

Suresh Chandra Satapathy  
P. S. Avadhani  
Siba K. Udgata  
Sadasivuni Lakshminarayana *Editors*

# ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India - Volume I

Hosted by CSI Vishakapatnam Chapter

# **Advances in Intelligent Systems and Computing**

Volume 248

*Series Editor*

Janusz Kacprzyk, Warsaw, Poland

For further volumes:

<http://www.springer.com/series/11156>

Suresh Chandra Satapathy · P.S. Avadhani  
Siba K. Udgata · Sadasivuni Lakshminarayana  
Editors

# ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India - Volume I

Hosted by CSI Vishakapatnam Chapter

*Editors*

Suresh Chandra Satapathy  
Anil Neerukonda Institute of Technology  
and Sciences, Sangivalasa  
(Affiliated to Andhra University)  
Vishakapatnam  
Andhra Pradesh  
India

P.S. Avadhani  
College of Engineering (A)  
Andhra University  
Vishakapatnam  
India

Siba K. Udgata  
University of Hyderabad  
Hyderabad  
Andhra Pradesh  
India

Sadasivuni Lakshminarayana  
CSIR-National Institute of Oceanography  
Vishakapatnam  
India

ISSN 2194-5357

ISSN 2194-5365 (electronic)

ISBN 978-3-319-03106-4

ISBN 978-3-319-03107-1 (eBook)

DOI 10.1007/978-3-319-03107-1

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

This AISC volume contains the papers presented at the 48<sup>th</sup> Annual Convention of Computer Society of India (CSI 2013) with theme 'ICT and Critical Infrastructure' held during 13th –15th December 2013 at Hotel Novotel Varun Beach, Visakhapatnam and hosted by Computer Society of India, Vishakhapatnam Chapter in association with Vishakhapatnam Steel Plant, the flagship company of RINL, India.

Computer society of India (CSI) was established in 1965 with a view to increase information and technological awareness among Indian society, and to make forum to exchange and share the IT- related issues. The headquarters of the CSI is situated in Mumbai with a full-fledged office setup and is coordinating the individual chapter activities. It has 70 chapters and 418 students' branches operating in different cities of India. The total strength of CSI is above 90000 members.

CSI Vishakhapatnam Chapter deems it a big pride to host this prestigious 48<sup>th</sup> Annual Convention after successfully organizing various events like INDIA-2012, eCOG-2011, 28th National Student convention, and AP State Student Convention in the past.

CSI 2013 is targeted to bring researchers and practitioners from academia and industry to report, deliberate and review the latest progresses in the cutting-edge research pertaining to emerging technologies.

Research submissions in various advanced technology areas were received and after a rigorous peer-review process with the help of program committee members and external reviewer, 173 ( Vol-I: 88, Vol-II: 85) papers were accepted with an acceptance ratio of 0.43.

The conference featured many distinguished personalities like Dr. V.K. Saraswat, Former Director General, DRDO, Prof. Rajeev Sangal, Director, IIT-BHU, Mr. Ajit Balakrishnan, Founder & CEO Rediff.com, Prof. L.M. Patnaik, Former Vice Chancellor, IISc, Bangalore, Prof. Kesav Nori, IIIT-H & IIT-H, Mr. Rajesh Uppal, Executive Director & CIO, Maruti Suzuki-India, Prof. D. Krishna Sundar, IISc, Bangalore, Dr. Dejan Milojcic, Senior Researcher and Director of the. Open Cirrus Cloud Computing, HP Labs, USA & President Elect 2013, IEEE Computer Society, Dr. San Murugesan, Director, BRITE Professional Services, Sydney, Australia, Dr. Gautam Shroff, VP and Chief Scientist, Tata Consultancy Services, Mr. P. Krishna Sastry, TCS, Ms. Angela R. Burgess Executive Director, IEEE Computer Society, USA, Mr. Sriram Raghavan,

Security & Digital Forensics Consultant, Secure Cyber Space, and Dr. P. Bhanu Prasad, Vision Specialist, Matrix Vision GmbH, Germany among many others.

Four special sessions were offered respectively by Dr. Vipin Tyagi, Jaypee University of Engg. & Tech., Prof. J.K. Mandal, University of Kalyani, Dr. Dharam Singh, CTAE, Udaipur, Dr. Suma V., Dean, Research, Dayananda Sagar Institutions, Bengaluru. Separate Invited talks were organized in industrial and academia tracks in both days. The conference also hosted few tutorials and workshops for the benefit of participants.

We are indebted to Andhra University, JNTU-Kakinada and Visakhapatnam Steel plant for their immense support to make this convention possible in such a grand scale. CSI 2013 is proud to be hosted by Visakhapatnam Steel Plant (VSP), which is a Govt. of India Undertaking under the corporate entity of Rashtriya Ispat Nigam Ltd. It is the first shore-based integrated steel plant in India. The plant with a capacity of 3 mtpa was established in the early nineties and is a market leader in long steel products. The Plant is almost doubling its capacity to a level of 6.3 mtpa of liquid steel at a cost of around 2500 million USD. RINL-VSP is the first integrated steel plant in India to be accredited with all four international standards, viz. ISO 9001, ISO 14001, ISO 50001 and OHSAS 18001. It is also the first Steel Plant to be certified with CMMI level-3 certificate and BS EN 16001 standard.

Our special thanks to Fellows, President, Vice President, Secretary, Treasurer, Regional VPs and Chairmen of Different Divisions, Heads of SIG Groups, National Student Coordinator, Regional and State Student Coordinators, OBs of different Chapters and Administration staff of CSI-India. Thanks to all CSI Student Branch coordinators, Administration & Management of Engineering Colleges under Visakhapatnam chapter for their continuous support to our chapter activities. Sincere thanks to CSI-Vizag members and other Chapter Members across India those who have supported CSI-Vizag activities directly or indirectly.

We take this opportunity to thank authors of all submitted papers for their hard work, adherence to the deadlines and patience with the review process. We express our thanks to all reviewers from India and abroad who have taken enormous pain to review the papers on time.

Our sincere thanks to all the chairs who have guided and supported us from the beginning. Our sincere thanks to senior life members, life members, associate life members and student members of CSI-India for their cooperation and support for all activities.

Our sincere thanks to all Sponsors, press, print & electronic media for their excellent coverage of this convention.

December 2013

Dr. Suresh Chandra Satapathy  
 Dr. P.S. Avadhani  
 Dr. Siba K. Udgata  
 Dr. Sadasivuni Lakshminarayana

# Organization

## Chief Patrons

Shri A.P. Choudhary, CMD, RINL  
Prof. G.S.N. Raju, VC, Andhra University  
Prof. G. Tulasi Ram Das, VC, JNTU-Kakinada

## Patrons

Sri Umesh Chandra, Director (Operations), RINL  
Sri P. Madhusudan, Director (Finance), RINL  
Sri Y.R. Reddy, Director (Personnel), RINL  
Sri N.S. Rao, Director (Projects), RINL

## Apex (CSI) Committee

Prof. S.V. Raghavan, President  
Shri H.R. Mohan, Vice President  
Dr S. Ramanathan, Hon. Secretary  
Shri Ranga Rajagopal, Hon. Treasurer  
Sri Satish Babu, Immd. Past President  
Shri Raju L. Kanchibhotla, RVP, Region-V

## Chief Advisor

Prof D.B.V. Sarma, Fellow, CSI

## Advisory Committee

Sri Anil Srivastav, IAS, Jt. Director General of Civil Aviation, GOI  
Sri G.V.L. Satya Kumar, IRTS, Chairman I/C, VPT, Visakhapatnam

Sri N.K. Mishra, Rear Admiral IN (Retd), CMD, HSL, Visakhapatnam  
Capt.D.K. Mohanty, CMD, DCIL, Visakhapatnam  
Sri S.V. Ranga Rajan, Outstanding Scientist, NSTL, Visakhapatnam  
Sri Balaji Iyengar, GM I/C, NTPC-Simhadri, Visakhapatnam  
Prof P. Trimurthy, Past President, CSI  
Sri M.D. Agarwal, Past President, CSI  
Prof D.D. Sharma, Fellow, CSI  
Sri Saurabh Sonawala, Hindtron, Mumbai  
Sri R.V.S. Raju, President, RHI Clasil Ltd.  
Prof. Gollapudi S.R., GITAM University & Convener, Advisory Board, CSI-2013

