static model, as comes from the name, the mobile nodes cannot relocate themselves. In uniform mobility models, all nodes including SIP server must be moving in one defined direction but can have different speed ranges. In random mobility model, SIP server is stable and other nodes can have different mobility directions. In random all mobility model, all nodes including SIP

server can have different mobility directions with different speed ranges. The simulation parameters are: 12 number of simulations, 128 simulations seed number, with 25 nodes in an area of 1km*1km, having packet size 512 bytes, over 10 minutes time duration as described in Table 1.

TABLE1: SIMULATION PARAMETERS

A. MANET			
Simulation Duration:	10 Minutes = 600 Seconds		
MANET Reactive Routing Protocols	AODV, DSR, TORA		
Number of Simulations	12	Simulation Seed Number	128
Mobility Models:	Static, Uniform, Random, and Random All		
WLAN Physical Characteristic:	802.11 n	Data Rate:	13 Mbps
Node Speed Range:	Uniform Speed between 5.57 m/s (20 km/hr) and 12.5 m/s (45 km/hr)		
Packet Size:	512 Bytes	Buffer Size:	32 Kbytes
Maximum Transmission Range between Nodes:	from 100 meters to 250 meters		
Frequency Band:	2.4 GHz	Transmissi on Power:	0.001 W
Number of nodes:	25 nodes	Area Dimension:	1 km x 1 km

Among 25 nodes, the central node uses SIP with equally distant distribution of remaining 24 nodes around it in OPNET simulator for all four types of

mobility models. The simulation results for active calls managed by SIP server is shown below in Fig.2.

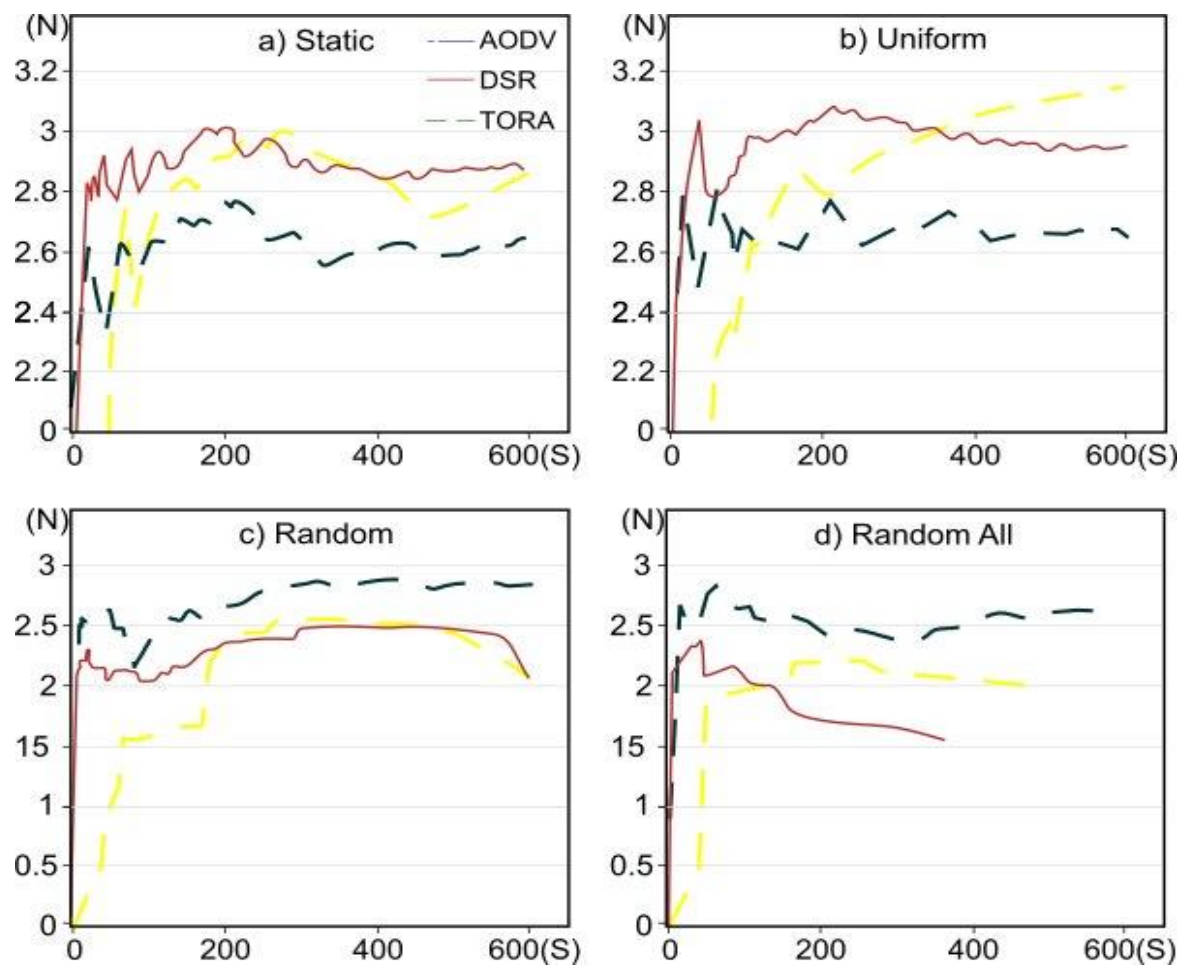


Fig2: Number of Active calls in SIP session

In fig 3. number of rejected SIP calls and comparative behavior of all three routing protocols is shown.

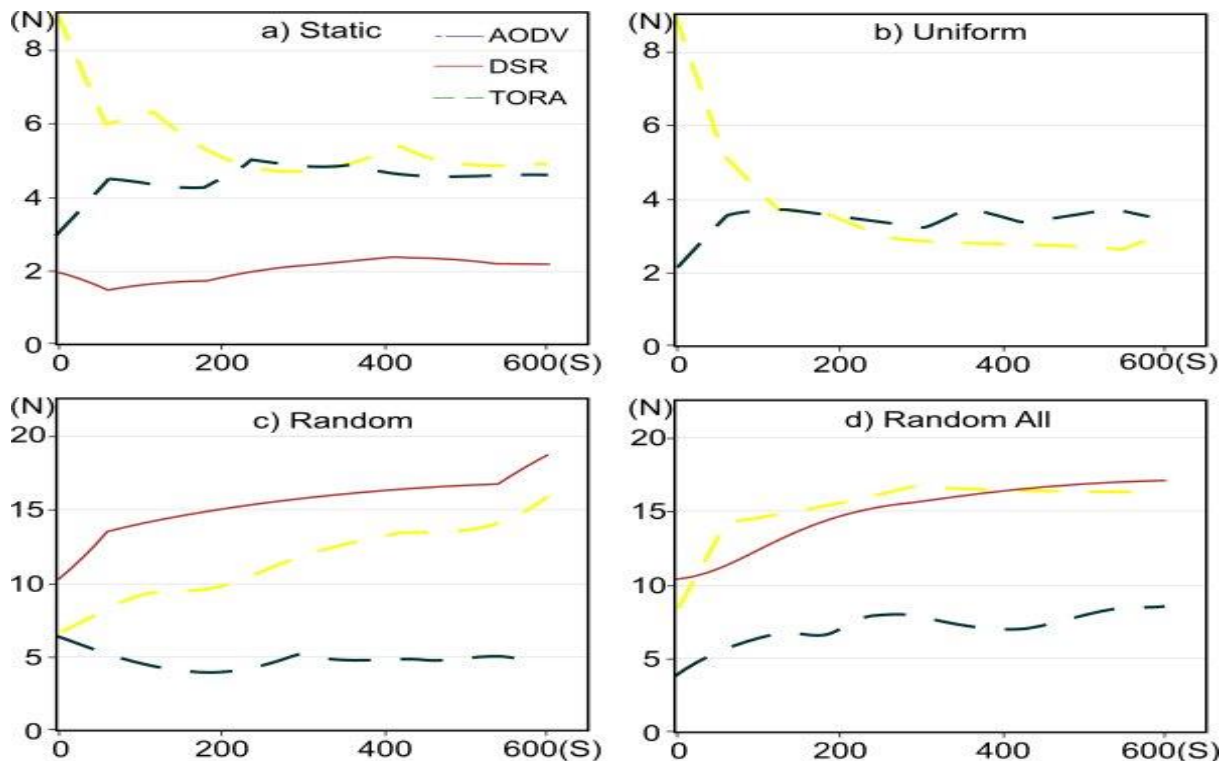


Fig.3: Number of rejected calls in SIP session

In Fig.4 average voice end to end delay in transmission is shown.

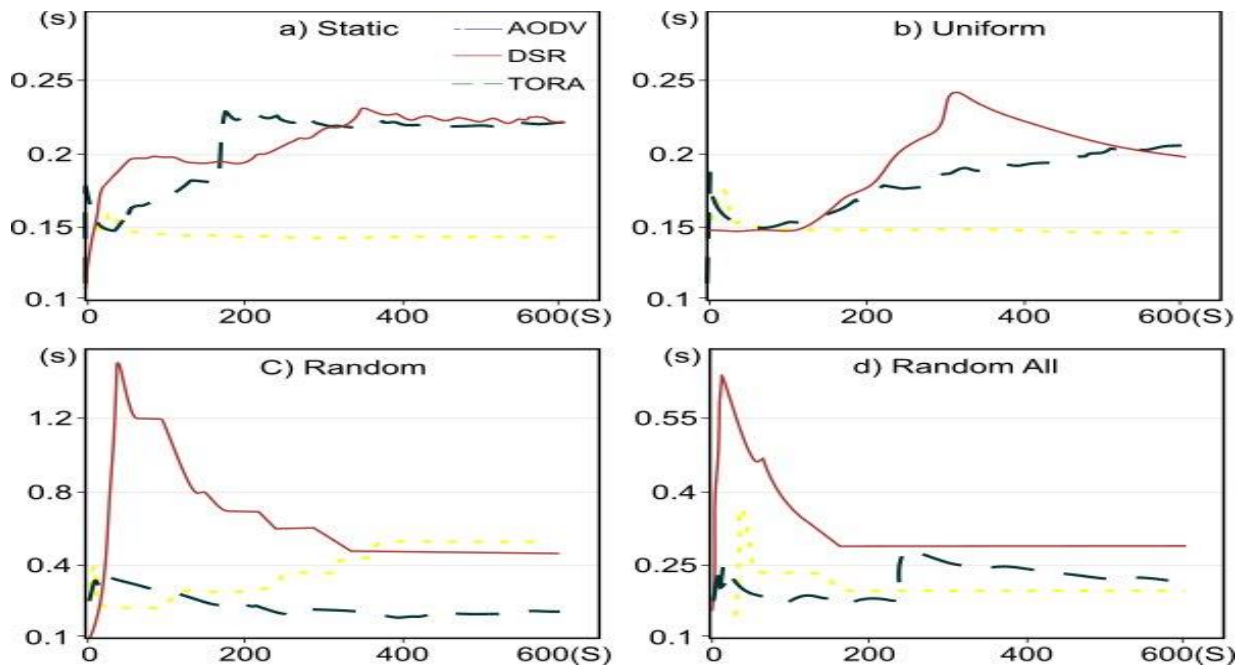


Fig.4: Average end to end delay in transmissions

IV. CONCLUSION

The above mentioned simulation parameters help us understand the voice quality of service and SIP signaling system between the caller and callee. The amount of data transferred or number of bytes is observed against simulation time. At the start of simulation all routing protocols showed bad results which make it difficult to analyze VoIP QoS before almost 120 seconds. This delay has occurred because of routing table formation and propagation in the whole MANET depending on the routing protocol being implemented. For example DSR showed longest end-to-end delays resulting in a greater amount of dropped packets and terminations of calls, which makes it unsuitable for random mobility models. In TORA, for Static and Uniform mobility models, end to end delays were fixed and for Random or RandomAll mobility models showed tolerable communication latencies. AODV showed acceptable delays for all types of mobility models including Random mobility models. For all mobility models, AODV showed a large number of successful calls per second. While on the other hand, DSR and TORA, received very small SIP requests for Random and RandomAll mobility models and for Static as well as Uniform models were somewhat satisfactory. So the results concluded that SIP server had not actively been contacted with moving nodes. So the higher average of successful voice calls gives AODV routing protocol a good protocol to be used when compared with DSR and TORA. Random and RandomAll models do not exhibit much high delays in AODV, DSR and TORA also, generally because of the less amount of traffic for VoIP sessions.

References:

- [1] S. Berger, H. Schulzrinne, S. Sidiroglou, X. Wu, Ubiquitous Computing Using SIP. In: Proc. of NOSSDAV'03, Monterey, California, USA, June 1-3, 2003.
- [2] Marcel C. Castro and Andreas J. Kasser, Optimizing SIP Service Provisioning in Internet Connected MANETs
- [3] K. Pandey and A. Swaroop, "A Comprehensive Performance Analysis of Proactive, Reactive and Hybrid MANETs Routing Protocols," IJCSI International Journal of Computer Science Issues, Vol. 8, No. 3, 2011.
- [4] Mazin Alshamrani, Haitham Cruickshank, Zhili Sun, Vahid Fami, and Basil Elmasri, Evaluation of SIP Signalling and QoS for VoIP Over MANETs Reactive Routing Protocols
- [5] P. CHOCHOL, "Qualitative factors that impact real implementing voip in private networks," ACTA ELECTROTECHNICA ET INFORMATICA, vol. 9, p. 61, 2009.
- [6] J. Fitzpatrick, S. Murphy, and J. Murphy, "An approach to transport layer handover of voip over wlan," Consumer Communications and Networking Conference, 2006.
- [7] U. Iqbal, Y. Javed, S. Rehman, and A. Khanum, "Sip-based qos management framework for ims multimedia," IJCSNS International Journal of Computer Science and Network Security, May 2010.
- [8] Bayrem TRIKI, Slim REKHIS, and Noureddine BOUDRIGA, Secure and QoS-aware SIP Handover for VoIP Communication in Vehicular Adhoc Networks
- [9] H. kaur and E. J. Singh, "Performance comparison of OLSR, GRP and TORA using OPNET," International Journal of Advanced Research in Computer Science and Software Engineering, 2 (10), Oct-2012, pp. 1-6.
- [10] P. Palta and S. Goyal, "Comparison of OLSR and TORA routing protocols using OPNET Modeler" in International Journal of Engineering Research and technology Vol. 1 Issue 5, July – 2012.
- [11] A. Shrestha and F. Tekiner, "On MANET Routing Protocols for Mobility and Scalability," In International Conference on Parallel and Distributed Computing, Applications and Technologies, p.p. 451-456, November 2009. IEEE Computer Society.
- [12] S. Gandhi, N. Chaubey, P. Shah, and M. Sadhwani, "Performance evaluation of DSR, OLSR and ZRP protocols in MANETs," Computer Communication and Informatics (ICCCI), 2012 International Conference on, pp. 1-5, 2012.
- [13] K. Jain, K. Somasundaram, B. Wang, J. Baras, and A. RoyChowdhury, "Study of OLSR for Real-time Media Streaming over 802.11 Wireless Network in Software Emulation Environment," 2010.
- [14] A. B. Johnston, "SIP Understanding the Session Initiation Protocol," third Edition, Artech House Publishers, 2009.
- [15] S. Ganguly and S. Bhatnagar, "VoIP: Wireless, P2P and New Enterprise Voice Over IP," Chichester, England: Wiley, 2008. Print.

Binary Search Optimization

Implementation and Amortized Analysis for Splitting the Binary Tree

Dhaval Kadia

Student, Department of Computer Science & Engineering
Faculty of Technology and Engineering, The Maharaja Sayajirao
University of Baroda
Vadodara, India
dhavalkadiamsu@gmail.com

Abstract—In this paper, I describe how binary tree can be manipulated well for accessing elements. Binary tree is famous data structure due to its geometry. Though having these well-organized properties, it has some demerits in some cases. If number of elements increases in binary tree, then those many elements placed in bottommost level, they have high time complexity to be accessed. One solution is to split main binary tree in multiple small binary trees and this is the way of making buckets for each of those. But approaching this solution, one question arises: How dense small (sub) binary trees should be extracted from main binary tree? An answer should explain even after splitting main binary tree, searching time complexity over should be decreased or should not be increased. These changes also should prevent sub tree to behave like a dense tree. Density of nascent sub tree should be in some range that will decrease searching time complexity. This paper describes proper relation among density of main binary tree i.e. depth, number of splits and density of sub binary tree i.e. depth. Explanation regarding code that I had implemented in Language C is mentioned in paper. Implementation is having manipulation of sub binary trees those are replica of one massive binary tree. Manipulation consists of balancing binary tree, merging light weighed binary tree and further splitting of dense sub tree.

Keywords—Data Structure; Binary Tree; Binary Search Tree; Tree Balancing; AVL Tree; Amortized Analysis; Optimization

I. INTRODUCTION

The motive of modifying binary tree is to get well-structured binary tree, for that main binary tree is split into proper ratio of depths of main binary tree and sub binary tree. Problem behind modification is groups of large number elements placed in and above the bottommost layer. Reason is those many elements have higher time complexity of searching than elements present in upper levels. They increase the amortized complexity of searching. If solution is made for bottommost elements, then amortized time complexity will be reduced. Two cases are shown how proper changes can be introduced without affecting original retrieving performance of main binary tree.

Secondly, methods are shown how individual sub binary tree can be organized. Sub trees are put in buckets having dynamic range of minimum and maximum value i.e. permitted to be inserted. On insertion, elements are inserted into its relative bucket having allowance for inserting element

value. As insertion goes on, skewness of sub tree is checked and balanced accordingly. It is also checked whether sub binary tree is becoming much denser or not. It is kept in mind that binary tree should not be skewed. On based of that, methods are implemented. Implementation very similar to AVL tree. Trees can be balanced recursively with proper control.

II. DEFINITIONS AND NOTATIONS

Let us assume depth of main binary tree T_p is t . So, there are t number of levels in T_p . After split, sub binary tree T_c has d number of levels. Number of sub trees extracted from main tree and are connected consecutively is l . For calculation, $l = 2^k$ is assumed.^a

Data structure used for storing binary tree is Linked List (obviously). Sub binary tree is assumed to be linked with next sub tree.

Defining Amortized Complexity as A.C.

III. ANALYSIS

Analysis is done on bases of condition of time complexity of acquiring sub tree in *constant time*.

In general, calculation of amortized complexity of searching in binary tree T having depth of n is as below:

A. Initialization

For searching element from whole set, that particular T_c is searched. For finding which T_c among set of sub binary tree, it is assumed that time taken is constant.

An assumption is acquired on bases of the static range of set division of T_p that says, each T_c is allotted particular and relative range of value to be allowed for insertion.

Once getting to that particular T_p , separate binary search can be performed on it. Benefit of this procedure is that less depth of tree is being searched. For binary search in binary trees, different nodes at different depth will have different access time.

^a. In T_p and T_c , p represent parent and c represent child. These are main or parent tree and sub and child tree.

It is shown number of elements at particular level and their searching complexity in Table 1. Defining Level which is Depth, Number of nodes at same level i as $N(i)$, Search-Time of node at Level i which is i itself. Total-Search-Time of all nodes at same level i is $TST(i)$.

TABLE I. SEARCHING TIME COMPLEXITY STATISTICS

Level	$N(i)$	Search-Time	TST(i)
0	2^0	1	$2^0 * 1$
1	2^1	2	$2^1 * 2$
2	2^2	3	$2^2 * 3$
3	2^3	4	$2^3 * 4$
.	.	.	.
n	2^n	n+1	$2^n * (n+1)$

For tree having maximum level of n (level starts from 0).

$$A.C. = \frac{\text{Sum of TST at every levels}}{\text{Total number of nodes per binary tree}}$$

$$= \frac{\sum_{i=0}^n 2^{i*} (i+1)}{2^n}$$

B. Derivation

$$A.C. \text{ for main tree having } t \text{ levels} = A.C._p = \frac{\sum_{i=0}^t 2^{i*} (i+1)}{2^t}$$

Now after splitting T_p of t number of levels into l number of identical subtrees T_c having d number of levels,

So, $2^d * l = 2^t$, suppose, $l = 2^k$

$$2^d * 2^k = 2^t$$

$$d + k = t$$

Where k represents 2^k number of buckets each having binary tree.

Now, A.C. searching element over whole set having l number of buckets

$$A.C._b = \frac{l * \text{Sum of TST at every levels of sub binary tree}}{\text{Total number of nodes in whole set}}$$

$$= \frac{l * \sum_{i=0}^d 2^{i*} (i+1)}{2^t}$$

Where 2^l is total number of nodes in all bucket which is same as number of elements in main binary tree because all buckets are extracted from main binary tree.

Next step is to find proper proportion of total number of elements contained by T_p and T_c . Comparing both cases,

$$\text{Ratio of main case over split case} = \frac{A.C._p}{A.C._b} \text{ (should } > 1)$$

$$= \frac{\sum_{i=0}^t 2^{i*} (i+1)}{2^t} \div \frac{l * \sum_{i=0}^d 2^{i*} (i+1)}{2^t}$$

$$= \frac{\sum_{i=0}^t 2^{i*} (i+1)}{l * \sum_{i=0}^d 2^{i*} (i+1)}$$

For $d + k = t$, taking $t = m * k$

$$d + k = m * k$$

$$d = (m - 1) * k$$

It represents for particular number of m with respect to n , what t should be. Later, it will lead us to result – for any value of k , how many number of elements should be allowed in each bucket i.e. T_c .

For different value of m , number of nodes in T_p and T_c will be decided and those values of m , Optimization Ratio will decide whether after splitting, searching becomes fast or not. That means, if Ratio is greater than 1, then $A.C._b$ is less than $A.C._p$. It means bucket phase decreases average searching time. Ratio is plotted versus value m . Ratio is greater than 1 so, we can say that average searching time reduced than previous case.

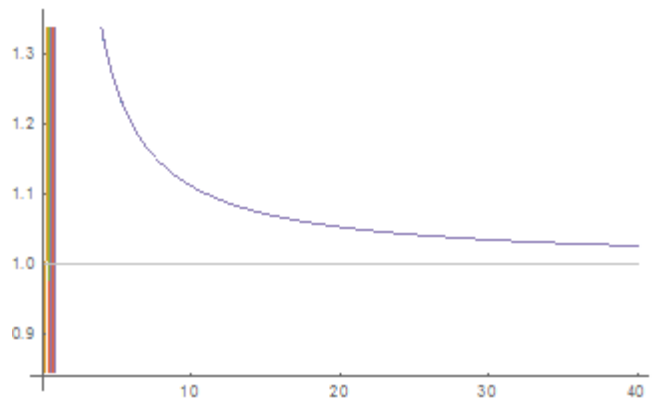


Fig. 1. Ratio versus m .^b

^b. Values only of m greater than or equal to 3 ($m \geq 3$) are meaningful.

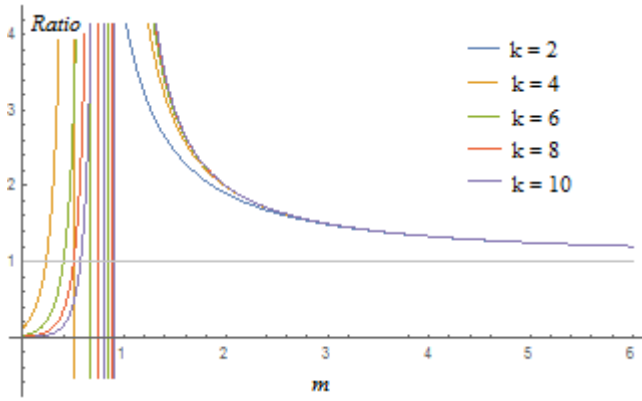


Fig. 2. Ratio versus m .^c

C. Result

Number of nodes in T_p and T_c are 2^t and 2^d respectively. As before, $t = m * k$ and $d = (m - 1) * k$. Number of bucket extracted from main tree is $l = 2^k$.

For $m = 3$, Ratio is 1.5. It explains that searching become 50% faster. It is shown in table 2.

TABLE II. RELATIVE NUMBER OF ELEMNTS IN MAIN AND SUB BINARY TREE WITH $M = 3$

k	Number of Buckets $l = 2^k$	Number of nodes in T_p 2^t	Number of nodes in T_c 2^d
2	4	2^6	2^4
4	16	2^{12}	2^8
6	64	2^{18}	2^{12}
8	256	2^{24}	2^{16}
10	1024	2^{30}	2^{20}

For $m = 4$, Ratio is 1.3. It explains that searching become 30% faster. It is shown in table 3.

TABLE III. RELATIVE NUMBER OF ELEMNTS IN MAIN AND SUB BINARY TREE WITH $M = 4$

k	Number of Buckets $l = 2^k$	Number of nodes in $T_p = 2^t$	Number of nodes in each $T_c = 2^d$
2	4	2^8	2^6
4	16	2^{16}	2^{12}
6	64	2^{24}	2^{18}
8	256	2^{32}	2^{24}
10	1024	2^{40}	2^{30}

Thus, if this proportion is maintained then average searching time will be reduced.

^c Values only of m greater than or equal to 3 ($m \geq 3$) are meaningful.

IV. PROGRAM IMPLEMENTATION

Current work is programmed in Language C using Linked List. In which main binary tree is split into buckets having sub binary tree. It is taken care if in any bucket if number of nodes which are being inserted increase beyond previously define limit, then that tree is split again and vice versa. If tree is unbalanced, it is balanced using left rotation or right rotation. At each insertion of element, balance of tree is checked. For less checking of it, it can be performed after certain intervals.

Binary tree is shown in figure 3. L is the rightmost node at bottommost level in left part of tree. L_up is parent node of L . R is the leftmost node at bottommost level in right part of tree. R_up is parent node of R .

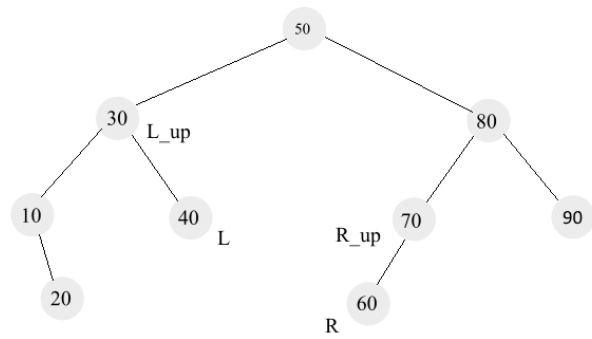


Fig. 3. Binary tree

While balancing particular binary tree, for rotating right or left, topmost node of tree is passed in parameter of function $shift_L()$ or $shift_R()$.

```

void shift_R(struct binary * target){
    get_R(target,target,target->right);
    get_L(target,target,target->left);

    if(get_L(target,target,target->left)==0)
    {
        tmpn=target->n;
        target->n=target->left->n;
        target->left=target->left->left;
        L->n=tmpn;
        L->left=NULL;   L->right=NULL;
        R->left=L;
    }
    else
    {
        tmpn=target->n;
        target->n=L->n;
        L_up->right=L->left;
        L->n=tmpn;
        L->left=NULL;   L->right=NULL;
        R->left=L;
    }
}

void shift_L(struct binary * target){
    get_R(target,target,target->right);
    get_L(target,target,target->left);

    if(get_R(target,target,target->right)==0)
    {
        tmpn=target->n;
        target->n=target->right->n;
        target->right=target->right->right;
        R->n=tmpn;
        R->left=NULL;   R->right=NULL;
        L->right=R;
    }
    else
    {
        tmpn=target->n;
        target->n=R->n;
        R_up->left=R->right;
        R->n=tmpn;
        R->left=NULL;   R->right=NULL;
        L->right=R;
    }
}

```

Fig. 4. Function Set 1 (Linked List using Language C)

```

int get_L(struct binary * home, struct binary *pre, struct binary *cur)
{
    if(cur -> right)
        get_L(home, cur, cur -> right);
    else
    {
        L = cur;
        L_up = pre;
    }

    if(home == L_up)
        return 0;
    else
        return 1;
}

int get_R(struct binary * home, struct binary *pre, struct binary *cur)
{
    if(cur -> left)
        get_R(home, cur, cur->left);
    else
    {
        R = cur;
        R_up = pre;
    }

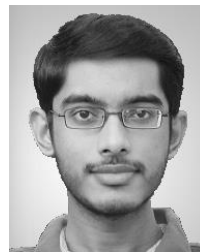
    if(home == R_up)
        return 0;
    else
        return 1;
}

```

Fig. 5. Function Set 2 (Linked List using Language C)

REFERENCES

- [1] DANIEL DOMINIC SLEATOR AND ROBERT ENDRE TARJAN
naming systems. Self-Adjusting Binary Search Trees
<http://dl.acm.org/citation.cfm?id=3835>.



Dhaval Kadia was born in Vadodara City in 1995. He is doing bachelor of engineering in computer science & engineering at Faculty of Technology and Engineering, The Maharaja Sayajirao University of Baroda, Gujarat, India.

He was Research Intern at Defence Research and Development Organisation (Robotics Research Center) - Ministry of Defence (INDIA). His research interest includes designing and analysis of algorithms, applications of image processing, computer vision, robotics, artificial intelligence, mathematical modelling, simulations, computer graphics, mixed reality & development of autonomous (self-driving) vehicles.

Key Selective Secure Privacy Framework for User Management and to Improvise Performance Parameters

Sandeep Kapur
Research Scholar
Guru Kashi University
Talwandi Saboo, India
sandeep.kapur82@gmail.com

Dinesh Kumar
Associate Professor
Guru Kashi University
Talwandi Saboo, India
kdinesh.gku@gmail.com

Abstract—Cloud Computing is one of the most practiced storage environment. Researchers and developers around the world are using the cloud architecture. With the increase in the demand of the cloud architecture, it is also required to put emphasis on the nature of load balancing architecture and speedup utilities. A well managed network provides better efficiency and more accurate allocation of the jobs. The proposed framework provides a hybrid structure of security and speedup utility. The proposed framework also concentrates over the user management and a way to response the user more quickly than the traditional approach. The proposed framework also aims to attain maximum CPU utilization with limited resources.

Keywords-Cloud Management; Resource Utilization; Security Framework; Speedup Utility

I. INTRODUCTION

Cloud [1] [2] [3] architecture encompasses three levels of services namely [6] [7] [8] [9] [10] [11],

- i. IAAS (Infrastructure as a Service)
- ii. PAAS (Platform as a Service)
- iii. SAAS (Software as a Service)

The "service" word refers to supply and fulfill the user demand. The proposed framework considers PAAS and SAAS and utilizes the services provided by IAAS.

Scheduling over a server refers to completing a set of jobs provided to the server. In order to complete a job sequence, the following terms should be referred.

A. Make span

It is the total time which is required to complete total listed jobs. Less make span refers to great scheduling performance.

B. CPU Utilization

The CPU utilization is the ease of accessing the processing unit wisely. High CPU Utilization refers to good scheduling results.

C. Balanced Load

A well balanced load prevents the system from any crash or from any extra consumption of energy.

D. Speed Up

Speed Up can be attained when jobs are arranged in appropriate manner or order. It refers to how quickly a job is scheduled and managed.

II. LITERATURE SURVEY

This section describes the survey on different author's research according to the used algorithm and methods in cloud computing. In this section drawbacks of existing work and their outputs are given in the tabular form.

W. E. Walshet.al [17] utilized the feedback algorithm for the management of virtualized machines. In this VM machines are grouped together into shared pool then as per SLO agreement, the VM allocation will takes place. Resource allocation is one of the fundamental technologies of cloud-computing domain, which utilizes the computing resources like bandwidth, energy delay and so on in the network to facilitate the execution of cumbersome tasks that require large-scale computation.

Jiayin Liet.al [18] proposed a general two-layer architecture that uses utility functions, adopted in the context of dynamic and autonomous resource allocation, which consists of local agents and global arbiter. Due to the vast population, natural resources are already on the verge of the extinction. Due to this, IT companies are moving in the direction of Green Computing as it leads towards decrease of the expenses, and decreases the utilization of the natural resources.

Shi J.Yet.al [19] proposed an adaptive resource allocation algorithm for the cloud system with preempt able tasks in which algorithms adjust the resource allocation adaptively based on the updated of the actual task executions. Resource allocation is one of the challenges of cloud computing since end-users could easily access resources from anyplace and at anytime. The resources present in a cloud could not be demanded straightly but it could be opened using SOAP/Restful web APIs which map requests for storage or computations purposes are plotted to virtualized ICT resources (which are servers, blob storage, elastic IP, and so on.

Goudarzi H et.al [20] reported a study of cloud HPC resource planning. In it author propose quantitative application dependent instrumentation method to investigate multiple important dimensions of a program's scalability. Cloud computing could be named as a new paradigm of computing services for dynamic provisioning which is upheld by state-of-the-art numerous datacenters that in general work using Virtual-Machine technologies for alliance as well as for environment isolation purposes.

A. A. Patel et.al [21] presented the resource allocation problem in which it is solved using SLA violations. . In it the upper bound of total profit is provided with the help of force-directed resource assignment (FRA) heuristic algorithm, in which initial solution is based on provided solution for profit upper bound problem. The term server consolidation is mainly used to maximize the usage of available resources. Presently, the main method of server consolidation is virtual migration. Traditional approaches usually determine a unique VMs placement before deploying a service in the production environment.

B.Weiet.al [22] implemented a framework for live migration and dynamic reallocation of VMs according to current utilization, while ensuring reliable QoS and minimizes power consumption and delay using two algorithms, Modified Best fit decreasing algorithm and genetic algorithm. The idea of Virtual Machines (VMs) is connected to diminish the energy utilization as it essentially decreases the rate of idle power in the general base.

D.Jayasingheet.al [23] proposed a dynamic threshold based approach for CPU utilization for host at data center. This consolidation will work on dynamic and unpredictable workload avoiding unnecessary power consumption.

K.Yeet.al [24] proposed a method, which is Dynamic Round-Robin method, to deploy the virtual machines for power saving purpose. Dynamic Round-Robin is an extension from original Round-Robin, in this First-Fit is used as baseline algorithm. In this method VMs are assigned to servers uniformly.

III. PROBLEM FORMULATION

Cloud computing suffers from a lot of problems. A well managed job sequence is one of the most complex problems of cloud computing. Consider a cloud model $C(U, S, sS, L)$ where U represents the total number of users in the network, S represents the servers in the framework, sS represents sub servers in the network and L the location of S , sS and U . The first problem is to assign user a sub server as per user's reach to that extra coverage cost can be ignored and high CPU utilization is attained. The second problem of this research work is to attain security if the user wants to keep any document over the cloud network. The third problem is to attain maximum scheduled jobs arranged according to their priority.

IV. PROPOSED METHODOLOGY

The above flow diagram represents the entire frame work. There are mainly four sections in the proposed architecture. The first one is the complexity finder based on type of the document which has been uploaded. Second section is the selection of the encryption algorithm. There is also a scheduling section which is a major part of the proposed framework. The final section is the storage section where the entire data is stored on the cloud. The entire framework is termed as Key Selective Secure Privacy (KSSP) framework. There are certain processes in the proposed framework. The framework looks for the User and Job complexity both. The proposed framework starts from the complexity finder block which is at the top of the entire architecture. The complexity finder is one of the key issues of the proposed KSSP architecture as it saves a lot of time when the data is sent for the encryption. Choosing a wrong encryption algorithm may lead to a huge time gap and loss in data identity. The complexity finder is applied to different type of multimedia files which is an attached block to the data type. All the attached data types have been considered. The algorithm is assigned according to the complexity of the document. There are four algorithms which have been considered for the encryption namely RSA, BLOWFISH, AES and NTRU. The proposed framework (KSSP) has wisely chosen these four algorithms as per encryption nature of the algorithm and the types of documents which are getting used in KSSP architecture. The encrypted bits are followed by the optimization block which is single handled by the Genetic Algorithm. Genetic Algorithm is a part of Natural Computing and can adjust according to the load given to the system. Hence, Genetic Algorithm can handle large number of bits provided to it. The population size for the genetic algorithm becomes equal to the total number of bits in the encrypted bits. There is cloud storage above the entire architecture

and the encrypted bits are stored at the top of the cloud storage. It has been always kept in mind that no single bit space is underutilized. The proposed architecture also actively participates in user management and it finds sub servers near to the user in order to prevent the time and costing. A sub server which is near to the user will definitely consume less cost and time as compare to the sub server which is

far. The proposed architecture has also a Job Management technique associated with it, it divided the entire framework in two sections namely high job complexity and low job complexity as per the demanding nature of the jobs. It has been always taken care of that any sub server does not sit idle or any other server is not over utilized.

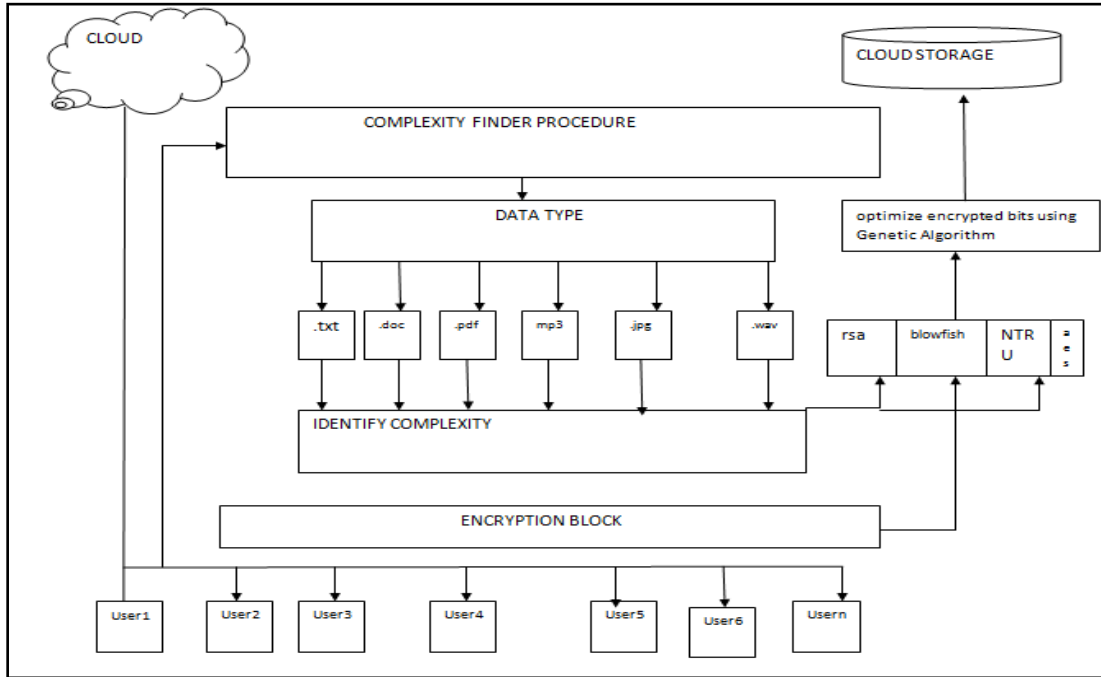


Figure 1. Key selective secure privacy framework

The proposed solution is divided into the following sections.

A. Location Based User Management

Cloud Server and user base management is one of the most complicated tasks. Maintaining the user location and the demands of the users requires a lot of attention. Let us suppose the following user model [12, 13].

$U (LU, SL, SSL)$ where (LU, SL, SSL) is a subset of L . LU is the location of the user which contains the x and y co ordinate and it can be represented as $LU(x,y)$. In the similar fashion, SL can be represented as $SL(x,y)$ and SSL as $SSL(x,y)$. Figure 1 represents the user and server model taken in the proposed methodology.

Figure 2 contains server, sub servers and users. If $U4, U5$ is considered, the nearest server to them is $ss1$ in the similar manner $ss2$ is more closed to $U3$ and $ss3$ is closer to U and $U2$. If the services to $U2$ are provided by $ss3$ then it would cost the network less as compared to the cost if the services are provided by $ss1$ or $ss2$.

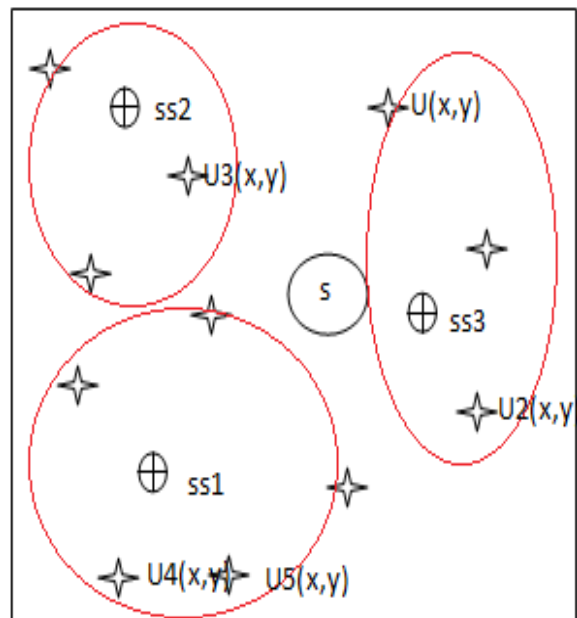


Figure 2. User / Server locations and their coverage range

The following algorithm illustrates the finding of suitable sub server as per the range in order to manage the load of the network and to minimize the cost.

Algorithm 1: Find closest Sub Server

```
Function SR = Find_closest( UL , SSL )
    for 1 = 1: SSL.count
        for j = 1: UL.count
            if (√( [(Xai - Xaaj)] ^2 + [(Xai - Xaaj)] ^2))
                ≤ Cov - limit
```

where Xa is subserver x co ordinate

Xaa is user x co ordinate.

Ya is subserver y coordinate Yaa

is user y coordinate

$-y_1)] ^2) < d\lambda$

Add sub servers to cov_limit

Find Cov_limit as

$$d\lambda = \frac{\text{width} * 35}{100}$$

There would be x co ordinates of the users and there would be y co ordinates of the users. In similar fashion there would be x coordinate of Sub servers and y co ordinate of Sub server. The following sort of table would be obtained after the implementation of the Algorithm 1.

TABLE 1. SAMPLE OF COVERAGE SET

Sub Server ID	COVERAGE ELEMENTS		
1	4	5	
2	3	10	12
3	0	2	

The proposed algorithm also includes a cache which after the completion of a job from a sub server, stores the record of sub server and the user. This helps in reduction of the computation cost. The computation cost includes the cost of search of a sub server and allocation of the job to the sub server. If the behavior of the user remains the same, this cache memory can be very helpful as the main server will not have to do the search again and again for the same type of jobs.

There are two constrains with the first part of the proposed algorithm.

- i. What if the sub server which is in range of the user is not compatible with the user requirement?
- ii. How any other sub server is going to be allocated if the sub server which is in the coverage set of the user is not compatible with the user requirements?

In order to solve above two problems the following solutions have been proposed.

Algorithm 2: Compatibility Testing [15]

```
Function flag_true= Compatibility_Test( Ur, Sr)
```

// Ur is the user requirement

// Sr is the sub server resources

```
for i=1:User.Count
```

```
If ( Sr> Ur(i))
```

```
flag_true(i)=1;
```

```
Else
```

```
flag_true(i)=0;
```

```
End if
```

```
End Function.
```

The steps of algorithm 2 explain the compatibility testing methodology. The compatibility testing refers to checking the resources whether they will be able to full fill the demands of the user's [14] requirement or not. This algorithm is a preliminary algorithm which must be applied once the coverage range of any sub server is identified.

The second solution comes when any sub server which is in range of user but not able to full fill the requirement of the user.

Algorithm 3: Find Sub server out of Coverage

```
p_ssno=find_SSOC(User_Req, SSF)
```

// SSF is the sub server //feature list , User_Req is the //list of the requirements //made by user

//p_ssno is possible subservers

```
mycount=1;
```

```
for i=1:sub_server.count
```

```
if (SSF(i) > User_req)
```

```
p_ssno(mycount)=sub_server_id(i);
```

```
mycount=mycount+1;
```

```
End
```

Algorithm 3 depicts which are the possible sub servers which can be used in order to complete the user requirement. Here, it is quite possible that the

algorithm returns more than sub server. In such a case the sub server which will provide the solution with least cost will be assigned as the working server.

B. Security Concerns Of User Data[16]

The security concerns refer to encrypt the document associated with any user and the user wants to keep the document on the cloud. Algorithm has already been proposed in predecessor research work of the proposed research work [16].

The key aspects of the security concerns proposed in predecessor research work were

- i. Selection of the encryption algorithm based on document (text, audio, video, image) complexity of the document determined by the complexity finder [16].
- ii. Optimization of the encrypted section using Genetic Algorithm in order to prevent the data space of storage at cloud.

One more security concern has been taken into contrast that is DDOS (Distributed Denial of Services) attack. The proposed architecture prevents the network from DDOS attack in the following manner.

The proposed algorithm has taken DDOS as a very serious threat for the cloud users and has made an attempt to prevent the datacenter from DDOS using Artificial Bee Colony Algorithm (ABC).

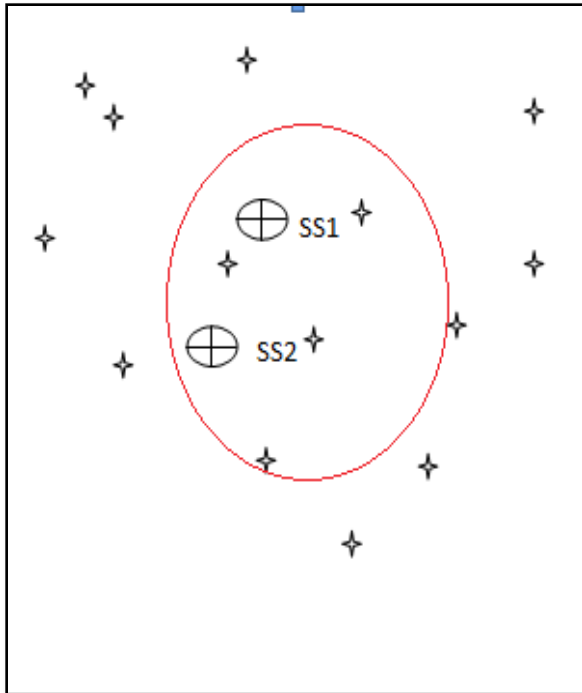


Figure 3. The situation where scheduling is required

The third section refers to scheduling the jobs on the right sub server as per the requirement. Now the question arises that when we require scheduling. When there are more than one sub servers in range of a user then scheduling is required to distribute the load on each sub server. For this purpose a high low priority scheduling algorithm has been proposed in which each and every server will get a high and low priority jobs in similar manner in order to avoid any kind of overload at any sub server. This not only saves the sub servers from getting overloaded but also prevents a lot of energy. A low priority job is basically a job which is low demanding in nature and a high priority job is a job which has high requirement demand. Consider the following Figure 2. Here there are two sub servers 1 and 2 available for the users in the same area. In such a situation the scheduling would be required.

V. SIMULATION ENVIRONMENT

In order to match the process environment which includes IAAS, PAAS and SAAS CLOUD SIM 8.02 has been utilized which is an open source cloud environment platform.

TABLE 2. SIMULATION ENVIRONMENT

SIMULATION TOOL	CLOUD SIM
Programming Language	Java
Ram Required	1 GB Minimum
HDD Requirement	50 GB
Processor	1.2 Ghz and above

In addition to the above the following key lengths has been utilized.

TABLE 3. UTILIZATION OF KEY LENGTH

ALGORITHM	KEY LENGTH
RSA	64 , 128 2-3 Divisions
BLOW FISH	128 2-3 Divisions
NTRU	128 , 256,512 3-6 Divisions
AES	256 3 Divisions

A. Performance Evaluation

1) Cracking Year

Security is one of major challenges in the modern world and another factor is how the security is attained. Encryption of data has become a very common trend in the cloud industry and cracking year is one of most suitable way to find out the strength of the encryption algorithm. The formula for the cracking year is as follows

$$IKS(y) = 56 + (y - DESTrust) * \left(\frac{12}{TechProgress} + \frac{BudgetDepend}{Budget} \right) - \log_2(SymCplrPerf)$$

Where SymCplrPerf is the ratio to number of cycles taken for the encryption to the number of cycles of cipher encryption

2) Encryption Time

It is the total time which is required to encrypt the data. The proposed algorithm encrypts a total of 7 types of file format including mp3, mp4, jpg, jpeg, txt, doc, pdf. The following score has been recorded.

TABLE 4. ENCRYPTION TIME W.R.T ALGORITHMS USED

ALGORITHM NAME	ENCRYPTION TIME IN SECONDS
NTRU	.65
RSA	.76
BLOWFISH	.74
AES	.75

3) Decryption Time

Decryption Time is the total time required to decrypt a file. Same as that of encryption, total 7 file formats have been considered. The following table illustrates the average decryption time for the algorithms used against all the file formats.

TABLE 5. DECRYPTION TIME W.R.T ALGORITHMS USED

ALGORITHM NAME	DECRYPTION TIME IN SECONDS
NTRU	.90
RSA	1
BLOWFISH	1.2
AES	.98

4) Back Up Compliance Rate

It is the rate through which early system recoveries can be attained. The proposed algorithm provides the servers a great edge to relax and hence there are always ready to provide backup to other system. In such a manner, system only consumes that time which the system requires in order to wake up. Following table has been observed at the time of the simulation.

TABLE 6. BACK UP COMPLIANCE RATE W.R.T ALGORITHMS USED

ALGORITHM NAME	BACK UP COMPLIANCE RATE
NTRU	.90
RSA	.75
BLOWFISH	.89

AES	.87
-----	-----

5) Delay

Delay is the time difference between one job to another with the addition to all the time which has been consumed already. Here with the proposed work, the delay would be least as the job scheduling has been done in such a manner that all the jobs are scheduled on time. The following table has been observed

TABLE 7. DELAY COMPLIANCE RATE W.R.T ALGORITHMS USED

ALGORITHM NAME	DELAY IN SECONDS
NTRU	1
RSA	2
BLOWFISH	1
AES	1.5

6) SLA Violation

SLA stands for "Service Level Agreement" [4, 5]. It is a kind of promise which the service provider supplies to the user. SLA can be termed to violate due to any reason, as for example, if a job has to be scheduled and it is not scheduled, then it refers to a SLA Violation. The proposed architecture only reflects a 15 % violation rate, even if the job flow is high.

7) Waiting Time

Waiting Time is the time duration which a file has to suffer if the scheduling processor is busy in performing some other task. As the proposed algorithm has managed efficiently scheduling of all the jobs, the waiting time is very less for the entire encryption or decryption algorithm.

TABLE 8. WAITING TIME W.R.T ALGORITHMS USED

ALGORITHM NAME	WAITING TIME IN SECONDS
NTRU	.10
RSA	.21
BLOWFISH	.15
AES	.25

8) Up-Time

The uptime is the total time for which the computer is still working. Here in the proposed architecture, the uptime will remain low as the work finishes in very timely fashion as the encryption of the data is done on the basis of the encryption selection algorithm. In addition to that, the jobs are also divided into low and high category and hence the system remains in rest mode as the work gets finished very early. The following table is designed to show the up time of different encryption algorithms.

TABLE 9. UP- TIME W.R.T ALGORITHMS USED

ALGORITHM	UPTIME IN SECONDS
AES	30
RSA	29
BLOWFISH	27
NTRU	23

9) Down Time

Down Time is the time for which the system remains off. Here the downtime will be more as compared to the architecture system as the system gets a lot of time to relax. The time of relaxation is increased due to the finishing nature of the scheduling algorithm and the time saving nature of the complexity finder.

TABLE 10. DOWN- TIME W.R.T ALGORITHMS USED

ALGORITHM	DOWN TIME IN SECONDS
AES	9
RSA	7.5
BLOWFISH	11
NTRU	14

10) Turnaround Time

The turnaround time is total time elapsed between a job submitted to the processing unit and the return of the whole output. The proposed algorithm is designed to minimize the turnaround time. The turnaround time depends upon the following factors [25, 26]

- a) How the jobs are scheduled
- b) Whether the processing units is satisfying the requirements of the jobs or not
- c) How efficient is the processing unit
- d) Whether the load is balanced or not

The proposed framework has managed to cover all the above mentioned points. If monitored, the architecture diagram shows that the jobs are divided into two sections namely high requirement jobs and low requirement jobs. The jobs are checked at servers before getting assigned so that their requirements are always full filled. The following table illustrates the total turnaround time of the proposed architecture under several algorithms.

TABLE 11. TURNAROUND- TIME W.R.T ALGORITHMS USED

ALGORITHM NAME	TURNAROUND TIME IN SECONDS
NTRU	1.2
RSA	1.1
BLOWFISH	1
AES	1.15

The average cost of the proposed architecture under every algorithm is as follows. The average cost is the sum of different costs like turnaround time, encryption time, decryption time and waiting time. All the times have been marked in seconds. The average cost of all algorithms is about 3 seconds. The table does not include the time which is elapsed in the allocation of the multimedia data to the appropriate encryption algorithm i.e. the complexity finder algorithm time has not been added.

TABLE 12. AVERAGE COST OF PROPOSED ARCHITECTURE

ALGORITHM NAME	TURN AROUND TIME IN SECONDS	ENCRYPTION TIME IN SECONDS	DECRYPTION TIME IN SECONDS	WAITING TIME IN SECONDS	TOTAL TIME
NTRU	1.2	.65	.90	.10	2.85
RSA	1.1	.76	1	.21	3.07
BLOWFISH	1	.74	1.2	.15	3.09
AES	1.15	.75	.98	.25	3.13

11) Total Security Score

The total security score is the target score which identifies the flexibility and ability of the security algorithm. Blow fish has highest security score 100 but is considered only for one environment. The proposed algorithm is seeking security from 3 algorithms and for multiple types of files and hence lacks by 10% in the average security score. The average security score can be calculate using the following formula [27]

$$\text{Security Score} = (\sum_{i=1}^n 5 * t + 5(1 - \frac{e^Q}{F})) / n$$

Where n is the total number of documents which has been encrypted, t is the total time of processing, Q is the time of encryption of the document and F is the time of Decryption

TABLE 13. SECURITY SCORES OF THE THREE ALGORITHMS

ALGORITHM	TOTAL SECURITY SCORE
CDPS	54
BLOW FISH	100
KSSP	90

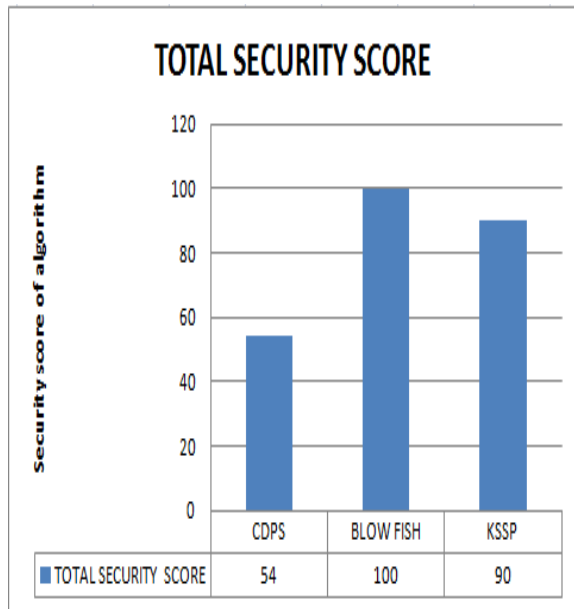


Figure 4. Total security score

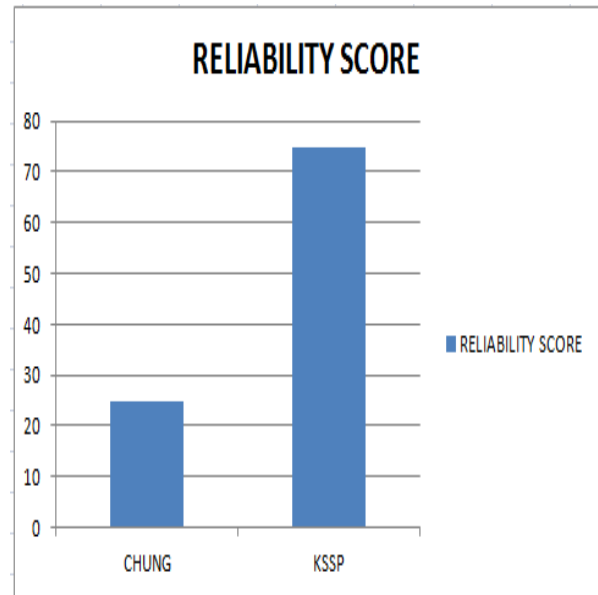


Figure 5. Reliability score

12) Reliability Score

The term reliability refers to a level of trust of using the proposed mechanism over different scenarios of data. The proposed model has tried and tested its efficiency over more than six types of files including PDF, MP3, MP4, JPEG, DOC, DOCX, TXT etc. The reliability of the proposed framework shows the reliability in providing the job scheduling solutions as well as the security solutions. It intends to save a lot of time by choosing the appropriate encryption algorithm depending upon the type of file which has been selected. The following graph illustrates the reliability score. The mathematical formula for the same is as follows [28]

$$\text{Reliability} = (\text{Pobserved} - \text{Pactual}) / \text{Pactual}$$

Where Pobserved is the probability of one document to be stored at cloud safely in the real time scenario where as Pactual=.50

TABLE 14. RELIABILITY SCORE

ALGORITHM	RELIABILITY SCORE
CHUNG	25
KSSP	75

13) Network Packet Rate

It is the total number of packet frames send in a given interval of time. The proposed algorithm even works better in these network parameters. The following table and figure illustrates that how different encryption algorithms perform when their encrypted packets are sent under a time frame. The formula for the same is as follows

$$\text{Network Packet Rate} = \left(\sum_{i=1}^n \text{PacketsTransferred} / T \right) / n$$

Where T is the total number of Packets and n is the total number of documents for which the transfer rate is being calculated.