### **International Technical Advisory Committee:**

Dr. L.M. Patnaik, IISc, India	Dr. D. Janikiram, IIT-M
Dr. C. Krishna Mohan, IIT-H, India	Dr. H.R. Vishwakarma, VIT, India
Dr. K. RajaSekhar Rao, Dean, KLU, India	Dr. Deepak Garg, TU, Patiala
Dr. Chaoyang Zhang, USM, USA	Dr. Hamid Arabnia, USA
Dr. Wei Ding, USA	Dr. Azah Kamilah Muda, Malaysia
Dr. Cheng-Chi Lee, Taiwan	Dr. Yun-Huoy Choo, Malaysia
Dr. Pramod Kumar Singh, ABV-IIITM, India	Dr. Hongbo Liu, Dalian Maritime
Dr. B. Biswal, GMRIT, India	Dr. B.N. Biswal, BEC, India
Dr. G. Pradhan, BBSR, India	Dr. Saihanuman, GRIET, India
Dr. B.K. Panigrahi, IITD, India	Dr. L. Perkin, USA
Dr. S. Yenduri, USA	Dr. V. Shenbagraman, SRM, India and many others

### **Organizing Committee**

#### **Chair**

Sri T.K. Chand, Director (Commercial), RINL & Chairman, CSI-Vizag

#### **Co-Chairmen**

Sri P.C. Mohapatra, ED (Projects), Vizag Steel

Sri P. Ramudu, ED (Auto & IT), Vizag Steel

#### **Vice-Chairman**

Sri K.V.S.S. Rajeswara Rao, GM (IT), Vizag Steel

#### **Addl. Vice-Chairman:**

Sri Suman Das, DGM (IT) & Secretary, CSI-Vizag

#### **Convener**

Sri Paramata Satyanarayana, Sr. Manager (IT), Vizag Steel



**Co-Convener**

Sri C.K. Padhi, AGM (IT), Vizag Steel

**Web Portal Convener**

Sri S.S. Choudhary, AGM(IT)

**Advisor (Press & Publicity)**

Sri B.S. Satyendra, AGM (CC)

**Convener (Press & Publicity)**

Sri Dwaram Swamy, AGM (Con)

**Co-Convener (Press & Publicity)**

Sri A.P. Sahu, AGM (IT)

**Program Committee****Chairman:**

Prof P.S. Avadhani, Vice Principal, AUCE (A)

**Co Chairmen**

Prof D.V.L.N. Somayajulu, NIT-W

Dr S. Lakshmi Narayana, Scientist E, NIO-Vizag

**Conveners:**

Sri Pulle Chandra Sekhar, DGM (IT), Vizag Steel – Industry

Prof. S.C. Satapathy, HOD (CSE), ANITS – Academics

Dr. S.K. Udgata, UoH, Hyderabad

**Finance Committee****Chairman**

Sri G.N. Murthy, ED (F&A), Vizag Steel

**Co-Chairmen**

Sri G.J. Rao, GM (Marketing)-Project Sales

Sri Y. Sudhakar Rao, GM (Marketing)-Retail Sales

**Convener**

Sri J.V. Rao, AGM (Con)

**Co-Conveners**

Sri P. Srinivasulu, DGM (RMD)

Sri D.V.G.A.R.G. Varma, AGM (Con)

X Organization

Sri T.N. Sanyasi Rao,  
Sr.Mgr (IT)

**Members** Sri V.R. Sanyasi, Rao  
Sri P. Sesha Srinivas

## **Convention Committee**

**Chair** Sri D.N. Rao, ED(Operations), Vizag Steel

### **Vice-Chairs**

Sri Y. Arjun Kumar  
Sri S. Raja  
Dr. B. Govardhana Reddy

Sri ThyagaRaju Guturu  
Sri Bulusu Gopi Kumar  
Sri Narla Anand

### **Conveners**

Sri S.K. Mishra  
Sri V.D. Awasthi  
Sri M. Srinivasa Babu  
Sri G.V. Ramesh  
Sri A. Bapuji  
Sri B. Ranganath

Sri D. Satyanarayana  
Sri P.M. Divecha  
Sri Y.N. Reddy  
Sri J.P. Dash  
Sri K. Muralikrishna

### **Co-Conveners**

Sri Y. Madhusudan Rao  
Sri S. Gopal  
Sri Phani Gopal  
Sri M.K. Chakravarty  
Sri P. Krishna Rao  
Sri P. Balaramu

Sri B.V. Vijay Kumar  
Mrs M. Madhu Bindu  
Sri P. Janardhana  
Sri G.V. Saradhi

### **Members:**

Sri Y. Satyanarayana  
Sri Shailendra Kumar  
Sri V.H. Sundara Rao  
Mrs K. Sarala  
Sri V. Srinivas  
Sri G. Vijay Kumar  
Mrs. V.V. Vijaya Lakshmi

Sri D. Ramesh  
Shri K. Pratap  
Sri P. Srinivasa Rao  
Sri S. Adinarayana  
Sri B. Ganesh  
Sri Hanumantha Naik  
Sri D.G.V. Saya

Sri V.L.P. Lal  
Sri U.V.V. Janardhana  
Sri S. Arun Kumar  
Sri K. Raviram  
Sri N. Prabhakar Ram  
Sri BH.B.V.K. Raju  
Sri K. Srinivasa Rao  
Mrs T. Kalavathi

Mrs A. Lakshmi  
Sri N. Pradeep  
Sri K. Dilip  
Sri K.S.S. Chandra Rao  
Sri Vamshee Ram  
Ms Sriya Basumallik  
Sri Kunche Satyanarayana  
Sri Shrirama Murthy

# Contents

## Session 1: Computational Intelligence and Its Applications

<b>Homogeneity Separateness: A New Validity Measure for Clustering Problems</b> .....	1
<i>M. Ramakrishna Murty, J.V.R. Murthy, P.V.G.D. Prasad Reddy, Anima Naik, Suresh C. Satapathy</i>	
<b>Authenticating Grid Using Graph Isomorphism Based Zero Knowledge Proof</b> .....	11
<i>Worku B. Gebeyehu, Lubak M. Ambaw, M.A. Eswar Reddy, P.S. Avadhani</i>	
<b>Recognition of Marathi Isolated Spoken Words Using Interpolation and DTW Techniques</b> .....	21
<i>Ganesh B. Janvale, Vishal Waghmare, Vijay Kale, Ajit Ghodke</i>	
<b>An Automatic Process to Convert Documents into Abstracts by Using Natural Language Processing Techniques</b> .....	31
<i>Ch. Jayaraju, Zareena Noor Basha, E. Madhavarao, M. Kalyani</i>	
<b>A Novel Method for CBIR Using Texture Spectrum in Wavelet Domain</b> . . . .	41
<i>G. Rosline Nesa Kumari, M. Sudheer, R. Tamilkodi</i>	
<b>Identification of Abdominal Aorta Aneurysm Using Ant Colony Optimization Algorithm</b> .....	49
<i>A. Dinesh Kumar, R. Nidhya, V. Hanah Ayisha, Vigneshwar Manokar, Nandhini Vigneshwar Manokar</i>	
<b>Color Image Watermarking Using Wavelet Transform Based on HVS Channel</b> .....	59
<i>G. Rosline Nesa Kumari, Syama Sundar Jeeru, S. Maruthuperumal</i>	
<b>Automatic Vehicle Number Plate Localization Using Symmetric Wavelets</b> .....	69
<i>V. HimaDeepthi, B. BalvinderSingh, V. SrinivasaRao</i>	