TABLE 15. PACKET TRANSFER RATE

ALGORITHM	PACKET TRANSFER RATE
NTRU	.76
BLOW FISH	.65
RSA	.79
AES	.65

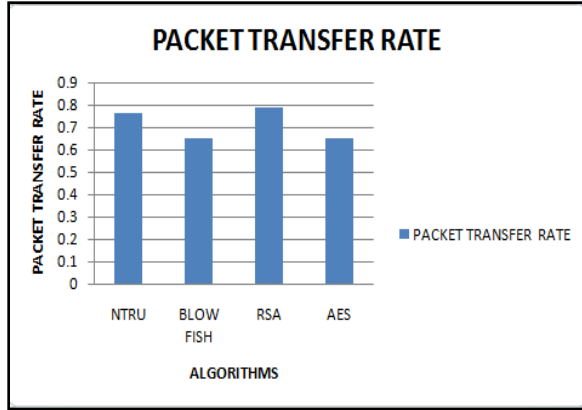


Figure 6. Packet transfer rate

14) Error Rate

It is the total error occurred while transferring the data from one end to another. Here it is observed that the RSA has the least error rate. It is because of the two way key utilization which makes it more efficient and less error prone but at the same time it consumes a lot of time and also it is not suitable for each type of file. The most efficient algorithm for the processing can be considered as NTRU as even for long bit pattern, it acts well. Yes for sure it is a little more error prone as compared to RSA and BLOWFISH due to its nature of processing. The formula for the same is as follows:

$$\text{Error Rate} = \sum_{i=1}^n \text{error}/i$$

Where n is the total number of packets sent.

TABLE 16. PACKET TRANSFER RATE

ALGORITHM	ERROR RATE
NTRU	.23
BLOW FISH	.15
RSA	.22
AES	.25

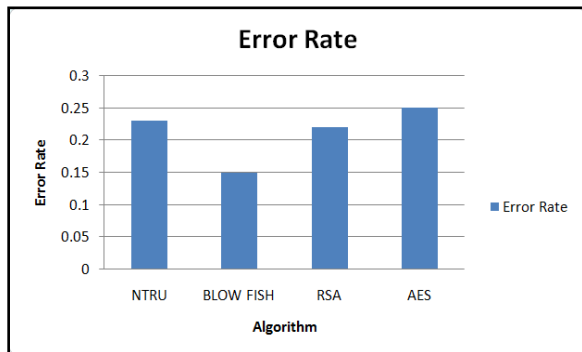


Figure 7. Error rate of algorithms

15) Total Time Cost

Total time cost is the time which is consumed for the successful encryption of the provided documents. The average time cost of the proposed framework is very less as the scheduling and management is strong enough to prevent extra time cost of the network. The simulation results shown in figure 3 demonstrate how efficient the proposed algorithm. Almost a total time gap of 7-8 seconds has been recorded.

TABLE 17. TOTAL TIME COST

ALGORITHM	TOTAL TIME COST
CDPS	8
BLOW FISH	12
KSSP	3
AES	10
NTRU	7.8
RSA	9.8

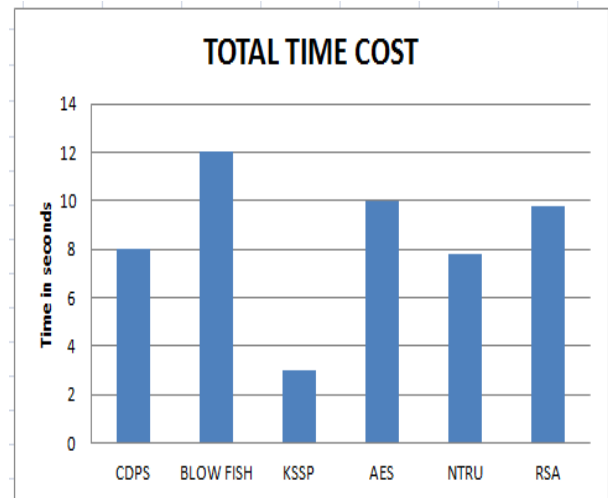


Figure 8. Total time cost

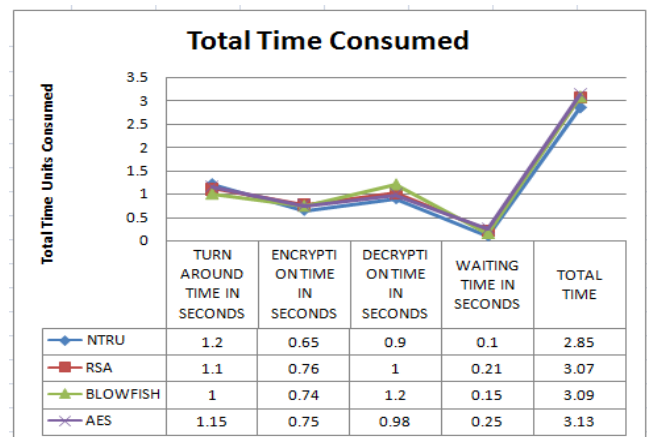


Figure 9. Different time intervals

16) Security Composition

Security Composition evaluates how efficient the security model would be. Here the security Compliance is high as the encryption is selected on the basis of the complexity finder which selects the most appropriate algorithm for the encryption. It also looks the security path in the scheduling also to prevent a high time which can be invested to some other process. The following table has been illustrated which indicates that the proposed algorithm possesses high security composition.

TABLE 18. SECURITY COMPOSITION OF PROPOSED ALGORITHM

COMPARE WITH	ENHANCED RATE	SECURITY LOSS
RC4 128	32 %	21%
RC4 256	35%	24%
BLOWFISH 245	12%	11%

VI. CONCLUSION

The current approach included 4 different algorithms in order to test their stability and reliability and other parameters which may affect any cloud performance including error rate and packet transfer rate. This paper has compared four algorithms on those parameters and has found that the proposed terminology works well and can be utilized for more algorithms as well. The proposed solution to prevent the data on cloud is not only a solution but an entire framework. The KSSP framework prevents the entire system from distributed denial of service attacks and concentrates on the maximum completion of the jobs.

The proposed approach opens a lot of gates for the future research workers. The proposed framework is only optimized for one single attack. In future, research workers may try preventing the network from different types of attacks. Hybridizing the encryption algorithm to improve the performance of the complexity finder can be another lucrative area of research.

REFERENCES

1. J. K. Liu, M. H. Au, X. Huang, R. Lu and J. Li, "Fine-Grained Two-Factor Access Control for Web-Based Cloud Computing Services", in IEEE Transactions on Information Forensics and Security, vol. 11, no. 3, pp. 484-497, March 2016.
2. Wang, K. Ren and J. Wang, "Secure Optimization Computation Outsourcing in Cloud Computing: A Case Study of Linear Programming", in IEEE Transactions on Computers, vol. 65, no. 1, pp. 216-229, Jan. 2016.
3. Ranabahu, E. M. Maximilien, A. Sheth and K. Thirunarayan, "Application Portability in Cloud Computing: An Abstraction-Driven Perspective", in IEEE Transactions on Services Computing, vol. 8, no. 6, pp. 945-957, 2013.
4. John Footen and Mohan Ananthanarayanan, "Service-Oriented Architecture and Cloud Computing in the

- Media Industry," SMPTE Motion Imaging Journal, pp.22-30,2012.
5. J. Mei, K. Li, A. Ouyang and K. Li, "A Profit Maximization Scheme with Guaranteed Quality of Service in Cloud Computing", in IEEE Transactions on Computers, vol. 64, no. 11, pp. 3064-3078, Nov. 1 2015.
6. Xinyu Lei, Xiaofeng Liao, Tingwen Huang, and Huaqing Li, "Cloud Computing Service: the Case of Large Matrix Determinant Computation", IEEE Transactions On Services Computing, pp1-14,vol. 10, no. 10, 2014.
7. Jungwoo Shin, Manseok Jo, Jongsu Lee, and Daeho Lee, "Strategic Management of Cloud Computing Services: Focusing on Consumer Adoption Behavior," IEEE transactions on engineering management, pp.1-9, 2013.
8. Srdan Dzombeta, Vladimir Stantchev, Ricardo Colomo-Palacios, Knud Brandis and Knut Haufe, "Governance of cloud computing services for the life sciences," IEEE Computer Societ y,pp.30-37,2014.
9. Jia-Lun Tsai and Nai-Wei Lo, "A Privacy-Aware Authentication Scheme for Distributed Mobile Cloud Computing Services," IEEE systems journal, vol. 9, no. 3, pp.805-815, 2015.
10. Zhen Xiao, Qi Chen, and Haipeng Luo, "Automatic Scaling of Internet Applications for Cloud Computing Services," IEEE transactions on computers, vol. 10, no. 10, pp.1-14, 2010.
11. Alexandru Iosup, Simon Ostermann, M. Nezhir Yigitbasi, Radu Prodan, Thomas Fahringer, and Dick H.J. Epema, "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing," IEEE transactions on parallel and distributed systems, vol. 22, no. 6, pp.931-945, 2011.
12. Zhaomeng Zhu, Gongxuan Zhang, Miqing Li, Xiaohui Liu, "Evolutionary Multi-Objective Workflow Scheduling in Cloud," transactions on parallel and distributed systems, 2014.
13. Kassian Plankensteiner, Radu Prodan, "Meeting Soft Deadlines in Scientific workflows using resubmission impact," IEEE transactions on parallel and distributed systems, vol. 23, no. 5, pp.890-901, 2012.
14. Abhishek Verma, Ludmila Cherkasova, Roy H. Campbell, "Orchestrating an Ensemble of MapReduce Jobs for Minimizing Their Makespan," IEEE Transactions On Dependable And Secure Computing, vol. 10, no. 5, pp. 314-327, 2013
15. Hongyan Cui, Xiaofei Liu, Nirwan Ansari and Yunjie Liu, "Cloud Service Reliability Modeling and Optimal Task Scheduling," IET Communications,pp.1-12,2014.
16. Sandeep Kapur, Dinesh Kumar, "Resource Utilization in Cloud Computing using Hybrid Algorithm", Indian Journal of Science and Technology, vol 9, issue. 43, 2016.
17. E. Walsh, G. Tesauro, J. O. Kephart, and R. Das, "Utility Functions in Autonomic Systems," in Proceedings of the First International Conference on Autonomic Computing. IEEE Computer Society, pp. 70-77, 2004.
18. Jiayin Li, Meikang Qiu, Jian-Wei Niu, Yu Chen, Zhong Ming, "Adaptive Resource Allocation for Preemptible Jobs in Cloud Systems," in 10th International Conference on Intelligent System Design and Application, pp. 31-36, 2011.
19. Shi J.Y., Taifi M., Khreishah A., "Resource Planning for Parallel Processing in the Cloud," in IEEE 13th

- International Conference on High Performance and Computing, pp. 828-833, 2011.
20. Goudarzi H., Pedram M., "Multi-dimensional SLA-based Resource Allocation for Multi-tier Cloud Computing Systems," in IEEE International Conference on Cloud Computing, pp. 324- 331, 2011.
 21. A. A. Patel, J. N. Rathod, "Reducing Power Consumption & Delay Aware Resource Allocation in Cloud Data centers", Journal of Information, Knowledge and Research in Computer Engineering, vol.2, pp.337-339, 2010.
 22. B.Wei, C.Lin, X.Kong, "Energy Optimized Modeling for Live Migration in Virtual Data Center", IEEE 4th International Conference on Cloud Computing, pp.736-737, 2011.
 23. D.Jayasinghe, C.Pu, T. Eilam, M.Steinder, I.Whalley, E.Snible, "Improving Performance and Availability of Services Hosted on IaaS Clouds with Structural Constraint-aware Virtual Machine Placement", IEEE International Conference on Services Computing, pp.72-79, 2011.
 24. K.Ye, D.Huang, X.Jiang, H.Chen, S.Wu, "Virtual Machine Based Energy-Efficient Data Center Architecture for Cloud Computing: A Performance Perspective", IEEE/ACM International Conference on Green Computing and Communications, pp.171-178, 2010.
 25. Chuang, I-Hsun, "An effective privacy protection scheme for cloud computing", Advanced Communication Technology (ICACT), 13th International Conference on. IEEE, 2011.
 26. Do Chung, Byung, Hangoo Jeon, and Kwang-Kyu Seo, "A Framework of Cloud service Quality Evaluation System—Focusing on Security Quality Evaluation", International Journal of Software Engineering and Its Applications, pp. 41-46,2014
 27. <http://math.stackexchange.com>
 28. <http://study.com/academy/lesson/reliability-coefficient-formula-definition-quiz.html>

Secure Metadata based Search over Encrypted Cloud data supporting Similarity Ranking

Mahreen Saleem, A.M.U, Department of Computer Engineering, Aligarh,UP, India
M.R.Warsi, A.M.U, Department of Computer Engineering, Aligarh,UP, India
Neelam Saleem Khan, NIT Srinagar, Department of IT, Srinagar,J&K, India
mehkhan27@gmail.com

Abstract- Cloud storage has become more and more popular in the recent trend since it provides various benefits over the traditional storage solutions. However, security and privacy issues arise by allowing a cloud service provider (CSP) to take the custody of sensitive data; hence making data encryption inevitable. Even though the encryption approach strengthens the data security it results in degraded data efficiency because of the reduced search ability. A major challenge confronting researchers especially in cloud computing environment, when sensitive data is stored over the cloud, is to store the data in hidden form and yet provide efficient retrieval of outsourced encrypted cloud data at the same time.

In this paper we present a scheme for secure Search over Encrypted Cloud Data (SECD) to improve the data discovery and user searching experience by searching metadata instead of original outsourced data. It supports multi-keyword, extends keywords to include synonyms and related terms, by incorporating the concept of domain dictionary. The necessary preprocessing of the keywords is done at the user end in order to reduce the overhead and processing time at cloud end. By incorporating XML based index file, we reduced the index file storage requirements and search time by significantly large amounts. The security analysis testifies that data security has been improved by searching the cloud using trapdoor instead of keywords. The performance analysis of the proposed scheme on the dataset concluded that time taken to update the index file has been reduced quite significantly over outsourced encrypted cloud data.

Keywords— Cloud computing, Domain dictionary, Index file, Metadata, Search scheme, Synonym search, XML

I. INTRODUCTION

In today's era of technology when massive quantities of data are generated every single day at an enormous pace, cloud storage has become a need of the hour. It allows data to be accessed on demand from any remote locations in a pay as you go manner. One of the most exciting features of Cloud computing is its flexibility of use, which attracts a wide population of users towards itself. With the rising popularity of Cloud computing, many organizations and data owners are motivated to outsource their large volumes of data to cloud servers like Amazon, Microsoft Azure, Google Drive, and so on. There is no doubt that Cloud computing has brought about a paradigm shift; but sadly along with many benefits provided by cloud storage, many security problems arise which inhibit enterprises from migrating their sensitive data to cloud storage. In order that many enterprises can draw the ubiquitous benefits that the Cloud computing in general and cloud storage in particular promises to offer, the underlying security issues associated with Cloud computing need to be addressed. The issue of confidentiality is taken care by applying cryptographic techniques prior to outsourcing of the sensitive data. But in order to implement efficient data utilization, search over encrypted cloud data has been a great challenge. In such a scenario, an efficient scheme that supports search over encrypted data in cloud computing becomes very significant. There is a need to retrieve the data efficiently, enhance system usability, reduce communication overhead and improve user searching capabilities and that too without compromising the security and privacy aspects of the outsourced data.

Various searchable encryption schemes have been developed so far to provide search capability over outsourced encrypted data. Mostly these schemes employ index based ranked search and include frequency of query keyword in the document; hence challenging the security and reducing efficiency. In order to address the aforementioned issues, in this paper we propose a Search over Encrypted Cloud Data scheme (SECD). In this proposed scheme we present a metadata based index generation scheme for search over encrypted data. We have proposed a novel concept of domain dictionary wherefrom a metadata based index file is to be generated that is to be queried while performing searching operations. The improvements in search results illustrated that our developed schemes are more efficient in terms of performance, and security than the former proposed schemes

The rest of the paper is structured as follows. We first present a review of the related works in Sect. II. Next we define the identified problem in Sect. III. We then discuss the detailed design of the proposed search scheme in Sect. IV. In Sect. V we analyze the performance of the defined search scheme and discuss the results, and finally conclude in Sect. VI.

II. RELATED WORK

To keep user's data confidential against an untrusted Cloud service Provider (CSP) many schemes were proposed in recent researches that enable keyword search over encrypted data in cloud computing. Few researchers have developed and implemented methods that retrieve data on the basis of attributes (e.g., country he lives in). In attribute based encryption, introduced by Sahai and Waters [1] [2], decryption of cipher text is possible only if the set of attributes of the user key matches the attributes of the cipher text. Keyword-based search techniques have been the most popular and user friendly search techniques and have been widely applied in plaintext scenarios, such as Google search [5]; thereby permitting users to selectively retrieve files of interest. But traditional keyword based search methods are not suitable for search over cloud environments when the data is in encrypted form. Data encryption leaves the data in unintelligible forms thus restricting users to perform keyword based search operations. In order to search securely over outsourced encrypted cloud data, searchable encryption techniques have been developed in recent years [7], [8], [11],[12]. These searchable encryption techniques allow search to be performed over the encrypted data and have been used in cloud environment.

D.Song et al.[3] proposed a two layered symmetric key cryptography to overcome security issues but the scheme was ineffective when searching was to be done on huge data. D.Boneh et al. [4] proposed public encryption keyword search. In this, file is encrypted using public key by the data owner, but the authorized users can search a file using their private key. For secure indexing, Cao et al. [6] proposed Privacy pre-serving multi-keyword ranked search (MRSE) which is one of the first works that proposed a viable solution to address a multi-keyword ranked query problem. In MRSE, all keywords are defined in a dictionary and the trapdoor is built using the inner product of two vectors to generate secure keyword queries and an internal ranking algorithm is used to determine the top k files to be returned to the data consumer. However, a major drawback to this approach is the use of a static dictionary that has to be rebuilt completely each time for every new keyword.

Z. Xu et al.[8] proposed Multi-Keyword ranked query on encrypted cloud data (MKQE) which is an improvement of MRSE. In MKQE there is a significantly great reduction in the amount of computations when the keyword

dictionary is expanded and also keyword access frequencies are taken into consideration. This scheme lacks functionalities like semantic query and fuzzy keyword query.

H.Kam et al. [9] developed three different schemes which are “Synonym-Based Keyword Search (SBKS)”, “Wikipedia-Based Keyword Search (WBKS)”, and “Wikipedia-Based Synonym Keyword Search (WBSKS)”. In SBKS for each user query a list of trapdoors is generated, hence for each keyword a lengthy trapdoor list is created. For WBKS the index is constructed using a set of Wikipedia articles. For performing search with a specific keyword, the TF-IDF is used to generate the similarity score. In WBSKS which is a hybrid of SBKS and WBKS the advantage of both the schemes is taken to expand the search to cover synonyms of the search keyword while maintaining a small index. In these schemes time of index construction was very high. Wang et al. [14] proposed an encrypted inverted index for providing search capabilities over encrypted data. In this scheme ranking of the retrieved documents was done according to the computed relevance score between documents and query.

Khan et al. [10] proposed secure ranked fuzzy multi-keyword search over out-sourced encrypted cloud data (RFMS). This scheme incorporated fuzziness and an internal ranking algorithm to rank the relevant files based on a relevance score. H.T.Poon et al. [13] investigate the problem of processing large amount of encrypted documents in XML-like formats. They presented a scheme which combines search and computations over encrypted documents to extract aggregated data from XML documents. The approach they proposed can be used for partially or fully encrypted XML documents. Vandana S. et al.[11] proposed Multi-Keyword Synonym based Fuzzy Ranked Search over Outsourced Encrypted Cloud Data (MSFRS). It supports multi-keyword, suggests synonyms if user forgot keyword & returns results after evaluating relevance scores based on frequency of keyword in the documents bearing same rank.

III. PROBLEM DEFINITION

Understanding the importance of the searching on encrypted outsourced cloud data, we propose a framework that includes new domain dictionary construction paradigm that employs a novel similarity index generation algorithm and uses an extended multi-keyword searching technique to retrieve relevant result sets. In SECD, we assume that the amount of data continues to increase from time to time; hence, the domain dictionary has to be expanded periodically. Our contribution can be summarized as follows:

1. The domain dictionary can be expanded dynamically with reduced overhead of the dictionary reconstruction when new keywords are added.
2. A novel similarity index generation algorithm is proposed in which each index entry is in the form of trapdoors, which are not dictionary words, thereby improving security.
3. Search query is converted into a trapdoor before being fed to cloud search engine that is in the form of a set of key-value pairs. Hence, making the search query more secure and precise.
4. Weights of search terms is considered, hence search results are more relevant.
5. All the required preprocessing of data and input query is done at the user end, thereby relieving the cloud end of preprocessing tasks, hence efficiency both in terms of time and performance is achieved.

6. The generated Similarity Index file to be uploaded to cloud server is stored in the XML format which provides interoperability, faster querying and a significant reduction in file size.

IV. SYSTEM DESIGN OF SECD SCHEME

We consider a system model involving three different entities; the cloud server (C), the data owner/super user (DO) and the data user (DU). The files containing sensitive data are owned by the DO and are to be outsourced to the cloud server. Fig. 1 shows the complete architecture of SECD scheme. In order to provide confidentiality, the sensitive information needs to be encrypted before outsourcing it to the cloud server. In order to build a searchable index for providing search capabilities the metadata needs to be extracted from the data files to be outsourced. Metadata construction begins with the creation of the Domain dictionary which can be viewed as a collection of data buckets. Each data bucket or domain in the domain dictionary contains all the keywords and their extended synonym set corresponding to the similar files out of whole set of documents. Each data bucket or domain is labeled with a unique domain id. A searchable index file *I* is constructed using the domain dictionary and is converted into XML format and along with the encrypted files is outsourced to the cloud.

In the search phase, a trapdoor is to be generated against the query keywords submitted by the data user. The cloud server will search the index *I* against the generated trapdoor and then return search results to the user. At the end of the search phase user receives the required encrypted files. These files are then decrypted at the user end and the intended original files are made available to the user.

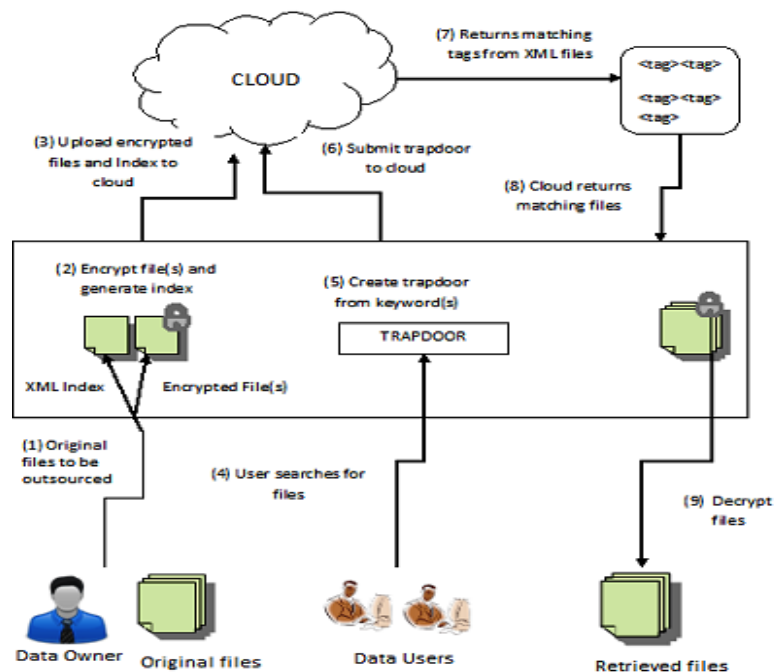


Fig. 1 Architecture of SECD

A. *Detailed design:*

The SECD system consists of following components:

– **Initialize ():** The DO determines the number of files containing sensitive data that are to be outsourced. But prior to outsourcing, keywords of interest need to be extracted from those files. The DO recognizes the distinguishable keywords and adds them to the domain dictionary. These keywords are later used for retrieving these files from cloud server C.

– **Build-Domain-Dictionary (D):** The user application expands each specified key-word into the keyword synonym set by the Keyword Set Extension (KSE) process. The WordNet® lexical database, provided by Princeton University, is used as the synonym dictionary to create the keyword synonym set. The keyword set is fed in-to the synonym dictionary to retrieve the extended keyword set and each domain is populated with the similar or related keywords which in effect represent the files that are conceptually related.

– **Encrypt (F, K_s):** The set of sensitive files F = (F₁, F₂ ...F_m) is encrypted using symmetric key encryption and an encrypted set of files E= (E₁, E₂...E_m) is created. The key K_s used for encryption can be any randomly generated key by DO.

– **Build-similarity-index (K_s,D,F):** In order to compute the similarity value between some file F and domain D_i all the keywords representing F and all the keywords present in D_i are represented as vectors with binary weights. Similarity value (S_i) for file i is calculated as in (1)

$$S_i = \frac{A \cdot B}{n_a} * W \quad (1)$$

n_a is the count of terms in vector A, w is the weight, and A.B is the dot product of two vectors A and B.

where, W=1 for exact match, and

W=0.66 otherwise

Likewise, S_{ij} is calculated for all i files with all j domains in D. The overall process of index generation is shown in the algorithm explained in Fig. 2.

– **Query-Trapdoor-creation (w_q):** When the user has to performs a search he sends a multi-keyword query w_q = (w_{q1}, w_{q2}, ...w_{qk}) to DO. Query trapdoor is calculated in a similar way as in case of index building. The similarity calculation between query Q and domain D is done using the formula given in (2) which is the modified version of (1) and is written as:

$$S = Sim(Q, D) = \frac{Q \cdot D}{n_q} * w \quad (2)$$

where, S is the similarity between extended query Q and domain D_i, Q.D is the dot product between query vector Q and domain vector D, n_q is the number of query terms

Algo 1:Build-index(k_s, F, D)

k_s : Set of file keywords, F : Original file, D : Domain Dictionary

Output I .

1. Initialization:
 - I. Select the keywords $K_w = (K_{w1}, K_{w2}, K_{w3}... K_{wn})$ from each file contained in the set of sensitive files $F = (F_1, F_2, F_3...F_m)$.
 - II. Build a domain dictionary $D = (D_1, D_2, D_3...D_n)$ for all extracted keywords.
2. Trapdoor set of each keyword
 - I. Extract the selected list of keywords $K_w = (K_{w1}, K_{w2}, K_{w3}... K_{wn})$ from domain in domain dictionary that match the file keywords.
 - II. Entire file is encrypted E_i and uploaded on Cloud Server C.
 - III. Similarity value of each keyword set and the matching domain(s) is calculated and stored in its corresponding trapdoor set $p = (p_1, p_2, p_3...p_m)$.

Element p_i in trapdoor set is of the form <File_ID, Similarity_value>.

3. Building the final index file
 - I. The corresponding trapdoor sets $p = (p_1, p_2, p_3...p_m)$ are stored against each Domain Id which constitutes the final index I .

Fig. 2 The details of Build-index for SECD

— **Rank-retrieval (R)**: The trapdoor set for keywords in query is extracted and the cloud server is queried with the trapdoor set generated by the user application. The generated trapdoor set is matched with the index file and the relevance score is calculated on the basis of similarity score between the trapdoor and the index file entries. On the basis of ranking score the most relevant files are sent to the user. The detailed description of rank based multi-keyword search is given below in Fig. 3.

Algo 2:Rank-retrieval (w_q)

Output: K ranked files.

1. Query received
 - I. The DO sends the multi-trapdoor query $w_d = (w_{d1}, w_{d2},... w_{dk})$ to cloud server C.
2. Parsing the query
 - I. Parse the received query to extract the domain Ids and similarity values.
3. Searching Index File
 - I. The search engine retrieves the tags matching the domain Ids from the Index I .
 - II. The retrieved tag values are parsed to extract relevant file names and similarity values.
 - III. Matching is done between the similarity values retrieved from the trapdoor query and those retrieved from tags.
 - IV. Relevant files are ranked according to the similarity score.
4. Returns list of *top K ranked files*.

Fig. 3. The details of Rank Retrieval for SECD.

V. PERFORMANCE ANALYSIS

In order to evaluate SECD scheme multiple experiments were conducted. The experiments were conducted over a dataset consisting of 1000 files of size approximately 7 KB and number of unique keywords ranging up to 2000. The necessary codes are written in PHP. The underlying platform consists of a CentOS server with Eucalyptus Cloud installed on it. Eucalyptus is open source software for building private and hybrid clouds. The Intel core i3 processor running at 3.20 GHz with 4GB RAM was used for the experiments. MSFRS from [11] is implemented and is used as the benchmark in our tests. We have used the symmetric-key encryption and decryption for the data files.

A. Security and Privacy Analysis

We evaluate the security of the proposed scheme by evaluating the fact that that cloud server C should not learn the original contents of sensitive data files uploaded by the DO. After applying the cryptographic techniques the cipher text generated for encrypted files resulted in special characters, instead of normal English text so, the dictionary attack is avoided. Also, the index file entries and the search queries are in the form of trapdoors. The trapdoor entry of an index file consists of similarity value paired with a particular file name against each domain in the domain dictionary, instead of the keywords the threat of brute force attack to know the secret key ks is eliminated completely. The example of Similarity Index entries is shown in Table 1. Also only the authorized data users are allowed to have access to the sensitive outsourced data thereby guaranteeing privacy.

Table1: Example of Similarity Index entries

Domain ID	Sim_Index value
1	<2.txt,0.66> <7.txt,0.33> <35.txt,1>
2	<phpfunctions.txt,1> <ITnotes.docx,0.333>
3	<quotes.txt,0.5> <25.txt,0.66>

B. Efficient Index Generation

An efficient feature of SCED over previous solutions is that the index field entries are in the form of XML tags. We compare the performance of our index construction with the previous scheme MSFRS. The results in Fig. 4 show that in both the schemes the time taken to construct the index file is almost same, but some hike in construction time was observed in case of MSFRS when the size of file to be uploaded increased. No such effect was shown by the SECD scheme. Hence in our scheme index construction time is independent of file size which is primarily due to not parsing the entire data file.

C. Improved User Searching Experience and Data Discovery

By incorporating the mechanism of fuzzy keyword search and also taking into consideration the synonyms and the related words to the search query in our scheme, the search results also include the documents containing related terms and are not restricted to exact keyword matching. Also user searching experience has been improved since

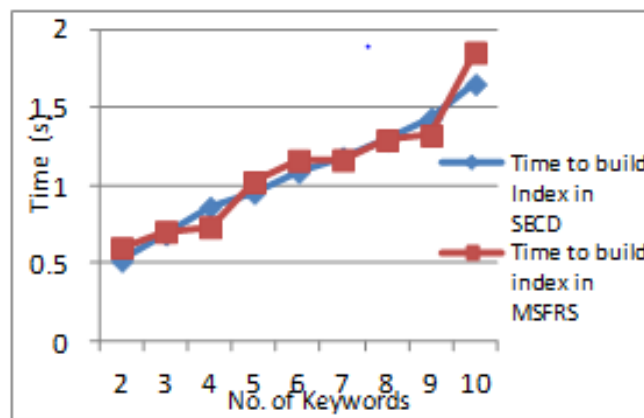


Fig. 4 Comparison of average time used to construct the index files with 100 files on C

fuzzy search makes room for the typing errors and format inconsistencies; hence the users need not to bother about that. For fuzzy search the Levenshtein function was used due to which the overhead of correcting the keyword was very low; of the order of a few microseconds.

D. Efficient Search

Our scheme involves only searching the metadata file which in fact is the XML index file, and eliminated the need of parsing of the encrypted files, hence, a large amount of time is saved while searching. Fig. 5 draws a comparison of time calculated for searching when 1000 files are listed on C using SECD and MSFRS schemes. Also, that the trapdoor entries created by our developed scheme are very short and precise, as they comprise of numeric similarity values, hence adding to computational efficiency. The trapdoor created against each user query consists of a similarity value of the query terms paired with a domain id. Table 2 shows an example of the trapdoor created against a user query.

Also ranking is done on the basis of similarity value between the query and the encrypted files. More the number of multiple keyword matches, more is the similarity, higher is its relevance to the search query.

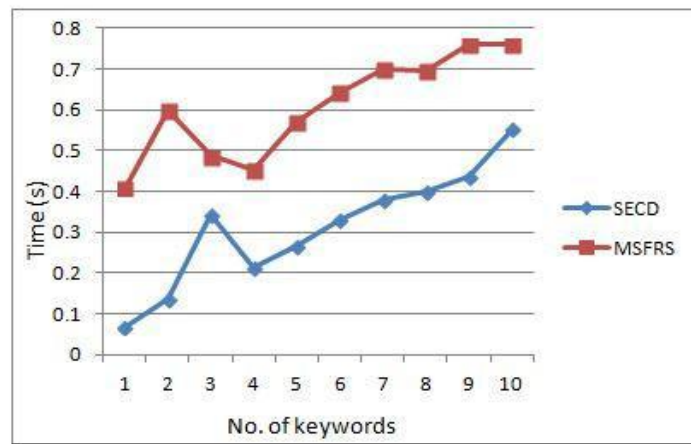


Fig.5 Comparison of time calculated for searching when 1000 files on C

Table 2 :Example of Trapdoor created against a user query

QUERY	TRAPDOOR
life, beautiful, nature, family	<1,0.25> <42,0.5> <44,0.25>

VI. CONCLUSION

In this work an attempt has been made to improve the data security, data discovery, and user searching experience by supporting secure Search over Encrypted Cloud Data, by searching metadata. We have incorporated the concept of similarity between the file and the domain dictionary and constructed an index file without involving the keywords. The test results proved that search time on cloud was reduced by large amounts by incorporating XML index file in our scheme.

The security and privacy analysis testify that the developed scheme is reliable in terms of security and privacy. The trapdoors generated in our scheme are not the dictionary words or cipher text of such words which hides users' query keywords against the server. In future, our searching scheme can be made more efficient by reducing the index generation time as much as possible.

REFERENCES

- [1] Sahai, A., Waters, B.: Fuzzy identity-based encryption. In: Proceedings of EUROCRYPT, pp. 457–473 (2005)
- [2] J. Bethencourt, A. Sahai, and B. Waters. Ciphertext-policy attribute based encryption. Proceed-ings of IEEE Symp. on Security and Privacy, 2007, pp: 321–334.
- [3] Song, D., Wagner, D., Perrig, A.: Practical techniques for searches on encrypted data. In: IEEE Symposium on Security and Privacy, pp. 44–55 (2000)
- [4] Boneh, D., Crescenzo, G.D., Ostrovskyand, R., Persiano, G.: Public key encryption with keyword search. In: International Conference on the Theory and Applications of Cryptographic Tech-niques, pp. 506–522 (2004)
- [5] Google, Britney spears spelling correction. Referenced online at <http://mooglem.com/threads/googlecache/64.233.187.104/jobs/britney.html>, June 2016.
- [6] Cao, N., Wang, C., Li, M., Ren, K., Lou, W.: Privacy preserving multi-keyword search over encrypted cloud data. J. IEEE Trans. Parallel Distrib. Syst. 25, 222–233 (2014)
- [7] Sharon. E, N. Saravanan, A Survey on Keyword Based Search over Outsourced Encrypted Data. International Journal of Engineering and Technology (IJET), Volume 5,pp: 905-909(2013)
- [8] Xu, Z., Kang, W., Li, R., Yow, K., Xu, C.: Efficient multi-keyword rank query on encrypted cloud data. In: 18th IEEE International Conference on Parallel and Distributed Systems, Singa-pore, pp. 244–251(2012)
- [9] Moh, T.S., Ho, K.H.: Efficient semantic search over encrypted data in cloud computing. In:International Conference on High Performance Computing and Simulation, pp: 382–390.Bologna (2014)
- [10] Khan Neelam, S., Challa, R.K.: Secure ranked fuzzy multi-keyword search over outsourced en-crypted cloud data. In: 5th IEEE International Conference on Computer and Communication Technology (ICCT), Allahabad (2014)
- [11] Saini V., Challa R.K., Khan N.S. : An Efficient Multi-keyword Synonym-Based Fuzzy Ranked Search Over Outsourced Encrypted Cloud Data. In: Advanced Computing and Communication Technologies. Advances in Intelligent Systems and Computing, vol 452. Springer, Singapore (2016)
- [12] Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., Lou, W.: Fuzzy keyword search over encrypted data in cloud computing. In: 29th IEEE International Conference on Computer Communications, pp. 1–5, San Diego (2010)
- [13] H. T. Poon, A. Miri, “Computation and Search over Encrypted XML Documents,” BigDataCongress,pp.1–4,(2015).
- [14] C. Wang, N. Cao, K. Ren, and W. J. Lou.: Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data.In: IEEE Transaction on Parallel Distributed System, vol. 23, no.8, pp. 1467-1479, (2012).

Rola: An Equi-Matrik Chhand of Hindi Poems

Kapil Kumar Kushwah

School of Electronics, Devi Ahilya Vishwa Vidhyalaya
Indore, India
kkushwah12@gmail.com

Brijendra Kumar Joshi

Military College of Telecommunication Engineering
Mhow, India
brijendrajoshi@yahoo.com

Abstract—The ancient Indian literature/text has now started becoming available in digital form. Hindi literature dominates such digitalization. Unfortunately, analysis of such text has not been attempted by many. Our aim is to detect Rola Chhand present in Hindi poems. Major portion of Indian literature in social and cultural domain is in the form of poems. There are many properties of these poems. The base of all such properties is Chhand. Without the proper knowledge of Chhand, one cannot understand and analyze poems. For the current younger generation Chhand is fast becoming outdated though it is glorious and scientific in nature. We have developed an algorithm that detects Rola Chhand present in any poem supplied as input. It will be very helpful for poem learners. It is the first attempt of its kind in Hindi poems. Though few poems are available in the digital form but their poetic properties are not aimed at. The algorithm developed aims at one such very important property.

Keywords—Hindi Poem; Rola; Chhand Shastra; Pingal; NLP; Information extraction.

I. INTRODUCTION

Today, NLP is one of the most talked about field of research. Every natural language has different properties, syntax and semantic structures. When it comes to processing by computers, natural languages pose different levels of challenges, Hindi is no different in this regard.

Usual techniques used in Natural Language Processing (NLP) do not apply here as ancient Indian poems are based on and around Chhand Shastra. Chhand Shastra was written by Maharshi (Sage) Pingal (Pinglak) before around 2850 B.C. All ancient Indian poems, written in any Indian language, follow the rules/techniques given in the Chhand Shastra.

In spite of being the third largest spoken language in the world and an official language of India, NLP of Hindi has not addressed many areas so far. Poems are one of them.

Basic syntactic structure of an ancient Hindi poem is called Chhand. Chhand is actually a Sanskrit poem property and inherited by Hindi like many other properties. In this paper we propose an algorithm for automatic detection of Rola Chhand.

II. LITERATURE SURVEY

Joshi and Selot [1] developed a Hindi-English-Hindi bilingual dictionary using hypertext technique. They represented each character of Hindi (along with modifier/matra) using a 2-byte unsigned integer. The dictionary was implemented under MS-DOS using C-language.

In his Ph.D. thesis, Joshi [2] described the development of an assembler, and a compiler (for a C-like language) in Hindi. In addition, he developed a Hindi-English bi-lingual text editor. The editor so designed does not allow the writing of wrong Hindi characters along with modifiers.

Hindi has 13 modifiers called Matras. Kushwah and Joshi developed an algorithm to detect modifiers [3]. The character image is first preprocessed and then wavelet transformed to minimize the features. Next, the resultant image is divided into some zones. According to these zones, the modifiers are identified.

In their paper, Joshi and Kushwah proposed two versions of Micro-Parsing algorithm [4]. Generally, a sentence is treated as a sequence of words and parsing is done to verify syntactic correctness. The authors parsed individual words with characters as tokens. Their algorithm rectifies the Google input tool. Google input tool does not restrict us to write syntactically wrong words.

Joshi and Kushwah [5] developed a rule based word formation technique for Hindi words, called Sandhi. They also developed a database to store these words for future use. They achieved 98% accuracy for compound words based on two-word Sandhi.

III. ROLA CHHAND

Rola Chhand was very popularly used in eras called Bhakti Kal (1300-1700) and Riti Kal (1700-1900) of Hindi. This Chhand is based on mathematics provided by Maharshi Pingal for Sanskrit, adopted by Hindi.

Every ancient Indian poem has some properties [6] [7][8][9] like Yati (very small pause), Gati (flow of poem), Gan (pair of matras) and Varn etc. Many Indian sages worked on these but the Shastra written by sage Pingal is the only available and considered authentic work. Pingal wrote all the properties of Chhand. Since Rola is an ancient Chhand of Prakrit language, these properties are also applicable to Rola.

The Rola Chhand belongs to Sam-Matrik Chhand category, in which poets focus on only matras. In each Charan of Rola the sum of matras is 24. When we read after first eleven matras, we take a very short pause called Yati. So, according to Yati, in a Rola Chhand every Charan has 24 matras in a group of eleven and thirteen. Since each Charan has equal sum of matras, we call this an Equi-Matrik Chhand. Moreover, the end of every pad has Antyanupras alankar. That means, the second last and the last matra both are Guru or Laghu.

IV. STRUCTURE OF ROLA CHHAND

The structure of this Chhand is shown in Figure 1. According to this, a Rola Chhand has 4 lines. In the first and second lines, the second last and last matras are same as well as in third and fourth lines also. For identification of Rola Chhand we use some standard. According to Figure 1 it is clear that when a line of Rola is completed, poets use a single purn viram (| stop sign in Hindi) and when second line is completed, they use double purn viram (|| is a sign of Hindi language used in only poem/chhand). Same technique is used in remaining lines of a poem.

V. PROPOSED ALGORITHM

For identification of a Rola Chhand, first we store the complete poem in a string called HindiPoem. After that we find the double purn viram sign in HindiPoem string. Since at double purn viram, half portion of Rola Chhand is completed, we take this for analysis. We store this portion in AnalyzePoem string, and set the location variables like startP and endP. Now the next task is to find the first and second line of Rola. For this we find single purn viram in analyzePoem string. This location is stored in a variable called midLoc and splits analyzePoem into two parts from this location and stores in pad1 and pad2 strings. Then we apply the Pingal mathematics^{7,8,9} on them, to convert them in mathematical form. This mathematical string is stored in strings called matraArray1 and matraArray2. Now we find the sum of matras in this string; if it is in the rage of Rola Chhand, we apply the other properties of Rola, which are explained in previous section. When all properties of Rola are satisfied, we say that the given Chhand is a Rola. The proposed algorithm for automatic detection of Rola Chhand is given in Figure 2.

Step 1: Start.

Step 2: Initialize all variables, strings and thresholds like HindiPoem, analyzePoem, pad1, pad2, matraArray1, matraArray2 and midLoc etc.

Step 3: if(HindiPoem[endP] == '|')

```
analyzePoem=HindiPoem.subStrin-g(startP,
endP-1);
startP=endP+1;
go to step 4.
```

else if(endP>HondiPoem.length)

गाधिराज को पुत्र, साधि सब मित्र शत्रु बल ।
दान कृपान विधान वश्य कीन्हों भुवमंडल ॥1॥
कै मन अपने हाथ, जीति जग इंद्रियगन अति ।
तप बल याही देह भये क्षत्रिय ते ऋषिपति ॥2॥

Figure 1. Rola Chhand from Ramchandrika written by Keshavdas.

```
go to step 8.
else
endP++;
go to step 3.
```

Step 4: if (analyzePoem[i] == 'l')

```
pad1=analyzePoem.subString(0,i-1);
pad2=analyzePoem.subString(i+1,
analyzePoem.length);
go to step 5.
```

else

```
i++;
go to step 4.
```

Step 5: Convert pad1 and pad2 in mathematical form according to the Pingal rules and store on matraArray1 and matraArray2 respectively than go to step 6.

Step 6: Apply Rola properties on pad1, pad2, matraArray1 and matraArray2 than go to step 7.

Step 7: if (all properties of Rola Chhand are satisfied)

```
Rola Chhand detected.
go to step 3.
```

else

```
Rola Chhand is not detected.
go to step 3.
```

Step 8 Stop.

Figure 2. An algorithm for aotomatic detection of Rola Chhand

VI. MANUALLY TESTING OF PROPOSED ALGORITHM

To test this algorithm we apply a Chhand shown in Figure 3 from the Chhand Prabhakar written by Jagannath Prasad, which was first published in 1894 [9].

In Figure 3 we have a Chhand, which has two pads. This Chhand is stored in HindiPoem string. First we find the double purn viram (||) in HindiPoem and store the contents before this in analyzePoem string. Now we find the Charan in analyzePoem. For this we find single purn viram (|) and store pre and post purn viran contents in pad1 and pad2 respectively. Now convert the text string into mathematical form of Pingal [6] [7] [8] [9], which is given in Figure 4.

रोला की चौबीस, कला यति शंकर तेरा ।
सम चरणन के आदि विषम सम कला बसेरा ॥1॥
राम कृष्ण गोविंद भजे पूजत सब आसा ।
इहां प्रमोद लहंत अंत बैकुंठ निवासा ॥2॥

Figure 3. Rola Chhand from Chhand Prabhakar written by Jagannath Prasad [9].

S S S S S I I S I I S I I S S
I I I I I S S I I I I I I S I S S
S I S I S S I I S S I I I S S
I S I S I I S I S I S S I I S S

Figure 4. Pingal string of Rola Chhand given in Figure 3.

We refer to this mathematical form as Pingal string. These strings are stored in MatraArray1 and MatraArray2 respectively. The S is equivalent to 2 and I is equivalent to 1, so the sums of MatraArray1 and MatraArray2 are 24 and the last matras of both pad1 and pad2 are same. Since all properties of Rola Chhand are satisfied, this Chhand is a Rola Chhand.

Now we apply the process explained above on second pad. After the analysis the second pad is also a Rola. Since both pads belong to Rola Chhand, the complete poem is a Rola Chhand.

VII. RESULTS

Rola is a type of Chhand, which is very melodious to reading as well as listening. Many poets used this to write their poetry, poetry sections and epics. But no single creation is available at present, which is based on this Chhand only. Most poets used them with other Chhands. We took around 100 Chhands from different poems as well as epics, in which 59 Chhands belong to Rola Chhand category. This algorithm detects 53 Chhands as Rola out of 59. We also checked each Chhand (analyzed by this algorithm) manually. On the basis of manual and automatic detection the accuracy of this algorithm is 89.83%.

VIII. CONCLUSION

The Indian culture is very decent; one of the reasons of this decency is the richness of the literature. To increase the richness of literature, Chhand worked as base of them. Chhand provides the base structure to poets, in which poets fill the related words according properties of poems. In this, the word meaning is also important. Proper meaning and the base structure selection enhanced the control of poet on language as well as the Chhand. In this paper we presented the first Chhand identification algorithm, which is based on their properties. This algorithm is 89.83 % accurate. Many types of Chhand are available in Hindi, and many possibilities are present. This algorithm does not detect those Rolas, in which poets make the assumption of last matra as Laghu to Guru and vice-versa. This assumption reduces the accuracy of algorithm. Moreover in some Rola Chhands, the sum of matras is greater than the threshold, so they are also not detected.

REFERENCES

- [1] B. K. Joshi and M. Selot, "Design and Implementation of a Bilingual Dictionary using Hypertext Techniques", Proceedings of the National Seminar on Computer Processing of Hindi Language (COMPHIL'93), Allahabad, April 10-11, 1993, pp TS1-3.1-TS1-3.4.
- [2] B. K. Joshi, "On the Design and Implementation of Computer Software and Software Development Tools in Hindi", PhD Thesis, Rani Durgavati University, Jabalpur, 1998.
- [3] K. K. Kushwah and B. K. Joshi, "Hindi Modifier Recognition based on Pixel Relationship", International Conference on ICT Business Industry Government, Indore, Nov 18-19, 2016.
- [4] B. K. Joshi and K. K. Kushwah, "Micro-Parsing of Hindi Words", International Journal of Computer Science and Information Security, Vol-14, No 11, November 2016, pp 261-265.
- [5] B. K. Joshi and K. K. Kushwah, "Sandhi: The Rule Based Word Formation In Hindi", International Journal of Computer Science and Information Security, Vol-14, No 12, December 2016, pp 781-785.
- [6] P. Pandye, "Manak Samanya Hindi", Arihant Publicationes (India) Ltd., Ed.-1, 2015, pp 28-36.
- [7] V. N. Prasad, "Adunik Hindi Vyakran aur Rachana", 23rd Edition, Bharti Bhavan Publication, 1993, pp 28-35.
- [8] H. Bahari, "Hindi: Shabda – Artha - Prayog", Abhivyakti Publication, 2015, pp 98-108.
- [9] J. Prashad, "Chhand Prabhakar", Uttar Pradesh Hindi Sansthan, Lucknow, 2008, pp 101-102.

AUTHORS PROFILE

Kapil Kumar Kushwah is a research scholar in Electronics science at School of Electronics, Devi Ahilya Vishwavidhyalaya Indore. He has obtained M.Sc. in Electronics and Communication stream from School of Electronics, Indore in 2011. He has qualified the UGC National Eligibility Test (NET) for Lectureship in June-2012. His research interests are Natural Language Processing, Hindi Language, Indian Literature in Poetic form and Programming Languages.

Dr Brijendra Kumar Joshi is a Professor of Electronics & Telecommunication and Computer Engineering at Military College of Telecommunication Engineering, MHOW (MP), India. He has obtained BE in Electronics and Telecommunication Engineering from Govt. Engg College, Jabalpur; ME in Computer Science and Engineering from IISc, Bangalore, PhD in Electronics and Telecommunication Engineering from Rani Durgavati University, Jabalpur, and MTech in Digital Communication from MANIT, Bhopal. He has more than 32 years of teaching experience. His research interests are programming languages, compiler design, digital communications, mobile ad-hoc and wireless sensor networks, image processing, software engineering and formal methods. He has number of research publications to his credit. He has supervised four PhD theses and currently supervising eight research scholars. He has authored two books on Data Structures and Algorithms in C/C++ published by Tata McGraw-Hill, New Delhi.

A Hybrid Approach for Compression of MPEG videos

Kanupriya
CSE. Chandigarh university
Mohali, Punjab (India)
Kanupriya1413@yahoo.com

Silky Narang
CSE. Chandigarh university
Mohali, Punjab (India)
silkynarang94@yahoo.com

Abstract - Data, Image and Video compression plays a significant responsibility in real-time exploration or video conferencing functionality. In any motion based video compression procedure, motion and variation evaluation is computationally more expensive and time consuming procedure. An imperative constituent of data, image and video compression systems is a transform. A transform is used to transform data, image and video intensities. The discrete wavelet transform of a video signal gives its time frequency representation, so, we can easily find the best pixel sets. This property of the discrete wavelet transform helps to analyze the different frequencies of video signals at different resolutions for each frame. So, in this research, we have presented a video compression system based on 5-level Discrete Wavelet Transform (5-Level DWT) using the H.264 video compression technique. By using H.264 compression technique, data compression is used to reduce the redundancy in video data and the probability of compression is increased. For this purpose, 5-Level DWT is used and for best compression ratio achievement, we have used Kalman filter in pre-processing phase. The goal of 5-Level DWT based compression is to store video data in a little space, so, that the transmission process will be easy. After that, we have analyzed the performance metrics of proposed research work based on 5-Level DWT using H.264 compression technique and the performance metrics are Compression Ratio, MSE, PSNR, BPP and Execution Time.

Keywords- Video Compression, 5-level Discrete Wavelet Transform, Kalman Filter, H.264 Compression technique and Image Processing.

I. Introduction

Compression is a reversible transformation (encoding) of data that contains fewer bits and allows more efficient storage and transfer of data. The inverse process of compression is called decompression (decoding). The software and hardware that can be encoded and decoded are called decoders. Compression can be lossy or lossless [1]. Lossless compression reduces the number of bits by identifying and eliminating statistical redundancy as shown in figure 1. In the lossless compression of the data, there will be no loss of any information.

LOSSLESS

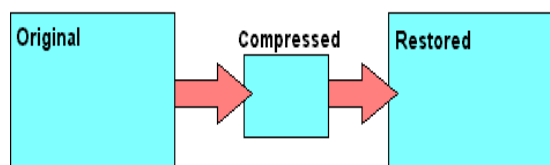


Figure 1: Lose-less Compression

Lossy compression reduces bits by removing the unwanted or less important information. The process

of reducing the data file size is called as data compression.

LOSSY

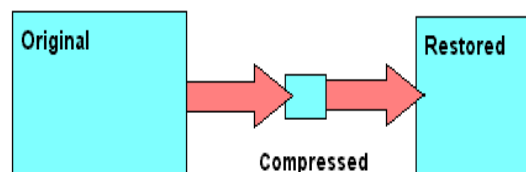


Figure 2: Lossy Compression

In this research, video compression system is utilized with 5-level DWT decomposition using H.264 video compression technique.

Video and images contain a large amount of redundant information. The main technique of video compression relies on discarding redundant information that the human eye is not sensitive to. When compressing video, there is a trade-off between video quality and the amount of compression achieved. If the compression achieved is high, the quality of the re-constructed video is low. User should also consider the delay of the compressed video [2]. Thus, a good compression

scheme is intended to achieve better compression while taking all the parameters into account. In today's world, video compression is a very important tool to save the amount of video used for storage space. The video just transmits a single frame / image at faster rate (typically 25 frames per second) movies). Instead of sending each frame one by one, the difference in the frame can be stored and transferred to reduce size and save time. This is the actual idea behind the video compression technology.

II. Video Compression

Video compression uses modern coding techniques to reduce redundancy in video data. Most video compression algorithms and codecs combine the spatial image compression and temporal motion compensation. Video compression is a practical implementation of the source coding in information theory. A variety of video standards are there which define the resolution and colors for display. For a PC, both the monitor and the video adapter determine the support for a graphics standard.

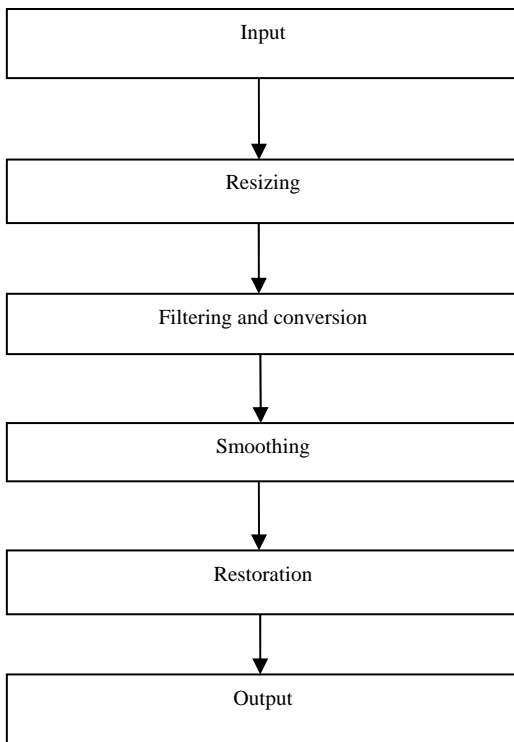


Figure 3: Pre-processing steps of video compression

Table 1 displays the uncompressed bit rates of some video formats. It can be clearly observed that even QCIF at 15 fps (i.e., relatively low quality video

suitable for video telephony) requires 4.6 Mbps for storage or transmission. Table 2 shows typical capacities of popular storage media and transmission networks.

Table 1 displays the uncompressed bit rates of some video formats. It can be clearly observed that even QCIF at 15 fps (i.e., relatively low quality video suitable for video telephony) requires 4.6 Mbps for storage or transmission. Table 2 shows typical capacities of popular storage media and transmission networks.

Table 1: Uncompressed bit rates of video formats

Video format	Color resolution	Intensity resolution	Bits per second (uncompressed)	Frames per second
QCIF	88 x 72	176*144	4.6Mbps	15
CIF	176 x 144	352*288	36.5 Mbps	30
ITU-R 601	429 x 525	858*525	216 Mbps	30

Table 2: Typical capacity of popular storage media and transmission networks

Media /network	Capacity
ADSL	Typical 1-2 Mbps (downstream)
ETHERNET LAN	Maximum 10 Mbps / Typical 1-2 Mbps
V.90 MODEM	56 kbps downstream / 33 kbps upstream
ISDN-2	128 kbps
CD-ROM	640 Mbytes
DVD-5	4.7-Gbytes

III. DWT (Discrete Wavelet Transform) Of 5-Level

The discrete wavelet transform is a valuable way for signal exploration as well as picture handling, chiefly in multi-resolution description. It can crumble the signal into different components in the frequency sphere. One-dimensional discrete wavelet transform (1-D DWT) decomposes an input into two components (the average component and the detail component). Two-dimensional discrete wavelet transform (2-D DWT) decomposes an input image

into four sub-bands, one average component (LL) and three detail components (LH, HL, HH). DWT has the characteristics of dividing the signal into low-frequency part and high-frequency part. The high frequency portion stores information about the edge and surface components whereas; low frequency portion is again divided into a high frequency portion and a low frequency portion.

H.264 Video Compression

This work considers the problem of video codec design and its solution until the evolution of the latest international standard known as H. 264 or MPEG-4 Advanced Video Coding (H. 264 / AVC). Video source coding Basal digital images or digital video frames are typically composed of three rectangular arrays of integer value samples, one for each of the three components of the three-color representation of the spatial region represented in the image. The video encoding software has three components called Y, Cb, and Cr. The component Y is called brightness and represents brightness.

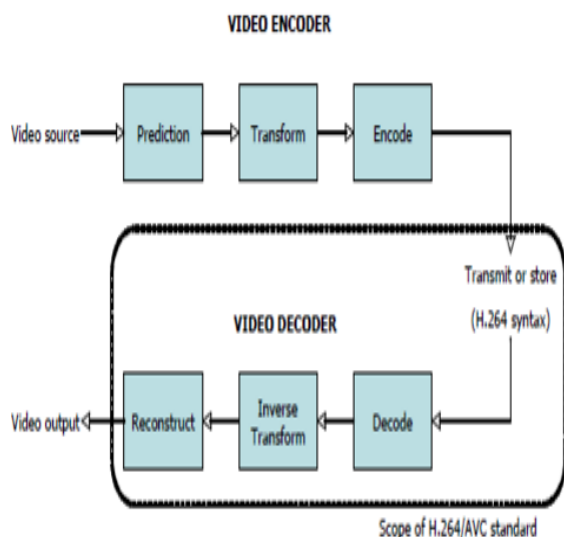


Figure 4: H.264 video encoding and de-coding process

The two color components Cb and Cr represent the degree of color deviation from gray to blue and red respectively. A frame array of video samples may be considered to contain two interlaced fields, a presence field, and a bottom field. The top contains the middle number line (0 is the top line number of the frame, the total number of rows), and the bottom field contains odd lines (starting at the second line of the frame). When using interlaced scanning, instead of capturing the entire frame at each sampling time, only one of the two fields is captured. Therefore, two

sample periods are required to capture each complete frame of the video.

IV. Related Work

Arif Sameh Arif et.al [2013], has proposed a technique for lossless compression of focused on parts inside of Fluoroscopy pictures, separating the region of interest (ROI) – for this situation the pharynx and esophagus, and utilizing modified relationship and the blend of Run Length and Huffman coding, to build compression proficiency. Neha Goel et.al [2015], concluded that data security and data authentication are achieved by applying watermarking techniques. Watermarking is a process that sensibly inserts information into digital data, with the ability to maintain the recoverability of digital information. Rajeshwar Dass et.al [2012], proposed that in the past three decades, video compression has been the object of in-depth study. Video compression technology is now mature and has been proven by the extensive use of this technology. This article gives ideas about the different technologies available for video compression. H. 264 / AVC showed better coding performance than its predecessor. T. Bernatin et.al [2014], concluded that in order to effectively deal with large amounts of data in video surveillance systems, efficient video retrieval techniques and advanced video compression techniques must be found. Soumaya et.al [2011], proposed an Object-based video compression method. It includes the movement of video objects throughout the scene. The neural network is used to perform the prediction step. A multi-step forward prediction is performed to predict the video object trajectory on the sequence. Mamta Sharma et.al [2010] has investigated Huffman calculation and contrasted it with other basic compression systems like Arithmetic, LZW and Run Length Encoding.

V. Simulation Work

Video compression reduces the amount of data required to represent a video by removing redundant information. Although, video looks like a continuous motion, but it is actually a series of several still images. The amount of space which is required for holding an uncompressed video, costs much more than a compressed one. Therefore, in this research work, we have compressed a normal video with the help of Compression technique and also applied Kalman filter in order to perform pre-processing on the Video. There are several steps involved in the process of video compression technique. These steps

are required to be followed in order to perform a video compression.

The methodology of the proposed work is briefly defined in the following steps:

- a) Design or develop a proper GUI of the proposed Video Compression Work.
- b) Develop a code for uploading a video for the compression from the database, on which testing will be performed.
- c) Apply Frame Extraction and Store the resultant frames in the database
- d) Perform Pre-Processing on the frames using Kalman Filter. In the pre-processing, the frames are processed in order to get the desired size and also it include color conversion. The Filtering of the frames is done by employing Kalman Filter.
- e) After the performing the pre-processing on the frames, 5-level DWT is performed on each pre-processed frames.
- f) After 5 levels DWT, the frame undergoes compression which is performed by using H.264 Process.

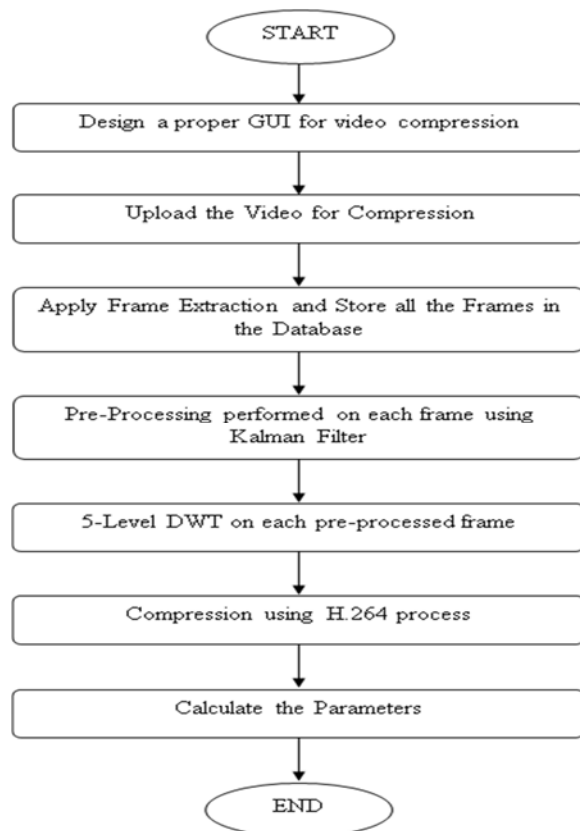


Figure 5: Proposed work Flowchart

- g) After the Compressing process of the frames the resultant video is compared with the actual video on basis of some computational parameters such as Compression Ratio (CR), MSE, PSNR, BPP and Execution Time

VI. Simulation Results

This section explains the results obtained after the implementation of the simulation work by using different parameters like Peak signal to noise ratio (PSNR), mean square error (MSE), compression ratio (CR), bit per pixel (BPP) and execution time (ET).



Figure 6: Main Figure Window

Above figure predicts the Main Figure window which comprises the title of our proposed work. This window has two push button named as START and EXIT. By clicking on START button a new working window appears for the further processing of our work, whereas, if we click on the EXIT button the Main Figure Window will shut close.

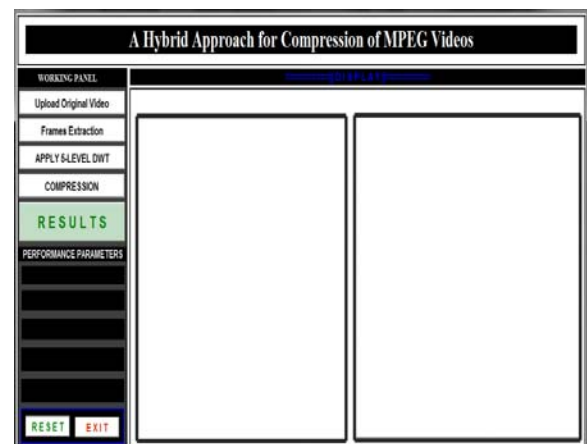


Figure 7: Working Figure Window

The above diagram shows the Working Figure Window. In the Working Window the proper GUI for

the video compression is designed and prepared for later use. On the Left side of this window, there is a working panel. This panel includes the further steps of our proposed work in the form of buttons. Each button performs differently in a proper sequence which will be defined later.



Figure 8: Uploading a Video

The above figure shows the window where the video is uploaded for compression process. For uploading a video click the upload original video button in the working panel present on the left of the working window. Then select the desired video on which the compression has to be performed.



Figure 9: 5-Level DWT Decomposition

After the uploading of a video, the frames are extracted from that particular window and stored in the database. Then 5 level DWT is applied on each frame that are extracted from the video. For applying the 5 level DWT, click on the 5-level DWT button present in the working panel on the left of the working window, as shown in the above figure.



Figure 10: Video Compression using H.264 Compression Technique

After applying the 5-level DWT on the extracted frames of the video, H.264 Compression technique is employed in order to compress the pre-processed video. This process is applied by selecting the compression button present in the working panel on the left side of the working window.



Figure 11: Result Window

The above figure represents the result window. After the completion of compression process, the resultant video is compared with the original video in order to check the performance on the basis of the performance metrics such as Compression Ratio (CR), Mean square Error (MSE), Peak Signal to Noise Ratio (PSNR), Bit per Pixel (BPP), Execution Time (ET). The Results of the performance metrics is

shown on the left side of the result window under the performance parameters.

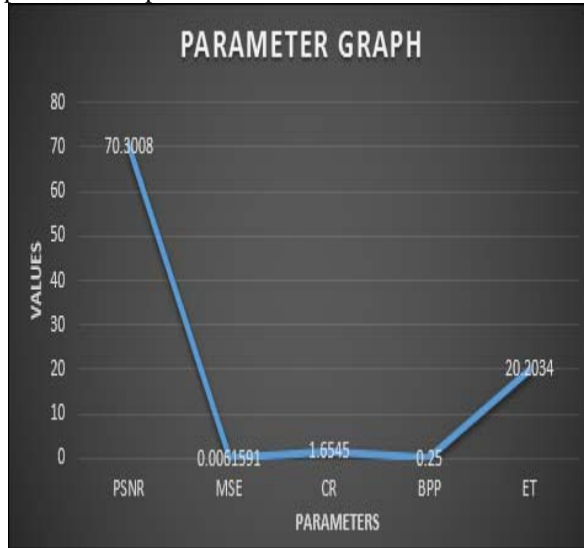


Figure 12: Comparative analysis w.r.t to different parameters

Table 1: Different parametric values

PARAMETERS (In ratio)	VALUES
Peak Signal to Noise Ratio (PSNR)	70.3008 (dB)
Mean Square Error (MSE)	0.0061591
Compression Ratio (CR)	1.6545
Bit Per Pixel (BPP)	0.25
Execution Time (ET)	20.2034 (secs)

Above figure is explaining the comparison graph of the parameters undertaken. It has been seen that the PSNR for the proposed work is 70.30 which is much better than the previous work. The value for MSE is 0.0061 which is less and it should be less for the reliability for the work being simulated. The compression ratio for the video is 1.6 whereas the BPP is 0.25 and the time taken for execution is 20.20 secs. X-axis defines the parameters taken and y-axis is explaining the values obtained after the simulation.

VII. Conclusion

In this research, we have presented a video compression system based on 5-level Discrete Wavelet Transform (5-Level DWT) using the H.264 video compression technique. By using H.264 compression technique, data compression is used to reduce the redundancy in video data and the probability of compression is increased. For this purpose, 5-Level DWT is used and for best compression ratio achievement, Kalman filter in pre-processing phase is being presented. Pre-processing phase is the most important phase for video compression system. The main goal of 5-Level DWT based compression technique is to store video data in a little space so that the transmission process will be easy. In the final phase, we are analyzing the performance metrics of proposed research work based on 5-Level DWT using H.264 compression technique and we observed that the Compression Ratio is near about 1.6, MSE is low and the average value is 0.006, the average PSNR is more than 70 dB, BPP is 0.25 and the Execution Time of video compression system is less than one minute. In future, video compression technique would be presented using the optimization technique to optimize the 5-Level DWT frequencies to achieve best compression ratio and PSNR.

References

1. Arif Sameh Arif, Rajasvaran Logeswaran, Sarina Mansor and Hezerul Abdul Karim, "Segmentation and Compression of Pharynx and Esophagus Fluoroscopic Images", In 2013 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 220-225.
2. Asaad Hakeem, Khurram Shafique, and Mubarak Shah. "An Object-based Video Coding Framework for Video", MM'05, November 6-11, 2005, Singapore. Pages 608-617
3. Bahadır Gunturk, K., Yucel Altunbasak, and Russell Mersereau. "Bayesian resolution-enhancement framework for transform-coded video." In Proceedings the IEEE International Conference on Image Processing, October-2001, vol. 2, pp. 41-44.
4. B. K. Gunturk, Yucel Altunbasak, and R. M. Mersereau, "Multi-frame Resolution-Enhancement Methods for Compressed Video", IEEE Signal Processing Letter, Vol. 9, No. 6, 2002, pp. 170-174.
5. Christoph H. Lampert, "Machine Learning for Video Compression: Macroblock Mode

- Decision”, In the 18th International Conference on Pattern Recognition (ICPR), 2006, Vol. 1, pp. 936-940.
6. Chung-Ming Kuo*, Chaur-Heh Hsieh*, Yue-Dar Jou*, Hsieh-Cheng Lin, Po-Chiang Lu, “Motion Estimation for Video Compression Using Kalman Filtering”, IEEE Transactions on Broadcasting 42, no. 2, 1996, pp. 110-116.
 7. Gary J. Sullivan and Thomas Wiegand, “Video Compression—From Concepts to the H.264/AVC Standard”, In the Proceedings of the IEEE, 2005, Vol. 93, No. 1, pp. 18-31.
 8. Jeffrey Ho, Kuang-Chih Lee, Ming-Hsuan Yang, David Kriegman. “Visual Tracking Using Learned Linear Subspaces”. Proc of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’04
 9. Ke Shen, “Video Preprocessing Techniques for Real-time Video Compression - Noise Level Estimation Techniques”, In IEEE 42nd Midwest Symposium on Circuits and Systems, 1999, Vol. 2, pp. 778-781.
 10. Mamta Sharma, “Compression Using Huffman Coding” In (IJCSNS) - International Journal of Computer Science and Network Security, Vol. 10, No. 5, May 2010, pp. 133-141.
 11. M. Atheeshwari, “Video Compression Techniques – A Comprehensive Survey”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, No. 1, January 2014, ISSN: 2277 128X.
 12. Neha Goel, Vinith Chauhan, Krishna Mohan Rai, “Study of 5 Level DWT and Comparative Performance Analysis of Digital Video Watermarking Techniques Using 3-L-DWT and 5-L-DWT”, In International Journal of Engineering and Technical Research (IJETR), ISSN: 2321-0869 (O) 2454-4698 (P), Volume-3, Issue-7, July 2015, pp. 331-338.
 13. Soumaya ghobrel, “Object-based Video compression using neural networks”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011 ISSN (Online): 1694-0814.
 14. Rajeshwar Dass, Lalit Singh, Sandeep Kaushik, “Video Compression Technique”, In the International Journal of Technology Enhancements and Emerging Engineering Research Vol. 1, No. 10, 2012, pp.114-119.
 15. Santanu Chaudhury, Subarna Tripathi, Sumantra Dutta Roy, “Parametric Video Compression using Appearance Space”, In IEEE 19th International Conference on Pattern Recognition (ICPR), 2008, pp. 1-4.
 16. S. Bhattacharya, T. Chattopadhyay, and A. Pal, “A survey on different video watermarking techniques and comparative analysis with reference to h. 264/AVC,” in IEEE Tenth International Symposium on Consumer Electronics, ISCE’06. IEEE, 2006, pp. 1–6.
 17. T. Bernatin and G. Sundari, “Video Compression Based on Hybrid Transform and Quantization with Huffman Coding for Video Codec”, In IEEE International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014, pp. 452-456.
 18. T. Hu and J. Wei, “A digital video watermarking scheme based on 1d-dwt,” in International Conference on Biomedical Engineering and Computer Science, april 2010, pp. 1 –3.
 19. Vinayak K Bairagi and Ashok M Sapkal, “ROI-based DICOM image compression for telemedicine” In 2011, Digital Image Processing, Vol. 3, No. 11, pp.662-666.
 20. W. Lu, H. Lu, and F. Chung, “Robust digital image watermarking based on subsampling,” Applied mathematics and computation, vol. 181, no. 2, pp. 886–893, 2006.
 21. X. Xia, C. Boncelet, and G. Arce, “A multiresolution watermark for digital images,” in International Conference on Image Processing, vol. 1. IEEE, 1997, pp. 548–551.
 22. ZhinoosRazaviHesabi, Mohsen Sardari, Ahmad Beirami, FaramarzFekri, Mohamed Deriche, and Antonio Navarro, “A memory-assisted lossless compression algorithm for medical images” 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2030-2034

AUTHORS PROFILE

1. Kamupriya, student of M.tech in Chandigarh University.
2. Silky Narang, Assistant Professor at Chandigah University.

Comparative performance analysis of Ad hoc routing protocol

Muhammad Usman Nasir

Department of Computer Sciences,
Pir Mehr Ali Shah, Arid Agriculture University
Rawalpindi Pakistan
usman_7670@hotmail.com

Yawar Abbas Abid

Department of Computer Sciences,
COMSATS Institute of Information Technology
Sahiwal Pakistan
yawar.abid@ciitsahiwal.edu.pk

Abstract—In this paper a rapid assessment and relative study of the adhoc routing protocols usage in MANET will be discussed. The exploration filed in Vehicular adhoc network (MANET) is emerging rapidly. There is a wide range of uses has been worked in variety of circumstances (highway, town and city). There are lot of protocols implemented to facilitate various types of topology and scenarios, these protocols are facing various kind of difficulties. MANETs gives statement among vehicles moving on roads. Vehicles fast speed can affect routing protocols In MANET that leads towards the regularly connections among the interconnected vehicles, so that to provide the information among vehicles in petite time MANET characteristics are adopted by adhoc routing protocols. The key objective of MANET is to construct a data network between vehicles which are running on the roads. Assemble a data system amongst vehicles that are moving on the roads that supports the vehicles to interconnect by individually additional for the protection measures. An exertion has been ended to differentiate six certainly implicit protocols ZRP, DSR, AODV, OLSR DSDV AND AODV. DSR, which are Used in MANET after the (traditional packages, end to end delay, packet deliver ratio, dropped packets, throughput, Average Jitter routing algorithm load, routing packets, ratio of packet loss,) facts.

Keywords—MANET, MANET, OMNET++, SUMO, DSR, DYMO, OLSR, DSD, ZRP, AODV.

I. INTRODUCTION

From the past two decades there has been a massive innovation happening in the area of wireless Adhoc networks. This invites a researcher in this field to do research and provide some other opportunity such as auspicious universal connectivity outside that is providing through the internet. The "mobile ad hoc networks" (MANETs) is a network that has no defined infrastructure and configures moving routers itself, linked through wireless frequencies. There's nonexistence of infrastructure because this is a fixed network like a mobile nodes, path also be change between the nodes, and because of all these constraint we can't apply the techniques of MANET on fixed network.

Optimized Link State Routing Protocol (OLSR) is a MANET is an adhoc network that have a no infrastructure and have the ability to increase their size by self. Even for connecting the modern devices like mobile and other using the

wireless channels for communication. They used the power each info needed to accurately route traffic [1]. The Vehicular Ad-Hoc Network, or MANET, is a superior kind of MANET that allows sending-data between nearby vehicles to vehicles and close the secure edge unit (RSU) from road-to-vehicle communication (V2I).

It is presented in fig.1 this communication help to avoid any mishap like accidents, serious injury, and traffic congestion to allow vehicles to communicate and distribute information safely to further nodes in MANETs to alarm drivers and provide information about the weather report, circulation and the actual period update [2]. MANETs give the opportunities for the numerous new solicitations that's help the user in real time.

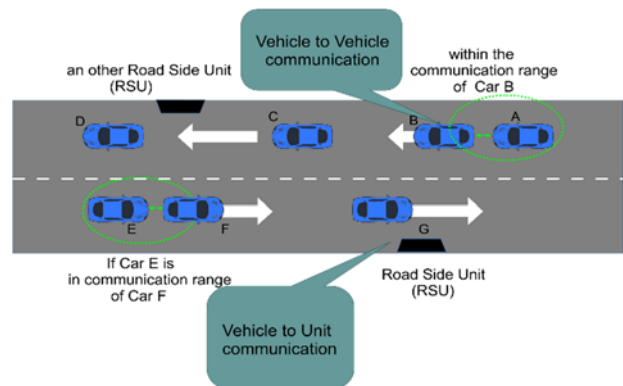


Fig. 1 MANETs Network

II. MANET ROUTING PROTOCOLS

In the vehicle to vehicle sending information is single the essential encounters in the enterprise of MANET as it designs like a dynamic routing protocol. When compare the routing of MANET with MANET, routing is different due to some exceptionally dynamic topologies.

The core point is to overcome the latency when travel the data V2V. In the routing perspective the MANET can be consider interested in about main category as shown in the fig. 2 [3]

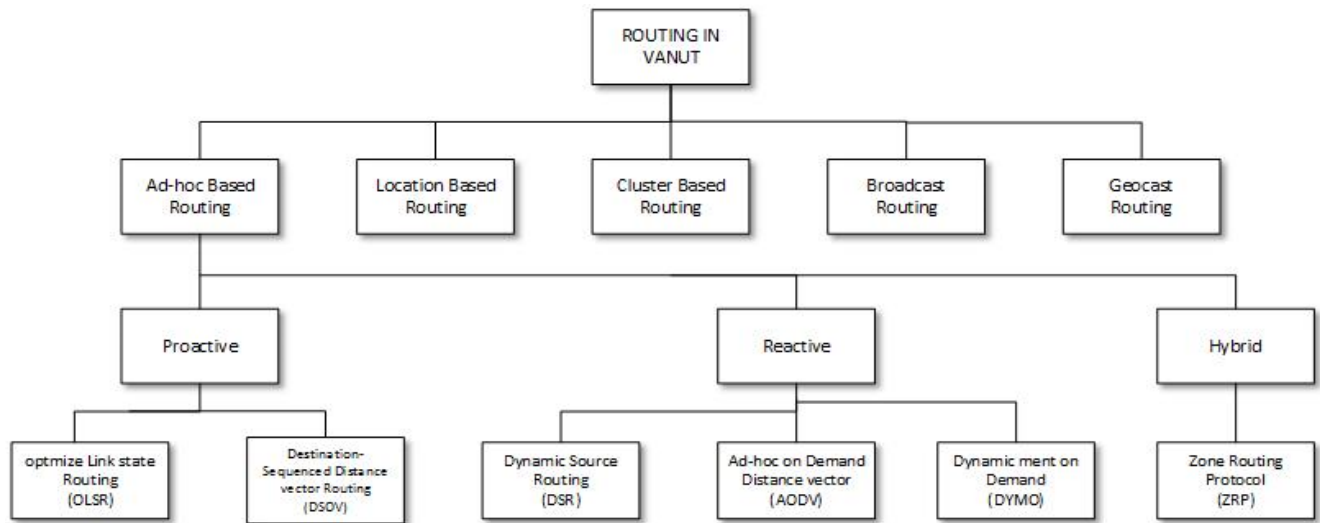


Fig. 2 MANET Routing Types

A. Adhoc Routing Protocols in MANET

Just like the other protocol the protocol for the routing is a level that control how the hubs select best tactic to track packet that coming from the devices in a domain of wireless and difference in various diversities. The protocol of routing is one for the encounters of MANET. Routing protocol performance for the MANET affected by the some internal and external factors. In the internal factors involve the highly bitty network, network commotion and vehicle mobility. Like the internal factors, external factor are environmental, roads traffic rules and the hindrances like trees and building. A competent protocol for routing is the single one protocol that send the packet of data in no time and not cost much amount of network bandwidth. The protocol of Adhoc network, uses information obtained about the link that was built when the packet send inside the network to the target hub from the basis hub. These characterized in three types [4]:

1. Proactive Routing

Proactive routing implies that routing information, such as the next transfer hop, is kept in the background despite communication requests. This type of protocol periodically exchanges the topological information between all the nodes of the network. As a result, the proactive routing protocol does not have a route discovery since the destination route is saved and sustained in a table and works effectively with the application that requires low latency. However, the drawback of this protocol is that it gives little sluggishness to the constant application [5]. An example of this protocol is the Optimized Link State Routing Protocol (OLSR) and the Destination Sequenced Distance Vector (DSDV).

2. Optimized Link State Routing (OLSR)

OLSR is the unique routing protocol used anywhere the courses area unit repeatedly offered once needed. In link state protocol the OSLR is the developed version of link state

protocol. The modification in the network will be lead to the change in topology of network, otherwise this thing go to problem in the network. For the shortest path finding OLSR usages Multipoint Relay (MPR) that is liable to promote the knowledge of link state and forwards the control massages periodically to get the smallest path. OLSR has two varieties of control messages referred to as hello and topology control. Hello messages area unit used to trace info regarding a host's neighbors and connection status. This protocol is effective for dense and larger wireless Adhoc networks [6].

3. DSDV

A modification in the standard Bellman-Ford algorithm, has been named as the Destination sequenced Distance-vector Routing (DSDV) protocol. This practical routing protocol is tabular in nature. In DSDV, every node retains an associated degree access of the table. It dress two main things: the Situation symbol of a target which is the concise parting metric to that particular destination and the Location symbol of the hub which is the shortcut to the destination.

B. Reactive Routing

These protocols work if and only if communication is required between two particular nodes. That is why they are sometimes known as on-demand routing protocols, which makes the use of periodic-flooding unnecessary. Examples include: DYMO, AODV, DSR protocols.

1. DSR (Dynamic supply Routing protocol)

This reactive routing protocol is typically designed to work in a multi-hop wireless adhoc-networking environment. It has a self-helping behavior which makes it not needing any existing system base. Course maintenance and Course disclosure are its two basic parts that enable the revelation and support of foundation passages in system [7].

2. Ad-hoc On-Demand Distance Vector (AODV)

This responsive routing protocol in which nodes behave as routers at the time of communication. AODV facilitate the regular steering tables, it have the unique section for the target. One of its main advantages is that it reduces the size of the overhead and uses the information measure with efficiency by decreasing the consumption of management messages [8].

3. Dynamic painter on Demand (DYMO)

This responsive routing protocol that has a little memory to

save and update the routing information. On receiving a data packet, a node creates the management packet. This protocol is designed for spontaneous networks that suffer from common disconnection. It works best in networks with thin traffic and enormous variety of nodes situations. The main difference in AODV and DYMO is that, AODV simply production the section of routing table for target hubs and following jump, whereas DYMO have the extra paths for each central road bound and the destination hub as shown in “Fig. 3” [9]

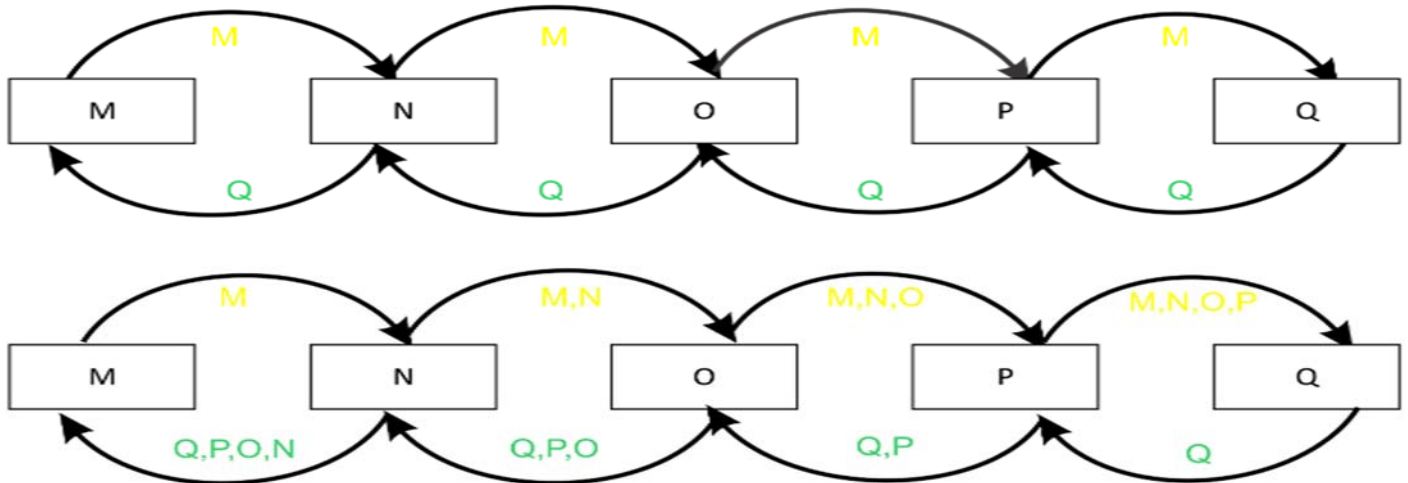


Fig. 3 Differentiate among DYMO and AODV

C. Hybrid Routing

Hybrid Routing connect close active and inclusive responsive routing protocols to composed and eliminate routing overhead. And stay because way to disclosure procedures. The pros of these protocols are larger effectiveness and capacity. Conversely, the cons exist is larger potential for detecting novel paths. Here are some of examples x Zone Routing Protocol (ZRP)[10].

Zone Routing Protocol keen on a hybrid strategy, captivating some advantage of active disclosure within a hub close and evaluating a responsive protocols for communicating among these capacities. Commonly these routing are area centered; the sum of node is distrusted into diverse areas for the purpose of creation of location and maintenances additional steadfast. Proactive approaches used for this protocols between near node functioning locally ; but, reactive routing is doing work worldwide to lactate the demanded nodes by introducing the request system somewhat than spreading the enquiry to each nodes in network . Elastic path detection and path upkeep usages “Intrazone” and “Interzone” routing. Interzone routing accomplishes worldwide route detection over sensitive transmitting protocol however intrazone navigation in sight of proactive guiding observance in concentration the conclusion objective to retain the development info nearby within of its particular close. The overall usual for ZRP is that it reduces the system above that is conceded on by dynamic foremost and it furthermore grasps the arrangement concede that is produced by receptive navigation agreements and achieve sequence expose all the additional efficiently [10].

III. RELATED WORKS

In 2006, Haerri et al [11] for the first time compared the enactment assessment of AODV, OLSR in MANETs, giving a concept of smart cities. He initiates the idea by settling traffic rate at the experiment spot is 0.2 to 2 Mbps, and gave vehicle's numbers 40-80 moving at a speed of 72km/s to 126km/s. The evaluation criteria was based on two parameters, data packets successfully delivered and end to end delay for a transmission. It was concluded by him that:

As OLSR exhibits lowermost end-to-end delay for transmission and directing slide.

For OLSR, packet delivery ratio, was less at low density than AODV, and then starts increasing at high density.

In 2009, Khan and Qayyum [12], examined the implementation of AODV and OLSR reactive protocols using NS2 simulator opposing the amount of hubs to 30, 50, 70, 90 and 120 vehicles, through usage of probabilistic Nakagami radio propagation model in smart city concept. The work designed mobility model in SUMO circulation simulator. The evaluation metrics dealt with the ratio of successful packets delivered and end to end transmission delay. It was observed that AODV outperforms OLSR.

In 2009, Ferreior-Lage et al[13], assessed the effectiveness of present Adhoc network routing protocols in NS2 and SUMO

mobility nodes simulators, by relating their on Vehicle-to-Vehicle networks to get the routing limitations in MANETs. They used two different scenarios, one is for cities with vehicle rapidity 50 km/s and the other is for freeways with vehicle rapidity 80 km/s. Three protocols named DSR, DSDV and AODV with altering amount of hubs (among 10, 20 and 40 nodes) had been tested for rerouting load, end to end transmission delay, successful ratio of packets delivered, and dropped packets ratio and the conclusion was :

a) It was revealed that AODV showed improved results than above mentioned routing protocols as the number of nodes increase in MANETs.

In 2011, Singh et al[14], showed a comparative examination of OLSR , AODV and DSR adhoc network routing protocols in city situations and highway situations by setting different amount of moveable vehicles. They designed the solution in NS-2.34 for vehicles from 20, 102, 254 and 610 in smart city concepts to 152, 340, 767 and 1216 for highway scenario and experiential package distribution proportion and end to end transmission delays. They concluded that:

The AODV is greater for effective package transfer proportion than further two routing protocols but it shows maximum end-to-end delay that's why not appropriate for transmission info which could be brought on interval.

The DSR achieves good speedy broadcasts but is not appropriate due to larger packages damage.

The OLSR exhibits midway enactment, so is not appropriate to be acceptable for speedy scenarios.

Prokop, in 2011 [15] established an easily obtainable simulator, to formerly assess these mobile Adhoc network routing protocols for vehicular Adhoc networks. These have been selected by AODV, DYMO and DSR. They have also been assessed in city and highway county settings using SUMO as a circulation simulator and OMNET++ as network simulator protocols.

The comparison was made w.r.t end to end transmission delay, successful delivery of packets and they concluded:

a) DYMO performance is high in case of throughput and minimization of end-to-end delay as compared with other protocols.

In [16] Jogendra Kumar et.al worked on performance comparison for MANETs considering different routing protocols DSR, DYMO, OLSR, ZRP and AODV in variable pause time. Performance metrics considered were, end to end transmission delay, total bytes received, successful packets received, average jitter, last packet received, and first packet received.

DSR demonstrations finest presentation in relations of total data received in bytes.

ZRP ishows finest strategies for primary package received, former piece received. Though, there was bad performance in case of transmission delay.

AODV achieves well for usual jitter and end to end transmission delay is disturbed.

In 2012, Evjola Spaho and Makoto Ikeda[17], executed DSDV and DYMO. This was an experimentation for MANETS and the evaluation made was rendering to two parameters: throughput and successful packets delivered. The result showed:

DSDV protocol consist on lesser successful packets delivered in use of a stale sequences if an increase of disconnected connections happen.

DYMO on the other hand showed satisfactory results when compared with DSDV.

In 2012, uma Nagaraj and Poonam P. Dhamal [18], gave analysis on constraints such as successful packets delivered and transmission delay for AODV, DSDV, OLSR, and DSR. They concluded that :

AODV contains an excellent enactment in expressions of packet delivery ratio of around 98% and DSDV showed around 97%.

OLSR consist on usual performance compared with packet delivery ratio

AODV consists of transmission delay.

Maltz D. et al [19]., experimented that in deployment of multi-hop mobile nodes, the reactive routing protocols projected good results. The study describes the impact of routing caching correctness, overhead price, and forward latency. A performance evaluation by varying number of nodes among reactive routing protocols is studied and all the transmission factors or trade-offs observed are being discussed.

Perkins C. et al. [20], has experimented two reactive wireless Adhoc routing protocols named, AODV and DSR and observed results in a different way. He concluded results for avg. delay of data transmission, successful packet delivery quantitative ratio, and routing overhead.

Gupta S. et al. [21], projected an advanced version of DSR routing protocol. Traditionally this routing protocol exhibits good successful transmissions as long as starting and sink nodes are placed within one network. However as soon as, starting and sink nodes get separated, the successful transmissions becomes unsuccessful. He experimented that putting these two nodes in different networks makes it impossible to do communication using DSR. Therefore there comes a need to design a new rule named "Smart DSR Protocol", abbreviated as "SDSR Protocol" to overcome this issue. This new design constructed a sensible node that act as external router to transfer the information between two acting networks. Other additional features are SRREQ and SRREP are added for that external router.

Velagaleti S. et al. [22], experimented SDSR routing protocol in NS-2 simulator for one thousand nodes. They experiment is a performance evaluation of SDSR with DSR and with AODV. Successful packet transmission, end-to end delay and data packets loss is observed and it was concluded

that SDSRE outperforms DSR with successful transmissions as the network size increases.

Al-Maashri A. et al. [23], has experimented an analysis of various wireless Adhoc routing protocols for MANETs. It is highly observed that some new routing protocols are designed and implemented with an idea of cosmic microwave background radiation or Poisson traffic models. Wireless adhoc routing protocols being implemented on those models are AODV, DSR, and OLSR to conclude best working evaluation for routing overhead, end to end transmission delay and packet transmissions. For above described particular case, it was observed that DSR outperforms AODV and OLSR for peak traffic loads. The end to end delay was much smaller for DSR for self-similar traffic.

Grossglauser M. et al. [24], observed that routing in massive MANETs is extremely difficult because each node is mobile. However it is partially easy for geographical routing because nodes can make native routing decisions using sending and receiving nodes' geographical coordinates. However there may exist need for associated degree of economical location services. It was analyzed that a node can broadcast sink location without a certain amount of overhead. This thing was ensured by having native information interaction time and geographical coordinates of each node being interacted. So a new algorithmic program, EASE (Exponential Age Search) was written and experimented. During the experimentation, wherever nodes did freelance stochastic process for an n sized sq. lattice, the algorithm calculated routes lengths for destination and source nodes. Primarily this algorithmic program relies upon the LER algorithmic program to calculate the routes by observing last encounters made.

Mohapatra S. et al. [25], experimented the behaviour of Adhoc routing protocols on NS2 simulator for some performance evaluation features. This experiment was done for AODV, DSR OLSR and DSDV to compare successful ratio, delay, and the route management overhead. It was concluded that DSR offers higher successful transmissions for smaller sized network upto 600*600sqm, as compared to AODV, DSDV and OLSR. packet delivery quantitative relation than different routing protocols. However for network size larger than this, OLSR outperforms all other above mentioned protocols in terms of high throughput and lowest delay in transmissions.

Sharma L. et al[26]., used network technology named Exata/Cyber 1.2 and observed the comparative results of DSR, AODV, and DYMO adhoc routing protocols simulations in MANETs . The evaluation criteria was set as turnout, jitter and transmission delay package distribution ratio. The simulation concluded that for PDR and end to end delay AODV did well and turnout result for DSR was better while DYMO showed worst results.

Boukerche A. et al [27]., implemented a new version of reactive routing protocol DSR using GPS system and named it GDSR. In this case, route paths may be a major issue as a result of often topology changes because of node quality in mobile networks. Opposite to DSR, GDSR routing protocol accommodates route request only for nodes that are distant from node being requested. This algorithmic program, is

simulated and results are observed in comparison with DSR. While on the other hand there is a limitation for successful routing in this protocol and it is a critical setting of GPS screening angle. As this protocol only sends route request for distant nodes, the overhead is much less than DSR and can be deployed in heavy networks.

IV.CONCLUSION

In the paper analyses the literature relating to some major routing protocols, AODV, DSR, OLSR, DYMO, ZPR and DSDV is used for data routing in MANET. It is specially designed for high density MANETs having high mobile nodes in addition to diverse topology and road map and scenarios.

According to literature different simulators are used i.e. OMNET++, NS2, and SUMO to evaluate routing protocols for different QoS parameters like ratio of packet loss, routing load, package provide percentage, from side to side put directing packages, end to end delay , dropped packets, Average Jitter, received packets.

These evaluations are accomplished as shown.

- OLSR and AODV
- DSDV ,AODV and DSR
- DSR, OLSR and AODV
- AODV, DYMO and DSR
- DYMO and DSDV.
- AODV, DSDV, DSR and OLSR
- AODV, DSDV and OLSR

In conclusion, we precise the evaluation assessment for the protocol of MANET in paper and has been revealed in “table I.” and “Fig. 4”.

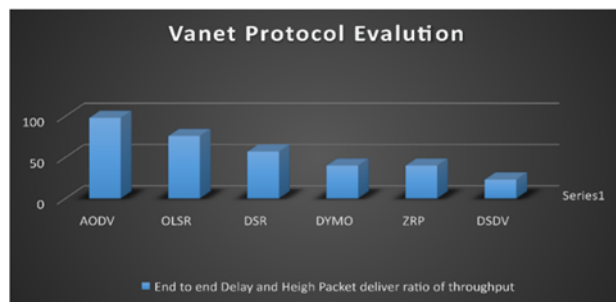


Fig. 4. Evaluation of VANET Protocols

FIG. 4. VANET PROTOCOL S PERFORMANCE EVALUATION

Protocols	Simulation	Results
AODV and OLSR	NS2	AODV higher as the density increases, and OLSR higher at low density.
DSR, AODV, OLSR and DSDV	NS2	AODV has shown superior performance than other
AODV, DYMO and DSR	OMNET++	The DYMO is the higher
DSDV and DYMO	OMNET++	DSDV superior
OLSR and AODV	NS3	OLSR is the superior
ZRP, OLSR and AODV	NS2	ZRP is superior to the others

Reference

[1] P. Panchal ,and M. Shroff, International Journal of Students’ Research in Technology & Management Vol 2 (05), August-September 2014, ISSN 2321-2543, pg 183187.

[2] K. C. Lee, U. Lee, and M. Gerla, “Survey of Routing Protocols in Vehicular Ad Hoc Networks,” in Advances in Vehicular Ad-Hoc Networks: Developments and Challenges, M. Watfa. IGI Global, 2010, ch. 8, pp. 149– 170.

[3] F. Li, and W. Yu, “Routing in Vehicular Ad Hoc Networks: A Survey,” Vehicular Technology Magazine, IEEE, vol. 2, no. 2, pp. 12–22, 2007.

[4] A. Indras, R. Murali, and Venkatesh, "Routing Protocols for Vehicular Adhoc Networks (MANETs): A Review ", Journal of Emerging Trends in Computing and

[5] M. Erritali, and B. EL Ouahidi, “Performance Evaluation of Ad hoc Routing Protocols in MANET”,

[6] K. Bala, and M. Sachdeva, “ENHANCEMENT OF OLSR ROUTING PROTOCOL IN MANET” CSE Department, SBS CET Ferozepur, Punjab (INDIA), January 23-24, 2016.

[7] M. B. Yassein, and A. Y. Al-Dubai, “Inspired Counter Based Broadcasting for Dynamic Source Routing in Mobile Networks” Department of Computer Science, Jordan University of Science and Technology, Irbid, Jordan, School of Computing Edinburgh Napier, University Edinburgh EH10 5DT, UK, 27 October 2015.

[8] A. Chandra, and S. Thakur, “Qualitative Analysis of Hybrid Routing Protocols Against Network Layer Attacks in MANET “Apoorva Chandra et al, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.6, June- 2015, pg. 538-543.

[10] N. Kumar, K. Vashishtha, and K. Babu, “A Comparative Study of AODV, DSR, and DYMO routing protocols using OMNeT++ “International Journal on Recent and Innovation Trends in Computing and Communication, September 2013. Third International Symposium on Automatic Amazigh processing, Vol. 3, Issue 2, pp: 33-40, 2013.

[11] R. Kumar, and M. Dave, “A Comparative Study of Various Routing Protocols in MANET” IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011.

[12] I. Khan, and A. Qayyum, “Performance Evaluation of AODV and OLSR in Highly Fading Vehicular Ad hoc Network Environments”, IEEE, Proceedings of the 13th international Multitopic conference, Islamabad, Pakistan, pp:1–5, 2009.

[13] J. Ferreior-Lage, C. Gestoso, O. Rubino, and A. Agelet, “Analysis of Unicast Routing Protocols for MANET”, Fifth International Conference on Networking and Services, Valencia, Spain, 2009.

[14] P. Singh, K. Lego, and T. Tuithung, “Simulation based Analysis of Adhoc Routing Protocol in Urban and Highway Scenario of MANET”, International Journal of Computer Application, Vol. 12, Issue 10, pp: 42 - 49, 2011.

[15] M. Prokop, “Routing Protocol Evaluation Development of a Fully Functional Simulation Environment for Vehicular Ad Hoc Networks”, M.Sc. Thesis, Rochester Institute of Technology Kate Gleason College of Engineering, New York, USA, 2011.

[16] J. Kumar, ” Comparative Performance Analysis of AODV, DSR, DYMO, OLSR and ZRP Routing Protocols in MANET Using Varying Pause Time”, (IJCCN), 2(3), pp 43–51 ,2012.

[17] E. Spaho, and M. Ikeda, “Performance Comparison of DSDV and DYMO Protocols for Vehicular Ad hoc Networks” 2012 26th IEEE International Conference on Advanced Information Networking and Applications.

[18] U. Nagaraj, and P. P. Dhamal, “Performance Evaluation of Proactive and Reactive Protocols in VANET”, International Journal of Information and Education Technology, Vol. 2, No. 5, October 2012.

[19] A. Moravejosharieh, H. Modares, R. Salleh, and E. Mostajeran, “Performance Analysis of AODV, AOMDV, DSR, DSDV Routing Protocols in Vehicular Ad hoc Network”, Research Journal of Recent Science, Vol. 2, Issue 7, pp: 66-73, 2013.

[20] David A. Maltz, Josh Broch, Jorjeta Jetcheva, and David B. Johnson, Member, IEEE, “The Effects of On-Demand Behavior in Routing Protocols for Multihop Wireless Ad Hoc Networks,” Ieee Journal On Selected Areas In Communications, VOL. 17, NO. 8, AUGUST 1999, pp-1439-1453, doi: 10.1109/49.779925.

[21] Charles E. Perkins, Elizabeth M. Royer, Samir R. Das and Mahesh K. Marina, “Performance Comparison of Two On-Demand Routing Protocols for Ad Hoc Networks,” IEEE Personal Communications • February 2001.

[22] Anupam Baliyan & Sanjeev Gupta, “Proposed Smart DSR(S-DSR) Protocol for Ad Hoc Network,” Global Journal of Computer Science and Technology Network, Web & Security Volume 12 Issue 10 Version 1.0 May 2012.

[23] Sessa Bhargavi Velagaleti, Dr.M.Seetha, Dr.S.Viswanadha Raju, “A Simulation And Analysis Of Secured S-DSR Protocol In Mobile Ad Hoc Networks,” International Journal of Computer & Organization Trends – Volume 3 Issue 9 – Oct 2013, pp-425-430.

[24] Ahmed Al-Maashri, A. and Mohamed Ould-Khaoua, M. (2006) “Performance analysis of MANET routing protocols in the presence of self-similar traffic,” In, Proceedings of the 31st IEEE Conference on Local Computer Networks, 2006, 14-16 November 2006, pages pp. 801-807, doi: 10.1109/LCN.2006.322040.

[25] Matthias Grossglauser, Member, IEEE, and Martin Vetterli, Fellow, IEEE. “ Locating Mobile Nodes With EASE: Learning Efficient Routes From Encounter Histories Alone”, IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 14, NO. 3, JUNE 2006, pp-457-469, doi: 10.1109/TNET.2006.876204.

- [26] S.Mohapatra, P.Kanungo , “Performance analysis of AODV, DSR, OLSR and DSDV Routing Protocols using NS2 Simulator,” International Conference on Communication Technology and System Design 2011 ,SciVerse Science Direct, Procedia Engineering 30 (2012), pp69-76, doi:10.1016/j.proeng.2012.01.835.
- [27] Lakhn Dev Sharma, Nirmal Roberts, “Effects of Velocity on Performance of DYMO, AODV and DSR Routing Protocols in Mobile Ad-hoc Networks,” C3IT-2012, SciVerse Science Direct, Procedia Engineering 4 , pp727-731, (2012), doi:10.1016/j.proty.2012.05.118.
- [9] Yih-Chun Hu, Adrian Perrig, David B. Johnson, “Aridane: A Secure On-Demand Routing Protocol for Adhoc Network,” Springer Science + Business media, Inc., Wireless Networks 11, pp 21-38,(2005), doi: 10.1007/s11276-004-4744-y.

A Comparative Evaluation Concept for Agent Oriented Methodologies.

Omankwu, Obinnaya Chinecherem, Nwagu, Chikezie Kenneth, and Inyama, Hycient

¹Computer Science Department,
Michael Okpara University of Agriculture, Umudike
Umuhia, Abia State, Nigeria
saintbeloved@yahoo.com

² Computer Science Department, Nnamdi Azikiwe University Awka,
Anambra State, Nigeria,
Nwaguchikeziekeneth@hotmail.com

³Electronics & Computer Engineering Department,
Nnamdi Azikiwe University, Awka
Anambra State, Nigeria.

Abstract

Multiple agent-oriented methodologies were introduced in recent years, however no systematic evaluation of these was offered. In this work we perform a comparative evaluation of three well-known agent-oriented methodologies: Gaia, Tropos, and MaSE. To perform this evaluation we use an existing framework that focuses on four major facets of a methodology: concepts and properties, notations and modeling techniques, development process, and pragmatics. Analyzing the results of our evaluation, we recognize several facets that need further improvements within the existing agent-oriented methodologies. Our study does not attempt to state what the right methodology is. Rather, it examines existing agent-oriented methodologies.

Keywords: Gaia, Tropos, MaSE, MAS, agentTool

1. Introduction

During the last decade, many methodologies for developing agent-based systems have been developed. A methodology is the set of guidelines for covering the whole lifecycle of system development both technically and managerially. A methodology, according to (Graham et al., 1997), should provide the following: a full lifecycle process; a comprehensive set of concepts and models; a full set of techniques (rules, guidelines, heuristics); a fully delineated set of deliverables; a modeling language; a set of metrics; quality assurance; coding (and other) standards; reuse advice; and guidelines for project management. The relationships between these components are shown in Figure 1.1. In that figure, we use the UML notations to depict the relationships between the components. As depicted in the figure, a methodology consists of a set of techniques, a modeling language and a lifecycle

process. The set of techniques consists of metrics, quality assurance (QA) activities, a set of standards and tools. The modeling language comprises notations and a meta model. The lifecycle process consists of project management, a number of roles (e.g., an analyst or a designer), a number of procedures (e.g., how to move between development stages), and a number of deliverables (e.g., a design document, source code). In addition, Figure 1.1 shows that the tools should be based on the meta model of the modeling technique and should represent the modeling technique's notations. The deliverables should use the modeling technique.

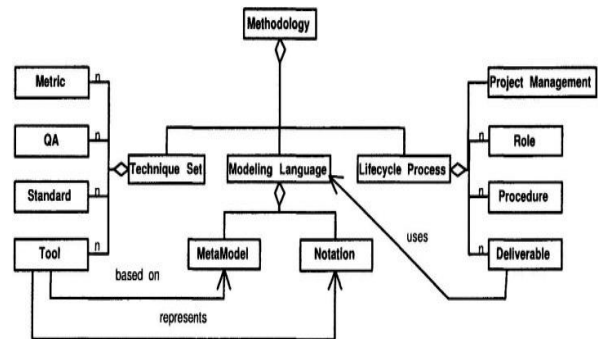


Figure 1.1. The components of a methodology and the relationships among them

At present, more than two dozen agent-oriented methodologies exist. The multiplicity and variety of methodologies result in the following problems:

- (i) industrial problem: selecting a methodology for developing an agent-based system/application becomes a non-trivial task, in particular for industrial developers which hold specific requirements and constraints;

- (ii) Standards problem: multiple different methodologies are counter-productive for arriving at a standard. With no standard available, potential industrial adopters of agent technology refrain from using it; and
- (iii) Research problems: excessive efforts are spent on developing agent-oriented methodologies, in times producing overlapping results. Additionally, as a result of allocating resources to multiple methodologies, no methodology is allocated sufficient research resources to enable addressing all facets and providing a full-fledged agent-based methodology.

A few evaluations of agent-oriented methodologies have been suggested. In (Yu and Cysneiros, 2002), the authors set a list of questions that a methodology should address. However, neither evaluation nor a comparison has been performed using that set. Another study (Cernuzzi and Rossi, 2002) suggests a framework for evaluating agent-oriented methodologies. That framework uses a set of evaluation criteria to examine methodologies' expressiveness; however it does not examine other properties encompassed within the methodology definition. In (Kumar, 2002), the author performs an evaluation of five agent-oriented methodologies, however, he refers only to some supported concepts such as organization design and cooperation and not to the broad set of attributes that constitute a complete methodology. In (Shehory and Sturm, 2001), the authors perform an evaluation of the modeling part within a methodology, however other parts are not evaluated. In (Dam and Winikoff, 2003), three methodologies were compared: MaSE), Prometheus and Tropos. The comparison was performed by gathering feedback regarding the properties of the methodologies from students that used them, and from the methodologies' developers. The gathered feedback included several inconsistent answers. This results in difficulty in analyzing methodology properties.

Other studies that deal with evaluating agent-oriented methodologies compared two or three methodologies, yet mainly with respect to the expressiveness of the methodologies and their supported concepts, and not with respect to other software engineering criteria.

In this paper we evaluate three well-known methodologies: Gaia, Tropos, and MaSE. Unlike previous research on the evaluation of agent methodologies, our evaluation examines multiple dimensions, possibly referring to all of the major facets relevant to methodology evaluation. We perform this evaluation relying on the evaluation framework suggest by (Sturm and Shehory, 2003).

The Evaluation Framework

The evaluation framework used in this chapter is based on a feature analysis technique. That is, the features of each of the examined methodologies are evaluated. The evaluation is performed based on information regarding the examined methodologies available in publications. The framework's four facets are: concepts and properties, notations and modeling techniques, development process, and pragmatics. These facets, and the metric used in conjunction with them, are introduced below.

Concepts and Properties

A concept is an abstraction or a notion inferred or derived from specific instances within a problem domain. A property is a special capability or a characteristic. This section deals with the question whether a methodology addresses the basic notions (concepts and properties) of agents and MAS. The following are the concepts according to which an agent-oriented methodology should be evaluated:

1 Autonomy: is the ability of an agent to operate without supervision;

Reactivity: is the ability of an agent to respond in a timely manner to changes in the environment;

3 Proactiveness: is the ability of an agent to pursue new goals; and

Sociality: is the ability of an agent to interact with other agents by sending and receiving messages, routing these messages, and understanding them.