<b>Response Time Comparison in Multi Protocol Label Switching Network Using Ant Colony Optimization Algorithm</b> .....	77
<i>E.R. Naganathan, S. Rajagopalan, S. Narayanan</i>	
<b>Software Effort Estimation Using Data Mining Techniques</b> .....	85
<i>Tirimula Rao Benala, Rajib Mall, P. Srikavya, M. Vani HariPriya</i>	
<b>Infrared and Visible Image Fusion Using Entropy and Neuro-Fuzzy Concepts</b> .....	93
<i>S. Rajkumar, P.V.S.S.R. Chandra Mouli</i>	
<b>Hybrid Non-dominated Sorting Simulated Annealing Algorithm for Flexible Job Shop Scheduling Problems</b> .....	101
<i>N. Shivasankaran, P. Senthilkumar, K. Venkatesh Raja</i>	
<b>DICOM Image Retrieval Using Geometric Moments and Fuzzy Connectedness Image Segmentation Algorithm</b> .....	109
<i>Amol Bhagat, Mohammad Atique</i>	
<b>Soft Computing Based Partial-Retuning of Decentralised PI Controller of Nonlinear Multivariable Process</b> .....	117
<i>L. Sivakumar, R. Kotteeswaran</i>	
<b>Multi Objective Particle Swarm Optimization for Software Cost Estimation</b> .....	125
<i>G. Sivanageswara Rao, Ch.V. Phani Krishna, K. Rajasekhara Rao</i>	
<b>An Axiomatic Fuzzy Set Theory Based Feature Selection Methodology for Handwritten Numeral Recognition</b> .....	133
<i>Abhinaba Roy, Nibaran Das, Ram Sarkar, Subhadip Basu, Mahantapas Kundu, Mita Nasipuri</i>	
<b>Brownian Distribution Guided Bacterial Foraging Algorithm for Controller Design Problem</b> .....	141
<i>N. Sri Madhava Raja, V. Rajinikanth</i>	
<b>Trapezoidal Fuzzy Shortest Path (TFSP) Selection for Green Routing and Scheduling Problems</b> .....	149
<i>P.K. Srimani, G. Vakula Rani, Suja Bennet</i>	
<b>A Fused Feature Extraction Approach to OCR: MLP vs. RBF</b> .....	159
<i>Amit Choudhary, Rahul Rishi</i>	
<b>Fusion of Dual-Tree Complex Wavelets and Local Binary Patterns for Iris Recognition</b> .....	167
<i>N.L. Manasa, A. Govardhan, Ch. Satyanarayana</i>	

<b>Evaluation of Feature Selection Method for Classification of Data Using Support Vector Machine Algorithm</b> .....	179
<i>A. Veeraswamy, S. Appavu Alias Balamurugan, E. Kannan</i>	
<b>Electrocardiogram Beat Classification Using Support Vector Machine and Extreme Learning Machine</b> .....	187
<i>C.V. Banupriya, S. Karpagavalli</i>	
<b>Genetic Algorithm Based Approaches to Install Different Types of Facilities</b> .....	195
<i>Soumen Atta, Priya Ranjan Sinha Mahapatra</i>	
<b>A Trilevel Programming Approach to Solve Reactive Power Dispatch Problem Using Genetic Algorithm Based Fuzzy Goal Programming</b> .....	205
<i>Papun Biswas, Bijay Baran Pal</i>	
<b>An Iterative Fuzzy Goal Programming Method to Solve Fuzzified Multiobjective Fractional Programming Problems</b> .....	219
<i>Mousumi Kumar, Shyamal Sen, Bijay Baran Pal</i>	
<b>An Efficient Ant Colony Based Routing Algorithm for Better Quality of Services in MANET</b> .....	233
<i>Abdur Rahaman Sardar, Moutushi Singh, Rashi Ranjan Sahoo, Koushik Majumder, Jamuna Kanta Sing, Subir Kumar Sarkar</i>	
<b>An Enhanced MapReduce Framework for Solving Protein Folding Problem Using a Parallel Genetic Algorithm</b> .....	241
<i>A.G. Hari Narayanan, U. Krishnakumar, M.V. Judy</i>	
<b>A Wavelet Transform Based Image Authentication Approach Using Genetic Algorithm (AWTIAGA)</b> .....	251
<i>Amrita Khamrui, J.K. Mandal</i>	
<b>Session 2: Mobile Communications and Social Networking</b>	
<b>A New Approach to Monitor Network</b> .....	259
<i>Hemant Kumar Saini, Anurag Jagetiya, Kailash Kumar, Satpal Singh Kushwaha</i>	
<b>Detection and Mitigation of Misbehaving Vehicles from VANET</b> .....	267
<i>Megha Kadam, Suresh Limkar</i>	
<b>Comparative Study of Hierarchical Based Routing Protocols: Wireless Sensor Networks</b> .....	277
<i>Pritee Parwekar</i>	
<b>Privacy Based Optimal Routing in Wireless Mesh Networks</b> .....	287
<i>T.M. Navamani, P. Yogesh</i>	

<b>Security Issues and Its Counter Measures in Mobile Ad Hoc Networks . . . . .</b>	<b>301</b>
<i>Pritee Parwekar, Sparsh Arora</i>	
<b>Finding a Trusted and Shortest Path Mechanism of Routing Protocol for Mobile Ad Hoc Network . . . . .</b>	<b>311</b>
<i>Rabindra Kumar Shial, K. Hemant Kumar Reddy, Bhabani Sankar Gouda</i>	
<b>Fusion Centric Decision Making for Node Level Congestion in Wireless Sensor Networks . . . . .</b>	<b>321</b>
<i>N. Prabakaran, K. Naresh, R. Jagadeesh Kannan</i>	
<b>An Intelligent Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks . . . . .</b>	<b>331</b>
<i>K.S.S. Anupama, S. Sri Gowri, B. Prabakara Rao, T. Satya Murali</i>	
<b>Energy Efficient and Reliable Transmission of Data in Wireless Sensor Networks . . . . .</b>	<b>341</b>
<i>Chilukuri Shanti, Anirudha Sahoo</i>	
<b>Inter-actor Connectivity Restoration in Wireless Sensor Actor Networks : An Overview . . . . .</b>	<b>351</b>
<i>Sasmita Acharya, C.R. Tripathy</i>	
<b>MANFIS Approach for Path Planning and Obstacle Avoidance for Mobile Robot Navigation . . . . .</b>	<b>361</b>
<i>Prases Kumar Mohanty, Krishna K. Pandey, Dayal R. Parhi</i>	
<b>The Effect of Velocity and Unresponsive Traffic Volume on Performance of Routing Protocols in MANET . . . . .</b>	<b>371</b>
<i>Sukant Kishoro Bisoy, Prasant Kumar Patnaik, Tanmaya Kumar Swain</i>	
<b>Improving the QOS in MANET by Enhancing the Routing Technique of AOMDV Protocol . . . . .</b>	<b>381</b>
<i>Ankita Sharma, Sumit Vashistha</i>	
<b>PRMACA: A Promoter Region Identification Using Multiple Attractor Cellular Automata (MACA) . . . . .</b>	<b>393</b>
<i>Pokkuluri Kiran Sree, Inampudi Ramesh Babu, S.S.S.N. Usha Devi Nedunuri</i>	
<b>Analyzing Statistical Effect of Sampling on Network Traffic Dataset . . . . .</b>	<b>401</b>
<i>Raman Singh, Harish Kumar, R.K. Singla</i>	
<b>Localization of Information Dissemination in Agriculture Using Mobile Networks . . . . .</b>	<b>409</b>
<i>Lokesh Jain, Harish Kumar, R.K. Singla</i>	
<b>Video Traffic in Wireless Sensor Networks . . . . .</b>	<b>417</b>
<i>Shilpa Pandey, Dharm Singh, Naveen Choudhary, Neha Mehta</i>	

<b>A Novel Pairwise Key Establishment and Management in Hierarchical Wireless Sensor Networks (HWSN) Using Matrix</b> .....	425
<i>B. Premamayudu, K. Venkata Rao, P. Suresh Varma</i>	
<b>An Agent-Based Negotiating System with Multiple Trust Parameter Evaluation across Networks</b> .....	433
<i>Mohammed Tajuddin, C. Nandini</i>	
<b>Enabling Self-organizing Behavior in MANETs: An Experimental Study</b> ...	441
<i>Annapurna P. Patil, K. Rajanikant, Sabarish, Madan, Surabi</i>	
<b>Secure Hybrid Routing for MANET Resilient to Internal and External Attacks</b> .....	449
<i>Niroj Kumar Pani, Sarojananda Mishra</i>	
<b>Compact UWB/BLUETOOTH Integrated Uniplanar Antenna with WLAN Notch Property</b> .....	459
<i>Yashwant Kumar Soni, Navneet Kumar Agrawal</i>	
<b>Session 3: Grid Computing, Cloud Computing, Virtual and Scalable Applications</b>	
<b>Open Security System for Cloud Architecture</b> .....	467
<i>S. Koushik, Annapurna P. Patil</i>	
<b>Client-Side Encryption in Cloud Storage Using Lagrange Interpolation and Pairing Based Cryptography</b> .....	473
<i>R. Siva Ranjani, D. Lalitha Bhaskari, P.S. Avadhani</i>	
<b>Analysis of Multilevel Framework for Cloud Security</b> .....	481
<i>Vadlamani Nagalakshmi, Vijeyta Devi</i>	
<b>Agent Based Negotiation Using Cloud – An Approach in E-Commerce</b> .....	489
<i>Amruta More, Sheetal Vij, Debajyoti Mukhopadhyay</i>	
<b>Improve Security with RSA and Cloud Oracle 10g</b> .....	497
<i>Vadlamani Nagalakshmi, Vijeyta Devi</i>	
<b>Negotiation Life Cycle: An Approach in E-Negotiation with Prediction</b> .....	505
<i>Mohammad Irfan Bala, Sheetal Vij, Debajyoti Mukhopadhyay</i>	
<b>Dynamic Scheduling of Requests Based on Impacting Parameters in Cloud Based Architectures</b> .....	513
<i>R. Arokia Paul Rajan, F. Sagayaraj Francis</i>	
<b>A Generic Agent Based Cloud Computing Architecture for E-Learning</b> .....	523
<i>Samitha R. Babu, Krutika G. Kulkarni, K. Chandra Sekaran</i>	



<b>Cache Based Cloud Architecture for Optimization of Resource Allocation and Data Distribution</b> .....	535
<i>Salim Raza Qureshi</i>	
<b>Implementing a Publish-Subscribe Distributed Notification System on Hadoop</b> .....	543
<i>Jyotiska Nath Khasnabish, Ananda Prakash Verma, Shrisha Rao</i>	
<b>Securing Public Data Storage in Cloud Environment</b> .....	555
<i>D. Boopathy, M. Sundaresan</i>	
<b>An Enhanced Strategy to Minimize the Energy Utilization in Cloud Environment to Accelerate the Performance</b> .....	563
<i>M. Vaidehi, V. Suma, T.R. Gopalakrishnan Nair</i>	
<b>An Efficient Approach to Enhance Data Security in Cloud Using Recursive Blowfish Algorithm</b> .....	575
<i>Naziya Balkish, A.M. Prasad, V. Suma</i>	
<b>Effective Disaster Management to Enhance the Cloud Stability</b> .....	583
<i>O. Mahitha, V. Suma</i>	
<b>An Efficient Job Classification Technique to Enhance Scheduling in Cloud to Accelerate the Performance</b> .....	593
<i>M. Vaidehi, T.R. Gopalakrishnan Nair, V. Suma</i>	
<b>A Study on Cloud Computing Testing Tools</b> .....	605
<i>M.S. Narasimha Murthy, V. Suma</i>	
<b>Session 4: Project Management and Quality Systems</b>	
<b>Co-operative Junction Control System</b> .....	613
<i>Sanjana Chandrashekar, V.S. Adarsh, P. Shashikanth Rangan, G.B. Akshatha</i>	
<b>Identifying Best Products Based on Multiple Criteria Using Decision Making System</b> .....	621
<i>Swetha Reddy Donala, M. Archana, P.V.S. Srinivas</i>	
<b>Design and Performance Analysis of File Replication Strategy on Distributed File System Using GridSim</b> .....	629
<i>Nirmal Singh, Sarbjeet Singh</i>	
<b>Extended Goal Programming Approach with Interval Data Uncertainty for Resource Allocation in Farm Planning: A Case Study</b> .....	639
<i>Bijay Baran Pal, Mousumi Kumar</i>	
<b>Design and Performance Analysis of Distributed Implementation of Apriori Algorithm in Grid Environment</b> .....	653
<i>Priyanka Arora, Sarbjeet Singh</i>	

<b>Criticality Analyzer and Tester – An Effective Approach for Critical Components Identification and Verification</b> . . . . .	663
<i>Jeya Mala Dharmalingam, Balamurugan, Sabari Nathan</i>	
<b>Real Time Collision Detection and Fleet Management System</b> . . . . .	671
<i>Anusha Pai, Vishal Vernekar, Gaurav Kudchadkar, Shubharaj Arsekar, Keval Tanna, Ross Rebello, Madhav Desai</i>	
<b>An Effective Method for the Identification of Potential Failure Modes of a System by Integrating FTA and FMEA</b> . . . . .	679
<i>Samitha Khaiyum, Y.S. Kumaraswamy</i>	
<b>Impact of Resources on Success of Software Project</b> . . . . .	687
<i>N.R. Shashi Kumar, T.R. Gopalakrishnan Nair, V. Suma</i>	
<b>Session 5: Emerging Technologies in Hardware and Software</b>	
<b>An Optimized Approach for Density Based Spatial Clustering Application with Noise</b> . . . . .	695
<i>Rakshit Arya, Geeta Sikka</i>	
<b>A Kernel Space Solution for the Detection of Android Bootkit</b> . . . . .	703
<i>Harsha Rao, S. Selvakumar</i>	
<b>Optimizing CPU Scheduling for Real Time Applications Using Mean-Difference Round Robin (MDRR) Algorithm</b> . . . . .	713
<i>R.N.D.S.S. Kiran, Polinati Vinod Babu, B.B. Murali Krishna</i>	
<b>A Study on Vectorization Methods for Multicore SIMD Architecture Provided by Compilers</b> . . . . .	723
<i>Davendar Kumar Ojha, Geeta Sikka</i>	
<b>A Clustering Analysis for Heart Failure Alert System Using RFID and GPS</b> . . . . .	729
<i>Gudikandhula Narasimha Rao, P. Jagdeeswar Rao</i>	
<b>Globally Asynchronous Locally Synchronous Design Based Heterogeneous Multi-core System</b> . . . . .	739
<i>Rashmi A. Jain, Dinesh V. Padole</i>	
<b>Recognition of Smart Transportation through Randomization and Prioritization</b> . . . . .	749
<i>Bhavya Boggavarapu, Pabita Allada, SaiManasa Mamillapalli, V.P. Krishna Anne</i>	
<b>Prioritized Traffic Management and Transport Security Using RFID</b> . . . . .	757
<i>V. Kishore, L. Britto Anthony, P. Jesu Jayarin</i>	

<b>A Novel Multi Secret Sharing Scheme Based on Bitplane Flips and Boolean Operations</b> .....	765
<i>Gyan Singh Yadav, Aparajita Ojha</i>	
<b>Large-Scale Propagation Analysis and Coverage Area Evaluation of 3G WCDMA in Urban Environments Based on BS Antenna Heights</b> .....	773
<i>Ramarakula Madhu, Gottapu Sasi Bhushana Rao</i>	
<b>Inclination and Pressure Based Authentication for Touch Devices</b> .....	781
<i>K. Rajasekhara Rao, V.P. Krishna Anne, U. Sai Chand, V. Alakananda, K. Navya Rachana</i>	
<b>Anticipated Velocity Based Guidance Strategy for Wheeled Mobile Evaders Amidst Moving Obstacles in Bounded Environment</b> .....	789
<i>Amit Kumar, Aparajita Ojha</i>	
<b>Reconfigurable Multichannel Down Convertor for on Chip Network in MRI</b> .....	799
<i>Vivek Jain, Navneet Kumar Agrawal</i>	
<b>Author Index</b> .....	807