In the following we present the building blocks that encompass the basic components of MAS. These building blocks are based on (Sturm and Shehory, 2003).

Agent: is a computer program that can accept tasks, can figure out which actions to execute in order to perform these tasks and can actually execute these

actions without supervision. It is capable of performing a set of tasks and providing a set of services.

2 Belief: is a fact that is believed to be true about the world.

3 Desire: is a fact of which the current value is false and the agent (that owns the desire) would prefer that it be true. Desires within an agent may be contradictory. A widely used specialization of a desire is a goal. The set of goals within an agent should be consistent.

Intention: is a fact that represents the way of realizing a desire. Some-times referred to as a plan.

5 Message: is a means of exchanging facts or objects between entities.

Norm: is a guideline that characterizes a society. An agent that wishes to be a member of the society is required to follow all of the norms within. A norm can be referred to as a rule.

Organization: is a group of agents working together to achieve a common purpose. An organization consists of roles that characterize the agents, which are members of the organization.

Protocol: is an ordered set of messages that together define the admissible patterns of a particular type of interaction between entities.

9 Role: is an abstract representation of an agent's function, service, or identification within a group.

Society: is a collection of agents and organizations that collaborate to promote their individual goals.

Task: is a piece of work that can be assigned to an agent or performed by it. It may be a function to be performed and may have time constraints.

Notations and Modeling Techniques

Notations are a technical system of symbols used to represent elements within a system. A modeling technique is a set of models that depict a system at different levels of abstraction and different system's facets (e.g., structural and behavioral facets). This section deals with the properties to which methodology's notations and modeling techniques should adhere. The list of these properties is adopted from (Shehory and Sturm, 2001).

1 Accessibility: is an attribute that refers to the ease, or the simplicity, of understanding and using a method. It enhances both experts and novices capabilities of using a new concept.

Analyzability: is a capability to check the internal consistency or implications of models, or to identify aspects that seem to be unclear, such as the interrelations among seemingly unrelated operations. This capability is usually supported by automatic tools.

Complexity management (abstraction): is an ability to deal with various levels of abstraction (i.e., various levels of detail). Sometimes, high-level requirements are needed, while in other situations, more detail is required. For example, examining the top level design of a MAS, one would like to understand which agents are within the system, but not necessarily what their attributes and characterizations are. However, when concentrating on a specific task of an agent, the details are much more important than the system architecture.

Executability (and testability): is a capability of performing a simulation or generating a prototype of at least some aspects of a specification. These would demonstrate possible behaviors of the system being modeled, and help developers determine whether the intended requirements have been expressed.

Expressiveness (and applicability to multiple domains): is a capability of presenting system concepts that refers to:

- The structure of the system;
- The knowledge encapsulated within the system;
- The system's ontology;
- The data flow within the system;
- The control flow within the system;
- The concurrent activities within the system (and the agents);
- The resource constraints within the system (e.g., time, CPU and memory);
- The system's physical architecture;
- The agents' mobility;
- The interaction of the system with external systems; and
- The user interface specification.

Modularity (incrementality): is the ability to specify a system in an iterative incremental manner. That is, when new requirements are added it should not affect the existing specifications, but may use them.

- 7 Preciseness: is an attribute of disambiguity. It allows users to avoid misinterpretation of the existing models.

Development Process

A development process is a series of actions that, when performed, result in a working computerized system. This section deals with the process development facet of a methodology. This facet is evaluated by examining the following:

- 1 Development context: specifies whether a methodology can be used in creating new software, reengineering or reverse engineering existing software, prototyping, or designing for or with reuse components.
- 2 Lifecycle coverage: specifies what elements of software development are dealt with within the methodology. Each methodology may have elements that are useful in several stages of the development lifecycle. Here, the lifecycle stages are defined as follows: requirements' gathering, analysis, design, implementation, and testing.

Having the development stages defined is not sufficient to render a methodology usable. A methodology should further elaborate the activities within the development lifecycle. Providing a detailed description of the activities included in the development lifecycle would enhance the appropriate use of a methodology and increase its acceptability as a well-formed engineering approach. To verify that a methodology provides detailed activity descriptions, we need to examine the details of the development process. This verification can be performed by answering the following questions regarding an evaluated methodology:

- 1 What are the activities within each stage of a methodology? For example, an activity can be the identification of a role, a task, etc. The methodology may consist of heuristics or guidelines helping the developer to achieve his/her system development goals.
Does the process provide for verification? This question checks whether a methodology has rules

for verifying adherence of its deliverables to the requirements.

- 3 Does the process provide for validation? This question checks whether a methodology has rules for validating that the deliverables of one stage are consistent with its preceding stage.
- 4 Are quality assurance guidelines supplied?
Are there guidelines for project management?

Pragmatics

Pragmatics refers to dealing with practical aspects of using a methodology. This section deals with pragmatics of adopting the methodology for a project or within an organization. In particular, the framework suggests examining the following:

Resources: are the (publicly available) publications describing in detail the methodology (e.g., textbooks and papers), users' groups, training and consulting services offered by third parties and automated tools (CASE tools) available in support of the methodology (e.g., graphical editors, code generators, and checkers).

- 2 Required expertise: is the required background of those learning the methodology. A distinguishing characteristic of many methodologies is the level of mathematical sophistication required to fully exploit the methodology. A criterion within the required expertise may check the required knowledge in some discipline.

Language (paradigm and architecture) suitability: is the level to which the methodology is coupled with a particular implementation language (e.g., object oriented programming language) or a specific architecture (e.g., BDI).

Domain applicability: indicates the level of suitability of a methodology to a variety of domains (e.g., information systems, real-time systems).

Scalability: is the ability of the methodology to be adjusted to handle various application sizes. For example, can it provide a lightweight version for simple problems.

Metric

To enable ranking of the properties examined in the evaluation process, the framework proposes a scale of 1 to 7 with the following interpretations:

- 1 Indicates that the methodology does not address the property.
2. Indicates that the methodology refers to the property but no details are provided.
3. Indicates that the methodology addresses the property to a limited extent. That is, many issues that are related to the specific property are not addressed.
4. Indicates that the methodology addresses the property, yet some major issues are lacking.
5. Indicates that the methodology addresses the property, however, it lacks one or two major issues related to the specific property.
6. Indicates that the methodology addresses the property with minor deficiencies.
- 7 Indicates that the methodology fully addresses the property.

Thus far, we have described the evaluation framework, its evaluation criteria, and its metric. Using these, we proceed with evaluating the Gaia, Tropos and MaSE methodologies.

Evaluating Gaia

In this section we evaluate Gaia according to the framework. We are fully aware of studies that extend Gaia in various facets such as expressiveness (Juan et al., 2002) and implementation (Moraitis et al., 2002). However, in this evaluation, we refer only to (Wooldridge et al., 2000b) and (Zambonelli et al., 2001b), written by the designers of the methodology.

Concepts and Properties

Below, we examine the extent to which Gaia addresses the concepts and the properties suggested by the evaluation framework. Gaia deals with all of the general concepts suggested, but lacks in depicting mental states of an agent.

Autonomy: in Gaia the autonomy is expressed by the fact that the role encapsulates its functionality (i.e., it is responsible for it). This functionality is internal and is not affected by the environment, thus represents the role's

autonomy. In addition, in Gaia there is an option to model alternative computational paths, which gives the role (and agents that consist of this role) autonomy in making decisions. The ranking grade is 7.

Reactiveness: in Gaia the reactiveness is expressed by the liveness properties within the role's responsibilities. The ranking grade is 7.

Sociality: in Gaia the sociality is expressed within the acquaintance model that defines the communication links among agent types. Further, some sociality aspects can be expressed using the organizational structure and rules. Yet, there is no explicit specification of relationships between organizations and roles and societies within MAS. The ranking grade is 4.

Examining the coverage of the framework's building blocks by Gaia, we found that Gaia covers most of them, as seen in Table 7.1. However, the BDI concepts, the social building blocks, and the knowledge representation are not dealt with within Gaia. The ranking grade is 4.

Table 1.1. The coverage of the framework building blocks by GAIA

Framework building block	GAIA Concept
Agent	Agent type
Belief	
Desire	
Intention	
Message	Protocol
Norm	Organizational Rule
Organization	Organization.
Protocol	Protocol
Role	Role
Society	Organization
Task	Activity, Responsibility

Notations and Modeling Techniques

Following, we examine the extent to which Gaia addresses the notations and modeling techniques'

properties suggested by the evaluation framework. Gaia has room for improvements with regards to these properties. In addition, Gaia does not define its entire set of notations.

- 1 Accessibility: Gaia models are easy to understand and use. Yet, the behavior of the system is introduced via a set of logic expressions. This might introduce difficulties in understanding the behavioral specification of a system. The ranking grade is 5.
2. Analyzability: this issue is not dealt with within Gaia. The ranking grade is 1.
- 3 Complexity management: in Gaia, there is no hierarchical presentation or another mechanism for complexity management. The system's description is flat. The ranking grade is 1.
4. Executability: this issue is not dealt with within Gaia. The ranking grade is 1.
- 5 Expressiveness: Gaia is expressive and can handle a large variety of systems due to its generic structure. However, Gaia is mostly suitable for small and medium scale systems. This is because of its flatness, which limits the ability to model a large amount of details. In the following we present our analysis regarding the expressiveness of Gaia according to the properties defined in the previous section:

Analysis Phase.

The MaSE Analysis phase produces a set of roles and tasks which describe how a system satisfies its overall goals. Goals are an abstraction of the detailed requirements and are achieved by roles. Typically, a system has an overall goal and a set of sub-goals that must be achieved to reach the system goal. Goals are used in MaSE because they capture *what* the system is trying to achieve and tend to be more stable over time than functions, processes, or information structures.

A *role* describes an entity that performs some function within the system. In MaSE, each role is responsible for achieving, or helping to achieve specific system goals or sub-goals. MaSE roles are analogous to roles played by actors in a play or by members of a typical company structure. The company (which corresponds to system) has roles such as "president," "vice-president," and "mail clerk" that have specific responsibilities, rights and relationships defined in order to meet the overall company goal.

The overall approach in the MaSE Analysis phase is straightforward: define system goals from a set of requirements and then define the roles necessary to meet

those goals. To help in defining roles to meet specific goals, MaSE uses Use Cases and Sequence Diagrams. The individual steps of the Analysis phase of Capturing Goals, Applying Use Cases, and Refining Roles are presented next.

3.1 Capturing Goals

The first step in the MaSE Analysis phase is Capturing Goals, whose purpose is to transform an initial system specification into set of structured system goals. The *initial system context*, the starting point for MaSE analysis, is usually a software requirement specification with a well-defined set of requirements. These requirements tell the analyst the services that the system must provide and how the system should or should not behave based on inputs to the system and its current state. There are two sub-steps in Capturing Goals: identifying goals and structuring goals. First, goals must be identified from the initial system context. Next, the goals are analyzed and put into a hierarchical form. Each of these sub-steps is described in detail below.

Identifying Goals. The goal of the step named Identifying Goals is to capture the essence of an initial set of requirements. This process begins by extracting scenarios from the initial specification and describing the goal of that scenario.

Goals embody the critical system requirements; therefore, an analyst should specify goals as abstractly as possible without losing the spirit of the requirement. This abstraction can be performed by removing detailed information when specifying goals.

Once the goals have been captured, they provide the foundation for the analysis model; all roles and tasks defined in later steps must support one of the goals. If, later in the analysis, the analyst discovers roles or tasks that do not support an existing system goal, either the roles or tasks are superfluous or a new goal has been discovered.

Structuring Goals. The final step in Capturing Goals is structuring the goals into a Goal Hierarchy Diagram, as shown in Figure 1.1. A Goal Hierarchy Diagram is a directed, acyclic graph where the nodes represent goals and the arcs define a sub-goal relationship. A goal hierarchy is not necessarily a tree as a goal may be a sub-goal of more than one parent goal.

To develop the goal hierarchy, the analyst studies the goals for their importance and inter-relationships. Even though goals have been captured, they are of various

importance, size, and level of detail. The Goal Hierarchy Diagram preserves such relationships, and divides goals into sub-goals that are easier to manage and understand.

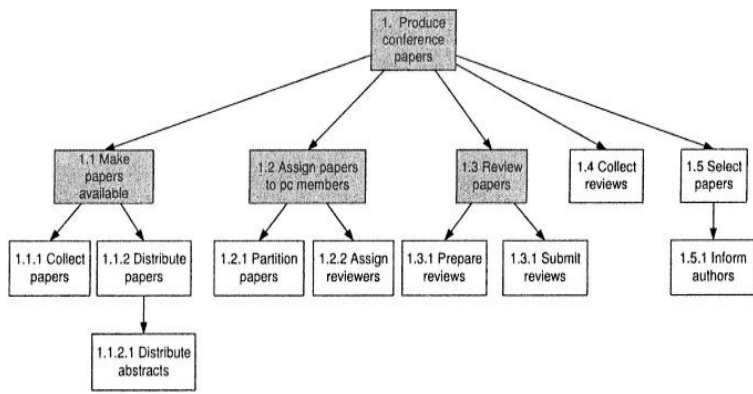


Figure 1.1. Example Goal Hierarchy Diagram

The first step in building is to identify the overall system goal, which is placed at the top of the Goal Hierarchy Diagram. However, it is often the case, as in our example above, that a single system goal cannot be directly extracted from the basic requirements. In this case, the highest-level goals are summarized to create an overall system. Once a basic goal hierarchy is in place, goals may be decomposed into new sub-goals. Each sub-goal must support its parent goal in the hierarchy and defines what must be done to accomplish the parent goal.

Although similar, Goal decomposition is not simply functional decomposition. Goals describe what, while functions describe how. Instead of a set of goals describing what the system will do, functional decomposition typically results in a set of steps prescribing how the system will do it. For example, functional steps for implementing the goal Assign papers to PC members might be to (i) group papers based on similar keywords; and (ii) select PC members whose expertise matches the paper groups. However, the appropriate sub-goals would be to: (i) Partition papers; and (ii) “Assign reviewers.” The fact that the papers are partitioned and PC members are assigned to papers are goals, how we divide the papers or on what basis we assign reviewers are immaterial at this point and will be decided on by the agents responsible for those goals. Goal decomposition continues until any further

decomposition would result in functions instead of a goal.

There are four special types of goals in a Goal Hierarchy Diagram. These are: summary, partitioned, combined, and non-functional. Goals can have attributes of more than one special goal type; however, they do not necessarily have to be one of these types at all.

A *summary goal* is derived from a set of existing peer goals to provide a common parent goal. This often happens at the highest levels of the hierarchy as was the case in the overall system goal in our example.

Some goals do not functionally support the overall system goal, but are critical to system operation. These *non-functional goals* are often derived from non-functional requirements such as reliability or response times. For example, if a system must be able to find resources dynamically, a goal to facilitate locating dynamic resources may be required. In this case, another “branch” of the Goal Hierarchy Diagram can be created and placed under an overall system level goal.

There are often a number of sub-goals in a hierarchy that are identical or very similar that can be grouped into a *combined goal*. This often happens when the same basic goal is a sub-goal of two different goals. In this case, the combined goal becomes a sub-goal of both the goals.

A *partitioned goal* is a goal with a set of sub-goals that, when taken collectively, effectively meet that goal. While this is always true of summary goals, it may be true of any goals with a set of sub-goals. By defining a goal as “partitioned,” it frees the analyst from specifically accounting for it in the rest of the analysis process. Partitioned goals are annotated in a Goal Hierarchy Diagram using a gray goal box instead of a clear box.

At the conclusion of Capturing Goals, system goals have been captured and structured into a Goal Hierarchy Diagram. The analyst can now move to the second Analysis step, Applying Use Cases, where the initial look at roles and communication paths takes place.

Applying Use Cases

The Applying Uses Cases step is crucial in translating goals into roles and associated tasks. Use cases are drawn from the system requirements and describe sequences of events that define desired system behavior; they are examples of how the system should behave. To help determine the actual communications in a MAS, the use cases are converted into Sequence Diagrams. MaSE Sequence Diagrams are similar to standard UML sequence diagrams except that they are used to depict sequences of events between *roles* and to define the

communications between the agents that will be playing those roles. The roles identified here form the initial set of roles used in the next step while the events are also used later to define tasks and conversations.

The first step in Applying Use Cases is to extract Use Cases from the initial system context, which should include both positive and negative Use Cases. A *positive Use Case* describes what should happen during normal system operation. However, a *negative Use Case* defines a breakdown or error. While Use Cases cannot be used to capture every possible requirement, they are an aid in deriving every communication paths and roles. Cross checking the final analysis against the set of derived goals and Use Cases provides a redundant method for deriving system behavior.

Refining Roles

The purpose of the Refining Roles step is to transform the Goal Hierarchy Diagram and Sequence Diagrams into roles and their associated tasks, which are forms more suitable for designing MAS. Roles form the foundation for agent classes and correspond to system goals during the Design phase. It is our contention that system goals will be satisfied if every goal is associated with a role and every role is played by an agent class.

The general case transformation of goals to roles is one-to-one, with each goal mapping to a role. However, there are situations where it is useful to have a single role be responsible for multiple goals, including convenience or efficiency.

Role definitions are captured in a MaSE Role Model as shown in Figure 1.2, which includes information on interactions between role tasks and is more complex than traditional role models, as described in (Kendall, 1998). Roles are denoted by rectangles, while a role's tasks are denoted by ovals attached to the role. Lines between tasks denote communications protocols with the arrow pointing from the initiator to the respondent. Solid lines indicate external communications while dashed lines denote communication between tasks in the same role instance.

Figure 1.2. MaSE Role Model

The tasks are generally derived from the goals for which a task is responsible.

While we could have specified all three goals in a single task, partitioning them in this way is modular and effectively encapsulates the actual approach used.

Roles should not share or duplicate tasks. Sharing of tasks is a sign of improper role decomposition. Shared tasks should be placed in a separate role, which can be combined into various agent classes in the Design phase.

Concurrent Task Model. After roles are created and tasks identified, the developer captures the role's behavior by defining the details of the individual tasks. A role may consist of multiple tasks that, when taken together, define the required behavior of that role. Each task executes in its own thread of control, but may communicate with each other. Concurrent tasks are defined in Concurrent Task Models (see Figure 1.3) and are specified as finite state automata, which consist of states and transitions. *States* encompass the processing that goes on internal to the agent while *transitions* allow communication between agents or between tasks.

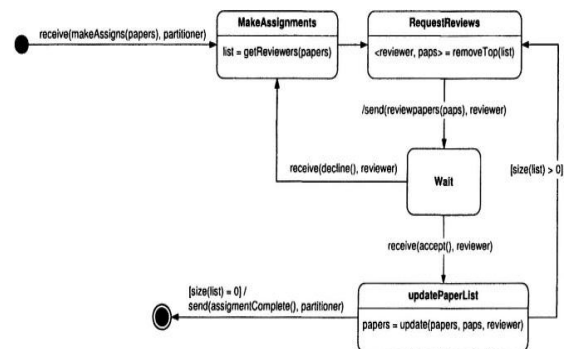
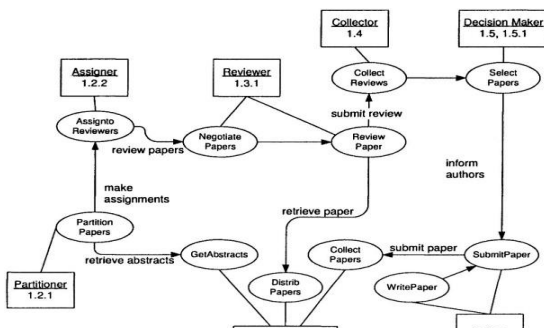


Figure 1.3 Concurrent Task Diagram

A transition consists of a source state, destination state, trigger, guard condition, and transmissions and uses the syntax *trigger [guard] ^ transmission(s)*. Multiple



transmissions may be separated with a semicolon (;), however, no ordering is implied. Generally, events on triggers or transmissions are to be associated with a task within the same role, thus allowing internal task coordination. However, two special events, send and receive, are used to indicate messages sent between agents. The *send* event (denoted *send(message, agent)*) is used to send a message to another agent while the *receive* event (denoted as *receive(message, agent)*) signifies the receipt of a message. The *message* is defined as a performative, which describes the intent of the message, along with a set of parameters that are the content of the message (i.e., *performative(p1 ... pn)* where *p1 ... pn* denotes *n* parameters). It is also possible to send a message to a group of agents via multicasting using a *< group-name >* versus a single agent name.

States may contain *activities* that represent internal reasoning, reading a percept from sensors, or performing actions via actuators. Multiple activities may be included in a single state and are performed in an uninterruptable sequence. Once in a state, the task remains there until the activity sequence is complete. The variables used in activity and events definitions are visible within the task, but not outside of the task or within activities. All messages sent between roles and events sent between tasks are queued to ensure that all messages are received even if the agent or task is not in the appropriate state to handle the message or event immediately.

Concurrent tasks have predefined activities to deal with mobility and time. The *move* activity specifies that the agent is to move to a new address and returns a Boolean value (*Boolean = move(location)*), which states whether the move actually occurred. The agent can reason over this value and deal with it accordingly.

To reason about time, the Concurrent Task Model provides a built in timer activity. An agent can define a timer using *t = setTimer (time)*, the *setTimer* activity. The *setTimer* activity takes a time as input and returns a timer that will timeout in exactly the time specified. The timer that can then be tested via the *timeout* activity, which returns a Boolean value, to see if it has “timed out”

(*Boolean = timeout(t)*).

Once a transition is enabled, it is executed instantaneously. If multiple transitions are enabled, the following priority scheme is used.

- 1 Transitions whose triggers are internal events.
- 2 Transitions whose transmissions are internal events.
- 3 Transitions whose trigger receives a message from another role.
- 4 Transitions whose transmissions are a message to another role.

5 Transitions with valid guard conditions only.

Figure 1.3 shows the *Assign to Reviewers* task for the Assigner role. The task is initiated upon receipt of a *makeAssigns* message from a Partitioner agent, which includes a list of papers to be assigned. After the message is received, the task goes to the *MakeAssignments* state where it computes a list of reviewers for the papers (a process that is as yet undefined). Once these list is defined, the task transitions to the *RequestReviews* state where the top reviewer/papers tuple is taken off the list.

Analysis Phase Summary

Once the concurrent tasks of each role are defined, the Analysis phase is complete. The MaSE Analysis phase is summarized as follows:

- 1 Identify goals and structure them into a Goal Hierarchy Diagram.

Identify Use Cases and create Sequence Diagrams to help identify roles and communications paths.

Transform goals into a set of roles.

Create a Role Model to capture roles and their tasks.

Define role behavior using Concurrent Task Models for each task.

Design Phase

There are four steps to the designing a system with MaSE. The first step is *Creating Agent Classes*, in which the designer assigns roles to specific agent types. In the second step, *Constructing Conversations*, the conversations between agent classes are defined while in the third step, *Assembling Agents Classes*, the internal architecture and reasoning processes of the agent classes are designed. Finally, in the last step, *System Design*, the designer defines the number and location of agents in the deployed system.

Creating Agent Classes

In the *Creating Agent Classes* step, agent classes are created from the roles defined in the Analysis phase. This phase produces an *Agent Class Diagram*, which depicts the overall agent system organization consisting of agent classes and the conversations between them. An *agent class* is a template for a type of agent in the system and is defined in terms of the roles they will play and the

conversations in which they may participate. If roles are the foundation of MAS design, then agent classes are the bricks used to implement MAS.

These two different abstractions manipulate two distinct system dimensions. Roles allow us to allocate system goals while agent classes allow us to consider communications and other resource usage.

The first step is to assign roles to each agent class. If assigned multiple roles, agent classes may play them concurrently or sequentially. To ensure that system goals are accounted for, each role must be assigned to at least one agent class. The analyst can easily change the organization and allocation of roles among agent classes during design, since roles can be manipulated modularly. This allows consideration of various design issues, which are based on standard software engineering concepts such as functional, communicational, procedural, or temporal cohesion.

During this step, we also identify the conversations in which different agent classes must participate. An agent's conversations are derived from the external communications of the agent's assigned roles. For instance, if roles A and B communicate with each other, then, if agent 1 plays role A and agent 2 plays role B, then there must be a conversation between agent 1 and agent 2.

The agent classes and conversations are documented via Agent Class Diagrams, which are similar to object-oriented class diagrams with two main differences. First, agent classes are defined by the roles they play, not by attributes and methods. Second, all relationships between agents classes are captured as conversations. A sample Agent Class Diagram is shown in Figure 1.4. The boxes in Figure 1.4 denote agent classes and contain the class name and the set of roles each agent plays. Lines with arrows identify conversations and point from the conversation initiator to the responder. In this design, the PC Chair agent plays the Partitioner, Collector, and Decision Maker roles while the PC Member agent plays both the Assigner and Reviewer roles. Outside of Authors, the only other agent is the DB agent, which provides an interface to the database containing papers, abstracts, and author information.

Figure 1.4. Agent Class Diagram

The Agent Class Diagram is the first design object in MaSE that depicts the entire MAS in its final form. If we have carefully followed MaSE to this point, the system represented by the Agent Class Diagram will support the goals and Use Cases identified in the Analysis phase. Of particular importance at this point is the system organization – the way that the agent classes are connected with conversations.

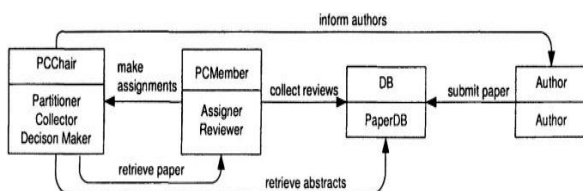
Constructing Conversations

Constructing Conversations is the next MaSE Design phase step. So far, the designer has only identified conversations; the goal of this step is to define the details of those conversations based on the internal details of concurrent tasks.

A *conversation* defines a coordination protocol between two agents and is documented using two Communication Class Diagrams, one each for the initiator and responder. A Communication Class Diagram is similar to a Concurrent Task Model and defines the conversation states of the two participant agent classes. The initiator begins the conversation by sending the first message. When the other agent receives the message, it compares it to its active conversations. If it finds a match, the agent transitions the appropriate conversation to a new state and performs any required actions or activities from either the transition or the new state. Otherwise, the agent assumes the message is a new conversation request and compares it to the conversations it can participate in with the sending agent. If the agent finds a match, it begins a new conversation.

As stated above, communication class diagrams use states and transitions to define the inter-agent communication. Transitions use the following syntax: *rec-mess(args1) [cond] / action ^ trans-mess(args2)*. This states that if the message *rec-mess* is received with the arguments *args1* and the condition *cond* holds, then the method *action* is called and the message *trans-mess* is sent with arguments *args2*.

Assembling Agents



Agent class internals are designed during the step Assembling Agents that includes two sub-steps: defining the architecture of agents and defining the architecture's components. Designers have the choice of either designing their own architecture or using predefined architectures such as BDI. Likewise, a designer may use predefined components or develop them from scratch. Components consist of a set of attributes, methods, and possibly a sub-architecture.

An example of an Agent Architecture Diagram is shown in Figure 1.5. Architectural components (denoted by boxes) are connected to either inner- or outer-agent connectors. *Inner-agent connectors* (thin arrows) define visibility between components while *outer-agent connectors* (thick dashed arrows) define external connections to resources such as agents, sensors and effectors, databases, and data stores. Internal component behavior may be represented by formal operation definitions or state-diagrams. The architecture and internal definition of the components must be consistent with the conversations defined in the previous step. At a minimum, this requires that each action or activity defined in a Communication Class Diagram be defined as an operation in one of the internal components. The internal component state diagrams and operations can also be used to initiate and coordinate various conversations.

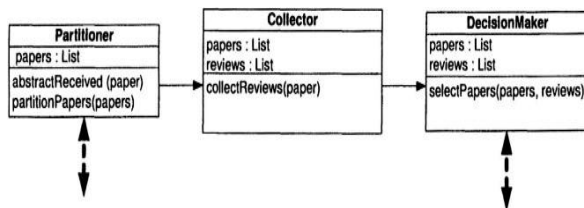


Figure 1.5: PCChair Agent Architecture.

The PCChair agent architecture is shown in Figure 1.5. The PCChair agent has three components, which basically implement pipeline architecture.

System Design

System Design is the final step of the MaSE methodology and uses Deployment Diagrams to show the numbers, types, and locations of agent instances in a system. System design is actually the simplest step of MaSE, as most of the work was done in previous steps. Figure 1.6 shows a Deployment Diagram for the conference management system. The three-dimensional boxes represent agents while the connecting lines represent actual conversations between agents. The

agents are identified by their class name in the form of *instance-name : class*. Dashed boxes define physical computational platforms.

A designer should define the system deployment before implementation since agents typically require Deployment Diagram information, such as a hostname or address, for communications. Deployment Diagrams also offer an opportunity for the designer to tune the system to its environment to maximize available processing power and network bandwidth. In some cases, the designer may specify a particular number of agents in the system or the specific computers on which certain agents must reside. The designer should also consider the communication and processing requirements when assigning agents to computers. To reduce communications overhead, a designer may choose to deploy agents on the same machine. However, too many agents on a single machine destroy the advantages of distribution gained by using the multiagent paradigm. Another strength of MaSE is that a designer can make these modifications after designing the system organization, thus generating a variety of system configurations.

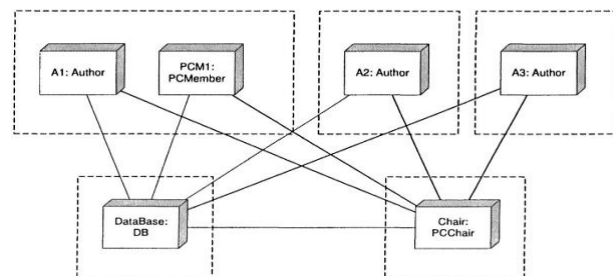


Figure 1.6: Deployment Diagram

Design Phase Summary

Once the Deployment Diagrams are finished, the Design phase is complete. The MaSE Design Phase can be summarized as follows:

- 1 Assign roles to agent classes and identify conversations.
- 2 Construct conversations, adding messages/states for robustness.
- 3 Define internal agent architectures. Define the final system structure using Deployment Diagrams.

agentTool

The agentTool system (DeLoach and Wood, 2001) has been developed to support and enforce MaSE. Currently agentTool implements all seven steps of MaSE as well as automated design support. The agentTool user interface is shown in Figure 6.8. The menus across the top allow access to several system functions, including analysis to design transformations (Sparkman et al., 2001), conversation verification (Lacey et al., 2000), and code generation. The buttons on the left add specific items to the diagrams while a text window displays system messages. The different MaSE diagrams are accessed via the tabbed panels across the top of the main window. When a MaSE diagram is selected, the designer can manipulate it graphically in the window. Each panel has different types of objects and text that can be placed on them. Selecting an object in the window enables other related diagrams to become accessible.

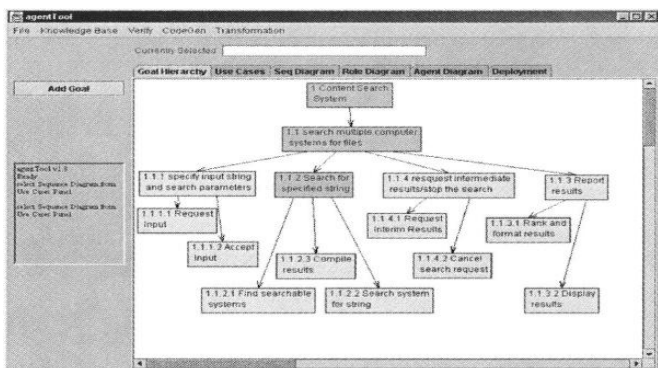


Figure 1.7: agentTool

While the designer may use existing architectures or design a new one from scratch, agentTool also provides the ability to semi-automatically derive the agent architecture directly from the roles and tasks defined in the analysis phase. This approach has the advantage of providing a direct mapping from analysis to design. Each task from each role played by an agent defines a component in the agent class. The concurrent task itself is transformed into a combination of the component's internal state diagram and a set of conversations. Activities identified in the concurrent task become methods of the component.

The transformation is actually a sequence of transformations that incrementally change roles and tasks into agent classes, components, and conversations. Before beginning the analysis-to-design transformation process, the Role Model and its set of concurrent tasks,

and the assignment of roles to agent classes must exist. During the first stage of the transformation process, agentTool derives agent components from their assigned roles and assigns external events to specific protocols. In the second stage, agentTool annotates the component state diagrams to determine where conversations start and end. During the last stage, agentTool extracts the annotated states and transitions and uses them to create new conversations, replacing them in the component state diagram with actions initiating the conversation.

A second set of transformations that is currently implemented in agentTool consists of transformations to add functionality required for mobility. In the analysis phase, mobility is specified using a *move* activity in the state of a concurrent task diagram. This *move* activity is copied directly into the associated component state diagram during the initial set of analysis-to-design transformation described above. During the mobility transformation, the existing design is modified to coordinate the mobility requirements between all components in the agent design. In the derived mobility design, the Agent-Component is responsible for coordinating the entire move and working with the external agent platform to save its current state and actually carry out the move.

The agentTool system also provides automatic verification of conversations. The verification process begins with the fully automated translation of system conversations into the Promela modeling language. Then, the Promela model is automatically analyzed using the Spin verification tool to detect errors such as deadlock, non-progress loops, syntax errors, unused messages, and unused states (Holzmann, 1997). Feedback is provided to the designer automatically via text messages and graphical highlighting of error conditions.

Applications

MaSE has been successfully applied in many graduate-level projects as well as several research projects. The Multiagent Distributed Goal Satisfaction project used MaSE to design the collaborative agent framework to integrate different constraint satisfaction and planning systems. The Agent-Based Mixed-Initiative Collaboration project also used MaSE to design a MAS focused on distributed human and machine planning. MaSE has been used successfully to design an agent-based heterogeneous database system as well as a multi-

gent approach to a biologically based computer virus immune system. More recently, we applied MaSE to a team of autonomous, heterogeneous search and rescue robots (DeLoach et al., 2003). The MaSE approach and models worked very well. The concurrent tasks mapped nicely to the typical behaviors in robot architectures. MaSE also provided the high-level, top-down approach missing in many cooperative robot applications.

References

- [1] G. Booch, J. Rumbaugh, and I. Jacobson. *The Unified Modelling Language User Guide.* Addison Wesley, 1998.
- [2] Jeffrey M. Bradshaw. *An Introduction to software agents.* In Jeffrey M. Bradshaw, editor, *Software Agents*, pages 3–46. AAAI Press / The MIT Press, 1997.
- [3] P. R. Cohen and H. J. Levesque. *Teamwork.* Nous, 25(4):487–512, 1991.
- [4] Jacques Ferber. *Multi-agent Systems: An Introduction to Distributed Artificial Intelligence.* Addison-Wesley, 1999.
- [5] Leonard N. Foner. *What's an agent anyway? -a sociological case study.* FTP Report - MIT Media Lab, May 1993.
- [6] S. Franklin and A. Graesser. *Is it an Agent, or just a Program? A Taxonomy for Autonomous*
- [7] *Agents.* In *Intelligent Agents III. Agent Theories, Architectures and Languages (ATAL'96)*, volume 1193, Berlin, Germany, 1996. Springer-Verlag.
- [8] N. R. Jennings and Michael Wooldridge. *Agent-Oriented Software Engineering.* In *San Francisco*
- [9] J. Garijo and Magnus Boman, editors, *Proceedings of the 9th European Workshop on Modeling Autonomous Agents in a Multi-Agent World: Multi-Agent System Engineering (MAAMAW-99)*, volume 1647, pages 1–7. Springer-Verlag: Heidelberg, Germany, 30–2 1999.
- [10] T. Juan, A. Pearce, and L. Sterling. *Roadmap: Extending the GAIA methodology for complex open systems.* In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, Bologna, Italy, July 2002.
- [11] Michael Luck, Peter McBurney, and Chris Preist. *Agent technology: Enabling next generation computing: A roadmap for agent-based computing.* Agent Link report, available from www.agentlink.org/roadmap, 2003.
- [12] H. S. Nwana. *Software agents: An overview.* Knowledge Engineering Review, 11(2):205–244, 1995.
- [13] James Odell. *Objects and agents compared.* Journal of Object Technology, 1(1):41–53, 2002.
- [14] M. Wooldridge and N. R. Jennings. *Intelligent agents: Theory and practice.* The Knowledge Engineering Review, 10(2):115–152, 1995.
- [15] Wooldridge, M; N.R. Jennings, and D. Kinny. *A methodology for agent-oriented analysis and design.* In *Proceedings of the third international conference on Autonomous Agents (Agents-99)*, 1999.
- [16] Wooldridge, M; Jennings, and D. Kinny. *The GAIA methodology for agent-oriented analysis and design.* Autonomous Agents and Multi-Agent Systems, 3(3), 2000.
- [17] Wooldridge, M; *An Introduction to Multi-Agent Systems.* John Wiley & Sons, 2002.

Constructing a Predictive Model for an Intelligent Network Intrusion Detection

Alebachew Chiche¹, Million Meshesha (PHD)²

¹Mizan-Tepi University, School of Computing and Informatics, Tepi, Ethiopia
alebachew.chz@gmail.com

²Addis Ababa University, School of Information Science, Addis Ababa, Ethiopia
meshe84@gmail.com

Abstract-With the wide use of Internet and network connectivity, it is important to prevent unauthorized access to system resources and data. In this study, we present a new Network Intrusion Detection System by integrating data mining and knowledge based system to detect a network attacks. Hybrid data mining process model is followed for data mining tasks to extract hidden knowledge from KDDCup'99 intrusion dataset. J48 decision tree, JRip rule induction, Naïve Bayes and Multilayer Perceptron (MLP) Neural Network are adopted to construct a predictive model on total datasets of 63, 661 instances. This study supports network administrators to fill the knowledge gap they have to detect network attacks efficiently and effectively. Experimental result shows that, the proposed system performs 91.43 percent and 83 percent accuracy and user acceptance, respectively. Further work is required to acquire and integrate prevention knowledge automatically with the predictive model.

Keywords: Intrusion Detection, Data Mining, Network Intrusions, Integration, Knowledge based system

1. Introduction

In the era of information society, a network-based computer system plays fundamental roles in modern society [1] [2]. Accordingly, Network based attacks are frequently increasing, resulting in a huge financial loss to the organizations and causing the network to be paralyzed for several hours. There are many methods to strengthen the network security at the moment, such as encryption, VPN, firewall, etc. [3], but all of these are too static to give an effective protection. However, intrusion detection is a dynamic one, which can give dynamic protection to the network security in monitoring, attack and counter-attack. Thus, Intrusion Detection Systems (IDS) have become a necessity in computer security systems because of the increase in unauthorized accesses and attacks. In spite of this, intrusion detection systems (IDSs) are designed to detect a wide range of security violations from attempted break-ins by outsiders to system penetrations and abuses by insiders [4].

Network intrusion has many types which falls into one of the following four main categories [2]: Denials-of Service (DoS), Probing, User-to-Root (U2R) and Remote-to-Local (R2L) attacks.

According to Mohammadreza et al [3] traditional intrusion protection paradigms such as firewall are too static for effective protection. However, intrusion

detection can give dynamic protection to the network security in monitoring attack and counter-attack. Hence, Intrusion detection systems (IDSs) play a major role in providing security to networks and have been applied to detect network intrusions [2].

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analysing them for signs of possible intrusions, which are violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices [5].

Generally, there are two techniques for IDSs, misuse detection and anomaly detection. Misuse detection aims to detect the attacks through the predefined signatures of them. Whereas, anomaly detection uses normal network behaviours to identify unknown attacks by detecting significance deviation from the established normal patterns [4]. In the case of detecting data target, intrusion detection system can be classified as host-based and network-based that employs one or both of the intrusion detection methods [6]. Host-based systems base their decisions on information obtained from a single host, while network-based intrusion detection systems obtain data by monitoring the traffic in the network to which the hosts are connected.

Most commercially available IDS systems today, such as SNORT[7] employ signature-base detection

techniques that works by matching the coming traffics with the knowledge encoded within it[8]. But it is very difficult to analyse huge network traffic and also specify intrusions using the rules as well.

Therefore, to overcome the limitations of the signature-based systems, a number of IDSs employ data mining techniques [9]. Data mining can efficiently extract patterns of intrusions for misuse (signature-based) detection, establish profiles of normal network activities for anomaly detection, and build classifiers to detect attacks, especially for the vast amount of audit data [10].

In this study, we proposed intrusion detection systems which integrate data mining with knowledge based system so as to come up with an adaptive network intrusion detection hat predict attacks as efficient and effective as possible for network administrators.

2. Related works

There are a number of research works to design an intrusion detection system using machine learning algorithms. Kalpana et al [11] have proposed a hybrid approach for intrusion detection using k-means, support vector machine and association rule mining algorithms. Also, Saale and Kale [12] proposed an intrusion detection method by combining Naïve Bayesian and support vector machine for intrusion detection based on KDDCUP99 dataset. The experimental result shows that support vector machine (SVM) has high detection rate than naïve Bayes. The result verified that the hybrid algorithm minimized false positives and maximizes detection rates. Further, Sandhya et al. [13] proposed a hybrid DT-SVM for intrusion detection and evaluated their performance based on the benchmark KDD Cup 99 Intrusion data. The empirical result shows that 99% accuracy is achieved by the hybrid approach. Based on experimental result, they concluded that the hybrid approach or model of any intrusion detection works better than the individual DT and SVM for normal class.

Panda and Patra [14] proposed a framework of NIDS based on Naïve Bayes algorithm on KDDcup'99 intrusion dataset. The proposed technique performs better in terms of cost, accuracy and computational time when applied to KDD'99 data sets compared to a back propagation neural network based approach. Sagane and Hande [15] proposed an anomaly based intrusion detection using naïve Bayes data mining algorithms based on KDDCup99 intrusion dataset. The proposed method has achieved 97% prediction accuracy on

KDDCup'99. The proposed approach achieve higher detection rate, less time consuming and has low cost factor with more false positives. Kumar and Gupta [9] presented an Intrusion Detection System by applying genetic algorithm with Neural Network to efficiently detect various types of network intrusions from KDD dataset. The result indicated that resilient back propagation with sigmoid function is the best one and used for classification. Kosamkar and Chaudhari [7] have used hybrid of C4.5 decision tree and Support Vector Machine (SVM) algorithms for developing the intrusion detection system. The experimental result shows that the combined algorithms increase the accuracy and detection rate and decrease false alarm rate. The finding implies that building an effective intrusion detection models with good accuracy and performance are very essential.

Tigabu [16] proposed a semi-supervised approach for modelling a network intrusion detection using decision tree and naïve Bayes as classification algorithm with a k-means clustering algorithm based on KDDCup'99 intrusion dataset. The finding verified that J48 decision tree algorithm is better than Naïve Bayes method in terms of both false positives and detection rate as well as prediction accuracy.

In this paper, we propose a Network Intrusion Detection System by integrating data mining and knowledge based system for detecting network intrusions.

3. Modelling

For this study the hybrid data mining process model [17] is followed for extracting hidden knowledge for constructing Network intrusion detection model. To construct a predictive model KDDCup99 intrusion data is collected from Massachusetts Institute of Technology (MIT) Lincon laboratory (URL), which is available for domain researchers. This data set was prepared by Stolfo et al. [18], and is built based on the data captured in DARPA'98 IDS evaluation program.

3.1. Data preparation

To understand the problem domain, the researcher used direct observation by closely working with domain experts, interviewing experts as well as domain researchers, reviewing documents, reports and literatures. This helped us to define the research problem and determine the research goals. After problem understanding, the KDDCup'99 dataset

were collected as per the objective set. During data understanding task we checked the syntax of the KDDCup'99 dataset, attributes and classes as well as the quality of the content.

To improve the quality of intrusion data set, pre-processing tasks, such as data cleaning, dataset reduction, attribute selection and balancing tasks are performed.

A higher sample size data require more computational resources, and hence strategically sampled data may work well to achieve once goal. Accordingly, representative sample data selected from KDDCup99 have been used in this study for constructing a model. Resampling method has been applied on the KDD dataset to select 63,661 instances. Table 1.shows that the pre-processed dataset for the five classes. This is followed by attribute selection. To select the best attributes for data mining, we used information gain ratio method which exists in WEKA data mining tool to get 14 out of 41 attributes.

Balancing is needed if one target class has much lower frequency than the other target class in the given dataset [19]. As shown in table 1, there is imbalance instance distribution for each class. DOS is almost taking higher share out of the four intrusion types. In order to balance the data set, we applied SMOTE (Synthetic Minority Over-sampling Technique), which is an over-sampling approach that generates synthetic examples in a less application specific manner [20]. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbours. For this study the class imbalance exists with the sampled dataset were balanced using SMOTE.

Table 1: Number of records after pre-processing

Classes	Share of each class Before balancing		Share of each class after balancing	
	Count	In percent (%)	Count	In percent (%)
Normal	21,352	52.99	21,352	33.54
DOS	18,469	45.23	18,469	29.01
Probe	626	1.53	10,016	15.73
U2R	7	0.22	7,168	11.26
R2L	104	0.028	6,656	10.46
Total	40,558	100	63,661	100

3.2. Experimentation

In this study, we conducted four experiments using four classification algorithms namely J48 decision tree, JRip rule induction, Naïve Bayes and MLP neural network with test mode, tenfold (10) Cross-Validation.

The summary of experimental results is shown below in table 2. The comparisons between the results of the four algorithms are done for attack prediction.

Table 2. Summery result of the four Algorithms

	Algorithms			
	J48	JRip	Naïve Bayes	MLP
Prediction accuracy	99.91 %	99.89 %	67.69 %	98.01 %
TP rate	99.9 %	99.9 %	67.7 %	98 %
FP rate	0 %	0 %	10.7 %	0.8 %

As shown in table 2, the accuracy of J48, JRip, Naïve Bayes and MLP classifiers is 99.91%, 99.89%, 67.69% and 98.01% respectively. From results of experiments, the average TP rates is 99.9%, 99.9 %, 67.7% and 98% for J48, JRip, Naïve Bayes and MLP algorithms respectively. The FP rate results of four algorithms are 0 %, 0%, 10.7% and 0.8% for J48, JRip, Naïve Bayes and MLP algorithms. From this evaluation, we understood that J48 decision tree have a better performance than the remaining three algorithms to predict attacks. Hence, J48 decision tree with 10-fold cross validation is selected as a working algorithm for this study. The confusion matrix for the selected classifier is presented in table 3 below.

Table 3. Confusion matrix for J48 decision tree algorithm before integration

Actual classes	Predicted classes				
	Normal	DOS	Probe	U2R	R2L
Normal	21331	7	8	1	5
DoS	10	18457	1	0	1
Probe	9	0	10,007	0	0
U2R	1	0	0	7165	2
R2L	11	0	0	2	6643

So, from the confusion matrix one can understand that selected algorithm classified most of the instances into their proper class. However, there are cases where normal instances are miss-classified with DOS or Probe instances, alternatively.

Finally, the selected algorithm has been integrated with knowledge based system to come up with an intelligent network intrusion detection system.

4. System architecture

The proposed architecture of the network intrusion detection system in this work consists of two major components, namely, data mining, and knowledge based system, as shown in Figure 1.

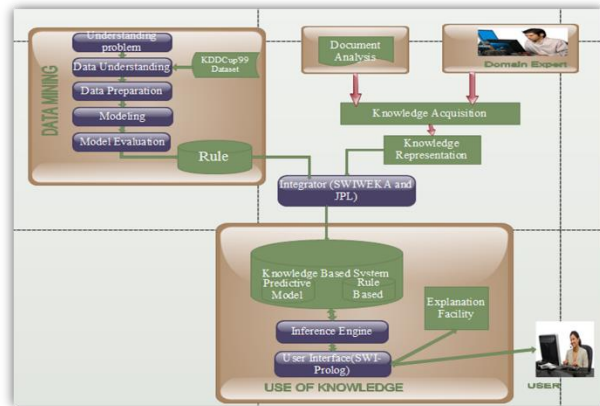


Figure 2. Architecture of the proposed system

The data mining technology is used to construct a predictive model using selected J48 decision tree algorithm. The knowledge based system (use of knowledge) is used on the other hand to predict network attacks automatically, map the numeric class value returned by the detection model to appropriate attack categories, and track log file and add this trails into the training dataset automatically. This makes our system to learn incrementally and adaptively. To do this, java programming was used to integrate WEKA result with the Knowledge Based System automatically. And also 'swiweka' is used as an interface that allows the use of WEKA API for classification; weka.jar, Weka_src.jar are used to construct a model when called from interface through swiweka package, jpl library to connect the Java layer with the Prolog layer.

The various modules of the proposed system performing detection of attacks are described below.

- Arff_writer module: this module accepts the attribute value form user interface and, writes the input values by changing into arff file format.
- Arff_reader module: The reader module reads the original ARFF file for training purpose and sample ARFF file created from user input for prediction. Beside this, the class index is set on both original and sample arff files for the classification task.
- Build classifier module: it is the responsibility of this module to construct a predictive model on the original dataset using the selected J48 decision tree algorithm.
- Detection module: this module integrates the predictive model with the knowledge based system for predicting network attack. Further, the knowledge based system provides the comprehensive description for detected attack.

A user interacts with the system through dialogue. The system forwards successive questions for user to predict an attack. Thereafter the system displays the predicted attack with available options for user.

5. System evaluation

The system has been evaluated in two ways; system performance testing and user acceptance testing.

System performance testing is done to verify whether the proposed system could work correctly in the absence of domain expert or not. For this study, the researcher prepared 35 test cases for testing performance of the proposed system. For testing purpose, the attributes and their corresponding value of the instances need not have a label so that the expert only feed attribute values of an instance for the system. As a result the proposed system performs in the absence of domain experts with 91.34% which indicates that the model is effective in detecting a network attack. The evaluation result is shown below.

Table 4. Confusion matrix of Intelligent Network Intrusion Detection System

Actual class labelled	System predicted value				
	Normal	DoS	Probe	R2L	U2R
Normal	10	0	0	0	0
DoS	0	7	0	0	0
probe	0	0	6	0	0
R2L	2	0	0	3	0
U2R	4	0	0	0	1

As we can observe from the table above, the integrated system performs better in all test cases. However, we can understand from the above table that there are cases that R2L and U2R are classified as normal by the integrated intrusion detection system. Since the sample size of class U2R attack and R2L attack is too small, the detection accuracy for U2R and R2L attack is low. The test result indicates that as samples in training data set and test data set do not have similar feature signature, the accuracy of the Intrusion Detection System less after the integration task.

6. Conclusion

In this study we present a new Intrusion detection system that can predict a network attack and provide description about predicted attacks. The systems are developed by integrating data mining model and knowledge based system. J48 decision tree algorithm, JRip rule induction, MLP neural network and the Naive Bayes algorithms are experimented as a classification algorithm for constructing a network intrusion predictive model. The dataset used in this study is collected from MIT Lincoln lab. Evaluation of the system shows that, the proposed system achieved an encouraging results with 91.43% system performance. Since the test cases is unknown for classifier and different in feature signature, the system confused to classify U2R and R2L correctly. We are now working towards adding knowledge for prevention mechanisms.

References

- [1] Subaira, A.S and Anitha, P. (2014) Efficient Classification Mechanism for Network Intrusion Detection System Based on Data Mining Techniques: a Survey. In Proceedings of 8th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, 2014(Coimbatore: IEEE), pp. 274-280.
- [2] Sandeep, D. and M.S.Chadhari. (2014) Review on Data Mining Techniques for Intrusion Detection System. International Journal of Innovative Research in Computer and Communication Engineering. Vol.2 (no.1):pp. 2587-2592.
- [3] Mohammadreza, E., Sara, M., Fatimah, S. and Lilly, S.A. (2010) Intrusion Detection Using Data Mining Techniques. IEEE: pp. 200-203.
- [4] Nannan, L., Shingo, M., Tuo, W. and Kotaro, H.(2012) Integrated Fuzzy GNP Rule Mining with Distance-based Classification for Intrusion Detection System. In Proceedings of 2012 IEEE International Conference on Systems, Man, and Cybernetics, Seoul, Korea, October 14-17, 2012(Korea: IEEE), pp. 1569-1574.
- [5] Karen, S. And Peter, M. (2007) Guide to Intrusion Detection and Prevention Systems (U.S Department of Commerce: Gaithersburg).
- [6] Shubhangi, S., Gujar and B.M.patil. (2014)Intrusion detection using Naive Bayes for real time data. International Journal of Advances in Engineering and Technology, vol. 7(no. 2): pp. 568-574.
- [7] Vaishali, K. and Sangita, S .C.(2014) Improved Intrusion Detection System using C4.5 Decision Tree and Support Vector Machine.(IJCSIT) International Journal of Computer Science and Information Technologies, vol. 5(no. 2): pp. 1464.
- [8] Ahmed, Y. and Ahmed, E. (2011) Network Intrusion Detection using Data mining techniques and network Behaviour Analysis. International Journal of Computer Science & Information Technology (IJCSIT), vol. 3(no. 6): pp. 87-98.
- [9] Praveen, K. and Nitin, G. (2014) A Hybrid Intrusion Detection System Using Genetic-Neural Network. In Proceedings of 29th National Conference on Advances in Engineering and Technology, Maharishi, March 2014(Maharishi: International Journal of Engineering Research and Applications), pp. 59-63.
- [10] Deepthy, K. D. and Anita, J. (2012) Survey on data mining techniques to enhance intrusion detection. In Proceedings of International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, January 10-12, 2012(India: IEEE),
- [11] Kalpana ,J.,Praveen, K. and Seema, R.(2015) Design and Development of a Prototype Application for Intrusion Detection using Data mining. In proceedings of 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), India,September 2-4. 2015(India: IEEE).
- [12] Amit, D. S. and Swati, G. K. (2014) Combining Naive Bayesian and Support Vector Machine for Intrusion Detection System. International Journal of Computing and Technology, vol. 1(no. 3):pp. 64.
- [13] Sandhya, P., Ajith, A., Crina, G. and Johnson, T. (2007) Modelling intrusion detection system using hybrid intelligent systems. Journal of Network and Computer Applications, vol. 30:pp. 144-152.
- [14] Mrutyunjaya, P. and Manas, R. P.(2007) Network Intrusion Detection using naive bayes.(IJCSNS)International Journal of Computer Science and Network Security, vol. 7(no. 1): pp. 258-263.
- [15] Anikt, p. S. and S.S.Dhande. (2014) Malicious Code Detection using Naive bayes classifier .International journal of application or innovation in engineering, vol. 3(no. 4): pp. 401-405.
- [16] Dagne, T. (2012) Constructing a Network Intrusion Detection System. Thesis, Addis Ababa University, Addis Ababa, Ethiopia.
- [17] Krzysztof, J. C., Witold, P., Roman W. S. and Lukasz, A. K. (2007) Data Mining: Knowledge Discovery Approach.(New York, USA: Springer).

- [18] Stolfo, S. J., Fan, W., Lee, W., Prodromidis, A. and Chan, P. K. (2000) Cost-based modelling for fraud and intrusion detection: Results from the jam project. Discex, vol. 2: p. 1130.
- [19] D. T. Larose. (2005) Discovering Knowledge in Data: An Introduction to Data Mining.(New Jersey,USA: John Wiley & Sons Inc).
- [20] Yetian, C. (2009) Learning classifiers from imbalanced, only positive and unlabeled data sets. Available via dialog.http://web.cs.iastate.edu/~yetianc/cs573/files/C573_ProjectReport_YetianChen.pdf. Accessed 15 Jan 2016.