# Homogeneity Separateness: A New Validity Measure for Clustering Problems

M. Ramakrishna Murty<sup>1</sup>, J.V.R. Murthy<sup>2</sup>, P.V.G.D. Prasad Reddy<sup>3</sup>,  
Anima Naik<sup>4</sup>, and Suresh C. Satapathy<sup>5</sup>

<sup>1</sup> Dept. of CSE, GMR Institute of Technology, Rajam, Srikakulam (Dist) A.P., India  
ramakrishna.malla@gmail.com

<sup>2</sup> Dept. of CSE, JNTUK-Kakinada, A.P., India  
mjonnalagedda@gmail.com

<sup>3</sup> Dept. of CS&SE, Andhra University, Visakhapatnam, A.P., India  
prasadreddy.vizag@gmail.com

<sup>4</sup> Majhighariani Institute of Technology and Sciences, Rayagada, India  
animanaik@gmail.com

<sup>5</sup> Dept. of CSE, ANITS, Visakhapatna, A.P., India  
sureshsatapathy@gmail.com

**Abstract.** Several validity indices have been designed to evaluate solutions obtained by clustering algorithms. Traditional indices are generally designed to evaluate center-based clustering, where clusters are assumed to be of globular shapes with defined centers or representatives. Therefore they are not suitable to evaluate clusters of arbitrary shapes, sizes and densities, where clusters have no defined centers or representatives. In this work, HS (Homogeneity Separateness) validity measure based on a different shape is proposed. It is suitable for clusters of any shapes, sizes and/or of different densities. The main concepts of the proposed measure are explained and experimental results on both synthetic and real life data set that support the proposed measure are given.

**Keywords:** Cluster validity index, homogeneity, separateness, spanning tree, CS Measure.

## 1 Introduction

The purpose of cluster analysis has been playing an important role in solving many problems in medicine, psychology, biology, sociology, pattern recognition and image processing. Clustering algorithms are used to assess the interaction among patterns by organizing patterns into clusters such that patterns within a cluster are more similar to each other than are patterns belonging to different clusters.

Cluster validity indexes correspond to the statistical–mathematical functions used to evaluate the results of a clustering algorithm on a quantitative basis. Generally, a cluster validity index serves two purposes. First, it can be used to determine the number of clusters, and second, it finds out the corresponding best partition. Validity

indices are used to evaluate a clustering solution according to specified measures that depend on measuring the proximity between patterns. The basic assumption in building a validity index is that patterns should be more similar to patterns in their cluster compared to other patterns outside the cluster. This concept has led to the foundation of homogeneity and separateness measures, where homogeneity refers to the similarity between patterns of the same cluster, and separateness refers to the dissimilarity between patterns of different clusters. However, each of homogeneity and separateness measures could have different forms depending on the clustering assumptions; some clustering methods assume that patterns of the same cluster will group around a centroid or a representative, others will drop this assumption to the more general one that there are no specific centroids or representatives, and that patterns connect together to form a cluster. In the more popular methods of clustering as K-Means, average and complete linkage clustering [1], the clustering preserves the common assumption about a globular shape of a cluster, where in this case a cluster representative can be easily defined. Since validity indices were built for the most common used algorithms, the resulting indices defined homogeneity and separateness in the presence of centroids, examples are Dunn's[15], Davies Bouldin indices[16]. However, moving from the traditional assumption of the globular shaped clusters, into the more general problem of having undefined geometrical cluster shapes, those algorithms are able to find clusters of arbitrary shapes, and where it is difficult to use cluster representatives as the case in globular-shaped clusters. In order to develop a validity measure for the more general problem of undefined cluster shapes other considerations for measuring homogeneity and separateness should follow. In this work, a validity measure that considers the general problem of arbitrary shapes, sizes and arbitrary densities clusters are proposed. It is based on minimizing the minimum spanning tree (MST) distances of the cluster, as a homogeneity measure, as the distance between two clusters have to be maximized, we have consider the average set distance between pairs of subsets in the data as a separateness measure.

The rest of the paper is organized as follows: we have discussed popular CS validity measures and introduce propose HS measure in II, Detailed simulation and results are presented in section 3. We conclude with a summary of the contributions of this paper in section 4.

## 2 Cluster Validity Measures

A comparative examination of thirty validity measures is presented in [2] and an overview of the various measures can be found in [3]. In [4] the performance of the CS measure is compared with five popular measures on several data sets. It has been shown that, CS measure is more effective than those measures for dealing with cluster of different densities and/or sizes. Since it is not feasible to attempt a comprehensive comparison of our proposed validity measure with each of these thirty measures we have compared only with CS measure. In this paper, the proposed **HS Measure** evaluates clustering results when the variance of the densities and/or sizes of clusters may be large and/or small and/or of different shapes. The performance of the

**proposed measure** is compared with popular CS measures on several both synthetic and real data sets.

Before going to discuss about the properties of each validity measures, the common characteristics have been defined in the following manner. Let  $P = \{P_1, P_2, P_3, \dots, P_N\}$  be a set of  $N$  patterns or data points, each having  $D$  features. These patterns can also be represented by a data matrix  $X_{N \times D}$  with  $N$   $D$ -dimensional row vectors. The  $i$ th row vector  $\vec{X}_i$  characterizes the  $i$ th object from the set  $P$ , and each element  $X_{i,j}$  in  $\vec{X}_i$  corresponds to the  $j$ th real-value feature ( $j = 1, 2, \dots, D$ ) of the  $i$ th partition ( $i = 1, 2, \dots, N$ ). Given such an  $X_{N \times D}$  matrix, a partition clustering algorithm tries to find a partition  $C = \{C_1, C_2, \dots, C_K\}$  of  $K$  classes. In the following, we use  $\vec{m}_i$  to denote cluster center of cluster  $C_i$ . Let  $u_{ij}$  ( $i = 1, 2, \dots, K; j = 1, 2, \dots, N$ ) be the membership of data point  $j$  in cluster  $i$ . The  $K \times N$  matrix  $U = [u_{ij}]$  is called a membership matrix. The membership matrix  $U$  is allowed to have elements with values between 0 and 1. The computation of  $u_{ij}$  depends on which clustering algorithm is adopted to cluster the data set rather than which validity measure is used to validate the clustering results. In addition, only fuzzy clustering algorithms (such as the FCM algorithm[5] and the Gustafson–Kessel (GK) algorithm [6]) will result in the membership matrix  $U = [u_{ij}]$ . Crisp clustering algorithms (e.g. the K-means algorithm), do not produce the membership matrix  $U = [u_{ij}]$ . However, if one insists on having the information of the membership matrix  $U = [u_{ij}]$  then the value of  $u_{ij}$  is assigned to be 1 if the  $j$ th data point is partitioned to the  $i$ th cluster; otherwise it is 0. For example, the FCM algorithm or the GK algorithm uses the following equation to compute  $u_{ij}$ :

$$u_{ij} = \frac{1}{\sum_{r=1}^K \left( \frac{d_{ij}}{d_{rj}} \right)^{\frac{2}{m-1}}}$$

Where  $d_{ij}$  represents the distance between the data point  $\vec{X}_j$  to the cluster center  $\vec{m}_i$  and  $m$  is the fuzzifier parameter.

### 1) CS Measure

The CS Measure is defined as

$$\begin{aligned} CS(K) &= \frac{\frac{1}{K} \sum_{i=1}^K \left[ \frac{1}{N_i} \sum_{\vec{X}_l \in C_i} \max_{\vec{X}_q \in C_i} \{d(\vec{X}_l, \vec{X}_q)\} \right]}{\frac{1}{K} \sum_{i=1}^K \left[ \min_{j \in K, j \neq i} \{d(\vec{m}_i, \vec{m}_j)\} \right]} \\ &= \frac{\sum_{i=1}^K \left[ \frac{1}{N_i} \sum_{\vec{X}_l \in C_i} \max_{\vec{X}_q \in C_i} \{d(\vec{X}_l, \vec{X}_q)\} \right]}{\sum_{i=1}^K \left[ \min_{j \in K, j \neq i} \{d(\vec{m}_i, \vec{m}_j)\} \right]} \\ \vec{m}_i &= \frac{1}{N_i} \sum_{\vec{X}_j \in C_i} \vec{X}_j, \quad i=1, 2, \dots, K \end{aligned}$$

Where distance metric between any two data points  $\vec{X}_i$  and  $\vec{X}_j$  is denoted by  $d(\vec{X}_i, \vec{X}_j)$ .

Where  $\vec{m}_i$  is the cluster center of  $C_i$ , and  $C_i$  is the set whose elements are the data points assigned to the  $i$ th cluster, and  $N_i$  is the number of elements in  $C_i$ ,  $d$  denotes a distance function. This measure is a function of the ratio of the sum of within-cluster scatter to between-cluster separation.

## 2) The Proposed Validity Measure

The proposed measure is based on human's visual cognition of clusters, where clusters are distinguished if they are separated by distances larger than the internal distances of each cluster. Here the view of a cluster, as a group of patterns surrounding a center, should be altered; algorithms that considered arbitrary shapes. The measure, proposed here, is formulated based on:

- a. Distance homogeneity between distances of the patterns of cluster
- b. Density separateness between different clusters.

### Distance Homogeneity

To be able to discover clusters of different density distributions, the distances between patterns of the cluster are used to reflect its density distribution. In this case relatively small distances reflect relatively high densities and vice versa. Thus, if internal distances of the cluster are homogenous, then the density distribution is consistent overall the cluster. This homogeneity will also guarantee that there is no merging of a different density region to a cluster, and thus the possibility of merging two natural clusters (naturally separated by low dense regions) is lowered.

To preserve distance homogeneity, the differences between the cluster's internal distances should be minimized. This can be done by minimizing the MST (Minimum spanning tree) of a cluster. The MST generalizes the method of selecting the internal distances for any clustering solution, and considers, only, the minimal set of the smallest distances that make up the cluster. How so, for that follow below some lines.

Given a connected, undirected graph  $G = (V, E)$ , where  $V$  is the set of nodes,  $E$  is the set of edges between pairs of nodes, and a weight  $w(u, v)$  specifying weight of the edge  $(u, v)$  for each edge  $(u, v) \in E$ . A spanning tree is an acyclic subgraph of a graph  $G$ , which contain all vertices from  $G$ . The Minimum Spanning Tree (**MST**) of a weighted graph is minimum weight spanning tree of that graph. Several well established **MST** algorithms exist to solve minimum spanning tree problem [9], [10], [11]. The cost of constructing a minimum spanning tree is  $O(m \log n)$ , where  $m$  is the number of edges in the graph and  $n$  is the number of vertices. A Euclidean minimum spanning tree (**EMST**) is a spanning tree of a set of  $n$  points in a metric space  $(E_n)$ , where the length of an edge is the Euclidean distance between a pair of points in the point set.

Clustering algorithms using minimal spanning tree takes the advantage of **MST**. The **MST** ignores many possible connections between the data patterns, so the cost of clustering can be decreased. The **MST** based clustering algorithm is known to be capable of detecting clusters with various shapes and size [12]. Unlike traditional clustering algorithms, the **MST** clustering algorithm does not assume a spherical shapes structure of the underlying data. The **EMST** clustering algorithm [13], [12] uses the Euclidean minimum spanning tree of a graph to produce the structure of point clusters in the  $n$  dimensional Euclidean space. Clusters are detected to achieve some measures of optimality, such as minimum intra-cluster distance or maximum inter-cluster distance [14]. The **EMST** algorithm has been widely used in practice.

All existing clustering Algorithm require a number of parameters as their inputs and these parameters can significantly affect the cluster quality. In this paper we want to avoid experimental methods and advocate the idea of need-specific as opposed to care-specific because users always know the needs of their applications. We believe it is a good idea to allow users to define their desired similarity within a cluster and allow them to have some flexibility to adjust the similarity if the adjustment is needed. Our Algorithm produces clusters of  $n$  –dimensional points with a given cluster number and a naturally approximated intra-cluster distance.

### Density Separateness

As the distance between two clusters has to be maximized, we have considered **the average set distance between pairs of subsets in the data** as separation. Similarity of two clusters is based on the two most similar (closest) points in the different clusters.

Hence validity measure is defined as

$$HS(K) = \frac{\frac{1}{K} \sum_{i=1}^K \{MST_i\}}{\frac{1}{K} \sum_{i=1}^K \left\{ \min_{j \in K, j \neq i} \left\{ \min_{\bar{X}_i \in C_i, \bar{X}_j \in C_j} \{d(\bar{X}_i, \bar{X}_j)\} \right\} \right\}}$$

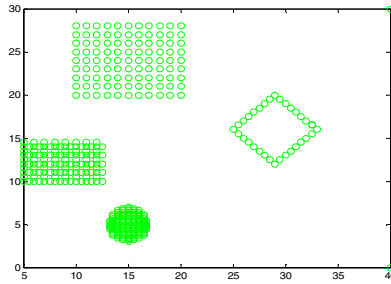
Where  $i = 1, 2, \dots, K$ ,  $C_i$ 's are clusters and  $d$  denotes a distance function,  $MST_i$  is the weight of minimum spanning tree of cluster  $C_i$ .

## 3 Experimental Results

To illustrate the effectiveness of the proposed validity measure we have tested on some data sets. For the comparison purpose, these data sets are also tested by the validity measures- the CS Measure (CS), the HS measure. For both the validity measures, the smallest value indicates a valid optimal partition. Either the FCM algorithm or the Gustafson-Kessel (GK) algorithm is applied to cluster these data sets at each cluster number  $K$  for  $K = 2$  to  $K = 10$ . The fuzzifier parameter,  $m$ , in the FCM algorithm and the GK algorithm is set to be 2. The Euclidean distance is adopted as the distance measure for the FCM algorithm. Since these two clustering algorithms are sensitive to the initialization, during the clustering procedures we have tried different initializations to cluster the data sets for each cluster number  $K$ . Then for each  $K$ , the clustering result that occurred with the highest frequency was chosen to be the clustering result for the cluster number  $K$  to be validated by the validity measures. Then all validity measures are computed from the same clustering results.

**Example 1:** The artificial data set1 shown in Fig.1 is used to illustrate that the HS measure can work well for the case of clusters with different shapes, sizes and densities. Total number of data point is 385. The performance of each validity measure is given in Table 1. In Table 1 and others to follow, the highlighted (**bold and shaded**) entries correspond to optimal values of the measures. Note that both the HS, CS validity measures find the optimal cluster number  $K$  at  $K=4$ .



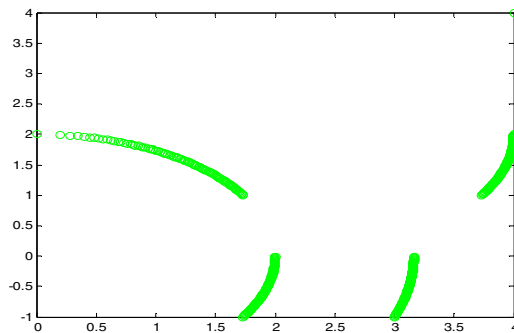


**Fig. 1.** Artificial dataset 1,  $k=4$

**Table 1.** Numerical values of the validity measures for example 1

	2	3	4	5	6	7	8	9	10
HS measure(min)	22.0023	47.4480	<b>11.06094</b>	13.4767	16.9958	16.3503	15.7359	24.2340	17.9038
CS measure(min)	3.4155	3.7116	<b>0.9453</b>	1.1238	1.2283	1.7655	1.9979	4.4811	2.1437

**Example 2:** The artificial dataset 2 shown in Fig.2 is used to illustrate that the HS measure can work well for the case of clusters with different shapes and sizes with same densities. Here total number of data points is 505. The performance of each validity measure is given in Table 2. In Table 2 and others to follow, the highlighted (**bold and shaded**) entries correspond to optimal values of the measures. Note that only the HS validity measures find the optimal cluster number  $K$  at  $K=4$ , whereas CS validity measures find the optimal cluster number  $K$  at  $K=3$ . Once again, this example shows that the proposed HS measure can work well for the data set with different shapes and sizes with same densities.