Fine Tuning of TCP/IP for Ad Hoc Networks

Sabahat Zahra

Department of Computer Sciences,
Pir Mehr Ali Shah, Arid Agriculture University
Rawalpindi Pakistan
sabahat.zahraa@gmail.com

Yawar abbas abid

Department of Computer Sciences,
COMSATS institute of Information Technology
Sahiwal, Pakistan
yawar.abid@ciitsahiwal.edu.pk

Muhammad Usman Nasir

Department of Computer Sciences,
COMSATS institute of Information Technology
Sahiwal, Pakistan
usmannasir@ciitsahiwal.edu.pk

Abstract—Transmission Control Protocol (TCP) is a reliable transmission protocol for the wired networks but as ad hoc networks are gaining popularity, here is a huge demand for TCP adaptation for ad hoc networks. Ad hoc networks are way much different as compared wired networks due to higher Bit-Error Rate (BER), less bandwidth, mobility and asymmetric behavior. So, to alter TCP/IP policies for ad hoc networks a lot of strategies are adapted to enhance its performance. To fine tune TCP/IP for ad hoc networks different variants of TCP are proposed. In this paper, we have evaluated the performance of those variations of Transport Control Protocol over a highly mobile' ad hoc network and we have also studied relative performance of three different variants of TCP -TCP Reno, TCP SACK and TCP Vegas. Relative performance of dual on-demand ad hoc routing protocol - AODV and DSR with respect to transport protocols is also observed. Our studies show that performance of TCP is variable and reliant on the movement patterns of the portable nodes.

Keywords: Mobile Ad Hoc Networks, TCP Reno, TCP SACK, TCP Vegas, AODV and DSR.

I. INTRODUCTION

Transport Control Protocol / Internet Control Protocol (TCP/IP) a well-known communication procedure for connection oriented, consistent and endwise robust communication [1].TCP/IP is especially designed for the traditional wired networks. The TCP/IP have been pitched to accomplish fine in outdated systems someplace the bit-error

degree is lesser, main foundation of packet loss is system overcrowding. Other hand a wireless mobile ad-hoc network [2] it is gathering of movable nodes accompanied by no pre-determined substructure, creating a short-term system. These networks have definite features that impose 'stringent require on the transmitting protocols. The furthestmost significant distinctive is the dynamic topology, that is the straight concern of node movement which might be a way failure or high probability of network failure. Along with mobility the other major characteristics of Wireless networks is higher Bit Error Rate (BER), less bandwidth as compared to wired networks, handoffs from one hop to another due to mobility and asymmetry due to which mobile host has partial authority and lesser handling competence. Consistent conveyance protocols like TCP are perfected to do glowing in wired nets with stationary nodes. But, when used in wireless networks, these protocols give poor performance [3]. The decrease in the performance is due to the violation of the main assumption of the TCP, which states that somewhat damage of the container is due to the overcrowding in the system only. As TCP/IP has a traditional way of working for wired networks, it deals with packet drop as a network congestion problem in wireless infrastructure [4]. It retransmits the unrecognized containers, after that droplets the scope of TCP overcrowding window, so subsequent now transmission rate decrease. Another, it stimulates the congestion control instruments which contain exponential back-off of the retransmission clock and reduction of the slow-start brink value characteristically, partial extent of the prior overcrowding window). As a final point, in which that is arrives a slow-start repossession stage to certify that the congestion has condensed [5-7].

This paper focuses on different TCP variants to check relative performance for TCP Reno, TCP Sack, and TCP Vegas [8]. Relative enactment of dual on-demand ad hoc routing protocol - AODV and DSR with respect to transport protocols is

also observed. Our studies show that enactment of TCP is flexible and reliant on the movement patterns of the mobile nodes. We have specifically sought to learn the enactment of TCP over ad hoc networks, namely, the impact on TCP caused by the mobility of the nodes. Using two familiar directing protocols - Ad hoc On require and Distance Vector (AODV) [9] and Dynamic Source Routing (DSR) [10] comparison has been done between measured throughput and expected throughput [3].

The organization of this paper is as follows. The afterward segment gives the literature review about TCP fine tuning for Ad Hoc networks along through the detailed overview of TCP variants and routing protocols used for communication. In Section III we discuss the relative enactment of two on demand ad hoc routing protocols - AODV and DSR. In Section IV, we compare the performance of different conveyance protocols over wireless mobile ad hoc networks - TCP Reno, TCP Sack and TCP Vegas. Section V discusses the conclusion about the performance of TCP- deviations for dissimilar mobility patterns then mobile nodes.

II. LITERATURE REVIEW

In this section, we discuss the TCP fine tuning for Ad Hoc networks along with the detailed overview of TCP variants and routing protocols used for communication along with the methodologies opted by different researchers for TCP fine tuning. In [11] has concentrated on the enactment of TCP by varying the ad hoc routing protocols. In [12] examined the interface among TCP and MAC level. They argued that linkage level guard; back-off strategy and choosy crocodile planning are dangerous features for effective and impartial process of ad-hoc networks under TCP. Holland and Vaidya [3] have investigated the influence of linkage breakage on TCP enactment. They introduced new metric expected throughput, which we have used in this paper. Explicit Link Failure Notification (ELFN) is used to better the reliability of TCP. In [13] K. Cliandran et.al argues an alike pattern. The TCP basis after receiving such package arrives a snooze formal that is comparable to the TCP Persist formal. In [14] Jain Liu et.al proposed an approach towards the improvement of TCP in wireless ad hoc networks. They presented TCP for Mobile Ad Hoc Networks (ATCP), which gives the better TCP's throughput by a factor of 2-3. Proposed solution implements a thin cover among IP and TCP layers of the protocol stack, which ensures appropriate TCP behavior while maintaining high throughput.

A. TCP Variants

Proposed technique different variants of TCP: , TCP SACK , TCP Reno and TCP Vegas.

1) TCP Reno

TCP Reno [15] is a variant of TCP used as a fast recovery technique which permits the transmitter to avoid difficult the pipe, and to transfer on or after $cwnd$ to $cwnd/2$ in the space of a solitary RTT. TCP Reno is TCP Tahoe through the adding of Fast Recovery. The impression is to usage the incoming duplicate ACKs to step retransmission.

2) TCP SACK

Selective Acknowledgment (SACK): the receiver evidently tilts which sets, messages, or wreckages in a creek are recognized (either negatively or positively). Optimistic selective reaction is an excellent in TCP that is appreciated in Satellite Internet access [16].

The Selective Acknowledgement (SACK) selection distinct in RFC 2018 is used by the TCP data receiver to acknowledge non-contiguous chunks of data not enclosed by the Cumulative Acknowledgement field. However, RFC 2018 does not stipulate the use of the SACK decision once identical sections are established.

3) TCP Vegas

TCP Vegas is a TCP congestion prevention procedure that highlights packet deferral, relatively as compared to package loss, as an only one to assistance control the rate at which to send packets. It was established at the University of Arizona by Lawrence Brakmo and Larry L. Peterson and introduced in 1994 [17].

B. TCP Mechanisms for Ad-Hoc Networks

Dissimilar cellular systems, at which individual the previous hop is founded on a wireless medium, ad hoc networks are collected absolutely of wireless links, at which multi-hop networks might be in place. Moreover, all nodes can travel easily and arbitrarily in an ad hoc scenario, it can creates the TCP congestion regulator moderately tough meanwhile this is a clock founded instrument. Subsequently, an error-detection and error-recovery policies intrinsic in ordinary TCP essential to stay modified in order to fitting this setting. Specially, subsequently the faults in this setting happens not only because of congestion nonetheless also due to intermediate restraints and movement, TCP wants to differentiate the creation of the inaccuracy so as to it can take the greatest suitable act for separately situation. Furthermore, the developing link and system cover algorithms for this kind of system can play a main part on TCP performance. Similarly, aspects like pathway irregularity (it might affected through inferior covers approaches, between more components) and congestion window dimension may also damage the accomplishment of this protocol [1]. Entirely these matters are addressed in aspect in the following segments.

TCP congestion control instrument has capability to decline their broadcast degree while system expressions overcrowding and sort practice of all leftover means else. The main indication of TCP is to analysis the system in imperative to regulate the obtainability of assets. It inserts packs at a growing degree obsessed by the system up until a pack loss is discovered, by which that assumes the network is covering congestion. At that moment the TCP correspondent shorten its congestion window (CWND), retransmits the lost packet and restarts broadcast at an inferior growing degree. If the sufferers persevere (no timely ACK received), at each retransmission the correspondent couples (up to 64s) its delay regulator, called Retransmission Time Out (RTO), and the situation can interval extensive for the ACK of the present packet existence transmitted. This instrument is known as exponential back off approach.

In Bit Error Rate (BER) Exponential Back off approach works well in a wired network. It is characteristically abstemiously little allowing some misdirected bundle to be conserved and used as a symbol of network congestion.

C. Network layer Control

Ad hoc networks encompass of an exceptionally active state and recurrent means differences are expectable, on TCP routing rules join in a major part performing as. Unlike the MAC layers, neither IEEE 802.11 protocol remained taked as the ordinary, I has been focus on extreme research scraps on ad hoc networks zone near calibration. Now, around have been a lot of deliberate guiding measures for this state and typically collectively of them has different significances on TCP.

1) Dynamic Source Routing (DSR):

DSR is an on-demand routing protocol in which host adapts new routes and changes in network as node moves and learns new routes [18]. It updates the correspondent node by a specific pathway by sending data packets for a list of nodes to update their routes. It keeps a cache of all the learned routes which causes overhead routes. Consequently, the middle nodes do not necessity to retain an upgrade period table of ways, thus evading periodic route announcements that source significant overhead. Problem with this method distress. The problem with this approach the huge chances of stale route in setting which mobility have middle restraints are usually exist. It can occur, for occasion, once an route response to back of sender but in which reply route can be when a route reply message is in its way back to the sender but the replied route is cannot be authorized due to other involved node it has been disturbed. The problematic is intensified via the element other nodes are eavesdrop the unauthorized path response and can rush their barriers through stale info. If the stale route can be perceived and improved in a speedy means, TCP can be run tack-off state, which will decline its presentation unsympathetically.

2) Ad Hoc On-Demand Distance Vector (AODV):

AODV is a reactive protocol for ad-hoc networks which provides routes for nodes on request [19]. Message takes place only when needed. AODV have a hop-to-hop procedure it is a mixture of on request and detachment route. Ob require demands means communication occurs only as they want detachment vector that means a link state protocol. In AODV a RREQ (Route Request) is guide to respectively and each node in the network. As they all have a middle nodes also valid then itself the route reply send to destinations. As they could not find any authorized route after that it send to the source node.

3) PATH asymmetry:

As TCP depend on spell subtle response info to make his movement controller, irregular trails can extremely cooperation its presentation. The fact is that, if TCP does not obtain appropriate ACKs it cannot enlarge its CWND to type use of the complete obtainable volume of the frequency in place, thus degenerative bandwidth. Hereafter, the frontward path features are also dissimilar as of the ones of the backward path, it would

reasonably possible look presentation difficulties. In an ad hoc network, anywhere the topology besides the setting circumstances can differ rather regularly and randomly, asymmetry can happen by dissimilar details, counting lesser layer approaches [20].

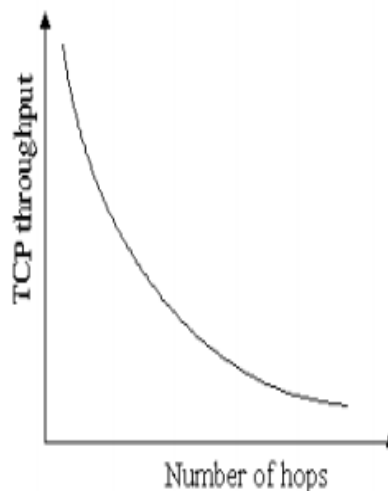


Figure 1: TCP Throughput vs Number of Hops

III. PROPOSED SOLUTION

Proposed network model consists of 30 mobile nodes moving in a random fashion in a 1500 X 300 m' rectangular flat area. Random Walk mobility model is followed by nodes in which all nodes are in signal through the imitation historical. We have measured 10. Different mobility patterns. For a given mobility pattern at different speeds, precisely same order of activities and link disappointments happens. The rapidity of each node is consistently dispersed in an interval of $0.9 \sim 1.1v$ for some mean speed v where $Y \in \{5, 10, 15, 20, 25\}$.

A. PERFORMANCE OF DIFFERENT VARIANTS OF TCP:

We associate the presentation of dissimilar transport protocols over wireless mobile ad hoc networks – TCP Reno, TCP Vegas and TCP SACK. Fig. 3 shows the Throughput vs. Speed graph for all variants of TCP when DSR is used. TCP Reno shows the best results as compared to Sack and Vegas. Initially, there is not much difference in the throughput but as the speed increases this difference becomes significant. TCP Sack is not able to match the performance of TCP Vegas. On the other hand, TCP Vegas doesn't show as good results as TCP Reno but for some speeds, it achieves nearly equal performance as TCP Reno. To summarize, TCP Reno shows the best performance whereas TCP Sack shows the worst performance. TCP Vegas achieves a good throughput but it is not able to match TCP Reno performance.

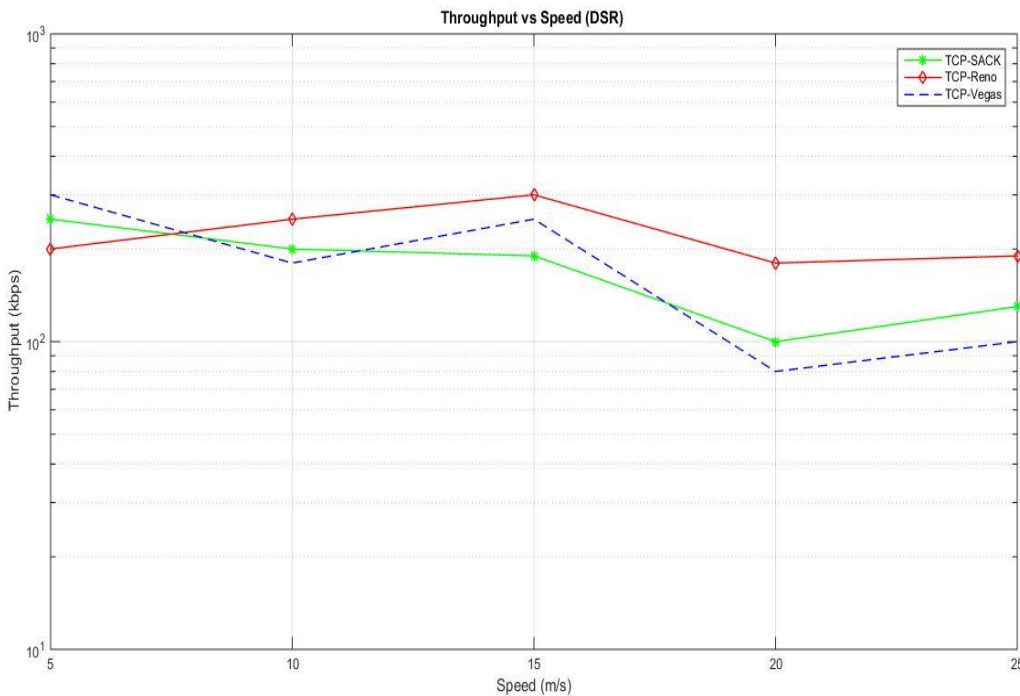


Figure 2: Throughput for all variants of TCP with DSR

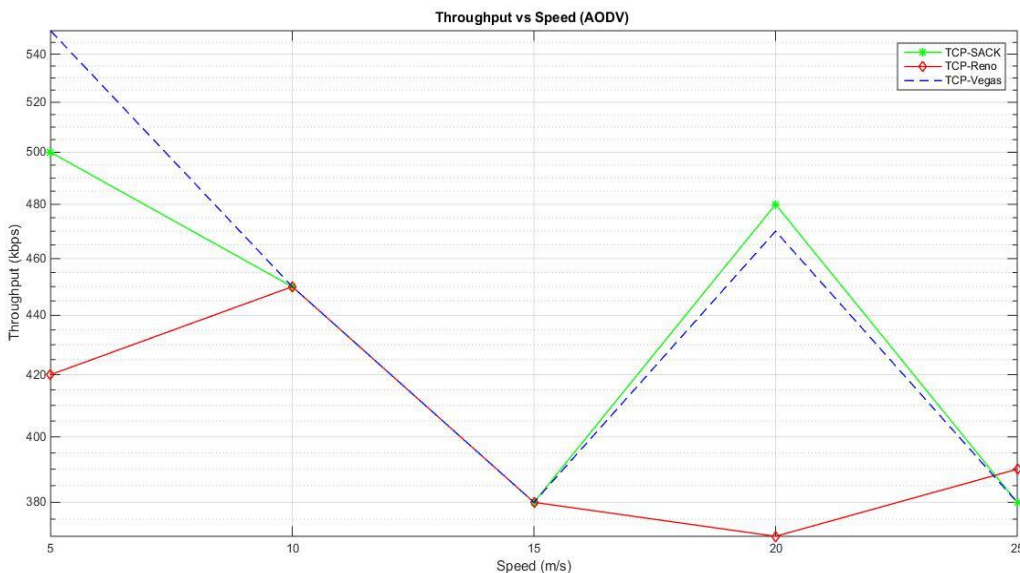


Figure 3: Throughput for all variants of TCP with AODV

Fig. 4 shows a similar graph when AODV is used. The improvement in the performance of all variants of TCP can be

easily seen in this case. Here, all three protocols achieve almost equal throughput with TCP Reno deviating only slightly. TCP

Vegas and TCP Sack show almost similar results for all speeds. At the highest speed of 25 m/s, slight improvement in the performance of TCP Reno is noted. Comparing the performance of DSR and AODV, there is nearly 100% improvement in the TCP performance when AODV is used as the underlying routing protocol. All three protocols show tremendous improvement in the throughput with TCP Reno not achieving throughput as good as throughput of TCP Sack and TCP Vegas. This clearly shows the improvement when AODV is used.

B. Comparison of AODV and DSR

In this section, we discuss the relative performance of two on-demand ad hoc routing protocols - AODV and DSR when the nodes are mobile. AODV outperforms DSR at all speeds. We have varied the speed from 5 m/s to 25 m/s i.e. 18 km/h to 90 km/h. This is fairly high speed for nodes in ad hoc networks, thus signifying the high stressful conditions. At 5 m/s, AODV obtains nearly 100% improvement in throughput. The best performance is achieved at 25 m/s. At such a high speed, AODV shows performance gain of 130% over DSR. DSR maintains all route entries in its route cache.

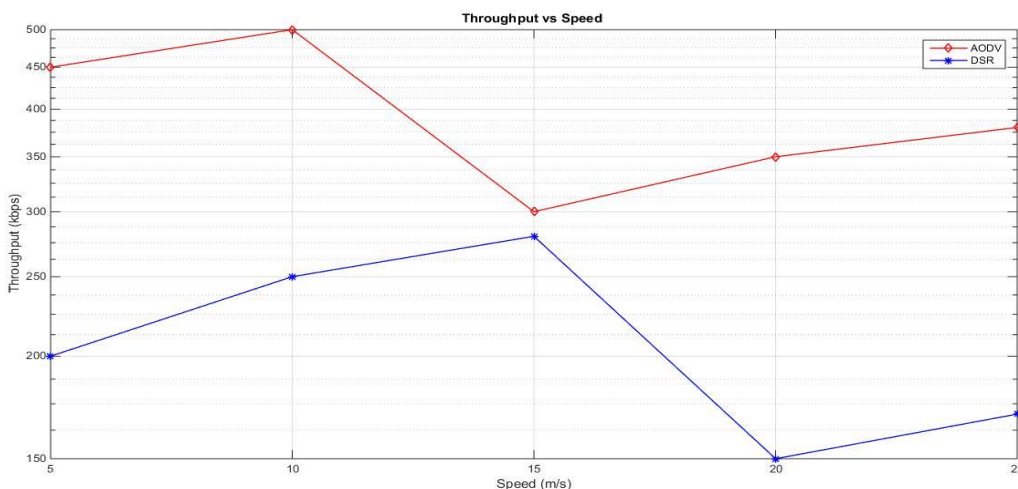


Figure 4: Comparison of AODV vs DSR

At great hustles, the ways in the route cache can go stale with a justly extraordinary incidence and mistaken routes might get shaped via route explorations introduced by the source [2]. Occasionally, stale routes in the cache may lead to routing loop creation. This can lead to degradation of TCP throughput especially at high mobility. If we consider the fraction of expected throughput for both routing protocols, AODV fares better here also. Although, expected throughput for both routing protocols for the static topology is almost equal, but under mobility, AODV outperform is DSR. So, it can be concluded that AODV is an improved protocol to be used in extraordinary mobile circumstances. Reason behind DSR poor performance is its excessive use of cache and storage of expire routes which causes problem to determine fresh routes in presence of multiple options. Although DSR generates much fewer routing packets overall, it generates more uni-cast routing packets which are not feasible in IEEE 802.11 MAC layer. Thus, DSRs apparent saving on routing load doesn't translate to an expected reduction on the actual load on the network. A constant byte overhead in each DSR packet further brings down the throughput. In AODV, way cache running is done finished cache entrance timeout. It confirms that only routes that are recently used are preserved in the route cache. By using this strategy the problem of a stale

route entry in the route cache is solved. AODV packets don't contain full route to the destination. So, a lot of byte overhead is prevented in this case.

IV. CONCLUSION

This paper discusses TCP in mobile ad hoc networks and the issues regarding mobility of nodes. It also explains why TCP performs poor for mobile nodes as it is unable to distinguish the packet drop due to mobility or congestion. Simulations performed for throughput and mobility for different variants of TCP. TCP Vegas however performs better as compared to TCP SACK or TCP Reno with AODV routing protocol whereas TCP Reno performs better when simulations are performed with DSR. In conclusion AODV has far more better performance as compared to DSR because DSR is source driven protocol which causes byte overhead in each packet and route cache also has detrimental effects on TCP performance.

REFERENCES

- [1] Ruy de Oliveira, Torsten Braun, TCP in Wireless Mobile Ad Hoc Networks, Institute of Computer Science and Applied Mathematics, University of Berne, Switzerland, July 2002.

- [2] Internet Engineering Task Force (IETF) Mobile AdHoc Networks (MANET) Networking Group Charter. <https://www.ietf.org/html.charters/manetcharter.html>.
- [3] G. Holland and N. H. Vaidya, "Analysis of TCP performance over mobile ad hoc networks," in Proc. of the Fifth Annual ACM / IEEE Conference on Mobile Computing and Networking (MOBICOM). August 1999.
- [4] D. Comer, Interworking with TCP/IP, vol.1, 3rd edition, Englewood Cliffs, N. J., Prentice Hall, 1995.
- [5] B. Postel, "Transmission Control Protocol," Internet Engineering Task Force, September 1981, RFC 793.
- [6] W. Stevens, "TCP slow start congestion avoidance, fast retransmit, and cut recovery algorithm," Internet Engineering Task Force, January 1997, RFC 2001.
- [7] W. Richard Stevens, TCP/IP Illustrated, Volume 1. The Protocol. Addison Wesley, 1994.
- [8] S.W. O' Malley and L. Peterson, "TCP Vegas: End-to-End Congestion Avoidance on a Global Internet" IEEE Journal of Selected Areas in Communications, 13(8): 1465-1480, Oct. '95.
- [9] E. Royer and C. Perkins, "Ad Hoc on demand distance vector routing" - IETF Internet Draft
- [10] David B. Johnson and David A. Maltz, "Dynamic source routing in ad hoc wireless networks" T. Imielinski and H. Korth. Editors, Mobile Computing, Kluwer Academic, 1996.
- [11] M. Gerla, K. Tang and R. Bagrodia, "TCP Performance in wireless multi-hop networks," in Proc. Of IEEE IVMCSA '99, New Orleans, LA, February 1999.
- [12] R. Sliorey et al, "Performance of TCP over different routing protocols in mobile ad hoc networks." Proc. Of IEEE Vehicular Technology Conference, 2000. Tokyo.
- [13] K. Cliandran et.al, "A Feedback-based Scheme for Improving TCP Performance in Ad Hoc Wireless Networks". Proc. Of CDCS '98. Amsterdam. May '98.
- [14] Jain Liu and Suresli Singli, "ATCP: TCP for Mobile Ad Hoc Networks", IEEE Journal of Selected Areas in Communications, July 2001.
- [15] M. Mathias, J. Mahdavi, S. Floyd and A. Romanow, TCP Selective Acknowledgement Options, IETF RFC 2018, October 1996.
- [16] M. Allman, V. Paxson and W. Stevens, TCP Congestion Control, IETF RFC 2681, April 1999.
- [17] L. Brakmo and L. Peterson. TCP Vegas: End to End Congestion Avoidance on Global Internet. IEEE Journal on Selective Areas of Communication, October 1995.
- [18] D. Johnson and D. Maltz. Dynamic source routing in ad hoc wireless networks, In T. Imielinski and H. Korth, editors, Mobile computing, Kluwer Academic. 1996.
- [19] S. Liu, Y. Yang and W. Wang, "Research of AODV Routing Protocol for Ad Hoc Networks, ELSEVIER, and AASRI Conference on Parallel and Distributed Computing and Systems, 2013.
- [20] H. Balakrishnan and V. N. Padmanabhan. How Network Asymmetry affects TCP. IEEE Communication Magazine, pages 2-9, April 2001.

Critical Issues and Challenges in Developing Mobile-Based Health Systems: Case of Tanzania

Frederick Henri Chali¹

PhD Cand: School of CoCSE,
Nelson Mandela African Institution of Science and
Technology, NM-AIST
Arusha, Tanzania
chalif@nm-aist.ac.tz

Eng. Dr. Zaipuna O. Yonah²

Senior Lecturer: School of CoCSE
Nelson Mandela African Institution of Science and
Technology, NM-AIST
Arusha, Tanzania
zaipuna.yonah@nm-aist.ac.tz

Dr. Khamis Kalegele³

Researcher:
Commission for Science and Technology, COSTECH
Dar es Salaam, Tanzania
Kalegs03@gmail.com

Abstract - Utilization of mobile devices worldwide is increasing every day. This has created an increasing demand for mobile apps, the development of which is not seamlessly without challenges. We conducted a research survey to identify and confirm the known challenges and issues that are faced by mobile app developers in their development work in developing countries particularly in Tanzania. This survey was done in three regions of Tanzania: Dar es Salaam, Dodoma and Arusha; because most of the mobile apps developers are located in these regions and eHealth information systems were used as a case for this purpose. Methodology used was literature, interview, questionnaire, and observation. This paper presents an overview of the identified challenges and issues facing mobile app developers and the proposed data exchange solution. These findings have justified our ongoing research on interoperability of multiple databases through a single mobile application.

Keywords-component; *Multiple Databases; Data Exchange; eHealth Systems; mHealth; mobile app.*

I. INTRODUCTION (HEADING 1)

One surveyed reports that about 40% of 7 billion people in the world are using internet [1]. Another study recently reported that, 80% among the internet users globally, own smart phones and 47% of them own Tablets [2]. Due to the increasing numbers of utilization of mobile devices, the world counts more than 170 million mobile apps in use by 2015 [3].

Said is differently, the potential to utilize mobile phones in developing countries is high. In 2012, about 70% of the global mobile subscribers out of more than 5 billion, belonged to low and mid income communities [4]. Banking and healthcare sectors are leading the way on harnessing this available potential in developing countries.

In mobile applications (mobile apps) development, challenges such as testing and supporting, security, device proliferation, development approach, etc. are some of the known challenges that continue to face apps development works [5], [6] and [7].

To confirm if these challenges are real, a survey was conducted in the health sector environment as a case specifically to identify and confirm the known issues and challenges in mobile apps development process focusing on connecting to medical record systems implemented with multiple databases integrated environment. In the end, we proposed a data exchange architecture that demonstrates the need for connection/integration between single mobile app and multiple heterogeneous databases. In section II we reviewed status of eHealth and mHealth systems in developing countries. Methodology used in our study is explained in section III and we have reviewed major issues and challenges that are generally facing mobile app development process worldwide in section IV. Section V contain survey results from our study and then we proposed data exchange component (DEC) architecture between single mobile app and multiple heterogeneous databases in section VI where we present our discussion in section VII. Finally in section VIII we provide conclusion and suggest further works.

II. EHEALTH AND MHEALTH SYSTEMS IN DEVELOPING COUNTRIES

eHealth systems have revealed a greater potential to reach large population for quality service provisions. It would easily reach and serve the population of 55% of the developing countries where the availability of enough health facilities and health care services is too low [8]. WHO define e-health as the transfer of health resources and health-care by electronic means [9]. In [10], It has been described that, e-health is the application of Internet and other related technologies in the healthcare industry to improve access, efficiency, effectiveness, and quality of clinical and business processes utilized by

healthcare organizations, practitioners, patients, and consumers in an effort to improve the health status of patients. It is also specifically described by Eysenbach, as an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies [11]. There are numerous definitions of eHealth from different dimensions worldwide; others have tried to extend the definition to any application of electronic means into the healthcare services.

Delivery of eHealth by use of mobile app services results into mobile health or mHealth. It is the utilization of mobile apps for the purpose of enhancing provision of health services. In [12], describes the term as the use of mobile phones for the purpose of improving the quality of care and efficiency of health services. Further Characteristics of eHealth and mHealth systems are described in the following subsections.

A. Infrastructure and Benefits of eHealth and mHealth in Developing Countries

Typically, Infrastructure that support eHealth system architecture include the following [13]:

- a) Internet
- b) Extranet
- c) Intranet
- d) Core Data Systems
- e) E-Mail
- f) Telecommunications
- g) Hardware

Mobile health can provide relevant benefits to health workers, health facilities, healthcare services etc. Modi, identifies the following benefits [12].

- a) Real time receipt of health-care information
- b) Timely addresses health issues
- c) Easy to track the prevalence of diseases and its rate globally
- d) Helps healthcare officials to be more proactive on addressing health issues rather than being reactive.
- e) Feasibly addresses gaps in health-care in remote areas.
- f) Provides Health Education Awareness.
- g) Raises treatment support and medication compliance.
- h) Improves the performance of healthcare workers.
- i) Supports diseases surveillance.

B. Status of eHealth in Some African Countries

Table 1 shows the results summary of the status of eHealth in some African countries. The summary shows the evidence of activity, planning or implementation of eHealth in those countries in terms of national ownership, foundation (Info-structure; the basic physical and

TABLE I. SUMMARY OF THE STATUS OF EHEALTH IN SELECTED AFRICAN COUNTRIES. (SOURCE: [14] - [15])

S N	Country	National Ownership	Foundation		Health Process Domain Component
			Info-structure	Infrastructure	
1	Angola	E _v P&I	E _v P	NoE _v	E _v I
2	Botswana	SE _v P	E _v P&I	E _v I	E _v I
3	Ethiopia	E _v N _s	E _v P	E _v N _s	E _v I
4	Ghana	SE _v P&I	SE _v P&I	E _v P	E _v I
5	Kenya	SE _v N _s	E _v P	E _v P	E _v I.
6	Nigeria	E _v P.	LE _v I.	NoE _v	LE _v I
7	Tanzania	SE _v N _s	E _v P&I	E _v P&I	E _v I.
8	Uganda	SE _v N _s	LE _v I.	SE _v P&I	E _v I.
9	Zambia	LE _v P	NoE _v .	E _v I	E _v I
10	Zimbabwe	E _v N _s	E _v P	E _v P&I	E _v I

Key:

- Evidence of = E_v
- Strong Evidence of = SE_v
- Little Evidence of = LE_v
- Planning = P
- Implementation = I
- National/Government Support = N_s
- No Evidence of Planning or Implementation = NoE_v

organizational structures and Infrastructure) and also health process domain component where it describes the availability of already running eHealth systems in a particular country.

C. eHealth and mHealth Systems in Tanzania; its Application

1. eHealth Systems

The application of computer systems in the health sector in Tanzania started in the 1990s. Deployment of information technology has brought revolution of the uses of computer applications from simple formats to complex systems.

However, eHealth is not well and completely implemented in developing countries; it has some challenges that technically need special attention to address. In their survey, Busagala and Kawono established challenges that are facing e-Health systems in Tanzania. The challenges include poor infrastructure for supporting health-care services, insufficient budget for ICT, unreliable electricity supply and lack of ICT skills for health-care workers [16].

2. mHealth Systems

In Tanzania more than 31 mHealth applications are currently providing health services countrywide [17]. Among them, the

TABLE II. MHEALTH APPLICATIONS ADDRESSED IN TANZANIA
(SOURCE: [17])

Application	Total Number
Data collection and reporting	18
Client education and behaviour change communication	15
Electronic decision support	14
Provider work planning and scheduling	8
Registries and vital events tracking	7
Electronic health records	7
Provider training and education	7
Provider to provider communication user groups, consultation	6
Service use supply chain management	5
Financial transactions and incentives	4
Sensors and point of case diagnostics (and Monitoring)	3
Human resource management	3
Tele consultation	3

most deployed applications are for data collection and reporting. Currently there are 18 such applications, approximated to be 58% of all mHealth applications in Tanzania. Table 2 lists other applications that are also deployed in Tanzania with their total number.

Some of the currently available mHealth projects in the country are briefly described in the following paragraphs.

- **SMS for Life**
It is an innovative mHealth project piloted in Tanzania under public-private partnership. It helps to reliably support availability of Malaria medication using mobile phone services. The project is led by Novartis and supported by several partners, each bringing specific skills [18].
- **Birth Registration by Mobile Phone**
This is a countrywide mHealth platform in Tanzania that gives parents ability to register their children's births by using mobile phones. It allows a health worker to send birth registration information such as baby's name, sex, date of birth and family details by phone to a central data base and finally a birth certificate is issued free of charge in a few days [19].
- **m4RH**
Known as Mobile for Reproductive Health (m4RH), is a low cost innovative system running in Tanzania, Kenya and Rwanda providing information on family planning. It is accessible from every mobile phone through short message service (SMS) or "text messaging" [20].

From the foregoing, it is hereby acknowledged that, mHealth can have a profound positive impact in health sector. It shows to have the ability to improve healthcare services even in remote areas and can free other resources for developing the country economy.

III. METHODOLOGY

A. Literature Review

In depth literature review was conducted to enhance more understanding of issues and challenges on mobile apps development of e-Health systems from multiple database integration point of view. More information was collected from previous researches to gather a wide knowledge about these issues and challenges, as presented in section II.

B. Sampling Technique and Data Collection

Data collection (including questionnaire, interview and observation) was done in three regions: Dodoma, Arusha and Dar es Salaam. Data were collected in Dodoma from the HPSS government project based in Dodoma [21]. In Dar es Salaam and Arusha data were collected from 78 programmers in three mobile development companies in total.

As a stimulus, mobile app utilization in health sector brings about the question concerning the technical aspect on the development of mobile apps for health care delivery. We intend to examine issues and challenges that facing mobile apps development in their integration with medical records. The following section describes major issues and challenges that are facing mobile app development process.

IV. MAJOR ISSUES AND CHALLENGES THAT ARE GENERALLY FACING MOBILE APP DEVELOPMENT PROCESS

From the literature review, we found a number of issues and challenges that are facing mobile application development processes worldwide. The identified issues and challenges are concerned with designs, security, user interface, networking, development skills, context etc. Among all these, security issues, user experiences, Vendor and device fragmentation are ranked to be the major ones [6]. These are further described below.

A. Security Issues

Security in mobile app development is speedily becoming a major concern. The primary concerns are: data upload and download wirelessly using insecure hotspots and from potentially insecure locations and the sizes of devices make them easily misplace-able. The degree level of security in mobile development in enterprises or organization also depends on the nature of a business. Bank application for instance, with bill payment

feature needs more security measures than an information data application [6].

B. The User Experience

Mobile user experience could be described as the perception an end user has on a mobile product or service [22]. Mobile devices are characterized by the smaller display and different styles of user interactions. These characteristics have an impact on interaction design for mobile apps. The user experience has a strong influence on application development. The developer has to consider user demands such as user interface, running speed of the application, etc in order for the end product to be well accepted. During development, there are some user experience factors that influence the process of development; these factors are such as widgets, touch, whether through stick or direct by finger, physical motion, keyboard (physical and Virtual) for designing user interface, and the context dependencies, such as proximity to other mobile devices and physical location [23]. Device characteristics such as size and its version are having a greater impact on the design of the application. Developers need to consider on how to build up an application that could highly meet user's expectations. The user experience could help make an application more successful by maximizing acceptance and usage.

C. Vendor and Device Fragmentation

Existence of multiple mobile application platforms such as iOS, Android, Windows Phone etc., is one of the big challenge in the development of mobile apps [5], [7], [24]. These mobile Platforms, rather than Unification, they are heading towards fragmentation instead [5]. Fragmentation in mobile apps could be grouped into two categories;

- 1) Fragmentation across platforms (Vendors/Platforms Fragmentation), this includes a variety of mobile OS such as Android, iOS, BlackBerry, Windows Phone etc.
- 2) Fragmentation within the same platform (Device Fragmentation) This includes different versions of the same platform [7].

Each platform requires separate skills to use in order to develop an application successfully. Different skills are needed for the developer to be able to develop and deploy the application successfully. This poses as a challenge during development and force developers to base their developing skills mostly on one platform.

V. SURVEY RESULTS

This section presents the results of the survey which in general confirm issues and challenges in mobile development in the connection/integration with remote multiple heterogeneous databases, Tanzania particularly as a case study.

The survey was conducted in three different regions in Tanzania. These regions are Dodoma, Dar es Salaam and Arusha. A total number of 78 mobile app developers in two regions, Dar es Salaam and Arusha were contacted to respond to the data collection tools (interview and questionnaire). Statistics revealed that country wide there are approximately 541 businesses whose services are IT related services such as Network Installation, hardware and software maintenance, IT equipment supplies, IT consultancy, web hosting and mobile app development. Among them, nearly 70% reside in Dar es Salaam and Arusha (Fig. 1).

Approximately 70 businesses out of 541 (about 12.9%) are doing mobile apps development (see Fig. 2). Since a large proportion of these businesses are small and medium enterprises (SMEs), it has been estimated that each has at least 3 developers on average. This accounts for 210 developers countrywide; the number does not include freelancer developers scattered in big towns and cities in the country who are mostly fresh university and college graduates. Table 3 shows a summary of results that came out of the survey.

A. Platform Consideration and Device Fragmentation

Mobile applications run in mobile operating system (mobile OS) platforms such as Android, iOS, Windows Phone, Black Berry etc. Developers mentioned that they normally choose a particular type of platform when they start the development of the application. This is very important to them since developer's choice is based on own/individual competencies. About 14% of them mentioned this.

Device fragmentation is one of the biggest challenges facing the Tanzanian developers today. Developers, about 27%, show this as their challenge. Because Android has the biggest share in sub-Saharan countries which include Tanzania [17], most of the developers are developing mobile app based on Android, 85% of them. The presence of different versions of Android OS creates difficulty in developing a multiplatform application. Vendor/Platform fragmentation challenge is not faced by most of developers in Tanzania since most of them are working on Android rather than other platforms, Figure 4 shows the details.

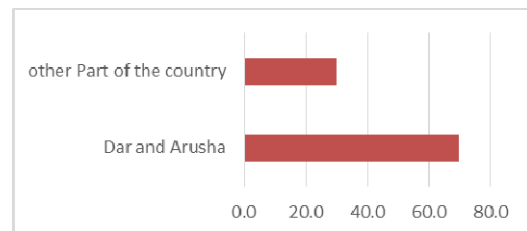


Figure 1. Availability of IT related businesses in Tanzania

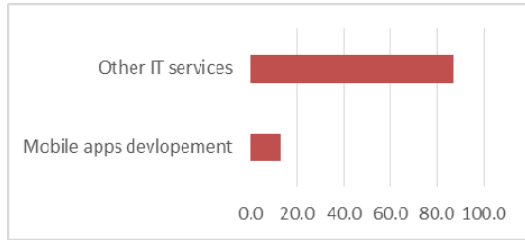


Figure 2. Distribution of IT related service businesses in Tanzania

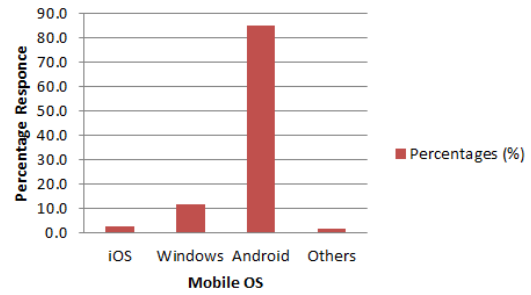


Figure 3. Mobile OS choices of mobile app developers in Tanzania.

B. User Experience and Mobile Device Capabilities

User experience was also mentioned by developers as one of the issues and challenges to the mobile app development process. From the survey, 24% of developers mentioned this. They utilized different application design tools and techniques to build up an app. They commented that they sketch the UI, database model, and User models etc as their techniques of designing good and interactive mobile apps. Others said they normally visit design samples from websites such as materialup, dribbble, material pallet etc to get a clue on the current graphic, web designing etc. It is difficult to understand the users taste and experience on the mobile applications.

They usually use adobe Photoshop for User Interface (UI) design, UML tools such as use case, activity, state diagrams etc as the tools for design. For simplicity and rapidness, they use Java programming language since this language is used by a wider community of mobile developers.

Mobile capabilities in terms of size, capacity, speed etc, are also challenges mentioned on the development of mobile apps. Developers also confirmed that, they consider capability of mobile device such as limited power and size during development.

TABLE III. ISSUES AND CHALLENGES FACING MOBILE APPS DEVELOPMENT PROCESS WHEN CONNECTING THE APP WITH REMOTE DATABASE(S).

S/N	THEME/ CATEGORY	Percentage (%)
1	Database adapters and incompatibility	2.6
2	Handling BLOB objects/data	6.4
3	Multiple databases integration	20.5
4	Platform consideration	14.1
5	User Experience and Mobile Device Capabilities	24.4
6	Device fragmentation	26.9
7	Vendor/Platform fragmentation	1.3
8	Other Challenges	3.8
Total		100

C. Multiple Heterogeneous Databases Integration

Databases are available in different types and technologies eg. Oracle, MySQL, SQL Server, etc. For the developer to link with only one among these external databases at a time with mobile app is very possible and in normal practice is always done. For instance when a developer intends to connect the app with MySQL database, he should configure HTTP POST library for the connection to be successful, when the intention is to utilize SQL server database, then s/he must download jTDS (JarFile, JDBC) Libraries, configure it and then connection will be established.

However, a challenge arises when the need is to link two or more databases of both the same or different technologies, with a single mobile app so as they can share information all together. Here, heterogeneity of the databases (Oracle, MySQL, SQL Server etc) was pointed out to be a big challenge. 21% among the developers said this.

The following sub-section presents current practices taken by mobile app developers when connecting single mobile app with multiple database systems.

D. Current Practice on the Connection Between Single Mobile App to Multiple Heterogeneous Databases

From the results of the conducted survey, we realized that, there is an existing challenge in mobile application development on the integration between a single mobile app with multiple databases (multiple database integration). We observed that, mobile app developers are comfortably and easily connecting mobile apps with remote databases in one to one connection fashion.

Our observation in this study revealed that, currently, developers connect a mobile app separately when they have a need to connect it to multiple databases. Fig. 4 illustrates the connection of single mobile application to two different databases of two different technologies, MySQL and SQL Server. Developer will not be able to connect it (at the same time) in both databases with a single connection. He must establish two different connection (two pipelines for data flow) using two different components in order to succeed on this demand.

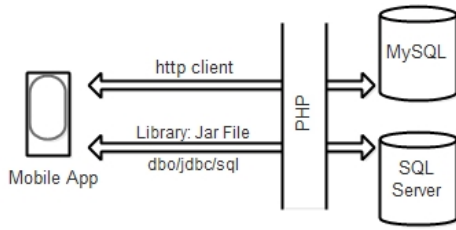


Figure 4. Connection between mobile app and two heterogeneous database systems

HTTP client, a client side HTTP transport library that provides basic functions for accessing resource via HTTP protocol [25], is used in the connection between mobile apps with MySQL database. Jar file library [26] is used in the connection between mobile application with SQL server, other libraries which serves the purpose for this connection are such as DBO (database Object) and JDBC (Java Database connector) [27]. All these connections regardless of technologies connect to each database by utilizing PHP functions to enable their connections.

Problems in mobile apps development stem from versioning, Incompatibility and Open source [28]. Our focus was to build up a mechanism where connectivity between these components (mobile apps and multiple databases) will become more convenient and seamless regardless of the technology the database is using. The developer should be able to create mobile app on the client side using same approach without being concerned about the kind of database is connecting to. Fig. 5 provides the illustrated concept of our focus.

E. Existing Data Exchange Approaches for Single Mobile App - Multiple Heterogeneous Databases/Information Systems

There is much advancement in programming that currently provide services similar to our focus i.e to support development of mobile apps working with remote databases/information systems. These approaches do not deal with legacy/existing databases rather they provide space to create or connect to the newly created databases [29], [30], and [31]. Firebase is among these approaches [32].

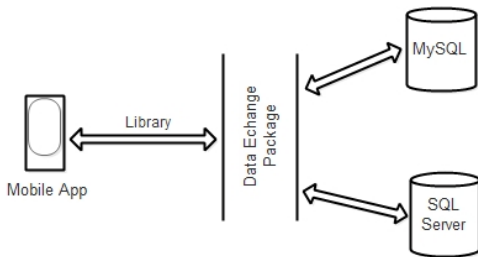


Figure 5. Architecture model for data exchange between single mobile app and multiple databases

Some of these development efforts are such as, the REST API (Application Programming Interface), a lightweight and scalable API to support mobile development which is extremely easy to build, integrate, test, extend and maintain [29]. The RapidSMS for extending system servers to sms applications [30]. Another one is the HIE API which resides at the mid of Health Information Mediator (HIM) in the Rwandan NHIS [31]. Firebase, a cloud based database service that allows mobile app developer to store and sync data across multiple clients. It provides client libraries that enable integration with android, IOS etc. Data sync across all clients in real time using NoSQL database [35]; [32].

Joseph et al, 2012 have implemented a cross-platform enterprise mobile framework to provide development environment for mobile applications at Intel enterprises. Other framework such as MADP and MEAP are designed to provide support on the same goal [36]. These efforts are already taking places worldwide.

VI. PROPOSED DATA EXCHANGE COMPONENT (DEC) ARCHITECTURE BETWEEN MOBILE APP AND MULTIPLE HETEROGENEOUS DATABASES

As alluded to in the previous sections, multiple database integration is one of the challenges facing mobile app development such that mobile app developers are not able to connect single mobile app with multiple databases.

In our approach we propose to build up an eHR-DE/ data exchange component (DEC) that will provide a connection interface to mobile application on one side and on the other side to multiple databases. The DEC provides a solution toward multiple system interoperability and also provides interface to the already developed mobile app to connect to the already connected multiple databases.

Three basic interoperability dimensions need to be considered when developing interoperability solutions for different systems, [33] and [34]. The three dimensions are; barriers (*conceptual, technological, organization*), levels (*legal, organizational, semantic, technical*) and concerns (*data, process, service, business*).

From these interoperability dimensions, our work is addressing the technological barrier by establishing information sharing between heterogeneous electronic health records and single mobile app through data exchange framework. In the interoperability levels, we are focusing on the technical aspects and in the interoperability concerns dimension we are sharing information from different bases i.e data interoperability. We present our focus in Fig. 6.

The design architecture (DEC) shown in Fig. 7 where we create a data exchange package/component to connect multiple heterogeneous databases with single mobile app.

Barriers	CONCEPTUAL	TECHNOLOGICAL	ORGANISATIONAL
Concerns			
BUSINESS			
PROCESS			
SERVICE			
DATA		eHR-DE	

* Level: Technical Interoperability Level.

Figure 6. EHealth data exchange (DEC) deal with technological barrier, data interoperability concern and technical interoperability level.

Utilization of developed DEC will be such that: a mobile app developer will download the DEC then connect a developed mobile app into it through an interface provided by the component. The component contains PHP and PHP/ODBC connectivity technologies that provide connection interface to the developed mobile application. The mobile app is connected to the DEC through http/url connect. The provided interface on the other side also has provision to connect multiple databases.

VII. DISCUSSION

Interoperability is one of the challenges existing in the healthcare domain worldwide. “Interoperability is the ability of different ICT systems and software applications to communicate, to exchange data accurately, effectively, and consistently and to use the information that has been exchanged” [37]. In [38], it is reported that, lack of coordination at all levels of systems development contribute about 62.5% on the source/causes of interoperability problem among eHealth systems in Tanzania. Another source is the common use (in hospitals and other healthcare facilities) of open source systems which were not specifically designed according to our context and environment and were not created with integration focus in mind.

Achieving interoperability from mobile developers’ side is also an existing challenge. The process of integrating two or more health information systems through single mobile app is not easily achievable. Multiple database integration through mobile app is one of the ways to achieve interoperability of existing eHealth systems. This is among the challenges facing mobile app developers confirmed by our study.

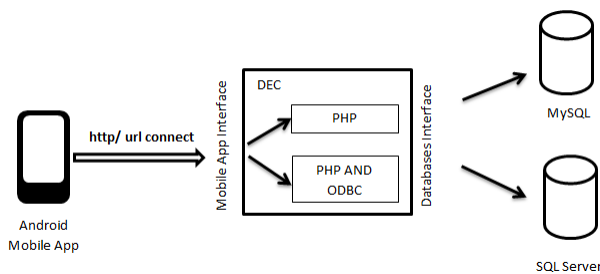


Figure 7: Data exchange (DEC) architecture connecting single mobile app with multiple databases.

Our main focus was to establish/enable interoperability of existing fragmented eHealth systems through mobile apps. In [39], a lists of solutions is provided that can be adapted to achieve complete interoperability in healthcare. Connection of legacy (existing) systems to healthcare network is among the solutions listed. This is lining up with our hypothesis that “mobile app development could support interoperability of existing/legacy multiple eHealth systems”.

The architectural model presented in Section V provides a possibility toward achieving interoperability in eHealth systems. We have proposed the DEC architecture in Section VI to show how the integration between single mobile app and multiple databases could be evidently achieved.

Our reviews shows that, health sector in Tanzania is characterized by fragmented electronic health information systems (eHealth Systems) with significant barriers to the effective sharing of information between healthcare participants [40]. The eHealth strategic plan of 2013-2018, describes that, infrastructure building blocks is one of the strategic objective for eHealth development. The report provides a reason for researching on the kind of required infrastructure to enable electronic sharing of health information across the Tanzanian health sector. The strategic plan shows the need and importance of implementing health information exchange that leverages mobile technologies [40].

VIII. CONCLUSION AND FURTHER WORKS

In this paper, we identify and confirm issues and challenges that are facing mobile application developers in developing countries focusing on connecting multiple eHealth systems with a single mobile application. Benefits and infrastructure of eHealth in developing countries also reviewed. We have identified and confirm the issues and challenges that are facing mobile app developers in developing their apps for eHealth systems. We have proposed the DEC architecture for solving a single mobile app to multiple databases. Also we discussed the interoperability challenge among eHealth systems and suggested that multiple database integration through single mobile app is a possible solution towards mitigating the interoperability challenge. Some issues and challenges in mobile app development confirmed in Tanzania are; Platform Consideration and Device Fragmentation, User Experience and Mobile Device Capabilities, and Multiple databases integration. These results are subsets of issues and challenges that are facing mobile developers in developing countries particularly in Tanzania. They serve to justify the need for wider efforts to be undertaken by mobile app developers and other stakeholders especially government to provide better environments to the mobile apps development processes in Tanzania.

REFERENCES

- [1] B. Sanou, “ICT Facts & Figures. The world in 2015.,” *Itu 150 Años (1865 - 2015)*, p. 6, 2015.
- [2] M. Jason, “80% of internet users own a smartphone,”

- Global Web Index, 2015. .
- [3] Statista, "Number of mobile app downloads worldwide from 2009 to 2017 (in millions)," 2015. [Online]. Available: <http://www.statista.com/statistics/266488/forecast-of-mobile-app-downloads/>.
- [4] SocialCubix, "Scope of Mobile Apps in Developing Countries – A Bright Future," 2015. [Online]. Available: <http://www.socialcubix.com/blog/scope-of-mobile-apps-in-developing-countries-a-bright-future>. [Accessed: 19-Mar-2016].
- [5] P. K. Mona Erfani Joorabchi, Ali Mesbah, "Real Challenges in Mobile App Development," 2013 *ACM/IEEE International Symposium*, 2013. [Online]. Available: <http://www.ece.ubc.ca/~amesbah/docs/mona-ese13.pdf>. [Accessed: 04-Sep-2015].
- [6] M. Katz, "Top Three Challenges in Enterprise Mobile App Development in 2015 - DZone Mobile," 2015. [Online]. Available: <https://dzone.com/articles/top-three-challenges>. [Accessed: 19-Sep-2015].
- [7] L. Shevchik, "Top 3 Challenges Facing Mobile Development - New Relic blog," 2012. [Online]. Available: <https://blog.newrelic.com/2012/11/15/top-3-challenges-facing-mobile-development/>. [Accessed: 22-Sep-2015].
- [8] IFAD, "Facts & Figures," *Rural Povert Report 2011*, 2011. [Online]. Available: <https://www.ifad.org/documents/10180/c1bbf5fabdc3-4ea6-9366-d163b95b1180>. [Accessed: 12-Apr-2016].
- [9] WHO, "WHO | E-Health," 2015. [Online]. Available: <http://www.who.int/trade/glossary/story021/en/>. [Accessed: 14-Dec-2015].
- [10] M. Jennifer, "E-Health: Navigating the Internet for Health Information Healthcare," *Healthc. Inf. Manag. Syst. Soc.*, 2002.
- [11] G. Eysenbach, "What is e-health?," *J. Med. Internet Res.*, vol. 3, no. 2, p. E20, Jan. 2001.
- [12] S. Modi, "Mobile health technology in developing countries: The case of Tanzania," *Pepperdine Policy Rev.*, vol. 6, no. 5, 2013.
- [13] J. M. Deluca and R. Enmark, "E-Health: The Changing Model of Healthcare," *Front. Health Serv. Manage.*, vol. 17, no. 1, p. 3, Oct. 2000.
- [14] R. Foster, "Review of Developing Country Health Information Systems," 2012.
- [15] E. Afarikumah, "Electronic health in ghana: current status and future prospects," *Online J. Public Health Inform.*, vol. 5, no. 3, p. 230, 2014.
- [16] L. Busagala and G. Kawono, "Underlying Challenges of E-Health Adoption in Tanzania," *Int. J. Inf. Commun. Technol. Res.*, vol. 2, no. 1, pp. 34–41, 2013.
- [17] GSMA, "Mobile for Development," 2014.
- [18] Novartis, "Novartis Malaria Initiative: SMS for Life," 2014. [Online]. Available: <http://malaria.novartis.com/innovation/sms-for-life/index.shtml>. [Accessed: 11-Dec-2015].
- [19] UNICEF, "Tanzania rolls out birth registration by mobile phone – Stories of UNICEF Innovation," 2015. [Online]. Available: <http://www.unicefstories.org/2015/10/16/tanzania-rolls-out-birth-registration-by-mobile-phone/>. [Accessed: 11-Dec-2015].
- [20] FHI 360, "PROGRESS: Mobile for Reproductive Health (m4RH) | FHI 360," 2015. [Online]. Available: <http://www.fhi360.org/projects/progress-mobile-reproductive-health-m4rh>. [Accessed: 11-Dec-2015].
- [21] Swiss TPF, "Swiss TPH : Project Details," 2015. [Online]. Available: http://www.swisstph.ch/resources/projects/project-details.html?tx_x4euniprojectsgeneral_pi1%5BshowUID%5D=514. [Accessed: 14-Dec-2015].
- [22] M. Rouse, "What is mobile experience ? - Definition from WhatIs.com," 2014. [Online]. Available: <http://searchcio.techtarget.com/definition/mobile-experience>. [Accessed: 19-Sep-2015].
- [23] A. I. Wasserman, "Software Engineering Issues for Mobile Application Development," 2010. [Online]. Available: <http://www.cmu.edu/silicon-valley/wmse/wasserman-foser2010.pdf>. [Accessed: 04-Sep-2015].
- [24] Propelics, "Avoid These 6 Mobile App Development Challenges," 2014. [Online]. Available: <http://www.slideshare.net/propelics/propelics-six-challenges-formobileappdevelopmentwebinar032014>. [Accessed: 22-Sep-2015].
- [25] Apache, "HttpClient," 2011. [Online]. Available: <http://hc.apache.org/httpclient-3.x/>. [Accessed: 10-Aug-2016].
- [26] Oracle, "JAR File Overview," 2016. [Online]. Available: <https://docs.oracle.com/javase/8/docs/technotes/guides/jar/jarGuide.html>. [Accessed: 11-Aug-2016].
- [27] B. Hopwood, M. Mellor, and G. O'Brien, "Sustainable development: Mapping different approaches," *Sustain. Dev.*, vol. 13, no. 1, pp. 38–52, 2005.

- [28] J. Stine, "Mobile application Development Platforms," pp. 23–25, 2013.
- [29] F. Andry, L. Wan, and D. Nicholson, "A Mobile Application Accessing PatientsHealth Records Through a REST API," ... *Conf. Heal.* ..., 2011.
- [30] F. Carl, "RAPIDSMS.," *Jembi Project*, 2014. [Online]. Available: <https://jembiprojects.jira.com/wiki/display/RHEAPILOT/RapidSMS>. [Accessed: 16-Oct-2014].
- [31] J. Ryan, C., Deshendran, M., Anban, P., & Christopher, "An Interoperability Architecture for the Health Information Exchange in Rwanda.," Cape Town, South Africa, 2014.
- [32] J. Poyzner, "Risebird Blog — The end of back-end as we know it!!! (AKA...)," 2014. [Online]. Available: <http://risebird.tumblr.com/post/98293128515/the-end-of-back-end-as-we-know-it-aka>. [Accessed: 16-May-2015].
- [33] D. Chen and N. Daclin, "Framework for enterprise interoperability 1," *Proc. IFAC Work. EI2N*, no. August 2016, pp. 77–88, 2006.
- [34] W. Guedria and E. Lamine, "HEALTH SYSTEMS INTEROPERABILITY : ANALYSIS AND COMPARISON Wided Guedria , Elyes Lamine , Hervé Pingaud To cite this version :," 2015.
- [35] Google, "Firebase Realtime Database | Firebase," 2016. [Online]. Available: <https://firebase.google.com/docs/database/>. [Accessed: 09-Sep-2016].
- [36] S. Joseph , D Aideen , M John, S & Ian, "Building a Mobile Application Development Framework.," *IT Best Practices Cloud Computing and Compute Continuum.*, 2012.
- [37] D. Asuman, N. Tuncay, O. Alper, L. Gokce, K. Yildiray, and E. Marco, "Key Issues of Technical Interoperability Solutions in eHealth and the RIDE Project," *Softw. Res. Dev.*, pp. 1–11, 2006.
- [38] A. Kajirunga and K. Kalegele, "International Journal of Computer Science & Information Security," *October*, vol. 9, no. 10, pp. 49–52, 2015.
- [39] O. Iroju, A. Soriyan, I. Gambo, and J. Olaleke, "Interoperability in Healthcare: Benefits, Challenges and Resolutions," *Int. J. Innov. Appl. Stud.*, vol. 3, no. 1, pp. 262–270, 2013.
- [40] URT-MOHSW., "Tanzania National eHealth Strategy.," Dar es Salaam, 2013.

Perceptions of crime behavior and Relationships: Rough Set Based Approach

Luai Al Shalabi
Arab Open University
Kuwait city, Kuwait
lshalabi@aou.edu.kw

Abstract— As the crime rate increases, the focus is on prediction or prevention the crimes with minimum cost and effort. The source targets are to find: firstly, the minimal reduct that reflects the interesting relationships about the crime which save the investigator effort as well as the time of discovering or preventing the crime, secondly, the best classifier for crime data. A comparison between two important feature selection methods were conducted, the Chi-square and the Rough set. An experiment to obtain better supervised classification learning algorithms to predict crime status by using the two studied feature selection methods was employed and tested against the crime real dataset. Five classifiers have been conducted to find such prediction and four metrics are measured during testing the classifiers, accuracy, precision, recall, and AUC. Results indicate that rough set method of feature selection is more promising than chi-square method and the performance of mining results is improved by using Rough set feature selection method. Results also show that SVM is the best classifier among the five tested ones in term of accuracy and AUC. The importance of these findings in terms of identifying interested relationships in crime data was highlighted. Empirical results were presented in order to guide intended audience about future decisions or research regarding this study.

Key words: Crime data; Chi-square; Rough set; Reducts; Classifiers; RapidMiner;

I. INTRODUCTION

Data mining is the process for dealing with problems such as classification, association, and prediction. The goal of data mining process is to predict a data mining model with some certainty while analyzing a small specific representative part of the data. Rough Set, Neural Networks, Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Naïve Bayesian, and Decision Tree are some examples of classifiers that discover knowledge.

One of the data which needs data mining techniques to discover and predict underlying patterns are crime data [1]. Data mining techniques are able to identify patterns rapidly for detecting future criminal actions [2]. Data mining methods accelerate crime analytics, provide better analysis and produce

real-time solutions to save considerable resources and time [3]. Data mining provides data preprocessing techniques, data analysis techniques, and attractive predicted patterns which respond to the difficult crime cases as well as the simple ones.

To analyze crimes, there are several characteristics such as different races in a society, income groups, age groups, family structure (single, divorced, married), level of education, the locality where people live, number of police officers allocated to a locality, number of employed and unemployed people among others [4].

II. PREVIOUS WORK

Buczak and Gifford [3] stated that local, regional, national, and international crime play important roles in allocating law enforcement resources and influencing investigative priorities across jurisdictions. For example, an aggravated assault is a local jurisdiction matter, whereas drug trafficking and terrorism exhibit regional and global implications. Local crime patterns may differ from surrounding communities, creating localized trends of criminal activity which are unique to a community. Similarly, certain crimes may be more probable in locations with higher populations and dense housing. Regional crime patterns can be discovered which enable law enforcement personnel and criminal investigators to address large-scale trends.

There exist several applications of crime data analysis which have been studied by the research community. Most efforts focus on crime nature, severity, location, duration, and frequency. Following are several current trends in crime analysis [3]:

- Geospatial, map-based visualization.
- Geographical clustering of crime activity, such as identifying hot spots.
- Serial criminal behavioral pattern profiling and criminal career analysis.
- Gang criminal network analysis.
- Data stream anomaly, novelty, or outlier detection.

- Temporal analysis of crime patterns, such as crime sprees (temporal association of crime from an individual or group).
- Linking threats to risk of critical infrastructure based on vulnerability assessments.

Association mining is one of the methods that some researchers have used to discover the underlying novel patterns on a large volume of crime data. Anna and Christopher proposed a fuzzy association rules mining application for community crime pattern discovery. The application produced interesting and meaningful rules at regional and national levels and, to extract novel rules, a relative support metric is defined [3].

Shojaee et al. conducted an experiment to obtain better supervised classification learning algorithms to predict crime status by using chi-square feature selection method tested on real dataset [5]. They stated that the performance of mining results is improved by using chi-square feature selection technique. Chen et al. presented a framework for crime data mining [6]. The authors categorize and discuss levels of implication for various types of crime established via consult from an experienced local detective. General techniques for analyzing crime data are summarized, such as entity extraction, clustering, association rule mining, and sequential pattern analysis. Many researchers were studied the semantic analysis and text mining for entity extraction from free-text narratives, police reports, and FBI bulletins including [7-9]. Bagui performed knowledge discovery by mining association rules, training a classifier for prediction, and utilizing clustering methods using a US State database [10]. Yu, Ward, Morabito, & Ding employ an ensemble of data mining classification techniques for crime forecasting [11]. Several classification methods that are included in the study are One Nearest Neighbor (1NN), Decision Tree (J48), Support Vector Machine (SVM), Neural Network (Neural) with 2-layer network, and Naïve Bayesian (Bayes). Nath uses a geospatial approach at discovering crime patterns by clustering and displaying regional crime on a map [12]. He applies clustering by considering the geographical approach which shows regional crimes on a map and clusters crimes according to their types by using a combination of *K*-means Clustering Algorithm and Weighting Algorithm. Phillips and Lee work on clustering and graph representations for identifying similar crimes and for visualization purposes [13,14]. They stated that cluster size, shape, and distribution can aid in inferring details about related crimes. They also stated that clustering is also utilized to group classes of criminals. Brown creates the Regional Crime Analysis Program (ReCAP) which provides crime analysts with both data fusion and data mining, to aid Virginia law enforcement for capturing professional criminals in their own region [15]. Ng et al. developed an incremental mining algorithm, called ITAR, for crime pattern discoveries via temporal association rules [16]. Redmond and Baveja propose a Crime Similarity System (CSS) to assist police departments for developing a decision-making strategic view

point [17]. The system makes a list of communities by using the cities' enforcement profiles, crime and socioeconomic for obtaining knowledge from past experiences. Abraham and De Vel propose a method to realize criminals' behaviors by using computer log files for seeking some relationship among data and produced profiles which are used to understand the behaviors [2].

III. DATA PREPROCESSING

Data pre-processing is an important step in the data mining process. Data pre-processing includes data cleaning which is important since data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g. Age: 500) and impossible data combinations (e.g. Sex: Male, Pregnant: Yes), missing values, normalization, transformation, feature extraction, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis [18]. In this work, missing data, normalization, and feature selection (reduction of the data set) will be highlighted and resolved.

A. Missing values

Missing values in incomplete data sets is a challenge [19]. Missing values usually affect the accuracy of a data mining system. It has been stated in many studies that the representation of imputed dataset may no longer be good and it may lead to the solutions that are far from optimal [20]. Missing data might occur because the value is not relevant to a particular case, could not be recorded when the data was collected, or is ignored by users because of privacy concerns [21].

Dempster, Larid, & Rubin proposed methods for solving missing data in datasets [22]. They summarized these methods in three categories. First: Ignoring and discarding data by deleting all records that have missing data. Second: Parameter estimation which is used to find the parameters for the complete data. This method is use the Expectation-maximization algorithm for handling the parameter estimation of the missing data. Third: Imputation technique which replaces the missing values based on estimated values that are the most probable values. The First method will be used in this study as it will be discussed later.

B. Normalization

Normalization may improve the accuracy and efficiency of mining algorithms involving distance measurements such as neural networks, nearest neighbor, and clustering classifiers. Such methods provide better results if the data to be analyzed have been normalized, that is, scaled to specific ranges such as [0.0, 1.0] as in min-max method. If using the neural network back propagation algorithm for classification mining,

normalizing the input values for each attribute measured in the training samples will help speed up the learning phase. For distanced-based methods, normalization helps prevent attributes with initially large ranges from outweighing attributes with initially smaller ranges [23].

There are many methods for data normalization includes min-max normalization, z-score normalization, and normalization by decimal scaling. Normalization can change the original data and it is necessary to save the normalization parameters (the mean and the standard deviation if using the z-score normalization and the minimum and the maximum values if using the min-max normalization) so that future data can be normalized in the same manner. Min-max normalization was used in order to transform the crime dataset values from large scale values into small scale values between [0.0, 1.0].

Min-max normalization performs a linear transformation on the original data. Suppose that \min_a and \max_a are the minimum and the maximum values for attribute A. Min-max normalization maps a value v of A to v' in the range $[\text{new-min}_a, \text{new-max}_a]$ by computing [19]:

$$v' = ((v - \min_a) / (\max_a - \min_a)) * (\text{new-max}_a - \text{new-min}_a) + \text{new-min}_a$$

C. Feature selection (Reduction of Data Set)

One common challenge in the data mining field is to find the minimal reduct by using a process of feature selection which selects a subset of the attributes occurring in the training set and using only this subset as features in classification. This subset should have a quality rate similar to the original dataset. According to Al-Shalabi et al., the data in the information system can be used to discern classes only to a certain degree. Not all attributes may be required in order to be able to do so. Therefore, discovering dependencies between attributes enables the reduction of the set of attributes [24]. The main advantages for using feature selection process comes from the fact that it makes training and applying a classifier more efficient by decreasing the size of the effective attributes. Also, it makes the training faster and it often increases classification accuracy by eliminating noisy features. Chi-square and rough set reduction methods are used and compared.

Chi-Square: Chi-square (χ^2) is one common feature selection method. It is used to test whether the occurrence of a specific term and the occurrence of a specific class are independent. High scores on χ^2 indicate that the null hypothesis (H_0) of independence should be rejected and thus that the occurrence of the term and class are dependent. If they are dependent then the feature will be selected for classification. The null hypothesis (H_0) is tested by chi-squared formula:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

where O_i is an observed frequency and E_i is an expected (theoretical) frequency, asserted by the null hypothesis (Accept or Reject).

From statistical point, the Chi-Square feature selection is quite hard to reach statistical significance. Thus we should expect that out of the total selected features, a small part of them are independent from the class.

Rough Set: Rough set theory which was introduced by Pawlak in 1982 is an important theory for classification problems [25]. The theory is more powerful in solving problems of data reduction, discovery of data dependencies, and dealing with missing values. Reduction based on rough set theory will be conducted in this study.

In rough set theory, an information system is denoted in 4-tuple by $S = (U, A, V, f)$, where U is the universe of discourse, a non-empty finite set of N objects $\{x_1, x_2, \dots, x_N\}$. A is a non-empty finite set of attributes such that $a : U \rightarrow Va$ for every $a \in A$ (Va is the value set of the attribute a).

$$V = \bigcup_{a \in A} Va$$

$f : U \times A \rightarrow V$ is the total decision function (also called the information function) such that $f(x, a) \in Va$ for every $a \in A, x \in U$. The information system can also be defined as a decision table by $S = (U, C, D, V, f)$. For the decision table, C and D are two subsets of attributes. $A = \{C \cup D\}$, $C \cap D = \emptyset$, where C is the set of input features and D is the set of class indices. They are also called condition and decision attributes, respectively.

Let $a \in C \cup D, P \subseteq C \cup D$. A binary relation $IND(P)$, called an equivalence (indiscernibility) relation, is defined as follows:

$$IND(P) = \{(x, y) \in U \times U \mid \forall a \in P, f(x, a) = f(y, a)\}$$

The equivalence relation $IND(P)$ partitions the set U into disjoint subsets. Let $U/IND(P)$ denote the family of all equivalence classes of the relation $IND(P)$. For simplicity of notation, U/P will be written instead of $U/IND(P)$. Such a partition of the universe is denoted by $U/P = \{P_1, P_2, \dots, P_i, \dots\}$, where P_i is an equivalence class of P , which is denoted $[x_i]P$. Equivalence classes U/C and U/D will be called condition and decision classes, respectively.

1. Lower Approximation:

Given a decision table $T = (U, C, D, V, f)$. Let $R \subseteq C \cup D, X \subseteq U$ and $U/R = \{R_1, R_2, \dots, R_i, \dots\}$. The R -lower approximation set of X is the set of all elements of U which

can be with certainty classified as elements of X , assuming knowledge R . It can be presented formally as

$$R-(X) = U\{Ri \mid Ri \in U/R, Ri \subseteq X\}$$

2. Positive Region:

Given a decision table $T = (U, C, D, V, f)$. Let $B \subseteq C$, $U/D = \{D1, D2, \dots, Di, \dots\}$ and $U/B = \{B1, B2, \dots, Bi, \dots\}$. The B -positive region of D is the set of all objects from the universe U which can be classified with certainty to classes of U/D employing features from B , i.e.,

$$POS_B(D) = \bigcup_{Di \in U/D} B^-(Di)$$

3. Reduct:

Given a decision table $T = (U, C, D, V, f)$. The attribute $a \in B \subseteq C$ is D - dispensable in B , if $POS_B(D) = POS_{(B-\{a\})}(D)$; otherwise the attribute a is D - indispensable in B . If all attributes $a \in B$ are D - indispensable in B , then B will be called D - independent. A subset of attributes $B \subseteq C$ is a D - reduct of C , iff $POS_B(D) = POS_C(D)$ and B is D - independent. It means that a reduct is the minimal subset of attributes that enables the same classification of elements of the universe as the whole set of attributes. In other words, attributes that do not belong to a reduct are superfluous with regard to classification of elements of the universe.

4. Reduced Positive Universe and Reduced Positive Region:

Given a decision table $T = (U, C, D, V, f)$. Let $U/C = \{[u'_1]_C, [u'_2]_C, \dots, [u'_m]_C\}$, Reduced Positive Universe U' can be written as:

$$U' = \{u'_1, u'_2, \dots, u'_m\}$$

and

$$POS_C(D) = [u'_{i1}]_C \cup [u'_{i2}]_C \cup \dots \cup [u'_{it}]_C$$

Where $\forall u'_{is} \in U$ and $|[u'_{is}]_C/D| = 1$ ($s = 1, 2, \dots, t$).

Reduced positive universe can be written as:

$$U'_{pos} = \{u'_{i1}, u'_{i2}, \dots, u'_{it}\}$$

and $\forall B \subseteq C$, reduced positive region

$$POS'_B(D) = \bigcup_{X \in U/B \wedge X \subseteq U'_{pos} \wedge |X/D|=1} X$$

where $|X/D|$ represents the cardinality of the set X/D . $\forall B \subseteq C$, $POS_B(D) = POS_C(D)$ if $POS'_B = U'_{pos}$ [38]. It is to be noted that U' is the reduced universe, which usually would reduce significantly the scale of datasets. It provides a more efficient method to observe the change of positive region when we search the reducts. We didn't have to calculate U/C , U/D , U/B , $POS_C(D)$, $POS_B(D)$ and then compare $POS_B(D)$ with $POS_C(D)$ to determine whether they are equal to each other or not. We

only calculate U/C , U' , U'_{pos} , POS'_B and then compare POS'_B with U'_{pos} .

IV. DATA SET DESCRIPTION

Crime dataset was obtained from UCI Machine Learning Repository. The dataset consisted of 2215 records and 147 attributes for communities, 125 predictive, 4 non-predictive and 18 potential goals. The states are represented in the form of number which representing its respective American state. Attributes include information across a variety of crime-related aspects, ranging from the percentage of officers assigned to drug units, to population density and percent considered urban and to median household income. Also included are measures of crimes considered violent, which are murder, rape, robbery, and assault [5].

The normalization preserves rough ratios of values within an attribute. For example, double the value for double the population within the available precision except for extreme values (all values more than 3 SD above the mean are normalized to 1.00; all values more than 3 SD below the mean are normalized to 0.00).

V. RAPIDMINER AND THE SELECTED CLASSIFIERS

A. RapidMiner

RapidMiner, formerly known YALE (Yet Another Learning Environment), is software widely used for machine learning, knowledge discovery and data mining. RapidMiner is being used in both research and also in practical data mining fields. It will be used here to discover useful relationships from crime data.

The Java programming language is used in Rapid Miner. RapidMiner can handle many formats of input such as CSV, Arff, SPSS, Xrff, Database example sources, and attributes that are described in XML file format. Different types of attributes that are present are Input, Output, data preprocessing and visualization.

RapidMiner contains more than 500 operators. The nested operator can be described through graphical user interface XML files which are created with RapidMiner. Individual RapidMiner functions can also be called directly from command line. It is used easily to define analytical steps and to generate graphs more effectively. It provides a large collection of data mining algorithms for performing classification. Many visualization tools such as overlapping histogram, 3D scatter plot and tree charts are present.

RapidMiner can handle any type of tasks like classification, clustering, validation, visualization, preprocessing, post processing etc. It also supports many kinds of preprocessing steps such as discretization, outlier (detection and removal), filter, selection, weighting, normalization, and

so on are available. All modeling and attribute evaluation methods from Weka are available within RapidMiner.

RapidMiner consists of two views, design view and result view. The design view is used to generate the process and run the process. The result view is used to generate the results.

B. Selected classifiers

Classification as a famous data mining supervised learning techniques is used to extract meaningful information from large datasets and can be efficaciously used to predict unknown classes [26]. The predictive accuracy of the classifier is measured by using the training set and the accuracy of classifier on a given test set is the percentage of test set tuples which are classified correctly. If the accuracy is acceptable, the classifier can be used for future data tuples for which the class label is unknown [27]. In this section, the five different classifiers that were used in this work are explained below.

1. Naïve Bayesian classifiers:

Naïve Bayesian classifiers, by adopting a supervised learning approach, have the ability to predict the probability of a given tuple dependency to a specific class. This classifier is very simple to construct, and it may be easily applied to huge data sets [27]. Naive Bayes considers attributes as independent of each other in terms of contributing to the class attribute [28]. For example a fruit may be considered as apple if it round, red, and 4'' in diameter. Although these features depend on each other, Naive Bayes considers these features independently to consider it as apple [29].

2. Decision trees:

Decision trees are a supervised learning technique commonly used for tasks like classification, clustering and regression. Each node refers a test on an attribute value. The leaves symbolize classes or class distributions which predict classification models. The branches show coincidences of features, which go to classes. Input to a decision tree is the set of objects described by the set of properties, and creates output as yes/no decision, or as one of several different classifications [30]. Decision tree creation involves dividing the training data into root node and leaf node divisions until the entire data set has been analyzed.

3. Support Vector Machines (SVM):

Support Vector Machines (SVM) is a group of supervised learning methods that can be employed for classification or regression [31-33]. In a two-class learning task, the SVM goal is to discover the best classification function to differentiate between members of the two classes in the training data. For that purpose, SVM construct a hyper plane or a set of hyper planes in a high or infinite dimensional space for separating dataset and SVM find the best function by maximizing the margin between the two classes.

4. K-Nearest Neighbor (k-NN):

K-Nearest Neighbor (k-NN) classifiers are based on learning by comparison of given test tuple with training tuples. For an unknown tuple, a k-NN classifier seeks to detect a group of k objects in the training set that are closest to the unknown tuple and the unknown tuple labels based on the predominance of a specific class in this neighborhood [27]. In this study, k is set to 10 looking for the highest results of accuracy.

5. Neural Networks:

The human brain serves as a model for Neural Networks. Neural Networks are useful for data mining and decision support applications. They are also useful for pattern recognition or data classification through the learning process. It is able to model real world complex relationships. It could estimate the posterior probabilities, which give the basis for setting up classification rules and conducting statistical analysis [34].

A neural network contains the neurons and weight building blocks. The strength of the network depends on the interaction between the building blocks. The MultiLayer Perceptron (MLP) Neural Network Model is mostly used, with networks that consist of three layers: input, hidden and output. The values of the input layer come from values in a data set. The input neurons send data via synapses to the hidden layer and through output layer through synapses [35].

VI. METHODOLOGY

Crime data mining give an important attention to discover unseen patterns in crime data [6]. Real-time solutions are important since they minimize significant resources and improve performing the law instantly against criminals. Volume of crimes data, suspicious activities, and suspect records recorded is considered high. Also, relationships between various crime attributes can be misinterpreted by human analysts. So, it is very difficult to discover significant hidden patterns manually. Limited researches are examined and investigated on crime data. Crime data is considered sensitive and has high dimension of attributes that express all possible features that may help to discover the criminal. It is known that criminals follow repetitive behavior patterns, so analyzing their behaviors can help to capture relations among events from past crimes [36].

Attributes contain a high number of missing values (such as the *pctPolicBlack* and the *pctPolicWhite*) were deleted because it is difficult to expect their values. Although they are imputed, the error rate will be very high and this could affect the accuracy of the dataset. Similar attributes were also deleted in order to remove any duplication so that the dataset will be reduced. Examples of such attributes are *Divorced (%)*, *Male Divorced (%)*, and *Female Divorced (%)*, and the attribute *Divorced (%)* is only kept. The remains number of attributes is 125 including the class attribute. All instances that contain missing value for their classes (Crime status) were deleted in order to avoid any risk of wrongly expecting the

missing class value (221 instances). The remains number of instances is 1994.

For all attributes except *state*, min-max normalization to [0.0, 1.0] is used to avoid the large value issue as it has the advantage of protecting exactly all relationships in the data and to prevent any bias injection [27]. Using min-max normalization, attributes retain their distribution and skew. For example, the *population* attribute has a mean value of 0.06 because most communities are small. Also, an attribute described as *mean people per household* is actually the normalized [0.0-1.0] version of that value.

In this study, rough set theory for community crime pattern discovery was studied deeply. It is introduced as a novel means for knowledge discovery in the crime domain. Both the size of features (reduct) and the quality of reduct are the core of this study. Rough set theory is pioneer in this means. This work is supported by experimental results on the crime data set.

Chi-Square and rough set theory as feature selection algorithms were studied and the reduced data set was obtained from each. The comparison between rough set feature selection technique and chi-square feature selection technique was made to determine which of them is the more useful to infer better classification accuracy and performance. The comparison also compared number of features resulted by each technique (the reduced data set) which usually affects the processing time and at same time it minimizes the number of relationships in the data set and this will help in putting more focus on the important features and consequently minimizes the time for predicting the result. Five different classification algorithms are used to classify dataset based on the given class. They are Decision Tree (J48), Support Vector Machine (SVM), Naïve Bayesian, k-Nearest Neighbor, and Neural Networks (Multilayer Perceptron). The comparison between the given five classifiers for the two given reduced datasets was studied and compared. Also, a comparison between different classifiers in terms of AUC for choosing more accurate algorithms to classify crime status in the United States of America for finding a deeper perception into crime. The comparison between classifiers was applied to each reduced data set. Rapidminer is the data mining tool which is used throughout this study. It is freely available from the website www.rapidminer.com.

Rough set and chi-square reduction methods were applied to the two datasets in order to find the most significant features that are able to classify any new object correctly.

VII. RESULTS

This study will be compared with the previous study [5] in term of the size of the reduct and the performance of the classifiers that were applied to crime data. The more efficient feature selection method and the more efficient algorithms in

crime status prediction will be determined. These classifiers are: Naïve Bayesian (Bayes), Decision Tree (J48), Support Vector Machine (SVM), Neural Network (Multilayer Perceptron), and k-Nearest Neighbor (k-NN).

Cross-validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. In data mining and machine learning 10-fold cross-validation (k=10) is the most common. 10 fold cross-validations is known to be a good way to get unbiased or nearly unbiased estimates of the error rates for classification / prediction based on a training set of a given size. It randomly divided database into 10 separate blocks of objects and then data mining algorithm was trained using 9 blocks. Meanwhile, the rest was used for testing the algorithm's performance and the process repeats the k times. Finally, the average of the results was calculated [5].

Two different data sets were considered in this study. One data set was introduced by the previous work [5] which consists of 94 features and the other data set which is derived (in this study) by the proposed model which consists of 63 features. The first evaluation was made based on the size of reduct (number of features). It is clear to notice that the proposed model was pioneer comparing to the previous work of Shojaee et al. The proposed model minimizes the original data set by 61 features (49.2%) while the previous work minimized the original data set by 30 features (24.2%). The quality of the proposed reduct was better that that of the previous work of Shojaee et al. The quality of the two different data sets (set1: 94 attributes and set2: 63 attributes) was evaluated by comparing the findings on precision, recall, accuracy, and AUC. The evaluation on the five explained classification algorithms on the two different data sets was also conducted. Accuracy is a relative number of correctly classified examples by classifiers or the percentage of correct predictions. AUC is the Area Under the Curve of the Receiver Operating Characteristics (ROC) graph which is a technique for visualizing, organizing and selecting classifiers based on their performance. Precision shows that the proportion of data is classified correctly and it represented by the following formula:

$$\text{Precision} = \frac{\text{Positives Correctly Classified}}{\text{Total Predicted Positives}}.$$

The total Predicted Positives is the sum of True Positives and False Positives. Recall is the percentage of information which is relevant to the class and is correctly classified. It is also called true positive rate. Recall is represented by the following formula:

$$\text{Recall} = \frac{\text{Positives Correctly Classified}}{\text{Total Positives}}.$$

Table 1 illustrates the precision and recall for both different sets (the two reducts).

Table I. precision and recall.

Method	Precision (%)		Recall (%)	
	Set 1: 94 attributes	Set 2: 63 attributes	Set 1: 94 attributes	Set 2: 63 attributes
Naïve Bayesian	59.29	63.3	83.98	83.73
Decision Tree	100	100	87.34	87.34
Support Vector Machine (SVM)	95.48	95.67	85.01	87.07
Neural Network (Multilayer Perceptron)	75.92	78.45	68.45	68.43
k-Nearest Neighbor (K=10)	80.36	80.58	46.39	47.17

As shown in Table I, precision and recall have been enhanced after using rough set feature selection technique comparing to chi-square technique. The precision of the proposed reduct is always better than the reduct generated by chi-square technique. Also, the recall shows interested results even it is not the best as for Naïve Bayesian and NN, however, the difference is not significant. Results prove that the feature selection helps in achieving better classification result.

The experiment is finding the best classifier among the five given classifiers based on set 1 and set 2. For the proposed reduct (set 2) and as shown in Table I, the precision results of Decision Tree (100%) is the best among other classifiers in both data sets followed by SVM (95.67%), KNN (80.58%), NN (78.45%), and Naïve Bayesian (63.3%) respectively. The recall value of Decision Tree (87.34%) is the best among other classifiers in both data sets followed by SVM (87.07%), Naïve Bayesian (83.73%), NN (68.43), and KNN (47.17%).

Table II. Accuracy and AUC

Method	Accuracy (%)		AUC (%)	
	Set 1: 94 attributes	Set 2: 63 attributes	Set 1: 94 attributes	Set 2: 63 attributes
Naïve Bayesian	85.7	87.41	90.9	92.8
Decision Tree	97.54	97.54	50	50
Support Vector Machine (SVM)	96.29	96.69	99.4	99.5
Neural Network (Multilayer Perceptron)	89.62	89.97	94.2	94.6
k-Nearest Neighbor (K=10)	87.41	87.51	90.7	91

As shown in Table II, accuracy and AUC have been also enhanced after using rough set feature selection technique comparing to chi-square technique. The accuracy of the proposed reduct is always better than the reduct generated by chi-square technique except for Decision Tree where both are equivalent in term of accuracy. AUC shows interested results since all results are better when the proposed reduct was used except for Decision Tree since it gave same results for the both reducts (set1 and set 2). Results prove that rough set feature selection technique helps in achieving better classification results comparing to chi-square technique.

For the proposed reduct (set 2) and as shown in Table II, Decision Tree algorithm outperformed other four algorithms due to highest accuracy (97.54 %), followed by SVM (96.69%), NN (89.97%), KNN (87.51%), and Naïve Bayesian (87.41%) respectively.

A classifier with a greater AUC is said to be better than a classifier with a smaller AUC. Results of AUC conclude that SVM (99.5%) is indeed the best classifier among the five classifiers under study for both reducts (set 1 and set 2) followed by NN (94.6%), Naïve Bayesian (92.8%), k-NN (91%), and Decision Tree (50%) respectively.

VIII. CONCLUSION

In this paper, comparisons between two feature selection techniques (Chi-square and Rough set) based on dimensionality and quality of each reduct generated by each of them were performed. Comparisons between five different classifiers applied to each reducts (set 1 and set 2) generated by chi-square and rough set techniques were also performed using RapidMiner. Comparisons were performed on the two data sets in terms of precision, recall, accuracy, and AUC. Results shows that rough set feature selection technique is the best in term of dimensionality of the generated reduct. The reduct generated by rough set feature selection technique consist of 63 features while the reduct generated by chi-square feature selection techniques consists of 94 features. Increasing values of accuracy confirm the impact of feature selection based on rough set. Results also show that SVM is the best classifier among the five tested ones (AUC, Naïve Bayesian, Neural Networks, k-Nearest Neighbor, Decision Tree, and Support Vector Machine) in term of accuracy and AUC, specifically by using rough set feature selection technique. Results demonstrated that feature selection is an important phase to enhance the mining quality.

REFERENCES

- [1] R. Wortley and L. Mazerolle, "Environmental Criminology and Crime Analysis," Willan Publishing, UK, 2008.
- [2] T. Abraham and O. De Vel, "Investigative Profiling with Computer Forensic Log Data and Association Rules," In Proceedings of the IEEE International Conference on Data Mining (ICDM'02), 2002, 11-18.
- [3] L.A. Buczak and C.M. Gifford, "Fuzzy Association Rule Mining for Community Crime Pattern Discovery," In ACM SIGKDD Workshop on Intelligence and Security Informatics (ISIKDD '10), 2012.
- [4] A. Malathi and S. Baboo, "Enhanced Algorithms to Identify Change in Crime Patterns" International Journal of Combinatorial Optimization Problems and Informatics, Aztec Dragon Academic Publishing, 2011, 2(3), 32-38.
- [5] S. Shojae, A. Mustapha, F. Sidi, and M.A. Jabar, "A Study on Classification Learning Algorithms to Predict Crime Status," International Journal of Digital Content Technology and its Applications (JDCTA), 2013, 7(9), 361-369.
- [6] H. Chen, W. Chung, J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime Data Mining: A General Framework and Some Examples. Computer," Los Alamitos, CA, IEEE Computer Society Press, 2004, 37(4), 50-56.
- [7] J. De Bruin, T. Cox, W. Kusters, J. Laros, and J. Kok, "Data Mining Approaches to Criminal Career Analysis," In Proceedings of the International Conference on Data Mining, Washington, D.C., IEEE Computer Society Press, 2006, 171-177.
- [8] M. Chau, J. Xu, and H. Chen, "Extracting Meaningful Entities from Police Narrative Reports," In Proceedings of the National Conference on Digital Government Research, 2002, 1-5.
- [9] C. Ku, A. Iriberry, and G. Leroy, "Crime Information Extraction from Police and Witness Narrative Reports," In Proceedings of the IEEE International Conference on Technologies for Homeland Security, Boston, MA, 2008, 193-198.
- [10] S. Bagui, "An Approach to Mining Crime Patterns," International Journal of Data Warehousing and Mining, 2006, 2(1), 50-80.
- [11] C. Yu, M.W. Ward, M. Morabito, and W. Ding, "Crime Forecasting Using Data Mining Techniques," In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW '11), 2011, 779-786.
- [12] S. Nath, "Crime Pattern Detection Using Data Mining," In Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology, Washington, D.C., IEEE Computer Society Press, 2006, 41-44.
- [13] P. Phillips and I. Lee, "Mining Top-k and Bottom-k Correlative Crime Patterns through Graph Representations," In Proceedings of the IEEE International Conference on Intelligence and Security Informatics, Dallas, TX, 2009, 25-30.
- [14] P. Phillips and I. Lee (2011). "Crime Analysis through Spatial Areal Aggregated Density Patterns," *GeoInformatica*, Springer, 2011, 15(1): 49-74.
- [15] D.E. Brown, "The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals," In Proceedings of the International Conference on Systems, Man, and Cybernetics, 1998, 2848-2853.
- [16] V. Ng, S. Chan, D. Lau, and C. Ying, "Incremental Mining for Temporal Association Rules for Crime Pattern Discoveries," In Proceedings of the Australasian Database Conference, 2007, 123-132, Ballarat, Victoria, Australia.
- [17] M. Redmond and A. Baveja, "A Data-driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments," *European Journal of Operational Research*, Science Direct, 2002, 141(3), 660-678.
- [18] D. Pyle, "Data Preparation for Data Mining," Morgan Kaufmann Publishers, Los Altos, California, 1999.
- [19] L. Al Shalabi Z. Shaaban and B. Kasasbeh, "Data mining: A Preprocessing Engine," *Journal of Computer Science*, 2006, 2(9), 735-739.
- [20] M.R. Berthold, and K.P. Huber "Missing values and learning of fuzzy rules," *Int J Uncertain, Fuzziness Knowl-Based Syst*, 1998, 6(2), 171-178.
- [21] A. Agrawal and R. Srikant, "Privacy Preserving Data Mining," In Proceedings of ACM SIGMOD conference, Dallas, Texas, 2000, 439-450.
- [22] A.P. Dempster, N.M. Larid, and D.B. Rubin, "Maximum likelihood from incomplete data via the Em Algorithm (with discussion)," *Journal of Royal Statistical Society*, 1977, B39, 1-38.
- [23] L. Al-Shalabi, R. Mahmod, A.A. Abd Ghani, and Y.M. Saman, "A New Model for Extracting a Classfatory Knowledge from Large Datasets Using Rough Set Approach," In World Engineering Congress (WEC'99), Kuala Lumpur, Malaysia, 1999, 105-108.
- [24] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information Sciences*, 2007, 177, 3-27.
- [25] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, USA, 2001.
- [26] E.W.T.Ngai, L. Xiu, and D.C.K. Chau, "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification," *Expert Systems with Applications*, Elsevier, 2009, 36(2), 2592-2602.
- [27] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, USA, vol. 5, 2012.
- [28] H. Zhang, "The Optimality of Naive Bayes," *American Association for Artificial Intelligence*, 2004.
- [29] P. Bhargavi and S. Jyothi, "Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils," *IJCSNS International Journal of Computer Science and Network Security*, 2009, 9(8), 117-122.
- [30] M.J. Aitkenhead, "A co-evolving decision tree classification method," *Expert Systems with Applications*, 2008, 34(1), 18-25.
- [31] O. Ivanciuc, "Applications of Support Vector Machines in Chemistry," *Reviews in Computational Chemistry*, 2007, 23, 291-400.
- [32] Y. Wang, F. Chen, and X. Qu, "Research and Application of Large-Scale Data Set Processing Based on SVM," *Journal of Convergence Information Technology (JCIT)*, AICIT, 2012, 7(16), 195-200.
- [33] C. Zhenzhou, "Local Support Vector Machines with Clustering for Multimodal Data," *Advances in information Sciences and Service Sciences (AISS)*, AICIT, 2012, 4(17), 266-275.
- [34] G.P. Zhang, "Neural Networks for Classification: A Survey, Systems, Man, and Cybernetics." Part C: Applications and Reviews. *IEEE Transactions*, 2000, 30(4), 451-462.
- [35] I.H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques," 2005, 233-237.
- [36] JS. De Bruin, TK. Ccoos, WA. Kusters, JF. J. Laros, and JN. Kok, "Data mining approaches to criminal career analysis," In the sixth International Conference on Data Minin (ICDM'06), 2006, 171-177.
- [37] K.L. Priddy and P.E. Keller, "Artificial Neural Networks: An Introduction," SPIE Press, USA, 2005.
- [38] Z. Xu, Z. Liu, B. Yang, and W. Song, "A Quick Attribute Reduction Algorithm with Complexity of $\max(O(|C||U|), O(|C|^2|U/C|))$," *Chinese Journal of Computers*, 2006, 29: 391-399

AUTHOR PROFILE

Al-Shalabi L. was born in Jordan in 1971. He received the B.S. in computer science from Yarmouk University, Jordan, in 1992, M.S. degrees in image interpretations from Universiti Sains Malaysia, Malaysia, in 1996, and the Ph.D. degree in data mining from University Putra Malaysia, Malaysia, in 2000.

From 2000 to 2008, he was an Assistant Professor with Computer Science Department, Applied Science University, Jordan. Since 2008, he has been an Assistant Professor with the Information Technology Department, Arab Open University, Kuwait. His research interests include data mining, knowledge discovery, databases, machine learning, and image interpretation.

VoIP Security: Common Attacks and their Countermeasures

Abdullahi Mohammed Jingi

Department of Computer Science
Adamawa State University (ADSU), Mubi
Mubi, Adamawa State
Nigeria.

modijingi@yahoo.com, modijingi@gmail.com

Murtala Muhammad

Department of Computer Science
Adamawa State University (ADSU), Mubi
Mubi, Adamawa State
Nigeria.

alferio1349@gmail.com

Abstract—Voice over Internet Protocol (VoIP) is one of the fastest growing technologies; it is receiving more attention from industries and general public due to its cost reduction and flexibility. However, the technology being deployed on IP data network is faced with a lot of security issues in its deployment. Common threats and attacks within VoIP network will be discussed in this paper. Also, some mitigation strategies that can be used in order to ensure security of VoIP network will be presented.

Keywords—component; VoIP security; threats; attack; SIP; denial of service

I. INTRODUCTION

Voice over Internet Protocol (VoIP) is a technology that allow users of the internet to transmit voice communications across IP network, the technology is growing at a faster rate and it assumes to replace the conventional Public Switch Telephone Network (PSTN)[1], this is due to its greater cost advantage and flexibility compared to PSTN[2]. It is possible to use VoIP and call any PSTN line or mobile phone located anywhere in the world. Although some services can only work in a computer or a specific VoIP phone, others support the use of traditional phone connected with adapter [3]. Originally, VoIP application was built to use on PCs and give users with microphones the ability to call other users who have the same application installed on their machines when both are connected to the internet. Today, with the advancement in network technologies, it is possible to use 3G, 4G and wireless networks to enjoy VoIP services on mobile phones [4]. Having VoIP services on phones makes it mobile and more convenient to use, which attracts more people to use it and thus slowly replacing the conventional phone. Most telecommunications companies are planning to deploy VoIP services. Ray Stanton stated that “There’s no doubt that VOIP is the future of telephony” [5]. This statement indicates that VoIP technology will replace the existing conventional Public Switch Telephone Network (PSTN) in years to come; this is of

course no doubt as most companies and organizations are deploying it.

The security issues raised against VoIP begins to emerge [3][6]. Combination of voice and data traffic on the same network provides a lot of opportunities as well as introducing security risks; VoIP can never be installed without having security in mind. A lot of factors are responsible for these security risks; these are (i) the nature of the VoIP application itself tends to attract more intruders as they are eager in viewing/hearing things and the open nature of the internet (ii) The VoIP network infrastructure and (iii) the IP network on which VoIP application are deployed [7].

In order to facilitate our understanding of security issues associated with VoIP network, a brief description of VoIP networks architecture is presented. There are three layers in a typical VoIP infrastructure, these are:

- End-User equipments which consists of “hard phones” which resembles conventional telephones or “soft phones” that uses some applications software which imitate telephone. The security of these equipments depends on how they are installed; in most cases they are vulnerable to attacks [3].
- Network components: VoIP inherits all vulnerabilities of IP network since it is deployed on the same network [8]. The network component is made up of devices like routers, switches and firewalls.
- Gateways: They integrate the IP network with the Public Switch Telephone Network (PSTN). Some of the main functions of the gateways are controlling signal, voice compression/depression and call routing. Careful attention should be taken in ensuring that gateway security policies do not give any chance of introducing vulnerability [3].

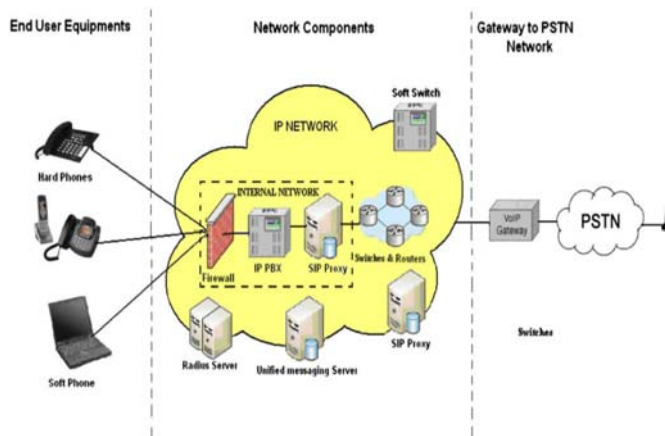


Figure1: VoIP Network[3]

Unlike data networking where majority of standards are quite stable for some time – For example TCP/IP has been stable for over 20 years. Implementation of VoIP related standards is inconsistent due to the rapid development of the standards themselves. The implication of this is that immature applications which are vulnerable to various threats arrive at the market. VoIP is not just another data protocol; it uses the same internet architecture with other data applications, however, VoIP is different due to: its real-time nature, separation of data and signalling and also voice conversation can be initiated from outside the firewall while most clients driven protocols begin requests from inside the firewall [9]. No dedicated network is used for VoIP[10], this shows that users are subjected to different attacks such as Denial of Service (DoS), eavesdropping, Spam over Internet Telephony (SPIT), vishing e.t.c. Information as a valuable asset of an organization, its security is very important. However, with VoIP the issue of privacy does not arise due to vulnerabilities from many ways. The aim of this paper is to explore some common vulnerability of VOIP and their countermeasures.

II. VoIP PROTOCOLS

Before going ahead and discuss VoIP security issues, we will discuss some basics of the technology. This will help us in understanding the security impacts of all the protocols associated with the VoIP. VoIP protocols can be classified either as signaling or media transport protocols, when designing these protocols security was not taken into consideration.

Signaling Protocols: these protocols are used for controlling signaling services such as setting up a call and call termination. Example of such protocols are Session Initiation Protocol (SIP), skinny client control protocols (SCCP) and Inter Asterisk eXchange (H.323, IAX), The SIP and H.323 protocols set up the call connection and pass it to media transport protocols. They perform the same set up services but in different approaches, SIP uses a simple approach similar to HTTP in which methods like BYE, FORWARD, REGISTER

are used to initiate a call. While the H.323 make use of group of subprotocols like H.239 and H.245 in session set up. In addition, both the two protocols use supporting servers to set up a call between two end points. After setting up a call, the two protocols use RTP to transfer audio between two or more end points [11]. The IAX which is not common like SIP or H.323 is normally used between two asterisk servers. It can also use for both call set up between two end points and the voice transfer. Therefore, IAX does not require the use of RTP to transfer media [11]. The skinny client control protocol (SCCP) is used between Cisco call manager and Cisco VoIP phone [3]

Media Transport Protocols: These protocols are used in decoding control and ordering of voice scripts for real-time communication. These are Real-time Transport Protocol (RTP) and Real-time Transport Control Protocol (RTCP). RTCP is used for delivering services like payload-type identification, monitoring of delivery and time stamping, it is also used to define the standard format of a packet for sending audio and video across IP networks. The main function of the RTCP is to provide information regarding the quality of service provided by RTP [3]

III. VULNERABILITIES/THREATS FACED BY VoIP NETWORK

Combination of two entirely different types of communications systems also integrate their threats and results to lots of vulnerabilities never witnessed before in both the data and voice systems [12]. VoIP network faces a lot of vulnerabilities that can be related to one of the following three categories: VoIP network infrastructure, the VoIP application itself and IP network on which it is deployed.

A survey sponsored by French National Research Agency on published and common vulnerabilities found that there were a total of 221 problems from 1999 through Nov 2009[6]. Fig2a shows the vulnerabilities by type using VoIP Security Alliance (VOIPSA); this figure indicates that most problems contribute to DoS attacks. Fig 1b analyse the vulnerabilities based on their locations and causes, it indicates that most vulnerabilities occur at the implementation stage.

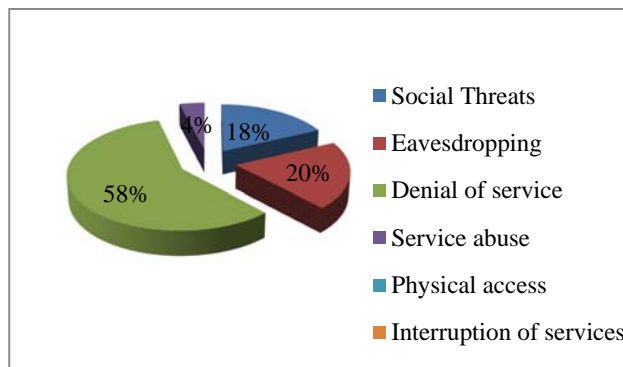


Fig.2a: Vulnerabilities of VoIP network using VOIPSA [6]

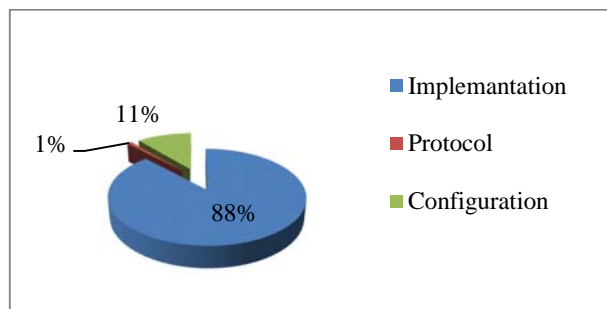


Fig. 2b: Vulnerabilities according to locations and causes [6]

Vulnerabilities of VoIP Network Infrastructure

VoIP network infrastructure is similar to a data network but with a different type of application traffic. VoIP network needs to have a unique design and specific components.

VoIP signalling protocols SIP and H.323 might not be compatible with the existing firewalls. A VoIP call with these standards is carried out in two phases; call setup and voice call, the call setup locates the person intended to call and allocates ports for the voice call. Eventually, this makes the VoIP end points to know which IP ports will be used for communication. However, firewalls will only allow voice call from recognised ports to pass, otherwise they will be blocked. To overcome this issue, we either use firewalls that can recognise signalling protocols and allow voice traffic to pass or pass the voice traffic using proxy server located outside the secure area of the firewall of both the calling parties [7].

Processing signalling protocols as mentioned above also invites some vulnerability. This is possible because the two protocols (SIP and H.323) are published and are well known by attackers, the attackers might generate attack with the intention of identifying common errors in their implementation. A study conducted by University of Oulu, Finland, on VoIP signalling devices found that majority of devices had vulnerability. This issue can be dealt with by constantly updating firewalls with upgrades and security patches [7].

VoIP Application Vulnerabilities

Apart from the VoIP network infrastructure components, VoIP itself is an application subject to vulnerabilities. Therefore, ensuring a secured VoIP comprises both the security of the network domain and the host application domain. Sometimes, an attacker might attack a VoIP network in order to enjoy toll-free call or have access to voice mail [7]. This indicates that toll fraud can motivate an intruder to hack into a VoIP network, and eventually places a long distance call after gaining access. This loophole can be closed using password protection to restrict unauthorised access.

IP Network Vulnerabilities

Both data and voice network transfers can be secured using encryption protocols like: Internet Protocol security (IPsec), Secure Socket Layer (SSL) and Transport Layer Security (TLS). IPsec is a set of protocols used to secure data packets

which are sent across IP layer and it is fully supported IPv6. Encryption and decryption takes longer time and therefore introduces latency/jitter [13]. Furthermore, firewalls will not be able to determine which ports will be allowed because the details of the signalling protocols are hidden as a result of end-to-end encryption of a voice stream.

Among the vulnerabilities of IP network is latency and jitter; this is due to longer time needed for encryption and decryption. Latency (time from when words are spoken at one end to when they are heard on the other end) is not tolerated in VoIP call, a delay of 150 milliseconds are recognized and is unacceptable[13].

VoIP Network Threats

VoIP threats refers to the means or ways through which attacker can launch attack. Some of these threats are:

- (1) **Denial of Service (DoS) threats:** this type of threat has the ability to deny legitimate users to access VoIP services. It is a serious problem in times of emergency situations or when the DoS attack disrupts all the user communication capabilities. Such attack could be as a result of physical disconnection of cables, shutting down of power or through traffic flooding attacks[14].
- (2) **Service abuse threats:** certain threats occur as a result of improper use of VoIP services by users, mostly where such services are rendered in a commercial setting. Example of this threat is toll fraud.
- (3) **Physical access threats:** these are threats that occur in VoIP services as a result of unrestricted or unauthorized physical access to VoIP facility. This is another chance which attackers use to attack VoIP system[14].
- (4) **Eavesdropping modification and Interception threats:** This refers to a situation whereby a person can unlawfully listen to conversation going on between two parties. This conversation might be secret conversation which involves business secrets, attacker can intercept and listen to the conversation and possibly sell the information to business competitors.
- (5) **Interruption of service threats:** this type of threat occurs as a result of unexpected problem that may affect VoIP services. For example loss of power.

IV. COMMON ATTACKS WITHIN VoIP NETWORKS

In[15]Dzubec stated that “the issue is IP itself. IP was never designed with security in mind.” A possible reason for that is; VoIP network is vulnerable to so many security threats due to the nature of how they are deployed. In this paper, we will focus on toll fraud, privacy concerns, Denial of service (DoS), spam/spit, eavesdropping, Middlemen attack, service theft etc.

A. Toll Fraud

This simply refers to accessing a legal VoIP network by an unauthorised user[15]. Toll fraud is considered a serious threat to VoIP networks because it increases the cost of enterprise by making use of its resources (bandwidth). According to the report presented by the French National Research Agency (ANR) VAMPIRE Project in 2009 on VoIP/SIP, toll fraud top the list of threats associated with the fraudulent usage of VoIP services. In 2010 30 members of criminal gang were arrested in Budapest and London following the allegation of stealing £11 million through VoIP toll fraud. The gang used stolen VoIP account details and make about half a million calls to premium rate numbers and they were paid a portion of high call charges[16]. Also, it was reported that a service provider in Miami (USA) hacked into the network of other service providers; he diverts his customer calls onto their network and receives the payment from his customers[3]. This indicates that once a hacker has the opportunity to access VoIP infrastructure, his/her target is to dial premium rate numbers which results in expensive call bills.

A. Eavesdropping

Eavesdropping is a situation whereby an attacker/intruder secretly monitors conversation of victims. The process involves receiving the voice data from both the caller and the recipient. Information is obtained from the data received and it is used for illegitimate purposes[17]. It is easy for an attacker to intercept VoIP traffic because most of the traffic are transmitted without encryption, this indicates that they are susceptible to eavesdropping attack by using a simple packet sniffer to capture the packets[15]. Confidentiality of information of an organisation is very important, but the issue of confidentiality is questioned as far as VoIP is concerned. In addition to the likelihood of hackers stealing sensitive information by monitoring voice traffic, they can also pretend as another VoIP caller. This is done by injecting a false ID into a VoIP call that makes the receiver think it is coming from authentic source. This may result in the victim supplying vital information like bank details, national insurance number and other forms of personal identification that can result in identity theft[16].

B. Vishing(VoIP Phishing)

This refers to the use of social engineering by criminals over the telephone system, mostly, using features facilitated by the VoIP; they try to manipulate people into disclosing confidential information for credit card and identity theft. A lot of people trust telephone lines more than the web, and the criminals are making use of this advantage to commit fraud. Although vishing is not an attack on a physical computer, it gives an attacker the opportunity to have access to system by applying social engineering concepts to obtain authentication details from the legitimate users. In 2006, users of PayPal received email messages requesting verification of their account details on a phone line, their bank accounts were

raided by the attackers after disclosing their accounts information[16]. A brief description on how vishing attacks are carried out is given in[18]. Firstly, the attacker must gain an unauthorized access to computer system with internet connection; this could be through physical access or by hacking on the victim's network. Secondly, after the attacker has control of the network, the next thing is to have a digital private branch exchange (PBX) installed on the host. A good example of such program is the Asterisk which is very powerful and flexible. Lastly, the attacker needs to be given the power to record victims phone conversation, the digital private branch exchange has an inbuilt functionality which can do that for the attacker, for example Asterisk is capable of recording a voice call and sending the recorded file to a desired destination. This feature allows the attacker to put his effort to only one server which serves the two functions (spam email and voice). Before launching any attack, attackers must possess real phone numbers of the victims or gained access through phreaking. After having computer system with internet connection, PBX installed on it, and successfully bridged SIP to PSTN; attackers can either manually make calls with victims to use their social engineering skills and obtain confidential information or send a fabricated email to potential victims[18].

C. Denial of Service attacks (DoS)

Denial of Service Attack is considered as one of the most critical attack. "A Denial of Service attack is defined as an assault on a network or computing system designed to cause a loss of service"[5]. The Session Initiation Protocol (SIP) is subject to DoS that can be carried out in two different ways: (1) the attacker can overload the SIP server by sending bulk SIP messages which deny access to legitimate users and (2) the intruder can send SIP requests that carry malformed SIP header fields which exploits several threats of SIP servers[19]. In this type of attack, attackers can overcome the resources of the system under target and utilise all the bandwidth by overflowing the network with traffic. If a main network is attacked, it may affect the rest of the network as routers get overloaded and fail to perform properly. Due to the distributed nature of the internet, it is very difficult to trace the origin of an attack, thus making denial of service attack as one of the most notorious problem regarding VoIP systems. Researchers are spending great amount of time and effort in order to come up with a solution that can stop or prevent denial of service attacks, but until now only a little success is achieved[5]. DoS attacks could result in inability to make or receive a call and can completely shutdown all operating applications which can lead to lost of phone service in businesses, this can cause a lot of disruptions to all the departments concerned. Attack on service availability is one of the greatest threats to VoIP security because it can affect customers, leads to system downtime, unproductivity and unplanned maintenance cost[20].