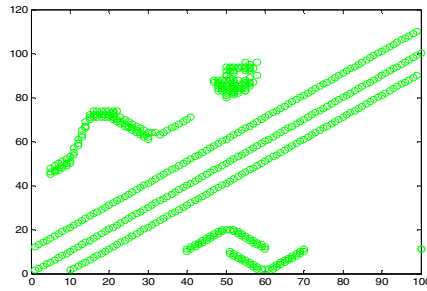


**Fig. 2.** Artificial dataset 2,  $K=4$

**Table 2.** Numerical values of the validity measures for example 2

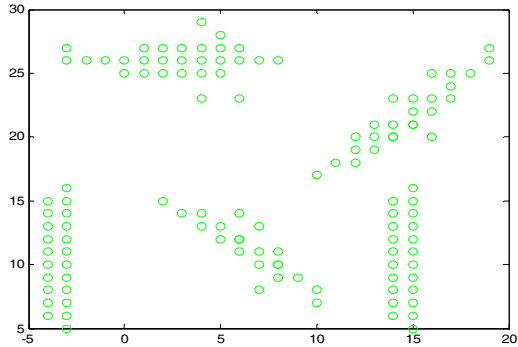
	2	3	4	5	6	7	8	9	10
HS measure (min)	3.9947	1.9536	<b>1.1762</b>	1.5217	2.2800	2.1507	4.1299	48.6929	41.9349
CS measure (min)	3.3024	<b>0.6350</b>	0.8247	1.1071	1.3757	1.8379	1.9563	2.3810	2.4130

**Example 3:** The artificial data set3 shown in Fig. 3. Here we have generated a mixture of lines, spherical, curve clusters with different densities. The total number of data points is 542. The GK algorithm is applied to cluster these data sets at each cluster number  $K=2$  to  $K=10$ . The performance of each validity measure is given in Table 3. As seen from table, both the HS and CS validity measures find the optimal cluster number  $K$  at  $K=7$ . This example demonstrates that the proposed HS measure can also work well for this case of clusters with different geometrical structures.

**Fig. 3.** Artificial dataset3,  $K=7$ **Table 3.** Numerical values of the validity measures for example 3

	2	3	4	5	6	7	8	9	10
HS measure (min)	380.4226	176.1192	158.5003	107.7843	88.2324	<b>30.6983</b>	32.7666	69.0382	62.7433
CS measure (min)	1.6396	1.5907	1.3201	1.4221	1.0540	<b>1.0090</b>	1.0453	1.1182	1.1671

**Example 4:** The artificial data set 4 shown in Fig. 4. The data set illustrate that HS measure can work well for the case of clusters with nearly similar shape and size with same densities. The size of the dataset is 121. The performance of each validity measure is given in Table 4 and others to follow, the highlighted (**bold and shaded**) entries correspond to optimal values of the measures. Note that only the HS validity measures find the optimal cluster number  $K$  at  $K=5$ .

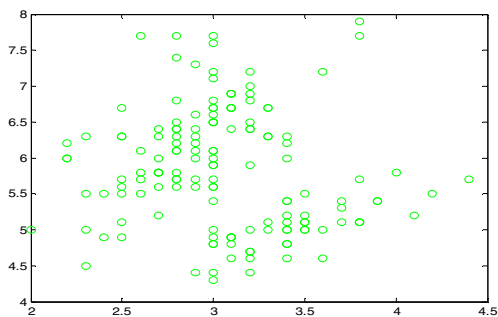


**Fig. 4.** Artificial dataset 4,  $K=5$

**Table 4.** Numerical values of the validity measures for example 4

	2	3	4	5	6	7	8	9	10
HS measure (min)	69.5208	15.3391	13.5281	<b>8.5638</b>	12.3394	10.8566	15.0347	13.2593	11.6142
CS measure (min)	4.1669	1.1655	<b>1.0223</b>	1.0698	1.4352	1.2124	1.2931	1.3414	1.6383

**Example 5:** Here we use the well-known iris data set to test the measures. The iris data set has three subsets (i.e. iris setosa, iris versicolor and iris virginica). There are a total of 150 data points in the data set and each class had 50 patterns. The FCM algorithm is applied to cluster the iris data set at each cluster number  $c$  for  $K=2$  to  $K=10$ . Table 5 shows the performance of each validity measure. Note that the HS and CS validity measures find the optimal cluster number  $K$  at  $K=3$ . This example demonstrates that the proposed CS measure can also be used on the high-dimensional real life data.

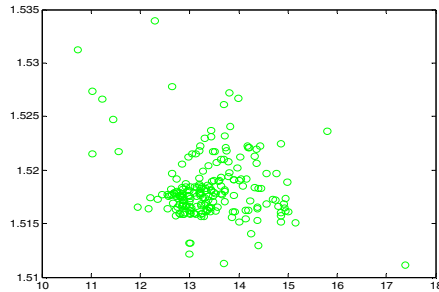


**Fig. 5.** Iris dataset,  $K=3$

**Table 5.** Numerical values of the validity measures for example 5

	2	3	4	5	6	7	8	9	10
HS measure (min)	60.5260	<b>19.2111</b>	57.7374	50.0703	30.4849	24.6778	21.0893	20.4061	19.3700
CS measure (min)	0.7029	<b>0.6683</b>	0.7589	0.7375	1.5693	1.7164	1.0297	1.1593	1.8513

**Example 6:** The have used another well known real life data set shown in Fig. 6 which is **glass** data. There are total of 214 data points in the data set. Glass data were sampled from six different types of glass: 1) building windows float processed (70 objects); 2) building windows non float processed (76 objects); 3) vehicle windows float processed (17 objects); 4) containers (13 objects); 5) tableware (9 objects); and 6) headlamps (29 objects). Each type has nine features: 1) refractive index; 2) sodium; 3) magnesium; 4) aluminum; 5) silicon; 6) potassium; 7) calcium; 8) barium; and 9) iron. The FCM algorithm is applied to cluster the glass data at each cluster number  $K=2$  to  $K=8$ . The performance of each validity measure is given in Table 6. In Table 6 and others to follow, the highlighted (**bold and shaded**) entries correspond to optimal values of the measures. Note that only the HS validity measures find the optimal cluster number  $K$  at  $K=6$ .

**Fig. 6.** Glass dataset,  $K=6$ **Table 6.** Numerical values of the validity measures for example 6

	2	3	4	5	6	7	8
HS measure(min)	58.5431	36.1847	39.5706	25.9539	<b>18.6575</b>	22.7788	19.1702
CS measure(min)	4.6484	2.5973	3.0886	1.9000	2.0532	2.1504	<b>1.1996</b>

It is shown that in every case **HS measure** shows the proper number of cluster.

## 4 Conclusion

In this paper, we first propose the HS measure to evaluate clustering results. The performance of the proposed HS measure is compared with popular CS measure on several data sets of both synthesis and real data sets. When HS index was compared with CS index, it always obtained better results. We hope that the procedure discussed here stimulates further investigation into development of better procedures for problems of clustering for data sets.