D. Spam Over Internet Telephony (SPIT)

As VoIP usage increases, the SPIT is becoming a nuisance to the users of VoIP, millions of unwanted spam messages are sent around the world every day. It is assumed that Spam over Internet Telephony (SPIT) will occupy voice mail boxes of users just as email spam messages does today. It is very difficult to combat SPIT because voice services are conducted in real-time[16]

E. Man-in-the-middle attack

The damage caused by Man-in-the-middle attack (MiTM) to VoIP network is quite enormous. The attackers insert themselves into an ongoing communication and obtain information from both parties; they could retrieve sensitive information regarding username and password of the victims [8]. Normally, the attackers listen to conversation between the two parties involved and alter it; they also try to convince the victims that they are actually communicating to each other, while the whole conversation is being controlled by them. The Man-in-the-middle either replaces one of the parties involved in the communication or act as an intermediary trying to make both parties believe that they are communicating to each other while they are not[21].

F. Caller identification Impersonation

Since every phone has its identity which is the phone number used in the device, it is very possible for attackers to impersonate the identity of others and use it as an attack to either place calls with the spoofed ID or receive incoming calls. In this type of attack, attackers set up their VoIP phone device in such a way that it will use the identity of the victim's device. Any call meant for the victim's number would be sent to the attackers' phone. The attacker can impersonate the victim in two ways either by receiving the victims call or by calling with the spoofed ID. When the attacker places a call, it is the victim's phone number that will appear on the phone receiving the call.

G. Fuzzing (Brute force) Attack

This technique is used by hackers to test how a protocol functions. They perform various tests on the protocol until they determine its weakness and attack it from that point [22].

H. Malicious Codes Attack

Viruses and Malwares can infect VoIP applications and OS that run on VoIP network. Viruses can degrade node's ability to handle traffic by infecting them.

B. COUNTERMEASURES OF VOIP ATTACKS

(1) Authentication and Encryption

There are list of protocols that can be used to secure VoIP communication. In [23]it was stated that protocols like Secure

Real Time Protocol(SRTP), IPSec and Datagram Transport layer security (DTLS) can be used to secure the media path. Using SRTP to secure voice communication can provide authentication, confidentiality and integrity. SRTP is widely used because it is designed with security in mind.

(2) S/MIME (Secure Multipurpose Internet Mail Extensions)

S/MIME provides confidentiality, integrity, and authentication for SIP. Both integrity and confidentiality can be attained with S/MIME.

(3) ISO27001 Compliance

VoIP services providers should conform to ISO27001 security policy in order to ensure all aspects of the organisation appropriately identify and deal with security risks. Being ISO27001 compliant helps to put in place security policy that can mitigate VoIP attacks that can occur as a result of physical access or service abuse threats. ISO27001 provides guidelines for security policy and hence guarantee security of VoIP facilities[24].

(4) Multi-layer Security

Multi layer security can be used to prevent vishingattacks. All people working in the IT must pay attention to their task; supervisors must ensure that employees are using all IT equipments in a secured manner to avoid vulnerabilities. Most importantly, all users must be enlightened about different types of threats and the best way to respond to such attacks[18].

(5) National Institute of Standards and Technology (NIST) Recommendations

Nine (9) recommendations were given by National Institute of Standards and Technology regarding the implementation of VoIP in[5]. Adhering to these recommendations will ensure confidentiality, integrity and availability of VoIP communications in an organisation.

(6) Separating VoIP and data Networks

Separating VoIP and data on to different network is one of the most important defence vectors in VoIP attacks. This is because attackers can no longer use PCs and workstations to get into the VoIP network. VLAN helps in the separation of the two networks [17].

(7) Intrusion Detection

Intrusion detection systems can be used to detect brute force attack[15].

(8) Antivirus software

Maintaining regular updates of patches will help to prevent malicious codes infections.

C. CONCLUSION

VoIP technology is gaining more popularity and it is expected to replace the traditional public switch telephone

network (PSTN) in years to come. However, as the technology grows, the issue of its security grows as well. Security of information is very important since it is the greatest asset of an organization. Confidentiality is not guaranteed with VoIP because it is vulnerable to different attacks. It has been shown in this paper that deployment of VoIP faces a lot of threats like denial of service, toll fraud, physical and service abuse threats, with denial of service having much number of attacks(58%) based on the research conducted by French National Research Agency.

Mitigation strategies which involve authentication and encryption of VoIP protocols, separating VoIP and data network, regular updates and upgrades of patches, complying with ISO27001 standard will help to strengthen security in a VoIP network.

References

- [1] J. K. Prasad and B. A. Kumar, "Analysis of SIP and realization of advanced IP-PBX features," in *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, 2011, pp. 218-222.
- [2] L. Liu, "Uncovering SIP vulnerabilities to DoS attacks using coloured petri nets," in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, 2011, pp. 29-36.
- [3] R. Dantu, S. Fahmy, H. Schulzrinne and J. Cangussu, "Issues and challenges in securing VoIP," *Comput. Secur.*, vol. 28, pp. 743-753, 2009.
- [4] Y. M. Li and S. W. Chiu, "VoIP Pricing in Competitive Telephony Markets," .
- [5] M. Garuba, J. Li and Z. Yi, "Security in the new era of telecommunication: Threats, risks and controls of VoIP," in *Information Technology: New Generations, 2008. ITNG 2008. Fifth International Conference on*, 2008, pp. 587-591.
- [6] A. D. Keromytis, "Voice-over-IP security: Research and practice," *Security & Privacy, IEEE*, vol. 8, pp. 76-78, 2010.
- [7] W. Chou, "Strategies to keep your VoIP network secure," *IT Professional*, vol. 9, pp. 42-46, 2007.
- [8] Y. Zhang and H. Huang, "VOIP voice network technology security strategies," in *Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011 2nd International Conference on*, 2011, pp. 3591-3594.
- [9] T. Porter, *Practical VoIP Security*. Syngress Publishing, 2006.
- [10] H. Abdelnur, I. Chrisment and C. Popi, "Assessing the security of VoIP services," in *Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on*, 2007, pp. 373-382.
- [11] H. Dwivedi, *Hacking VoIP: Protocols, Attacks, and Countermeasures*. No Starch Press, 2008.
- [12] F. Palmieri and U. Fiore, "Providing true end-to-end security in converged voice over IP infrastructures," *Comput. Secur.*, vol. 28, pp. 433-449, 2009.
- [13] V. Iossifov, T. Totev and A. Tochatschek, "Experiences in VoIP telephone network security policy at the university of applied sciences (FHTW) berlin," in *Proceedings of the 2007 International Conference on Computer Systems and Technologies*, 2007, pp. 3.
- [14] A. Keromytis, "A survey of Voice over IP security research," *Information Systems Security*, pp. 1-17, 2009.
- [15] E. A. Blake, "Network security: VoIP security on data network--a guide," in *Proceedings of the 4th Annual Conference on Information Security Curriculum Development*, 2007, pp. 27.
- [16] A. Bodhani, "Voicing concerns," *Engineering & Technology*, vol. 6, pp. 76-79, 2011.
- [17] D. Butcher, X. Li and J. Guo, "Security challenge and defense in VoIP infrastructures," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, pp. 1152-1162, 2007.
- [18] S. E. Griffin and C. C. Rackley, "Vishing," in *Proceedings of the 5th Annual Conference on Information Security Curriculum Development*, 2008, pp. 33-35.
- [19] M. A. Akbar and M. Farooq, "Application of evolutionary algorithms in detection of SIP based flooding attacks," in *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, 2009, pp. 1419-1426.
- [20] B. Materna, "VoIP insecurity," *Communications Engineer*, vol. 4, pp. 39-42, 2006.
- [21] Z. Chen, S. Guo, K. Zheng and H. Li, "Research on man-in-the-middle denial of service attack in SIP VoIP," in *Networks Security, Wireless Communications and Trusted Computing, 2009. NSWCTC'09. International Conference on*, 2009, pp. 263-266.
- [22] T. Wulff and R. Hunt, "New Approaches to Mitigation of Malicious Traffic in VoIP Networks," 2010.
- [23] D. Pérez-Botero and Y. Donoso, "VoIP eavesdropping: A comprehensive evaluation of cryptographic countermeasures," in *Networking and Distributed Computing (ICNDC), 2011 Second International Conference on*, 2011, pp. 192-196.
- [24] P. Thermos and A. Takanen, *Securing VoIP Networks*. Addison-Wesley Professional, 2007.

Authors Profile

Author 1:Abdullahi Mohammed Jingi holds B.Tech (Hons.) Operations Research (With First Class) in 2008 from ModibboAdama University of Technology, Yola, formerly known as Federal University of Technology, Yola. He also studied Msc. Applied Computing and Information Technology from University of Bedfordshire, Luton, United Kingdom. Currently, he is a lecturer in the Department of Computer Science, Adamawa State University, Mubi. His area of interest is Optimization and Operational Research.

Author 2:Murtala Muhammad is a PhD student in the School of Computing, Science, and Engineering, University of Salford, Manchester, United Kingdom. He holds B.Tech (Hons.) Computer Science in 2009 from ModibboAdama University of Technology, Yola, formerly known as Federal University of Technology, Yola. He also studied Msc. Computer Networking from The University of Greenwich, London, United Kingdom. Currently a lecturer in the Department of Computer Science, Adamawa State University, Mubi, Nigeria.

Improving Root Finding in Arabic Stemmers

Abstract—Analysis of Arabic language has become a necessity because of its big evolution; we propose in this paper a rule based extraction method of Arabic text, to solve some weakness founded on previous research works. Our approach is divided on preprocessing phase, on which we proceed to the tokenization of the text, and formatting it by removing any punctuation, diacritics and non-letter characters. Treatment phase based on the elimination of several sets of affixes (diacritic, prefixes, and suffixes), and on the application of several patterns. A check phase that verifies if the root extracted is correct, by searching the result in root dictionaries.

Keywords— rule based method; Arabic text; preprocessing; tokenization; dictionaries.

I. INTRODUCTION AND RELATED WORK

Arabic is a Semitic language spoken by more than 400 million people as a native language and ranked at the seventh position of Internet users in 2010. However, the task of performing the retrieve of information of Arabic language is very problematic, because of many aspects, such as: polysemy, irregular and inflected derived forms, various spelling of certain words, various writing of certain combination character, short vowels (diacritics) and long vowels, and the spectacular availability of affixes in the Arabic words [2, 3]. Different methods and approaches have been introduced to retrieve Arabic information [3, 4, 5, 6, 9].

To study Arabic morphology effectively, we divide words in Arabic into three self-contained categories as follows :

- ✓ اسم: It includes nouns, pronouns, adjectives, adverbs, etc
- ✓ فعل: Verbs
- ✓ حرف: Particles, articles, and conjunctions

Particles are completely unpredictable; they don't fall into the templatic system (i.e. they have no patterns) nor do they undergo any morphophonemic changes. They are what they are and must be memorized. The up side is that there are relatively few of them in the language – within one hundred.

Nouns and verbs do fall into the templatic system and have very systematic morphophonemic rules that govern them. This includes the study of how verbs are conjugated, how they move from pattern to pattern to enhance their meanings, how the participles and other nouns are derived, how nouns pluralize, etc...

Each declinable noun and each verb is made up of a certain set of base letters, called its root: جذر. (Nouns that are always indeclinable (such as pronouns) usually don't follow this system.)

Verbs can either have 3 base letters, or 4. Nouns can have 3, 4 or 5. Now these base letters can be augmented with extra letters, and they can be dropped or changed due to morphophonemic rules as well.

Particles - The third part of speech in Arabic mentioned above is the particle. The meaning of a particle is often understood in the context of the sentence and words before and after the particle. The sign of the particle is that it does not accept the signs of nouns or verbs.

Analyzing Arabic text was treated by many researchers, all of them tried to extract an exact root or stem from a word, there is two ways to treat a text; morphological analyze, which consist to find roots, and there is statistical stemmers that group word variants using clustering techniques.

The first approach of morphological analyze is manually constructed dictionaries based on roots, Kharashi and Evens worked with small text collections, for which they manually built dictionaries of roots for each word to be indexed [1]. Tim Buckwalter developed a set of lexicons of Arabic stems, prefixes, and suffixes, with truth tables indicating legal combination [2].

Nehar et al. [4] and Taghva et al. [17] introduces new stemming techniques which do not rely on any dictionary, the first one is based on the use of transducers. Nehar et al. [5] proposed also a heavy stemmer which does not use any dictionary of roots .S. Khoja and R. Garside [3] developed a dictionary based stemmer, and Larkey et al.[6] developed a Java program based on their own Arabic stemmer that we will develop and evolve, in order to take in count some nouns and verbs categories described in the previous paragraph. Taghva et al. [7] proposed IRSI Arabic Stemmer Algorithm which doesn't use a Root Dictionary. ISRI stemmer per-forms better than the other approaches on the shorter title queries. For the long texts and narrative queries, stemming made a difference: the Khoja, ISRI, and Light stemmers were significantly better than not stemming. Ghwanmeh et al. [14] presents an Arabic root-based algorithm based on patterns. This stemmer is restricted to native Arabic words that consist of four or more Arabic alphabets.

All algorithms mentioned before have some weaknesses. In this paper we will prove that the best way of stemming is the one that have a strong preprocessing phase, and it is based on both “patterns check” and “root list”.

This paper is an extension of work originally presented in conference [].

In fact, we present the weakness in Heavy and Light stemming Algorithms and we try to propose some new solutions for each point treated, then we will compare results of our new stemmer with other ones.

In section II we present the different areas for improvement in Arabic text classification, in section III we present our approach, and in section IV we present some tested examples and, compare our solution with others.

II. IMPROVEMENT AREAS IN ARABIC CLASSIFICATION

Light stemming algorithms removes suffixes, and prefixes from words, producing a form of word called “stem” [10], there was many versions of the light stemming algorithms and the last one is light10 [11]. This algorithm after removing punctuation and non letters, diacritics, Hamza from letter “أ”, he replaces final letter “ة” with letter “ه” and then replace final letter “ى” with letter “ي”. After that the algorithm search in irregular word list to find out if the word exists on this table or not. Then the algorithm removes the letter “و” from the beginning of the words if the length of the word is more than three characters, because it considers that this letter is usually a conjunction.

This step generates several errors on stem extraction, I give below some examples:

Word	Stem extracted by Light10
وَبَيْلٍ	بَيْلٍ
وَجُوبٍ	جُوبٍ
وَرَيْثٍ	رَيْثٍ

Table1. Morphology of Arabic word

As we can see, when removing the letter “و” from the beginning of those words, we change the meaning of the word, for the first word “وَبَيْلٍ” it means calamitous; disastrous, and when we remove the letter “و” the word means torch.

In Khoja’s Approach, and TC system proposed by M.Hadni, A.Lachkar and S. Alaoui Ouatik in [8], and also in Mohammed N. Al-Kab who proposed evolution of Khoja’s algorithm [13], we find this same issue, so in our algorithm we will take care of this point and we propose to check if the word doesn’t exist in the list of words that begins by ‘و’, and then remove diacritic ‘و’ (primarily weak vowels), this list is constituted manually and must be maintained regarding the evolution of Arabic language.

The second point we have improved is about the removing the letter “أ”, in Light Approach, Khoja’s Approach and M.Hadni’s one, this letter is deleted because it is considered as a prefix. The issue is when this letter is a part of word as for the word “أَبَاحٍ” which means “permit”, and when we delete it the word means “confide”, to solve that, we built a list of words that starts with letter “أ”.

The third point we involve in this paper is the stemming of five nouns (أب, أخ, حمور, فور, ذو), those nouns are excluded from the other single nouns according to the syntactic case. They have other marks to indicate them syntactic cases that the other doesn’t have. The single noun always depends on rules in order to indicate its syntactic cases but the five nouns are contradicting those rules. The five nouns aren’t depending on al Harakat (vowelization on system) rather than the letters. They have preconditions to be different from other single nouns:

- 1-It has to be adjunct to another noun in other words there must be a noun after it that is genitive noun.
- 2-The noun after must not be (ي) that indicates the speaker [12].

So for our algorithm, we treat the fives nouns separately.

It is true the orthography in Arabic is less ambiguous and more phonetic with the use of diacritics. For example, a word can be written using the same characters and be pronounced differently. The main purpose of diacritics including vowel marks, known as Harakat a phonetic is to provide, “حركات” aid to show the correct pronunciation. Arabic vowel marks include Fatha ف, “فتحة”, Kasra “كسرة”, Damma, «ضممة», Sukun “سكون”, Shadda and “شدة” Tanwin “تنوين”. The pronunciation of these vowel marks are represented in Table2 below:

Double Constant	No Vowel	Tanwin			Vowel		
		in	un	an	Kasrah	Dammah	Fathah
ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ
shadda	Sukun	in	un	an	Kasrah	Dammah	Fathah

Table1. Arabic Diacritics

However, in Modem Standard Arabic (MSA), vowel marks are not usually included in printed and electronic text, and the understanding and correct pronunciation of the word is determined within its context by the reader, so we decide not to remove (if it exists), the vowels as a step on preprocessing phase.

III. IMPORTANT STEPS

The method we proposed, is based on preprocessing step, treatment and check steps, here is a description of each one:

A. Preprocessing

In this step we proceed to:

- Divide text into words
- Format the word by removing any punctuation, diacritics and non-letter characters
- Check if the word is a stop word (3)

(3) This step consists on eliminating (very frequent) words that contain no or little information to help discriminate the text they occur in.

We use a list of large stop words, here an example:

ان	انما	إنما	بعد	ضد
يلي	الى	في	من	حتى
وهو	يكون	به	وليس	أحد
على	وكان	تلك	كذلك	التي
لكن	عن	مساء	ليس	منذ

This list should be maintained.

B. Searching in strange words list

In this step our algorithm will check if the word is a part of strange words (it is a word that comes from another language than Arabic, and used in the modern Arabic language especially), those words exists in a list of Strange words constituted manually.

Here an examples of strange words:

وكوفي	بلورة	خوجة	ألمانيا	أوروبا
شيراك	فرنسا	مانديلا	ديسمبر	تلفزة

Table2. Strange words in Arabic

If the word exists the algorithm returns the word, otherwise the treatment continues.

C. Check if the word exists in the list of words that begins by Waw:

The stemmer removes letter “و” (“and”) from the beginning of the words if the length of the word is more than three characters, and if the word doesn’t exist in the list of “Words_begins_by_Waw.txt”, because many common Arabic words begin with this character.

وشط	وشز	وشر	وشخ	وشح
وطر	طد	وشغ	وشع	وشظ
وطي	وطن	وطف	وطش	وطس

D. Check if the word exists in the list of words that begins by ‘AL’

The stemmer removes letter “ال” (“and”) from the beginning of the word if the length of the word is more than three characters, and if the word doesn’t exist in the list of “Words_begins_by_AL.txt”, because many common Arabic words begin with this character.

أبط	أبض	أبر	أبد	أبب
أبو	أبه	أين	أبل	أبق
أتي	أتو	أتن	أتم	أبي

E. Normalization

The third step in the stemmer is normalization of the words. Normalization process in the proposed stemmer is the similar to the normalization process in Light10 stemmer which runs as following:

- Remove Hamza from letter “أ” (Replace “أ | إ | ؤ” with “ا”)
- Replace final letter “ة” with “ه”.
- Replace final letter “ى” with “ي”.

F. Removing the prefixes and suffixes from the words

This step consists of removing the prefixes and suffixes from the words.

G. Applying rules : check if the word matches any of the patterns

The last step after deleting the prefixes and suffixes of the words is correcting any word that its meaning changed:

In some cases, a letter in the pattern of the word is deleted affecting the process of root extraction, like in the word

In (فعل) of the pattern (رأى) the present tense of the word, ‘ر’ is deleted, and ‘ال’ the letter (يفعل) which follows the pattern for morphological (يرأى) not (يرى) the word becomes of ‘ع’ reasons. So the letter the pattern is deleted, becoming (يفل) instead of (يفعل)

There are three rules which apply in the stemmer for correcting some words their meaning was affected.

- Adding “ي” to the end of the word if the suffix “يه” is deleted
- Adding “ه” to the end of the word if the suffix “ته” is deleted
- Replacing the letter “ى” to the end of the word by “ء” if the suffix of the word is deleted.

IV. DETAILED ALGORITHM

To implement our algorithm, we have used Java code program, based on Khoja’s one.

These are the schema of our algorithm:

```

{
// check if the word consists of two letters:
if ( word.length () == 2 )
if the word consists of two letters, we treat two cases:
- A root consisting of two letters (though I can't think of any!)
- A letter was deleted as it is duplicated or a weak middle or last letter.

// if the word consists of three letters
if( word.length () == 3 && !rootFound )
- If the last letter is a weak letter or a hamza, then check for last week words list.
- If the second letter is a weak letter then check for second week words list.

// if the word consists of four letters
if( word.length () == 4 )
// check if it's a root
- Check on the list of four letters root.

// if the root hasn't yet been found
if( !rootFound )
{
// check if the word is a pattern
- Try and find a pattern that matches the word
}
// if the root still hasn't been found
if ( !rootFound )
{
// check for a definite article, and remove it
word = checkDefiniteArticle ( word );
- look through the vector of definite articles search through each definite article, and try and find a match
- Check to see if the word is a root of three or four letters
- If the word has only two letters, test to see if one was removed
- If the root hasn't been found, check for patterns/ check for suffixes/ prefixes
}
}

```

```
// if the root still hasn't been found
if ( !rootFound && !stopwordFound )
{
    // check for the prefix waw
    word = checkPrefixWaw ( word );
    - Check to see if the word is a stopword
    - Check to see if the word is a root of three or four
      letters, that begin by 'waw'
    - If the word has only two letters, test to see if one was
      removed
    - if the root hasn't been found, check for patterns
    - Check for suffixes
    - check for prefixes
}
}
```

```
// if the root STILL hasn't been found
if ( !rootFound && !stopwordFound )
{
    // check for suffixes
    word = checkForSuffixes ( word );
}
}
```

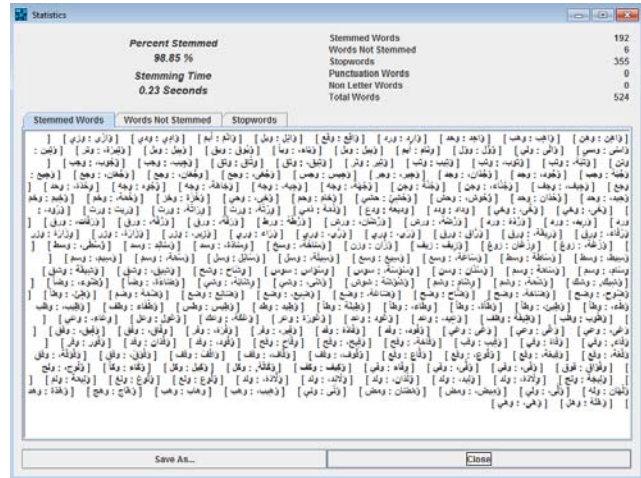
```
// if the root STILL hasn't been found
if ( !rootFound && !stopwordFound )
{
    // check for prefixes
    word = checkForPrefixes ( word );
    - Check to see if the word is a stopword
    - Check to see if the word is a root of three or four
      letters.
    - If the word has only two letters, test to see if one was
      removed
    - if the root hasn't been found, check for patterns
    - Check for suffixes
    - check for prefixes
}
}
return word;
}
```

V. SOME RESULTS :

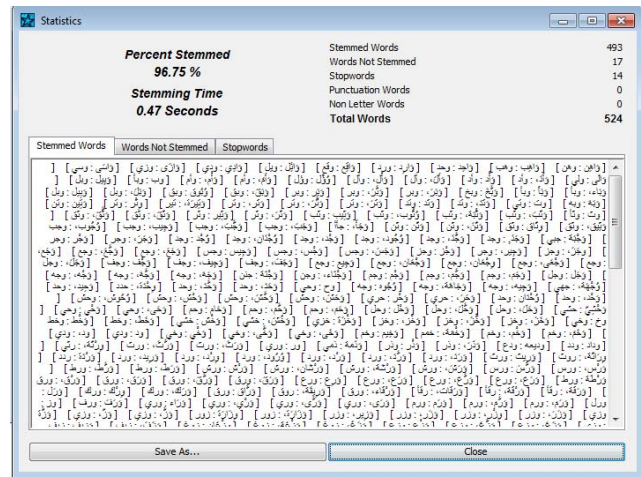
The stemming result of the word will be correct, if the output form of the word is the same as the target form of the word. Otherwise, the result of the word will be incorrect. We have used Khoja's java code and we applied our approach on it. This program take in input a text file and returns in output a list of words theirs stems and the type of the each word.

We have used as a first test a list of 524 words that begins by "waw" letter, we give below results of our stemmer comparing it to Khoja's one:

a. Our Stemmer's results:



b. Khoja's results:



c. Comparing results:

	Khoja's Stemmer	Our Stemmer: EST.Stemmer
Number of words used	524	534
Stemmed words	96,75 %	98,85 %
Not Stemmed words	3,25 %	1,15 %

A second test that we have done with an article that contained 1418 word, this are results:

	Khoja's Stemmer	Our Stemmer: EST.Stemmer
Number of words used	1418	1418
Stemmed words	93,16 %	93,23 %
Not Stemmed words	6,84 %	6,77 %

Since the text didn't contain words that begins by "waw", some strange words, and five nouns, the difference between the two results is not large.

Some particular cases that are solved with our Stemmer:

For example, if we take: وَسْمَةٌ، وَسَامَةٌ.

With Khoja's algorithm we have as stemmed text:

Word	Stem	Type
وَسْمَةٌ	سمي	ROOT
وَسَامَةٌ	سوم	ROOT

With our algorithm EST.Stemmer:

Word	Stem	Type
وَسْمَةٌ	وسم	ROOT
وَسَامَةٌ	وسم	ROOT

We have also review and modify stop word list to solve some issues detected in our tests, for example we have added 'ففي' and 'منهم' in this list.

In the future work we will test the accuracy of our algorithm and compare it with Light and Heavy ones.

References

- [1] Soufiane Farrah, Hanane El Manssouri, El Houssaine Ziyati. « Proposition of improvement areas in most heavy an light stemmer algorithms novel stemmer: EST.Stemmer». In *Information Science and Technology (CIST), 4th IEEE CiS'16 Program*. IEEE, 2016.
- [2] Buckwalter, T. Qamus: Arabic lexicography <http://members.aol.com/ArabicLexicons/>
- [3] Khoja, S. and Garside, R. Stemming Arabic Text. Computing Department, Lancaster University, Lancaster. Internet Home Page, 1-7 (1999). [Available at: <http://www.comp.lancs.ac.uk/computing/users/khoja/stemer.ps>]
- [4] H. AlSerhan, R. A. Shalabi, and G. Kannan, "New approach for extracting arabic roots," in Proceedings of The 2003 Arab conference on Information Technology (ACIT 2003), Alexandria, Egypt, Dec. 2003, pp. 42-59
- [5] Nehar, Attia, Djelloul Ziadi, Hadda Cherroun, and Younes Guellouma. "An Efficient Stemming for Arabic Text Classification." In *Innovations in Information Technology (IIT), 2012 International Conference on*, 328-32. IEEE, 2012.
- [6] Larkey, Leah S., et Margaret E. Connell. « Arabic Information Retrieval at UMass in TREC-10. » In *TREC*, 2001.
- [7] Taghva, Kazem, Rania Elkhoury, et Jeffrey S. Coombs. « Arabic Stemming Without A Root Dictionary. » In *ITCC (1)*, 152-157, 2005.
- [8] Hadni, Meryeme, Abdelmonaime Lachkar, et S. Alaoui Ouatik. « A new and efficient stemming technique for Arabic Text Categorization ». In *Multimedia Computing and Systems (ICMCS), 2012 International Conference on*, 791-796. IEEE, 2012.
- [9] Saudagar, Abdul Khader Jilani, Habeeb Vulla Mohammed, Kamran Iqbal, et Yasir Javed Gyani. « Efficient Arabic text extraction and recognition using thinning and dataset comparison technique ». In *Communication, Information & Computing Technology (ICCICT), 2015 International Conference on*, 1-5. IEEE, 2015.
- [10] M. A. H. Omer and S. Ma, Stemming Algorithm to Classify Arabic Documents , Journal of Communication and Computer, Vol. 7, No. 9, Sep. 2010
- [11] Elrajubi, Osama Mohamed. « An improved Arabic light stemmer ». In *Research and Innovation in Information Systems (ICRIIS), 2013 International Conference on*, 33-38. IEEE, 2013.
- [12] The Five Nouns In Arabic : <https://fr.scribd.com/doc/93320581/The-Five-Nouns-In-Arabic>
- [13] Mohammed N. Al-Kabi, Saif A. Kazakzeh, Bilal M. Abu Ata, Saif A. Al-Rababah, Izzat M. Alsmadi « A novel root based Arabic stemmer ». In *Journal of King Saud University - Computer and Information Science*, 94-103, Elsevier, April 2015.
- [14] Ghwanmeh, S., Kanaan, G., Al-Shalabi, R., Rabab'ah, 2009. "Enhanced Algorithm for Extracting the Root of Arabic Words". In: the Sixth International Conference on Computer Graphics, Imaging and Visualization, (CGIV '09). pp. 388-391.
- [15] Chen, A., Gey, F., 2002. "Building an Arabic stemmer for information retrieval". In: Proceedings of the Eleventh Text REtrieval Conference (TREC 2002). National Institute of Standards and Technology, pp. 631-639.
- [16] A.F.A. Nwesri, S.M.M. Tahaghoghi, F. Scholer "Stemming Arabic Conjunctions and Prepositions". Lect. Notes Comput. Sc., 3772 (2005), pp. 206-217 http://dx.doi.org/10.1007/11575832_23
- [17] Taghva, K., Elkhoury, R., Coombs, J., 2005. "Arabic stemming without a root dictionary". In: International Conference on Information Technology: Coding and Computing (ITCC 2005), pp. 152-157.

CHALLENGES THAT RESTRICT THE EFFICIENCIES OF SECURITY FRAMEWORKS IN E-COMMERCE: A REVIEW

Kenneth Longo Mlelwa
PhD Cand., School of CoCSE,
NM-AIST, Arusha, Tanzania
mlelwak@nm-aist.ac.tz

Eng. Dr. Zaipuna O. Yonah
Senior Lecturer, School of COCSE
NM-AIST, Arusha, Tanzania
zaipuna.yonah@nm-aist.ac.tz

Abstract - Most of the businesses all over the world have appeared on the Internet to offer everything possible. Some of these businesses have succeeded and some have failed spectacularly. The only thing that the successful organizations have in common is the reality that they understand that they are doing eCommerce to make money. Businesses who prefer to perform eCommerce are taking a risk. They are investing in new knowledge and latest ways of providing goods and services in expecting of generating a profit from the activity. The jeopardy to the organization come from numerous areas: the public may not accept the service, the new clients may not appear, or existing clients may not prefer the new service. Therefore these organizations which are performing eCommerce, a whole new set of threats and vulnerabilities must be considered. These new threats and vulnerabilities generate new risks that must be managed. Thus Security in eCommerce grow to be inevitable and hence this paper has investigate, identify, classify and present a different type of security challenges facing eCommerce transactions.

Index Terms — eCommerce Challenges, eCommerce Security, eCommerce Security Measure, Security Efficiencies.

I. INTRODUCTION

Electronic commerce is any economic or business activity that uses Information Communication Technology (ICT) based applications to facilitate the buying and selling of products and services and to facilitate the transaction of trade activities between and among merchants, individuals, governments or other organizations. This includes using ICTs to toughen a company's internal activities, such as logistics, procurement, and human resource and contracts management, information as well as data management, communication functions, and to assist the flow of products between merchants and customers, e.g. marketing, ordering, payment, delivery, and searching for suppliers [1]. From economic advantages eCommerce has a several benefits such as increasing market expansion, reduction of product source prices, promotion of productivity, reducing of operation costs and inflation, reducing uncertainty, sharing market information, and aiding in distribution channel efficiency and plays a fundamental role in an endogenous economic growth [2]. It can also be sources that develop domestic economic and fast globalization of production, and development of available technology [3]. Both Africa and the Middle East e experience from very specific issues that need to be integrated into world agenda and agreements taking place where, the obstacles are very well understood and have been researched by many [4, 5]

Holistically, e-commerce refers of using technological development to promote everything involving the exchange of business information among computers and humans or traders and customers [6] [7]. Due to that; everyone who is using eCommerce needs to be concern about the security of their personal information. Thus this study intends to determine the main constraints that restrict the efficiencies of security frameworks in e-commerce.

A. Currently Security Framework

Security is clearly defined as the state of being secure that is free from danger and threats, as well as to be protected from adversaries, which are those who would intentionally or unintentionally do harm. eCommerce security is the protection of information, systems and hardware that store, use and pass on information throughout digital transaction.

A big number of security frameworks concentrate on three areas namely, detection by using scanners, prevention by using tools such as proxy and firewall, and recovery using tools regarding cryptography techniques and proper planning. These frameworks intend to specifically address the problems of security and confidence from theoretical and practical perspective [8]. This includes those actors that interact with and play a role in Ecommerce or may contribute to its improvement. (Refer the Figure 1)

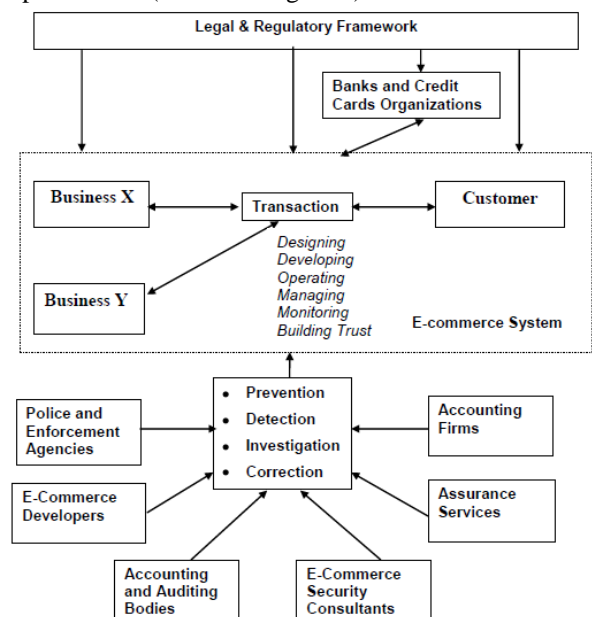


Figure 1: eCommerce Security framework
According to the security attacks or threats defined in the X.800 and RFC 2828 documents, security attacks are

classified into two, which is the passive attack that only involves eavesdropping with motives of obtaining information that is being transmitted, while the other being the active attack that involves modification of the data stream or creation of false stream with motives of obtaining authorization.

B. Security frameworks' requirements

Generally eCommerce consist of a chain of events; several products and techniques are used to secure parts of the chain, with that fact in mind, here under are descriptions of the main categories of security needs.

- *Authentication*, an assurance that the communicating entity is the one claimed to be.
- *Access Control*, prevention of unauthorized personnel who misuse resources.
- *Data Confidentiality*, protection of data from unauthorized disclosure.
- *Data Integrity*, the assurance that received data is as sent by an authorized entity.
- *Non-repudiation*, protection against denial by one of the parties in a communication.

The majority of eCommerce transactions frameworks consist of four parties; which are a client, a merchant, a respective bank as well as card issuing bank [9, 10]. A client, i.e. the cardholder, makes a payment using a card issued by the card issuing bank (issuer) for something bought from a merchant. The acquiring bank (acquirer) is the financial institution with which a business has a contractual arrangement for receiving (acquiring) card payments. The underlying payment model is shown in Figure 1.

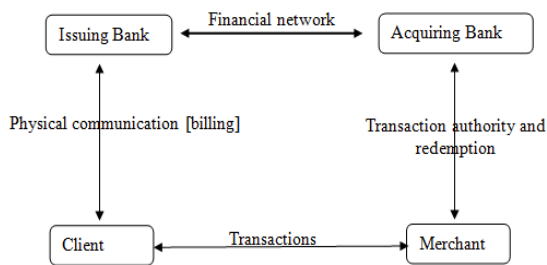


Figure 2: eCommerce payment framework

As the figure above shown; the security requirements for each party vary and as examined. But both acquires and issuers' requirements are combine simply because they're both financial institutions, they are both obliged to abide by the rules of the relevant payment system, and it can reasonably be assumed that they have a similar risk model.

1) Issuers and Acquires

Non-repudiation: Issuers and acquirers need to ensure that neither clients nor eCommerce merchants can reject their involvement in a transaction (especially when the transactions involve a reimbursement from merchant to client).

Authentication it has two major levels: high-level and standard. A "personal identifier" (username or name) and something you recognize (password) are the standard level. If a higher level of security than passwords is needed, user can be required to "have something" and "know something". The have-something part includes biometrics (e.g., fingerprints), smartcards and a private or public key infrastructure (PKI) key.

Here the Client authentication is required for the issuers and acquirers in order to prove that it is the client who authorized the payment and that is a legitimate cardholder. Or else, a client can deny making a transaction and the issuer may end up being liable for refunding the amount to the client [10].

Data integrity means that data are not changed in transit. It is important to ensure that once details of a transaction have been confirmed, no one can maliciously modify them. Merchants must not be able to alter the amount that a client has agreed to pay.

Replay protection (Privacy): A malicious merchant should not be able to use a once authorized transaction to obtain a repeat payment. Additionally, merchants should not be able to use an old transaction to request a new payment authorization no matter how many similar transactions the client has made with them. Issuers and acquirers need a mechanism to detect if a transaction has been replayed so that they do not authorize an illegitimate transaction [10].

2) Merchants

Non-repudiation: A merchant needs proof that a customer has agreed to pay the amount allied with a deal. A merchant also desires to verify that the client is the genuine cardholder; or else, the merchant can be accountable for refund. This occurs when a client tells his/her issuer that a certain transaction was not made. The card issuer then instantly submits a chargeback to the acquirer to recover the amount from the account of the merchant in question. Within a predefined period of time, the merchant can quarrel the chargeback by providing evidence of, such as, purchase or delivery. Therefore, it is important for merchants to have non-repudiable evidence of the transaction

Authentication: As stated before, merchants need client authentication to make sure that the client is the legitimate cardholder. Moreover, they need to be sure that they are communicating with the genuine acquirer. Otherwise, an adversary may masquerade as an acquirer and authorize an illegitimate transaction.

Integrity: No one should be able to change the particulars of a transaction once they have been settled upon. A merchant will not wish to be credited with payment for less than the amount agreed. In addition, an acquirer or issuer should not be able to amend a transaction that has been authorized.

Replay protection: A malicious client should not be able to present an old proof of purchase to claim for repeat delivery of goods. Likewise, it should not be possible for an acquirer to claim that a merchant has obtained a payment using an old transaction.

3) Clients

Confidentiality and privacy: Transaction confidentiality, especially card information confidentiality, may be the security service of most concern to users. It is important that cardholder account details are kept secret from any party except the issuer, since they are the main basis on which Internet payments are made. Moreover, some users may require confidentiality protection for the nature of their transactions

Integrity: As for the other parties, transaction integrity is important to the client. No one should be able to maliciously modify the transaction details once they have been confirmed. Clients will not want an adversary

to change a delivery address, the price, or the description of the merchandise after they have agreed a payment.

Authentication: A client needs to be sure that he/she is dealing with a trustworthy merchant. When shopping on the Internet, it is relatively easy to be lured into visiting a site which appears to sell something but is actually simply collecting card details. Even though a client may have made a purchase from a site before, it is not always obvious whether the page that is being fetched is authentic.

Replay protection: Clients need a mechanism to ensure that a malicious merchant or an adversary will not be able to reuse previously authorized payments to make a repeat charge.

Non-repudiation: Clients also require non-repudiation, for example a proof of payment so that no one involved in the transaction can repudiate that a payment has occurred.

- The protection of information in transit involving the customer’s system and the server
- The safety of information that is saved to the customer’s system and
- The protection of the fact that a particular customer made a particular order.

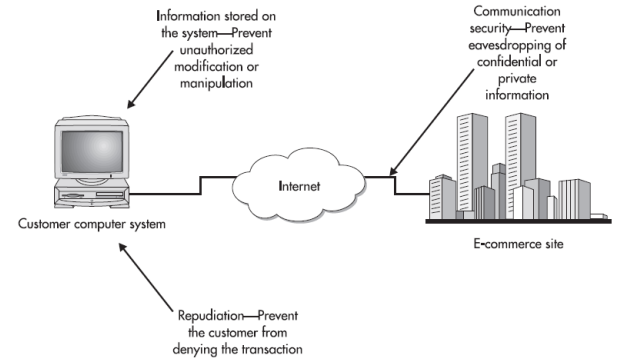


Figure 4: Client-level security components

DIMENSIONS	CUSTOMER'S PERSPECTIVE	MERCHANT'S PERSPECTIVE
Integrity	Has information I transmit or receive been altered?	Has data on the site been altered without authorization? Is data being received from customers valid?
Nonrepudiation	Can a party to an action with me later deny taking the action?	Can a customer deny ordering products?
Authenticity	Who am I dealing with? How can I be assured that the person or entity is who they claim to be?	What is the real identity of the customer?
Confidentiality	Can someone other than the intended recipient read my messages?	Are messages or confidential data accessible to anyone other than those authorized to view them?
Privacy	Can I control the use of information about myself transmitted to an e-commerce merchant?	What use, if any, can be made of personal data collected as part of an e-commerce transaction? Is the personal information of customers being used in an unauthorized manner?
Availability	Can I get access to the site?	Is the site operational?

Table 1: Customer and Merchant perspectives on the different dimensions of eCommerce Security

C. Most Security threats in EC environment

In eCommerce framework the key points of which is vulnerable for attack are; Client level, Server level and Communications pipeline sometime refers as Internet communication channels. These Client side, Server side and communication Channels collective in this study we refer as an eCommerce environment.



Figure 3: Vulnerable Points in an eCommerce Environment

Client-level security deals with the security from the consumer’s desktop system to the eCommerce server. This part of the system consists of the customer’s computer and browser software and the communications link to the server [11]. On this part of the system, there are several issues such as;-

Communications security for e-commerce

applications covers the security of information that is involved the client’s system and the e-commerce server. This may contain sensitive information such as credit card numbers or site passwords. It may also consist of confidential information that is sent from the server to the client’s system such as customer files.

A Server-level security, these consist of the Physical eCommerce server and the Web server software running on it. The eCommerce server itself must be available from the Internet. Access to the system may be limited or it may be open to the public. Again on this part of the system there are two issues related to server security;

- The security of information stored on the server
- The protection of the server itself from compromise

D. Other Security threat

1) Denial of Service attacks

A denial-of-service (DoS) attack is an effort to make a machine or network resource unavailable to its intended users, like temporarily or indefinitely interrupt or suspend services of a host connected to the Internet.

A denial-of-service attack is portrayed by an explicit attempt by attackers to prevent genuine users of a service from using that service. There are two common forms of DoS attacks: those that crash services and those that flood services.

The most severe attacks are distributed [12] and in many or most cases include falsify of IP sender addresses (IP address spoofing) so that the site of the attacking machines can neither easily be identified, nor can filtering be done based on the source address.

2) SQL Injection attack

This is a code injection technique, used to attack database-driven application, in which malicious SQL statement are inserted into an entry field for execution.

This attack permit attackers to spoof identity, tamper with existing data, cause repudiation issues like voiding transactions or changing balances, permit the complete disclosure of all data on the system, devastate the data or make it otherwise unavailable, including becoming an administrators of the database server.

3) Session Hijacking

Also known as cookie hijacking is the utilization of a valid computer session to achieve unauthorized access to information or services in a computer system. In particular, it is used to refer to the theft of a magic used to authenticate a user to a remote server. It has particular significance to web developers, as the HTTP cookies used to maintain a session on many web sites can be easily stolen by an attacker using an intermediary computer or with access to the saved cookies on the victim's computer.

A mostly preferred method is using source-routed IP packets. This allows an attacker at point X on the network to participate in a conversation between Y and Z by cheering the IP packets to pass through Y's machine.

If source-routing is turned off, the attacker can deploy a "blind" hijacking, whereby it guesses the responses of the two machines. Therefore, the attacker can send a command, but can never see the response. Though, a common command would be to set a password allowing access from somewhere else on the net.

4) Cross-site script (XSS)

Cross-site scripting (XSS) is a type of computer security vulnerability typically found in web applications. XSS enables attackers to inject client-side scripts into web pages viewed by other users. A cross-site scripting vulnerability may be used by attackers to bypass access controls such as the same-origin policy.

An attacker can use XSS to send a malicious script to an unsuspecting user. The end user's browser has no way to know that the script should not be trusted, and will execute the script. Since it thinks the script came from a trusted source, the malicious script can access any cookies, session tokens, or other sensitive information retained by the browser and used with that site. These scripts sometimes can even rewrite the content of the HTML page.

While XSS can be taken advantage of within VBScript, ActiveX and Flash (although now considered legacy or even obsolete), unquestionably, the most widely abused is JavaScript – primarily because JavaScript is fundamental to most browsing experiences.

E. TECHNOLOGY SOLUTION

1) Repudiation

One big risk allied with the client side (Client-level) of eCommerce is the possibility for a client to repudiate a transaction. Noticeably, if the client actually did not initiate the transaction, the organization should not let it. Nevertheless, how does the organization choose whether a client is really who he says he is? The respond is through authentication.

The category of authentication that is used to confirm the identity of the client depends on the danger to the organization of making an error. In the case of a credit card purchase, there are established procedures for performing a credit card transaction when the card is not present. These include having the client supply a proper mailing address for the purchase.

If the eCommerce site is providing a service that needs a verification of individuality to access certain information, a credit card may not be suitable. It may be better for the organization to use user IDs and passwords

or even two-factor authentication. In any of these cases, the terms of service that are sent to the client should detail the requirements for protecting the ID and password. If the correct ID and password are used to access customer information, it will be assumed by the organization that a genuine customer is accessing the information. If the password is lost, forgotten, or compromised, the organization should be contacted instantly.

2) Information Stored on the Server

The eCommerce server is open to access from the Internet in some way. As a result, the server is at most partly-trusted (un-trusted). An un-trusted system should not store sensitive information. If the server is used to accept credit card transactions, the card numbers should be instantly removed to the system that actually processes the transactions (and that is located in a more secure part of the network). No card numbers should be kept on the server.

a) Protecting the Server from Attack

If information has to be kept on the e-commerce server, it should be protected from unauthorized access. The way to do this on the server is employ the use of file access controls. Additionally, there are things that can be done to protect the server itself from successful penetration as follows:

Server location

Normal server location consists of physical location and its network location. Since server is more important to any organization thus, it should be located within a physical protected area where by physical access to the server should be protected by a locked cage and separated.

Like the server location, its network location is also more important the Figure 4 below illustrate the proper location of the server within the DMZ (a DMZ or Demilitarized zone is a physical or logical subnetwork that contains and exposes an organization's external-facing services to a larger and untrusted network, such as Internet). The firewall should be configured so as to only allow access to the eCommerce server on ports 80 (for HTTP) and (for HTTPS). No more services are required for the public to access the eCommerce server and therefore should be stopped at the firewall [11].

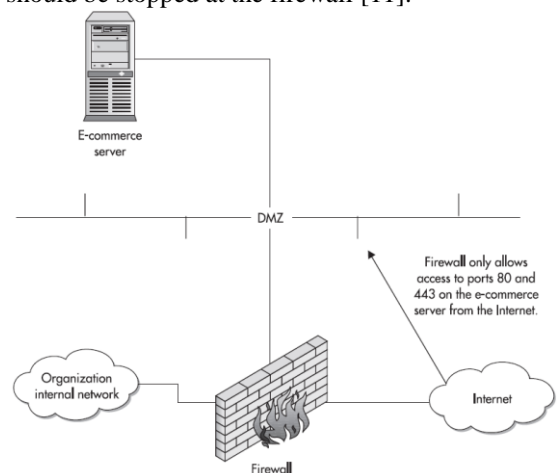


Figure 5: Location of network for the eCommerce server

Operating system configuration

Normal eCommerce server operating system is configured with security in mind. Choosing OS depends on a number of factors, such as the expertise of the administration staff of an organization, also another factor like performance requirements and fail-over capabilities must be considered. Again it is advisable to choose operating system that the administration staff is familiar with.

The most important step in configuring the server securely is to eradicate or turn off any unnecessary services. The system is chiefly a Web server and, as a result, it must run a Web server [11].

The second step is to patch the system. Ensure for the latest patches for the chosen operating system and load them. When the patches are loaded, configure the system to conform to organization policy with regard to password length and change frequency, audit, and other requirements.

Lastly sooner than the system is declared ready for production, you should scan it for vulnerabilities. Vulnerability scanners can be saleable or freely available, but they have to be recently.

Web server configuration

The last component of the server security is the web server itself. There is specific configuration requirements for web servers instead there are some common configurations that should be made despite of the web server.

Initial, the server software should be upgraded and patched according to the manufacturer's recommendations.

Never run the Web server as root or administrator. If the Web server is successfully penetrated, the attacker will have privileges on the system as the admin of the Web server. In its place, create a separate user who owns the Web server and run the server from that account.

Every Web server needs the administrator to define a server root directory. This directory informs the Web server where to find scripts and document files as well as limits the Web server in what files can be accessed via a browser. The Web server root is not supposed to be the same as the system root directory, and it should not comprise configuration and security files that are important to the operating system [11]

b) Protecting Internet Communication

There is one reasonable solution to this: **encryption**. A good number of standard Web browsers contain the facility to encrypt traffic. This is the default solution if HTTPS is deployed rather than HTTP. The encryption of HTTPS will guard the information from the time it leaves the client's computer until the time it reaches the Web server. Since the public has learned of the dangers of someone gaining access to a credit card number on the Internet HTTPS become as a preferred solution.

When HTTPS is used, a Secure Socket Layer (SSL) connection is made between the client and the server. All traffic over this tie is encrypted.

c) Other protecting techniques

Password policies

A password policy is a set of rules intended to enhance computer security by encouraging users to employ strong passwords and use them correctly. This policy is often part of an organization's official regulations and may be taught as part of security awareness training. Moreover the password policy is

simply advisory, or the computer systems enforce users to abide by it. Some governments have national authentication frameworks [13] that describe requirements for user authentication to regime services, including requirements for passwords.

Digital Signatures and Certificates

With the development of technology many people and organization are using online documents instead of tradition documents for their day-to-day activities, and due to that digital signatures and digital certificates support this phenomenon by providing assurance about the validity and authenticity of a digital document.

A **Digital signature** is a mathematical technique used to confirm and validate the authenticity and integrity of a message, software or digital document. It is equivalent of a handwritten signature or embossed seal, but offering extra inherent security, a digital signature is planned to solve the dilemma of tampering and impersonation in digital communications. Digital signatures can offer the added assurances of a proof to origin, identity and status of an electronic document, transaction or message, as well as acknowledging informed consent by the signer.

Digital Certificates is an electronic "passport" that permits a person, computer or organization to swap information securely over the With the development of technology many people and organization are using online documents instead of tradition documents for their day-to-day activities, and due to that digital signatures and digital certificates support this phenomenon by providing assurance about the legitimacy and authenticity of a digital document.

A **Digital signature** is a mathematical technique used to confirm the authenticity and integrity of a message, software or electronic document. It is equivalent of a handwritten signature or stamped seal, but offering extra inherent security, a digital signature is planned to solve the dilemma of tampering and imposture in digital communications. Digital/electronic signatures can offer the added assurances of a proof to origin, identity and status of an electronic document, transaction or message, as well as acknowledging informed consent by the signer.

Digital Certificates is an electronic "passport" that permits a person, computer or organization to exchange information securely over the Internet using the public key infrastructure (PKI). A digital certificate sometime referred to as a public key certificate. In order to provide proof that a certificate is legitimate and valid, it is digitally signed by a root certificate belonging to a trusted certificate authority. Operating systems and browsers preserve records of trusted CA root certificates so they can simply verify certificates that the CAs have issued and signed. When PKI is deployed internally, digital certificates can be self-signed.

Firewalls

This is a network security system that scrutinizes and manages the received and leaving network traffic based on predetermined security policy [14]. A firewall typically establishes an obstacle among a trusted, secure internal network and external network, like the Internet, that is assumed not to be secure or trusted. Here we categorized firewall as either personal firewall or Web Server Firewall.

A **personal firewall** (sometimes called a **desktop firewall**) is a software program used to protect a sole Internet-connected computer from intruders. Personal firewall protection is particularly useful for users with "constantly-on" connections such as DSL or cable modem. It also controls network traffic to and from a computer, allowing or disallowing communications based on a security policy. Typically it works as an application layer firewall.

Web Server Firewall or Web application firewall is a security policy enforcement point positioned between a web application and the client end point. This functionality can be implemented in software or hardware, running in an appliance device, or in a typical server running a common operating system. It may be a stand-alone device or integrated into other network components [15]

This protects web applications AS in the same technique a traditional firewall protects a network. It controls the input and output, as well as the access to and from the asset it is protecting.

In contrast to traditional firewalls that usually block access to certain ports or filter by IP address, web application firewalls look at every demand and response within the different web service layers such as HTTP, HTTPS, SOAP, and XML-RPC. The thorough scrutiny of web traffic that web application firewalls perform has also earned them the nickname "Deep Packet Inspection Firewalls" [15].

F. CONCLUSIONS AND RECOMMENDATIONS

Due to the massive development current technology pave the way for a secure site design. However no company can ever claim to be 100% covered by any security measure. Security matters are extremely important for the survival of any eCommerce solution, and for that reason must be constantly analyzed and taken care of. Security problems in eCommerce frameworks are caused by many factors hence to solve these problems from different aspects, offer a variety of countermeasures as elaborated here under;

Social ethical Norms More awareness should be paid to risks and security, which is a major issue why customers shy away from using eCommerce in general [16]; figure 5 below depict that those 47% who has never used this service their main reason was mainly safety of payment and low trust [16].



Figure 6: Showing the barriers for online shopping [16]

Security Strategy; In order to avoid and discourage future security threats, it is necessary for an organization to constantly develop a security policy related know-how and to invest in technology such as deploy the use of a proxy server, firewall, Virtual Private Network technology; in the identification and authentication, encryption and authentication techniques.

Legal Issues in eCommerce transaction and security issue should be protected by law. Lawmaking developments are facing a new challenge due to the fast improvement of technology and by the newly created difficulty of deploying existing regulations in a networked environment. Concurrently, technology offers more and more solutions to the threats created by the presence of this new technology. Technology and law must develop along each other's progress and integrate common input.

REFERENCES:

- [1] Dr. Wanjau K., Macharia R. and Ayodo E (2012). Factors Affecting Adoption of Electronic Commerce among Small Medium Enterprises in Kenya: Survey of Tour and Travel Firms in Nairobi: International Journal of Business, Humanities and Technology.
- [2] Molla, A., Taylor, R. & Licker, S., (2006). Ecommerce Diffusion in Small Island Countries: the Influence of Institution in Barbados. The Electronic Journal of Information Systems in Developing Countries, 28(2),pp. 1-15.
- [3] Sheth, J. N., & Sharma, A., (2005). International emarketing: Opportunities and issues. International Marketing Review, 22(6), pp. 611-622.
- [4] A. Molla and P.S. Licker, (2005). Perceived EReadiness factors in E-Commerce adoption: An empirical investigation in a developing country, International Journal of Electronic Commerce 10 (1) ,pp. 83-110.
- [5] Travica, B., (2002). "Diffusion of electronic commerce in developing countries: The case of Costa Rica, Journal of Global Information Technology Management, 5(1), pp. 4-24.
- [6] Kaetzel L.J E-commerce and IP on the Internet 1996 page 186
- [7] Curie Wendy (2000). "The Global Information Society". Chichester, John Wiley & Sons, Ltd.
- [8] R. Jamieson and N. Cerpa (2001) "A Research Framework for Risk, Security, Trust and Assurance within an Electronic Commerce Domain" CACS Conference September 2001
- [9] V. Hassler. Security Fundamentals for E-commerce. Artech House Publishers, 2001.
- [10] D. O'Mahony, M. Peirce, and H. Tewari. Electronic Payment System for E-Commerce. Artech House Publishers, 2001.
- [11] E. Maiwald, Network Security; A beginner's guide, 3rd Edition, Dreamtech press, 2008.
- [12] Taghavi Zargar, Saman (November 2013). "A Survey of Defense Mechanisms against Distributed Denial of Service (DDoS) Flooding Attacks" (PDF). IEEE communications surveys & tutorials. pp. 2046-2069. Retrieved 2016-02-07
- [13] <http://folk.uio.no/josang/papers/ATJK2012-SARSSI.pdf> as accessed on 10th February 2016
- [14] Boudriga, Nouredine (2010). Security of mobile communications. Boca Raton: CRC Press. pp. 32-33. ISBN 0849379423
- [15] <http://www.applicure.com/solutions/web-application-firewall> accessed on 10th January 2016
- [16] K. Mlrelwa, B. Chachage and Y. Zaipuna. Article: E-Commerce Trend in Developing Countries: A Case Study of Tanzania. International Journal of Computer Applications 125(1):27-33, September 2015. Published by Foundation of Computer Science (FCS), NY, USA. BibTeX

Pressure Oscillations in the Water Distribution Network before and behind the Solenoid Valve

Saeid Jafari, Ali Ghaffari*

Department of computer engineering, Tabriz branch, Islamic azad university, Tabriz, Iran

A.Ghaffari@iaut.ac.ir

Abstract

Leakage in water distribution networks in urban and rural is inevitable. Today, reducing leakage using hydraulic parameters management such as pressure on leak detection project is of particular importance. Intelligent control of pressure is a good way to control leakage and reduce damage caused by high pressure in water distribution networks and applying the results of conducted studies in this regard is needed to help new equipment. In this paper, the main objective is to determine the effect of pressure reduction on leakage in a network at low at peak times of consumption. Accordingly, a network has been modelled using flow measurement method using the minimum and maximum hydraulic analysis software and using the results of model, a time plan has been applied on valve of input network to changes the head of output pressure during the day and on the basis of minimum low pressure in the lowest point of pressure.

Keywords: Reduction of Leakage, Management of Hydraulic Parameter, Intelligent Control of Pressure, Changes of Pressure Head

1. Introduction:

Pressure is one of the most important factors affecting the rate of water without current in water distribution networks [1-4]. It also has the largest and fastest hydraulic effect on the amount of leakage[5]. Operational efficiencies generated using prevalent water industry methods and techniques are becoming more difficult to achieve. Monitoring incoming flow rate to networks indicates reduction the water consumption in the midnight hours and increasing the pressure in the water distribution network at the same time [6, 7]. Recently with advancement in ICT

(information and communication technology) based schemes such as wireless sensor networks (WSNs) [8-13], the managing and controlling of water pipe status such as failure or breaking is done in effective way. Researches and experiences have shown that in durations more than an hour, consuming rate at night is not a function of pressure and the only ground leakage and breakages change by pressure [14, 15]. In researches that conducted by the international water association in leakage management, pressure management has been determined as the easiest, quickest and probably the cheapest way to reduce leakage. Industrial control systems of pressure primarily depend on reducing valves with constant pressure output. High amplitude of pressure fluctuations between the minimum and the maximum level in these conditions causes to fatigue of network and increasing the risk of failure in it.

Experts believe that can preserve the network pressure at peak times at the maximum necessary by adjusting the flow and constantly changing of settings of output pressure of the reducing valves in addition to make lower the pressure in the hours of reduction consumption [16-19]. The idea of intelligent control pressure [20-23] has been implemented for a long time in the form of executive designs in some countries and Statistical results proved its authenticity. In this paper, the main objective is to determine the effect of pressure reduction on leakage in a network at low at peak times of consumption.

The rest of the paper is organized as follows: Section 2 presented related works. Section 3 indicated the proposed scheme. Section 4 presented the evaluation of the proposed scheme and finally, Section 5 conclude the paper.

2. Related works

Pressure oscillation has an important effect in water distribution networks and many related works exist in this area [24]. In [25], the authors have announced the expansion of water supply systems as a necessity for promoting health in communities. Therefore, planning for repair process and maintenance after that also is considered as one of the important pillars of the distribution network. In [26], the authors have dealt in an article on how automated measuring the flow and pressure by using precision instruments deployed in the distribution network. Review, analysis and the proposed method based on newly developed technique to generate information on the operation of the distribution network, work showed that this technology can be alarming and fully functional. So that in the result of creating this thinking in addition,

network management, to minimize the negative impact had been effective the pressure on customer service standards in achieving the goals of optimization. In recent decades the international communities and countries have paid special attention to the problem of limitation of water resources and finding solutions to deal with water shortages especially prevent water loss.

In [3], the authors presents a new logic algorithm for real-time control of regulation valves in water distribution networks. This method entails identifying in real time the appropriate closure setting of regulation valves in order to reach and keep the desired piezometric height at the control node(s), by making use of measurements concerning both the piezometric height at the control node(s) and the water discharge in the pipes fitted with regulation valves.

Water issue that is not accounted for in water supply systems and urban water distribution networks are of the cases that are taken into consideration in the past two decades in many countries and by scientific and systematic planning, good experiences have been acquired also about theoretical arguments and its executive cases of its reduction. Fortunately, this fact has been also welcomed more than a decade in Iran and first in the form of national study design, development and improvement of water distribution facilities during the second development program and then has been studied and reduced losses and unaccounted water by monitoring offices on the reduction of unaccounted water in Water and Wastewater Companies (Publication No. 380 A in 2010).

3. Automation concept

Automation or intelligent building means automatic control of a process or system that is done by analog and digital inputs and outputs with programming of CPU PLC (programmable logic control). These automatic control systems are used in factories, industrial powerhouse, and different kinds of machines and production lines to reduce costs and minimize manpower. Over this aim, increasing productivity automation (automated systems 24 hours a working day) and reduce costs associated with automation operator have been consisted of three basic parts. They are: (1). Measurements, (2). Controllers and (3). Stimulants.

Measurements such as types of sensors: temperature, pressure, motion, transmitter when at time of stimulating them, and they respond to this reaction and do not require external energy for this reaction. The controllers are parts of the two instruments of control part. Control is the lead and maintenance of one or more process with condition or desirable conditions, this concept involve quality and quantity controlling and safety that are the main objectives of control. Stimulus are the utensils that take output signal from the controller parts and act in accordance with the signals such as control valves, electric motors, contactors.

- stepper motor
- Structure of stepper motor
- How to control
- 1-bit control
- 2-bit control
- How to move the electric motors

3.1. Stepping Motor: it is one of the electric motors that its movement is quite accurate and predefined and it can be moved by sending bits 0, 1 to its coil. Figure 1 shows how to move all motors.

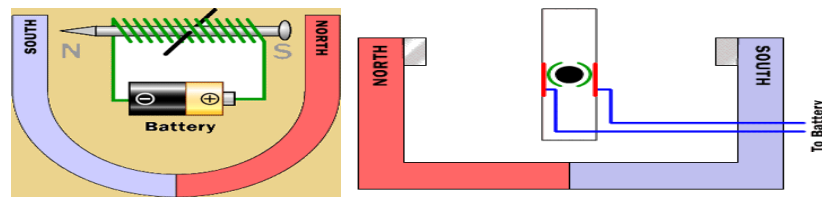


Figure 1. How to move all motors

3.2. Structure of stepper motor

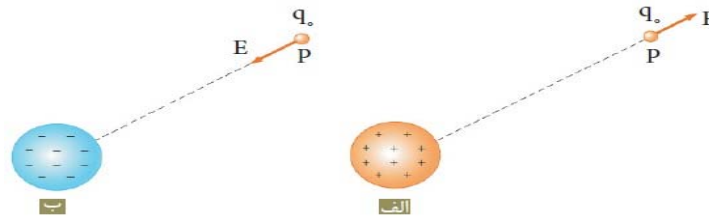


Figure 2. Structure of Stepping Motor OF Drives

This motor (Figure 2) generally has four poles that windings are wound on these four poles and you create magnetic field in fact by sending bits 0 and 1 to these windings that this field causes the rotor magnet movement within the stepper motor. Of course should be done these windings to the sequence of 0, 1, otherwise the motor will not turn according to your will and a characteristic of this engine is its angular momentum and each engine has its own movement angle.

For example, if the motor that its movement angle is 7 degrees, this motor at every time that its windings have voltage will turn 7 degrees in counter-clockwise direction or unlike its direction depending on the windings in what order will have the voltage. This 7 degree rotation for this stepper motor is considered an example of a step; you understand with this definition that a stepper motor in a full circle may have 100 to 200 steps less or more depending on the type of engine.

You can even use a stepper motor in half way of the stairs means the half angle of movement. These motors also move in micro step in fact the purpose to move are very superfine and accurate. When you see a stepper motor from near you understand some colored wires that come out of the stepper motor In fact, these wires are each connected to a winding and a wire is shared and common between all the wires.

4. How to control

This engine moves in the way of one bit or two bits in one bit form at any moment only one coil receives pulses 1 and In the case of two-bit two coils receive the pulse 1 at any moment. If this receiving the pulse is done on a regular basis and one after another engine also moves for the

correct way in counter-clockwise direction or it will move in the opposite direction. The way of stepper motor control is investigated in two forms of a bit or two-bit.

4.1. How to control 1-bit

In the 1-bit mode if first we stimulate the coil 1, and windings 2, 3 and 4 should be without stimulation. Direction of movement of stepper motor is in the counter-clockwise direction and after winding 1 is turn of winding 2 that to be stimulated.

And in this case also other windings are without stimulation. After that is the turn of 3 and 4 windings. Note that at any moment a winding stimulates if we stimulate winding 4 after winding 1 and then go to the third and second one, the motor will rotate in a clockwise direction.

4.2. How to control the 2-bit

In the mode of two bit at the moment two windings are charged for example if the first windings 1 and 2 are stimulated and then windings 2, 3, and then 3 and 4 and finally 4 and 1 to move the stepper motor should keep this order until that time that you want that motor has movement. Now if you change this order the motor moves in the opposite direction of current direction.

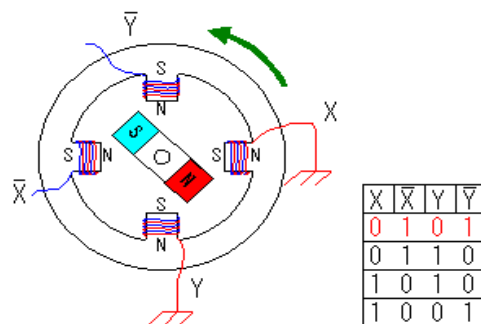


Figure 3. Moving in the opposite direction of clockwise (2-bit stimulation)

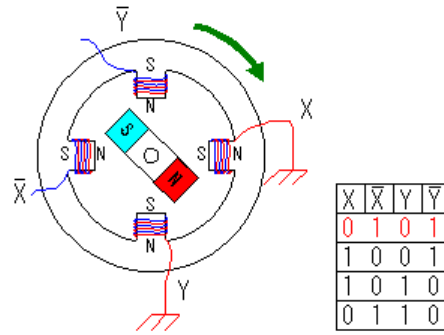


Figure 4. Moving in clockwise direction (2-bit stimulation)

5. The output of the results of operation of the device

Device built based on the accepted planning, controls pressure values and the subsequent of it distribution of water at different hours of the day and night. This can be achieved by developing the working program in C environment and also Java programs. The inverter converts AC current to DC and converts DC output to the AC frequency to make frequency to motor drive in order to be able to increase or decrease the frequency. Then it takes Feedback of command input terminals. In this case, we divide 0 to 10 volts by 1000 stairs to control the output frequency of inverter to direction of drive of motor from 0 to 60 (Hz).

Set width

In solenoid valve the direction of movement is between 0 to 100 percent that based on the pressure of solenoid valve to be able to control the pressure moves in different directions until it stops by reaching to the point where we want it for the pressure value. And we use the following formula to get the value of psi. (For transformation of flow of 4 to 20 mA to psi pressure is used this mode).

$$\blacksquare \text{Base} * 4 + \text{offset} \quad \longrightarrow \quad \text{offset} = 62$$

$$\text{Psi} = \blacksquare - 62.4$$

And to obtain the required pressure value is calculated from the following formula.

■ = Base * 4 + offset ➔ offset = 43

Psi = ■ - 43 / 3.8

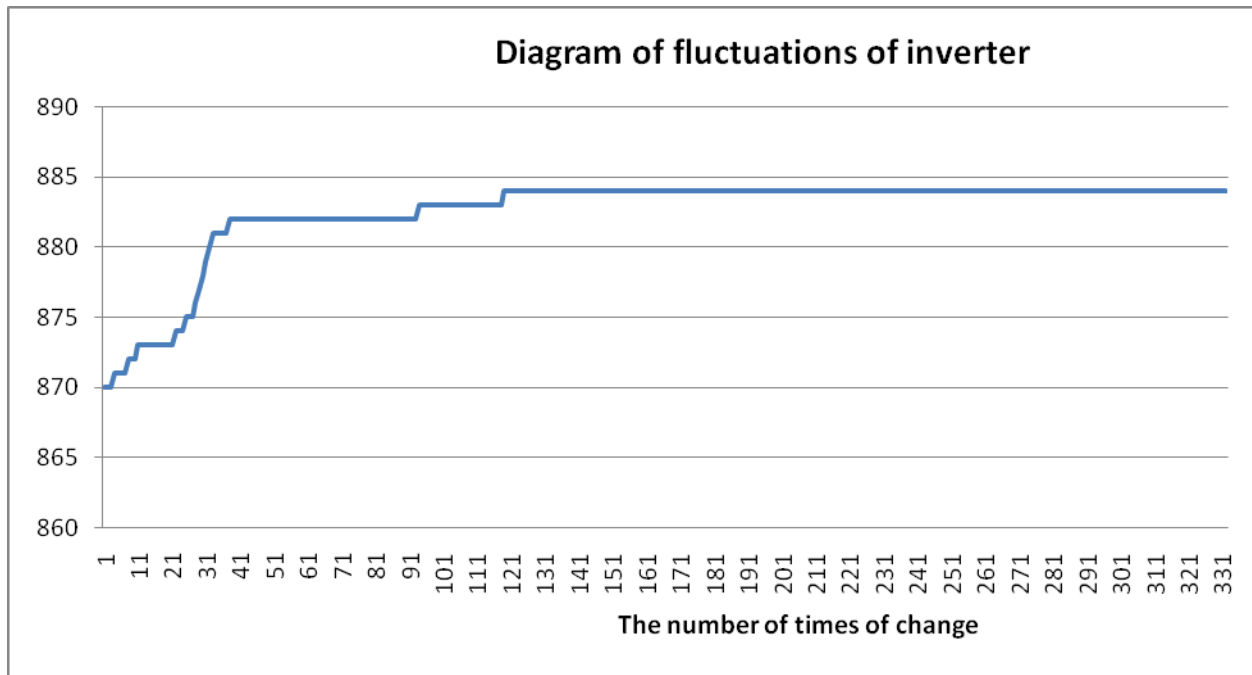


Figure 5 - Diagram of fluctuations of inverter

The above diagram represents the number of fluctuations of the inverter. These fluctuations have occurred for change in pressure range defined in the machine. As can be seen after fluctuations and consecutive stepper changes, finally in repeat of 122th, the amount of fluctuations is cut off and inverter operates linearly. In fact, the corresponding number to start of this stage is the set the pressure range in defined range in the network.

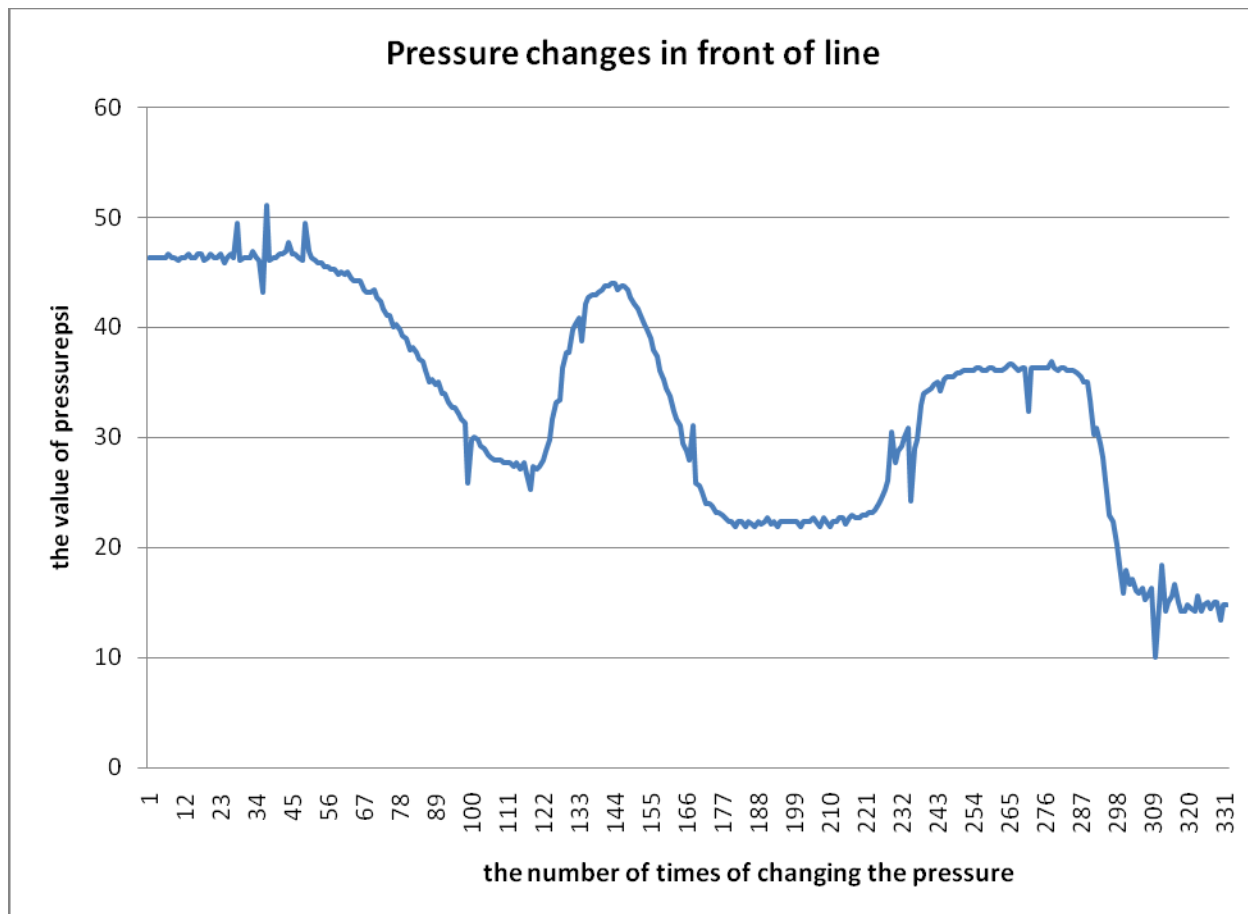


Figure 6. The diagram of Pressure changes versus solenoid valve

The above diagram shows the pressure fluctuations in network and before the solenoid valve. Lars line and thalweg are the times that consumptions are too high and too low. What is evident is that this at the peak time, the pressure has been at its defined lowest level. This interpretation would be contrary to the low consumption mode.

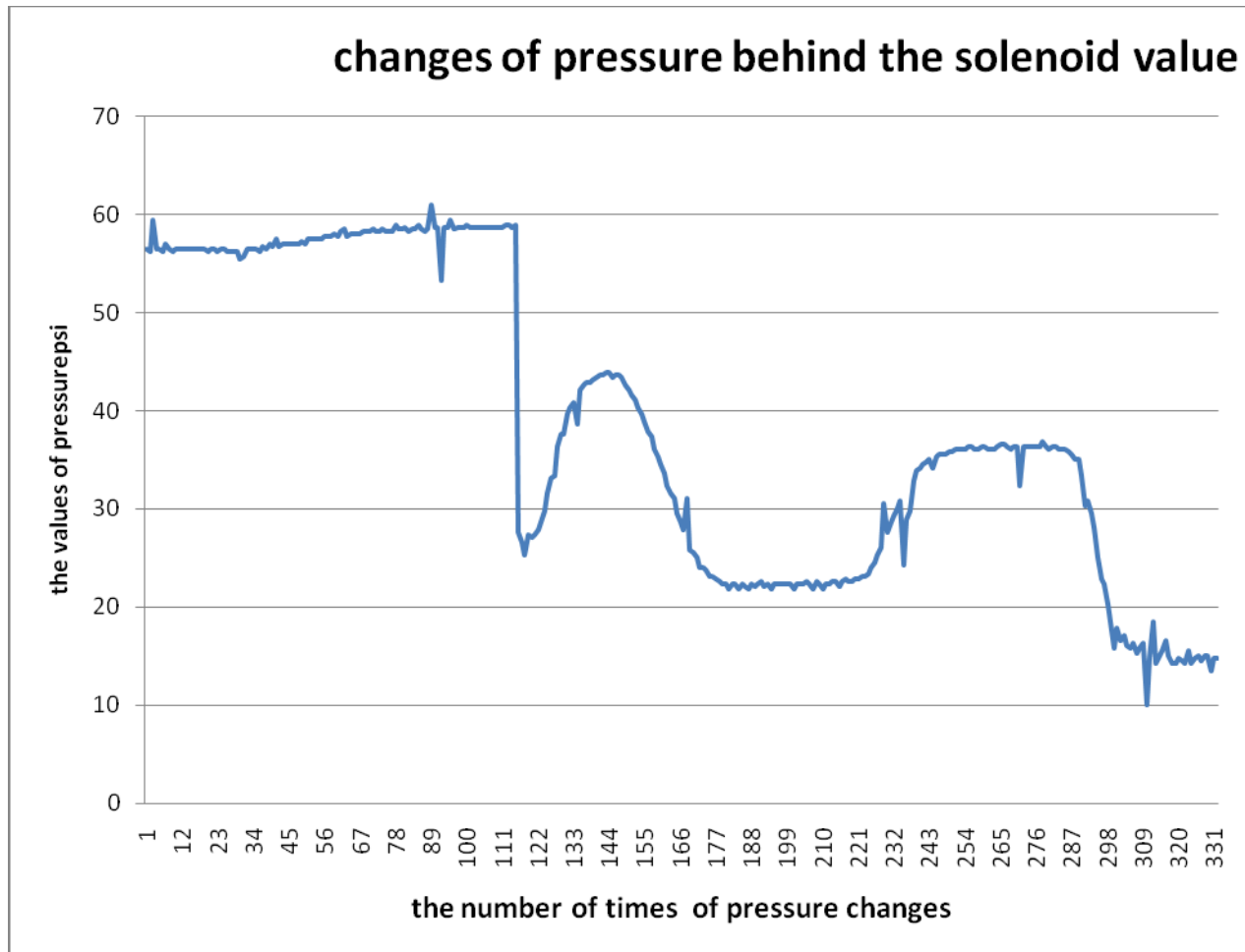


Figure 7. Diagram of pressure changes behind the solenoid valve

Interpretations offered to Fig. 6, are also true in Fig. 7. What that is important in the shape of above diagrams is important, is the existence some small fluctuations that have been recorded. In fact, these fluctuations reflect the imprecision enough used gas valve (in place of the solenoid valve).

Looseness in the valve is the main cause of these jumps because despite the movement of solenoid valve at time that is issued of the inverter the order of open or close, the gas valve does not behavior due to the existence the looseness in accordance with solenoid valve.

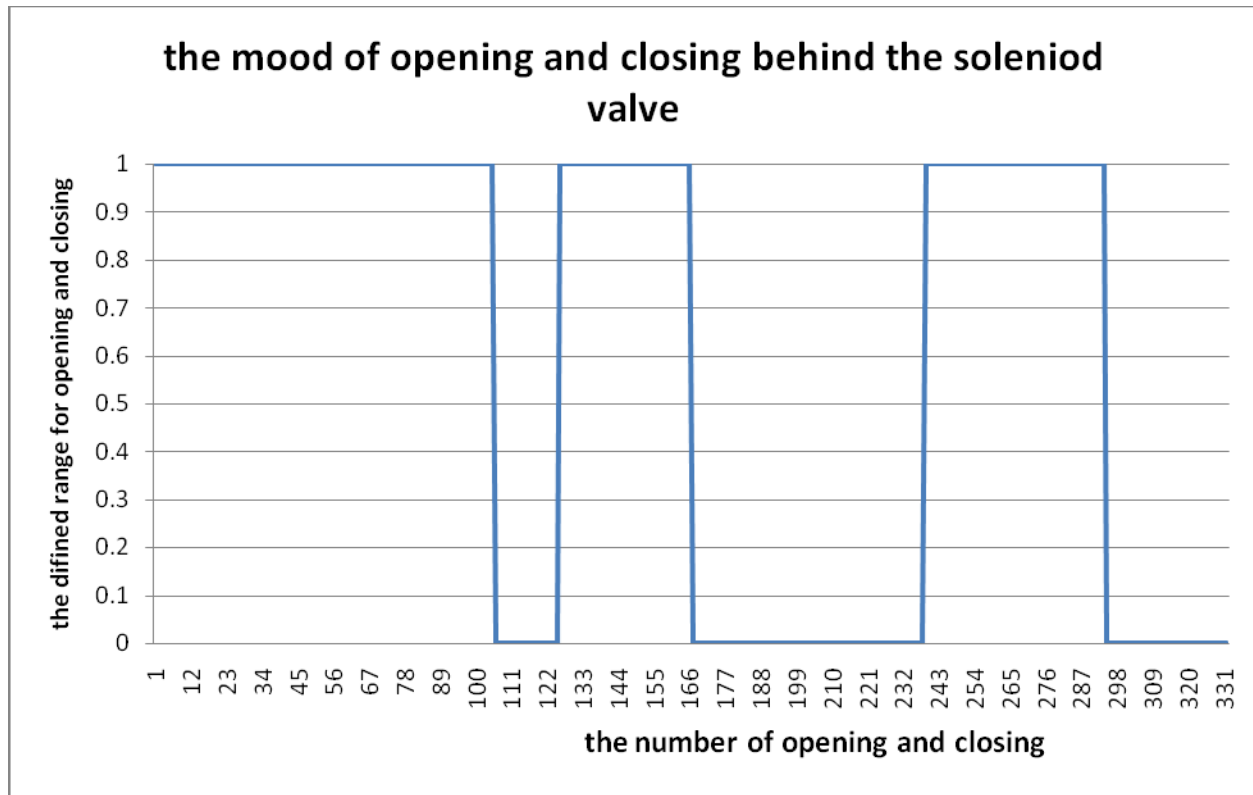


Figure 8. Diagram of pressure changes behind the solenoid valve

Fig. 8 shows that when the solenoid valve has received the command of opening or closing. The corresponding values of openness (become) is one, and for closing, is zero.

6. Conclusion:

Based on what is mentioned, in a comprehensive look, the advantages of this methodology can be expressed as follows:

- Conventional methods due to be fixed the slope of changes and in its consequent lack of their flexibility in different situations, will not be effective help in preventing water loss.
- Smart control method causes to balance the pressure in the whole network and provides the ability to control pressure in terms of water use in network and provides in the standard range. It prevents physical and even non-physical water losses.
- Intelligent control method makes adjusting the water pressure in the whole distribution network and consequently prevents the occurrence of events.

Reference list

- [1] C. D. Onal, X. Chen, G. M. Whitesides, and D. Rus, "Soft mobile robots with on-board chemical pressure generation," in *Robotics Research*, ed: Springer, 2017, pp. 525-540.
- [2] J. Gregory and M. N. Tomac, "A review of fluidic oscillator development," in *43rd AIAA Fluid Dynamics Conference*, 2013, p. 2474.
- [3] E. Creaco and M. Franchini, "A new algorithm for real-time pressure control in water distribution networks," *Water Science and Technology: Water Supply*, vol. 13, pp. 875-882, 2013.
- [4] A. F. Falcão and J. C. Henriques, "Oscillating-water-column wave energy converters and air turbines: A review," *Renewable Energy*, vol. 85, pp. 1391-1424, 2016.
- [5] C. M. Shanks, J.-B. Sérodes, and M. J. Rodriguez, "Spatio-temporal variability of non-regulated disinfection by-products within a drinking water distribution network," *Water research*, vol. 47, pp. 3231-3243, 2013.
- [6] P. van Thienen, "A method for quantitative discrimination in flow pattern evolution of water distribution supply areas with interpretation in terms of demand and leakage," *Journal of Hydroinformatics*, vol. 15, pp. 86-102, 2013.
- [7] A. Pham, M. Barisik, and B. Kim, "Pressure dependence of Kapitza resistance at gold/water and silicon/water interfaces," *The Journal of chemical physics*, vol. 139, p. 244702, 2013.
- [8] A. Ghaffari and L. Azari, "Proposing a novel method based on network-coding for optimizing error recovery in wireless sensor networks," *Indian Journal of Science and Technology*, vol. 8, p. 859, 2015.
- [9] A. Ghaffari, "Designing a Wireless Sensor Network for ocean status notification system," *Indian Journal of Science and Technology*, vol. 7, pp. 809-814, 2014.
- [10] R. Mohammadi and A. Ghaffari, "Optimizing reliability through network coding in wireless multimedia sensor networks," *Indian Journal of Science and Technology*, vol. 8, p. 834, 2015.
- [11] Z. Mottaghinia and A. Ghaffari, "A Unicast Tree-Based Data Gathering Protocol for Delay Tolerant Mobile Sensor Networks," *Information Systems & Telecommunication*, vol. 59, pp. 1-12, 2016.
- [12] A. Ghaffari, "An energy efficient routing protocol for wireless sensor networks using A-star algorithm," *Journal of applied research and technology*, vol. 12, pp. 815-822, 2014.
- [13] A. Ghaffari, "Congestion control mechanisms in wireless sensor networks: A survey," *Journal of network and computer applications*, vol. 52, pp. 101-115, 2015.
- [14] G. Freni, M. De Marchis, and E. Napoli, "Implementation of pressure reduction valves in a dynamic water distribution numerical model to control the inequality in water supply," *Journal of Hydroinformatics*, vol. 16, pp. 207-217, 2014.
- [15] E. M. Chiapero, M. Fernandino, and C. Dorao, "Experimental study of pressure drop oscillations in parallel horizontal channels," *International Journal of Heat and Fluid Flow*, vol. 50, pp. 126-133, 2014.
- [16] R. López-Roldán, S. Platikanov, J. Martín-Alonso, R. Tauler, S. González, and J. L. Cortina, "Integration of Ultraviolet-Visible spectral and physicochemical data in chemometrics analysis for improved discrimination of water sources and blends for

- application to the complex drinking water distribution network of Barcelona," *Journal of Cleaner Production*, vol. 112, pp. 4789-4798, 2016.
- [17] P. Sivakumar and R. Prasad, "Simulation of water distribution network under pressure-deficient condition," *Water resources management*, vol. 28, pp. 3271-3290, 2014.
- [18] J. El-Chakhtoura, E. Prest, P. Saikaly, M. van Loosdrecht, F. Hammes, and H. Vrouwenvelder, "Dynamics of bacterial communities before and after distribution in a full-scale drinking water network," *water research*, vol. 74, pp. 180-190, 2015.
- [19] M. Mair, W. Rauch, and R. Sitzenfrei, "Spanning Tree-Based Algorithm for Generating Water Distribution Network Sets by Using Street Network Data Sets," in *World Environmental and Water Resources Congress 2014*, 2014, pp. 465-474.
- [20] F. Zheng, A. Simpson, and A. Zecchin, "Improving the efficiency of multi-objective evolutionary algorithms through decomposition: An application to water distribution network design," *Environmental Modelling & Software*, vol. 69, pp. 240-252, 2015.
- [21] A. De Corte and K. Sørensen, "HydroGen: an artificial water distribution network generator," *Water resources management*, vol. 28, pp. 333-350, 2014.
- [22] F. De Paola, N. Fontana, E. Galdiero, M. Giugni, D. Savic, and G. S. Degli Uberti, "Automatic multi-objective sectorization of a water distribution network," *Procedia Engineering*, vol. 89, pp. 1200-1207, 2014.
- [23] A. Campisano and C. Modica, "Two-Step Numerical Method for Improved Calculation of Water Leakage by Water Distribution Network Solvers," *Journal of Water Resources Planning and Management*, vol. 142, p. 04015060, 2015.
- [24] S. Jafari and A. Ghaffari, "Designing Intelligent Device to Control Pressure in Water Distribution Networks," *International Journal of Computer Science and Information Security*, vol. 14, p. 35, 2016.
- [25] F. R. Da Fonseca, W. S. Silva, C. F. Fontana, and E. M. Dias, "Division in sectors technique applied to hydric supplying systems," in *Proceedings of the 2nd international conference on Circuits, Systems, Communications & Computers*, 2011, pp. 51-61.
- [26] J. Machell, S. Mounce, B. Farley, and J. Boxall, "Online data processing for proactive UK water distribution network operation," *Drinking Water Engineering and Science*, vol. 7, pp. 23-33, 2014.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Dr Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

Prof. Syed S. Rizvi, University of Bridgeport, USA
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Dr. S. Mehta, Inha University, Korea
Dr. Dilip Kumar S.M, Bangalore University, Bangalore
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia
Dr. Saqib Saeed, University of Siegen, Germany
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India
Dr. Muhammad Sohail, KUST, Pakistan
Dr. Manjaiah D.H, Mangalore University, India
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Dr. M. Azath, Anna University, India
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr Suresh Jain, Devi Ahilya University, Indore (MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Dr. Hanumanthappa. J. University of Mysore, India
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Dr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India
Dr. P. Vasant, Power Control Optimization, Malaysia
Dr. Petr Ivankov, Automatika - S, Russian Federation
Dr. Utkarsh Seetha, Data Infosys Limited, India
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore
Assist. Prof. A. Neela madheswari, Anna university, India
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh
Dr. Atul Gonsai, Saurashtra University, Gujarat, India
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand
Mrs. G. Nalini Priya, Anna University, Chennai
Dr. P. Subashini, Avinashilingam University for Women, India
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat
Mr Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah
Mr. Nitin Bhatia, DAV College, India
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia
Assist. Prof. Sonal Chawla, Panjab University, India
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology,
Durban, South Africa
Prof. Mydhili K Nair, Visweswaraiah Technological University, Bangalore, India
M. Prabu, Adhiyamaan College of Engineering/Anna University, India
Mr. Swakkhar Shatabda, United International University, Bangladesh
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
Mr. Zeashan Hameed Khan, Université de Grenoble, France
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India
Dr. Maslin Masrom, University Technology Malaysia, Malaysia
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE
Dr. Abdul Aziz, University of Central Punjab, Pakistan
Mr. Karan Singh, Gautam Budtha University, India
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institution of Engg. & Tech. CHD, India

Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim
Dr. Bin Guo, Institute Telecom SudParis, France
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore
Mr. Miliindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India
Dr. C. Arun, Anna University, India
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology
Subhabrata Barman, Haldia Institute of Technology, West Bengal
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand
Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India
Mr. Serguei A. Mokhov, Concordia University, Canada
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA
Dr. S. Karthik, SNS College of Technology, India
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain
Mr. A.D.Potgantwar, Pune University, India
Dr. Himanshu Aggarwal, Punjabi University, India
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai
Dr. Prasant Kumar Pattnaik, KIST, India.
Dr. Ch. Aswani Kumar, VIT University, India
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA
Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA
Mr. R. Jagadeesh Kannan, RMK Engineering College, India
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia
Mr. R. Mohammad Shafi, Madanapalle Institute of Technology & Science, India
Dr. F. Sagayaraj Francis, Pondicherry Engineering College, India
Dr. Ajay Goel, HIET, Kaithal, India
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India
Mr. Suhas J Manangi, Microsoft India
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India
Dr. Amjad Rehman, University Technology Malaysia, Malaysia
Mr. Rachit Garg, L K College, Jalandhar, Punjab
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan
Dr. Thorat S.B., Institute of Technology and Management, India
Mr. Ajay Prasad, Sir Padampat Singhania University, Udaipur, India
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA
Mr. Anand Kumar, AMC Engineering College, Bangalore
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India
Dr. V V S S S Baram, Sreenidhi Institute of Science and Technology, India
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India
Prof. Niranjana Reddy, P, KITS, Warangal, India
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India
Dr. A. Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India
Dr. Tossapon Boongoen, Aberystwyth University, UK
Dr. Bilal Alatas, Firat University, Turkey
Assist. Prof. Jyoti Praaksh Singh, Academy of Technology, India
Dr. Ritu Soni, GNG College, India
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.
Dr. Binod Kumar, Lakshmi Narayan College of Tech. (LNCT) Bhopal India
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan
Dr. T.C. Manjunath, ATRIA Institute of Tech, India
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan

Assist. Prof. Harmunish Taneja, M. M. University, India
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India
Mr. G. Appasami, Dr. Pauls Engineering College, India
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan
Mr. Yaser Miaji, University Utara Malaysia, Malaysia
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India
Dr. S. Sasikumar, Roever Engineering College
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India
Mr. Nwaocha Vivian O, National Open University of Nigeria
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia
Dr. Dhuha Basheer abdullah, Mosul university, Iraq
Mr. S. Audithan, Annamalai University, India
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad
Mr. Deepak Gour, Sir Padampat Singhania University, India
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India
Mr. Ali Balador, Islamic Azad University, Iran
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India
Dr. Debojyoti Mitra, Sir padampat Singhania University, India
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia
Mr. Zhao Zhang, City University of Hong Kong, China
Prof. S.P. Setty, A.U. College of Engineering, India
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India
Dr. Hanan Elazhary, Electronics Research Institute, Egypt
Dr. Hosam I. Faiq, USM, Malaysia
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India

Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India
Prof Anupam Choudhary, Bhilai School Of Engg.,Bhilai (C.G.),India
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.
Dr. Kasarapu Ramani, JNT University, Anantapur, India
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India
Dr. C G Ravichandran, R V S College of Engineering and Technology, India
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India
Dr. Nikolai Stoianov, Defense Institute, Bulgaria
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia
Dr. Nighat Mir, Effat University, Saudi Arabia
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India
Mr. P. Sivakumar, Anna university, Chennai, India
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia
Mr. Nikhil Patrick Lobo, CADES, India
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India
Assist. Prof. Vishal Bharti, DCE, Gurgaon
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India
Mr. Hamed Taherdoost, Tehran, Iran
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran
Mr. Shantanu Pal, University of Calcutta, India
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria
Mr. P. Mahalingam, Caledonian College of Engineering, Oman
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt

Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India
Mr. Muhammad Asad, Technical University of Munich, Germany
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran
Prof. S. V. Nagaraj, RMK Engineering College, India
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India
Mr. Sunil Taneja, Kurukshetra University, India
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia
Dr. Yaduvir Singh, Thapar University, India
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India
Prof. Shapoor Zarei, UAE Inventors Association, UAE
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India
Prof. Anant J Umbarkar, Walchand College of Engg., India
Assist. Prof. B. Bharathi, Sathyabama University, India
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore
Prof. Walid Moudani, Lebanese University, Lebanon
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India
Associate Prof. Dr. Manuj Darbari, BBD University, India
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India
Dr. Abhay Bansal, Amity School of Engineering & Technology, India
Ms. Sumita Mishra, Amity School of Engineering and Technology, India
Professor S. Viswanadha Raju, JNT University Hyderabad, India
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India
Dr. Sukumar Senthikumar, Universiti Sains Malaysia, Malaysia

Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India
Mr. Shervan Fekri Ershad, Shiraz International University, Iran
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India
Ms. Sarla More, UIT, RGTU, Bhopal, India
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India
Dr. M. N. Giri Prasad, JNTUCE, Pulivendula, A.P., India
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India
Assist. Prof. Navnish Goel, S. D. College Of Engineering & Technology, India
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan
Mr. Mohammad Asadul Hoque, University of Alabama, USA
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina
Dr S. Rajalakshmi, Botho College, South Africa
Dr. Mohamed Sarrab, De Montfort University, UK
Mr. Basappa B. Kodada, Canara Engineering College, India
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India
Dr . G. Singaravel, K.S.R. College of Engineering, India
Dr B. G. Geetha, K.S.R. College of Engineering, India
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India
Assist. Prof. Maram Balajee, GMRIT, India
Assist. Prof. Monika Bhatnagar, TIT, India
Prof. Gaurang Panchal, Charotar University of Science & Technology, India
Prof. Anand K. Tripathi, Computer Society of India
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India
Assist. Prof. Supriya Raheja, ITM University, India

Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India
Prof. Mohan H.S, SJB Institute Of Technology, India
Mr. Hossein Malekinezhad, Islamic Azad University, Iran
Mr. Zatin Gupta, Universti Malaysia, Malaysia
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India
Assist. Prof. Ajal A. J., METS School Of Engineering, India
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India
Md. Nazrul Islam, University of Western Ontario, Canada
Tushar Kanti, L.N.C.T, Bhopal, India
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh
Dr. Kashif Nisar, University Utara Malaysia, Malaysia
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan
Assist. Prof. Apoorvi Sood, I.T.M. University, India
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia
Mr. Swapnil Sonar, Truba Institute College of Engineering & Technology, Indore, India
Ms. Yogita Gigras, I.T.M. University, India
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India
Mr. Mehdi Golsorkhatabar Amiri, Islamic Azad University, Iran
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India
Assist. Prof. P.Oliver Jayaprakash, Anna University, Chennai
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India
Dr. Asoke Nath, St. Xavier's College, India
Mr. Masoud Rafiqhi, Islamic Azad University, Iran
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India
Mr. Sandeep Maan, Government Post Graduate College, India
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India
Mr. R. Balu, Bharathiar University, Coimbatore, India
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India
Prof. P. Senthilkumar, Vivekanandha Institue of Engineering and Techology for Woman, India
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran
Mr. Laxmi chand, SCTL, Noida, India
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad
Prof. Mahesh Panchal, KITRC, Gujarat
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhanian University, India
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India
Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India
Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India
Mr. G. Premsankar, Ericsson, India
Assist. Prof. T. Hemalatha, VELS University, India
Prof. Tejaswini Apte, University of Pune, India
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India
Mr. Vorugunti Chandra Sekhar, DA-IICT, India
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia
Dr. Aderemi A. Atayero, Covenant University, Nigeria
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India
Mr. Hassen Mohammed Abdulllah Alsafi, International Islamic University Malaysia (IIUM) Malaysia
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan
Mr. R. Balu, Bharathiar University, Coimbatore, India
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India
Prof. K. Saravanan, Anna university Coimbatore, India
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN
Assoc. Prof. S. Asif Hussain, AITS, India
Assist. Prof. C. Venkatesh, AITS, India
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan
Dr. B. Justus Rabi, Institute of Science & Technology, India
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India
Mr. Alejandro Mosquera, University of Alicante, Spain
Assist. Prof. Arjun Singh, Sir Padampat Singhanian University (SPSU), Udaipur, India
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India
Mr. Hassen Mohammed Abdulllah Alsafi, International Islamic University Malaysia (IIUM)
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu
Dr. K. Reji Kumar, N S S College, Pandalam, India

Assoc. Prof. K. Seshadri Sastry, EILM University, India
Mr. Kai Pan, UNC Charlotte, USA
Mr. Ruikar Sachin, SGGSIET, India
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt
Assist. Prof. Amanpreet Kaur, ITM University, India
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia
Dr. Abhay Bansal, Amity University, India
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA
Assist. Prof. Nidhi Arora, M.C.A. Institute, India
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India
Dr. S. Sankara Gomathi, Panimalar Engineering college, India
Prof. Anil kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India
Dr. Lamir LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India
Mr. Ram Kumar Singh, S.V Subharti University, India
Assistant Prof. Sunish Kumar O S, Amalijothei College of Engineering, India
Dr Sanjay Bhargava, Banasthali University, India
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India
Mr. Roohollah Etemadi, Islamic Azad University, Iran
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria
Mr. Sumit Goyal, National Dairy Research Institute, India
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur
Dr. S.K. Mahendran, Anna University, Chennai, India
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab
Dr. Ashu Gupta, Apeejay Institute of Management, India
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus
Mr. Maram Balajee, GMR Institute of Technology, India
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria
Mr. Jasvir Singh, University College Of Engg., India
Mr. Vivek Tiwari, MANIT, Bhopal, India
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India
Mr. Somdip Dey, St. Xavier's College, Kolkata, India

Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh
Mr. Sathyapraksh P., S.K.P Engineering College, India
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India
Mr. Md. Abdul Ahad, K L University, India
Mr. Vikas Bajpai, The LNM IIT, India
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai
Mr. A. Siles Balasingh, St. Joseph University in Tanzania, Tanzania
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India
Mr. Kumar Dayanand, Cambridge Institute of Technology, India
Dr. Syed Asif Ali, SMI University Karachi, Pakistan
Prof. Pallvi Pandit, Himachal Pradesh University, India
Mr. Ricardo Verschueren, University of Gloucestershire, UK
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India
Dr. S. Sumathi, Anna University, India
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India
Assist. Prof. M. Anand Kumar, Karpagam University, Coimbatore, India
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat
Mr. Sivakumar, Codework solutions, India
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA
Mr. Varadala Sridhar, Varadhman College Engineering College, Affiliated To JNTU, Hyderabad
Assist. Prof. Manoj Dhawan, SVITS, Indore
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India
Dr. S. Santhi, SCSVMV University, India
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh
Mr. Sandeep Reddivari, Mississippi State University, USA
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal
Dr. Hazra Imran, Athabasca University, Canada
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India
Ms. Jaspreet Kaur, Distance Education LPU, India
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India

Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India
Mr. Khaldi Amine, Badji Mokhtar University, Algeria
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India
Dr. Nadir Bouchama, CERIST Research Center, Algeria
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco
Dr. S. Malathi, Panimalar Engineering College, Chennai, India
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan
Dr. G. Rasitha Banu, Vel's University, Chennai
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India
Ms. U. Sinthuja, PSG college of arts & science, India
Dr. Ehsan Saradar Torshizi, Urmia University, Iran
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt
Dr. Nishant Gupta, University of Jammu, India
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India
Dr. Rahul Malik, Cisco Systems, USA
Dr. S. C. Lingareddy, ALPHA College of Engineering, India
Assistant Prof. Mohammed Shuaib, Interat University, Lucknow, India
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India
Dr. T. Thambidurai, Sun Univercell, Singapore
Prof. Anandkumar Telang, BKIT, India
Assistant Prof. R. Poorvadevi, SCSVMV University, India
Dr Uttam Mande, Gitam University, India
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India
Dr. Mohammed Zuber, AISECT University, India
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India

Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq
Dr. Urmila Shrawankar, G H Raison College of Engineering, Nagpur (MS), India
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India
Dr. Mukesh Negi, Tech Mahindra, India
Dr. Anuj Kumar Singh, Amity University Gurgaon, India
Dr. Babar Shah, Gyeongsang National University, South Korea
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India
Assistant Prof. Ankit Garg, Amity University, Haryana, India
Assistant Prof. Rajashe Karappa, SDM CET, Karnataka, India
Assistant Prof. Varun Jasuja, GNIT, India
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India
Dr. Faouzi Hidoussi, UHL Batna, Algeria
Dr. Naseer Ali Hussein, Wasit University, Iraq
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai
Dr. Ahmed Farouk Metwaly, K L University
Mr. Mohammed Noaman Murad, Cihan University, Iraq
Dr. Suxing Liu, Arkansas State University, USA
Dr. M. Gomathi, Velalar College of Engineering and Technology, India
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran
Dr. Thiyagu Nagaraj, University-INOU, India
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India
Dr. Shenshen Liang, University of California, Santa Cruz, US
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia
Mr. Snehasis Banerjee, Tata Consultancy Services, India
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia
Dr. Ying Yang, Computer Science Department, Yale University, USA
Dr. Vinay Shukla, Institute Of Technology & Management, India
Dr. Liviu Octavian Maftciu-Scai, West University of Timisoara, Romania
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India
Dr. Timothy Powers, University of Hertfordshire, UK
Dr. S. Prasath, Bharathiar University, Erode, India
Dr. Ritu Shrivastava, SIRTIS Bhopal, India
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania

Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India
Assistant Prof. Mallikarjun C Sarsamba Bheemna Khandre Institute Technology, Bhalki, India
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Dr. Parul Verma, Amity University, India
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India
Assistant Prof. G. Selvavinayagam, SNS College of Technology, Coimbatore, India
Assistant Prof. Madhavi Dhingra, Amity University, MP, India
Professor Kartheesan Log, Anna University, Chennai
Professor Vasudeva Acharya, Shri Madhwa vadiraja Institute of Technology, India
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia
Assistant Prof., Mahendra Singh Meena, Amity University Haryana
Assistant Professor Manjeet Kaur, Amity University Haryana
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India
Assistant Prof. Dharmendra Choudhary, Tripura University, India
Assistant Prof. Deepika Vodnala, SR Engineering College, India
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India
Assistant Prof. Chirag Modi, NIT Goa
Dr. R. Ramkumar, Nandha Arts And Science College, India
Dr. Priyadarshini Vydhialingam, Harathiar University, India
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka
Dr. Vikas Thada, AMITY University, Pachgaon
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore
Dr. Shaheera Rashwan, Informatics Research Institute
Dr. S. Preetha Gunasekar, Bharathiyar University, India
Asst Professor Sameer Dev Sharma, Uttaranchal University, Dehradun
Dr. Zhihan Iv, Chinese Academy of Science, China
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar
Dr. Umar Ruhi, University of Ottawa, Canada
Dr. Jasmin Cosic, University of Bihac, Bosnia and Herzegovina
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran
Dr. Ayyasamy Ayyanar, Annamalai University, India
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia
Dr. Murali Krishna Namana, GITAM University, India
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India

Dr. Sushil Chandra Dimri, Graphic Era University, India
Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam
Dr. S. Rama Sree, Aditya Engg. College, India
Dr. Ehab T. Alnfrawy, Sadat Academy, Egypt
Dr. Patrick D. Cerna, Haramaya University, Ethiopia
Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India
Associate Prof. Dr. Jiliang Zhang, North Eastern University, China
Dr. Sharefa Murad, Middle East University, Jordan
Dr. Ajeet Singh Poonia, Govt. College of Engineering & technology, Rajasthan, India
Dr. Vahid Esmaeaelzadeh, University of Science and Technology, Iran
Dr. Jacek M. Czerniak, Casimir the Great University in Bydgoszcz, Institute of Technology, Poland
Associate Prof. Anisur Rehman Nasir, Jamia Millia Islamia University
Assistant Prof. Imran Ahmad, COMSATS Institute of Information Technology, Pakistan
Professor Ghulam Qasim, Preston University, Islamabad, Pakistan
Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women
Dr. Wencan Luo, University of Pittsburgh, US
Dr. Musa PEKER, Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Gunasekaran Shanmugam, Anna University, India
Dr. Binh P. Nguyen, National University of Singapore, Singapore
Dr. Rajkumar Jain, Indian Institute of Technology Indore, India
Dr. Imtiaz Ali Halepoto, QUEST Nawabshah, Pakistan
Dr. Shaligram Prajapat, Devi Ahilya University Indore India
Dr. Sunita Singhal, Birla Institute of Technology and Science, Pilani, India
Dr. Ijaz Ali Shoukat, King Saud University, Saudi Arabia
Dr. Anuj Gupta, IKG Punjab Technical University, India
Dr. Sonali Saini, IES-IPS Academy, India
Dr. Krishan Kumar, MotiLal Nehru National Institute of Technology, Allahabad, India
Dr. Z. Faizal Khan, College of Engineering, Shaqra University, Kingdom of Saudi Arabia
Prof. M. Padmavathamma, S.V. University Tirupati, India
Prof. A. Velayudham, Cape Institute of Technology, India
Prof. Seifeidne Kadry, American University of the Middle East
Dr. J. Durga Prasad Rao, Pt. Ravishankar Shukla University, Raipur
Assistant Prof. Najam Hasan, Dhofar University
Dr. G. Suseendran, Vels University, Pallavaram, Chennai
Prof. Ankit Faldu, Gujarat Technological University- Atmiya Institute of Technology and Science
Dr. Ali Habiboghli, Islamic Azad University
Dr. Deepak Dembla, JECRC University, Jaipur, India
Dr. Pankaj Rajan, Walmart Labs, USA
Assistant Prof. Radoslava Kraveva, South-West University "Neofit Rilski", Bulgaria
Assistant Prof. Medhavi Shriwas, Shri vaishnav institute of Technology, India
Associate Prof. Sedat Akleylek, Ondokuz Mayıs University, Turkey
Dr. U.V. Arivazhagu, Kingston Engineering College Affiliated To Anna University, India
Dr. Touseef Ali, University of Engineering and Technology, Taxila, Pakistan
Assistant Prof. Naren Jeeva, SASTRA University, India
Dr. Riccardo Colella, University of Salento, Italy
Dr. Enache Maria Cristina, University of Galati, Romania
Dr. Senthil P, Kurinji College of Arts & Science, India

Dr. Hasan Ashrafi-rizi, Isfahan University of Medical Sciences, Isfahan, Iran
Dr. Mazhar Malik, Institute of Southern Punjab, Pakistan
Dr. Yajie Miao, Carnegie Mellon University, USA
Dr. Kamran Shaukat, University of the Punjab, Pakistan
Dr. Sasikaladevi N., SASTRA University, India
Dr. Ali Asghar Rahmani Hosseinabadi, Islamic Azad University Ayatollah Amoli Branch, Amol, Iran
Dr. Velin Krlev, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria
Dr. Marius Iulian Mihailescu, LUMINA - The University of South-East Europe
Dr. Sriramula Nagaprasad, S.R.R.Govt.Arts & Science College, Karimnagar, India
Prof (Dr.) Namrata Dhanda, Dr. APJ Abdul Kalam Technical University, Lucknow, India
Dr. Javed Ahmed Mahar, Shah Abdul Latif University, Khairpur Mir's, Pakistan
Dr. B. Narendra Kumar Rao, Sree Vidyanikethan Engineering College, India
Dr. Shahzad Anwar, University of Engineering & Technology Peshawar, Pakistan
Dr. Basit Shahzad, King Saud University, Riyadh - Saudi Arabia
Dr. Nilamadhab Mishra, Chang Gung University
Dr. Sachin Kumar, Indian Institute of Technology Roorkee
Dr. Santosh Nanda, Biju-Pattnaik University of Technology
Dr. Sherzod Turaev, International Islamic University Malaysia
Dr. Yilun Shang, Tongji University, Department of Mathematics, Shanghai, China
Dr. Nuzhat Shaikh, Modern Education society's College of Engineering, Pune, India
Dr. Parul Verma, Amity University, Lucknow campus, India
Dr. Rachid Alaoui, Agadir Ibn Zohr University, Agadir, Morocco
Dr. Dharmendra Patel, Charotar University of Science and Technology, India
Dr. Dong Zhang, University of Central Florida, USA
Dr. Kennedy Chinedu Okafor, Federal University of Technology Owerri, Nigeria
Prof. C Ram Kumar, Dr NGP Institute of Technology, India
Dr. Sandeep Gupta, GGS IP University, New Delhi, India
Dr. Shahanawaj Ahamad, University of Ha'il, Ha'il City, Ministry of Higher Education, Kingdom of Saudi Arabia
Dr. Najeed Ahmed Khan, NED University of Engineering & Technology, India
Dr. Sajid Ullah Khan, Universiti Malaysia Sarawak, Malaysia
Dr. Muhammad Asif, National Textile University Faisalabad, Pakistan
Dr. Yu BI, University of Central Florida, Orlando, FL, USA
Dr. Brijendra Kumar Joshi, Research Center, Military College of Telecommunication Engineering, India
Prof. Dr. Nak Eun Cho, Pukyong National University, Korea
Prof. Wasim Ul-Haq, Mathematics Department Faculty of Science, Majmaah University, Saudi Arabia

CALL FOR PAPERS

International Journal of Computer Science and Information Security

IJCSIS 2017-2018

ISSN: 1947-5500

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



© IJCSIS PUBLICATION 2017

ISSN 1947 5500

<http://sites.google.com/site/ijcsis/>