## References

1. Hartigan, J.A.: Clustering Algorithms. Wiley Series in Probability and Mathematical statistics (1975)
2. Milligan, G.W., Cooper, M.C.: An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50, 159–179 (1985)
3. Dubes, R., Jain, A.K.: Validity studies in clustering methodologies. *Pattern Recognition* 11, 235–253 (1979)
4. Chou, C.-H., Su, M.-C., Lai, E.: A new cluster validity measure and its application to image compression. *Pattern Anal. Applic.* 7, 205–220 (2004), doi:10.1007/s10044-004-0218-1
5. Gath, I., Geva, A.B.: Unsupervised optimal fuzzy clustering. *IEEE Trans. Pattern Anal. Machine Intell.* 11, 773–781 (1989)
6. Gustafson, D.E., Kessel, W.C.: Fuzzy clustering with a fuzzy covariance matrix. In: *Proceedings of the IEEE Conference on Decision and Control*, San Diego, pp. 761–766 (January 1979)
7. Halkidi, M., Vazirgiannis, M.: Clustering validity assessment: Finding the optimal partitioning of a data set. In: *Proc. IEEE ICDM*, San Jose, CA, pp. 187–194 (2001)
8. Babuška, R., van der Veen, P.J., Kaymak, U.: Improved Covariance Estimation for Gustafson-Kessel Clustering, 0-7803-7280-8/02/\$10.00 ©2002. IEEE (2002)
9. Prim, R.: Shortest connection networks and some generalization. *Bell Systems Technical Journal* 36, 1389–1401 (1957)
10. Kruskal, J.: On the shortest spanning subtree and the travelling salesman problem. In: *Proceedings of the American Mathematical Society*, pp. 48–50 (1956)
11. Nesetril, J., Milkova, E., Nesetrilova, H.: Otakar boruvka on minimum spanning tree problem: Translation of both the 1926 papers, comments, history. *DMATH: Discrete Mathematics* 233 (2001)
12. Zahn, C.: Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers* C-20, 68–86 (1971)
13. Preparata, F., Shamos, M.: *Computational Geometry: An Introduction*. Springer, Newyork (1985)
14. Asano, T., Bhattacharya, B., Keil, M., Yao, F.: Clustering Algorithms based on minimum and maximum spanning trees. In: *Proceedings of the 4th Annual Symposium on Computational Geometry*, pp. 252–257 (1988)
15. Dunn, J.C.: A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well-Separated Clusters. *Journal Cybern.* 3(3), 32–57 (1973)
16. Davies, D.L., Bouldin, D.W.: A cluster separation measure. *IEEE Trans. Pattern Analysis and Machine Intelligence* 1(4), 224–227 (1979)

# Authenticating Grid Using Graph Isomorphism Based Zero Knowledge Proof

Worku B. Gebeyehu\*, Lubak M. Ambaw\*, M.A. Eswar Reddy, and P.S. Avadhani

Dept. of Computer Science & Systems Engineering, Andhra University, India  
{workubrhn, eswarreddy143}@gmail.com,  
{rosert2007, psavadhani}@yahoo.com

**Abstract.** A zero-Knowledge proof is a method by which the prover can prove to the verifier that the given statement is valid, without revealing any additional information apart from the veracity of the statement. Zero knowledge protocols have numerous applications in the domain of cryptography; they are commonly used in identification schemes by forcing adversaries to behave according to a predetermined protocol. In this paper, we propose an approach using graph isomorphism based zero knowledge proof to construct an efficient grid authentication mechanism. We demonstrate that using graph isomorphism based zero knowledge proof provide a much higher level of security when compared to other authentication schemes. The betterment of security arises from the fact that the proposed method hides the secret during the entire authentication process. Moreover, it enables one identity to be used for various accounts.

**Keywords:** Zero Knowledge proof, zero knowledge protocol, graph isomorphism, grid authentication, grid security, graph, prover, verifier.

## 1 Introduction

The intention of authentication and authorization is to deploy the policies which organizations have devised to administer the utilization of computing resources in the grid environment. According to Foster in grid technology is described as a “resource-sharing technology with software and services that let people access computing power, databases, and other tools securely online across corporate, institutional, and geographic boundaries without sacrificing local autonomy”[1]. For example a scientist in research institute might need to use regional, national, or international resources within grid-based projects in addition to using the major network of the campus. Each grid-project mainly requires its own authentication mechanisms, commonly in the form of GSI based or kerberos based certificates so as to authenticate the scientist to access and utilize grid based resources.

---

\* Corresponding authors.

These days, computer scientists are exerting lots of efforts to develop different kinds of authentication mechanisms that provide strong Grid security. The currently existing grid authentication mechanisms are usually bound with only one, in most cases based on public key infrastructure (PKI). Such system unnecessarily limits users since they are required to use only the one mechanism, which may not be flexible or convenient.

Moreover, regardless of particular type of certificates or PKI, one of the key drawbacks of the PKI is that current tools and producers for certificate management are too complicated for users. This leads either to rejection of the PKI or to insecure private-key management, which dis-empowers all the Grid infrastructure [2]. This paper proposes a novel methodology to overcome the aforementioned limitations by implementing ZKP based grid authentication.

## 2 Basic Concepts

### 2.1 What Is Zero-Knowledge Proof?

A zero-knowledge proof (ZKP) is a proof of some statement which reveals nothing other than the veracity of the statement. Zero-Knowledge proof is a much popular concept used in many cryptography systems. In this concept, two parties are involved, the prover A and the verifier B. Using this technique, it allows prover A to show that he has a credential, without having to give B the exact number.

The reason for the use of a Zero-Knowledge Proof in this situation for an authentication system is because it has the following properties:

- Completeness: If an honest verifier will always be convinced of a true statement by an honest prover
- Soundness: If a cheating prover can convince an honest verifier that some false statement is actually true with only a small probability.
- Zero-knowledge: if the statement is true, the verifier will not know anything other than that the statement is true.

Information about the details of the statement will not be revealed. [3]. A common application for zero-knowledge protocols is in identification schemes, due to Feige, Fiat, and Shamir[4].

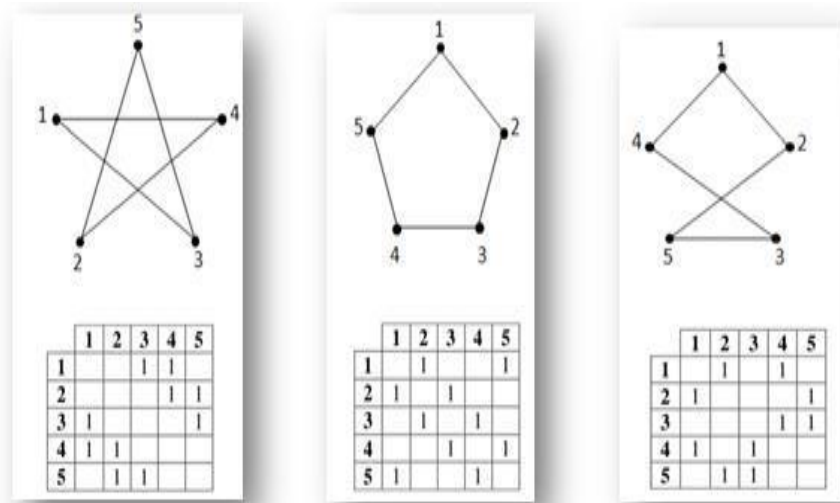
### 2.2 Graph

A graph is a set of nodes or vertices  $V$ , connected by a set of edges  $E$ . The sets of vertices and edges are finite. A graph with  $n$  vertices will have:  $V = \{1, 2, 3, \dots, n\}$  and  $E$  a 2-element subsets of  $V$ . Let  $u$  and  $v$  be two vertices of a graph. If  $(u,v) \in E$ , then  $u$  and  $v$  are said to be adjacent or neighbors.

A graph is represented by its adjacency matrix. For instance, a graph with  $n$  vertices, is represented by a  $n \times n$  matrix  $M=[m_{i,j}]$ , where the entry is “1” if there is an edge linking the vertex  $i$  to the vertex  $j$ , and is “0” otherwise. For undirected graphs, the adjacency matrix is symmetric around the diagonal.

### 2.3 Graph Isomorphism

Two graphs  $G_1$  and  $G_2$  are said to be isomorphic, if a one-to-one permutation or mapping exists between the set of vertices of  $G_1$  and the set of vertices of  $G_2$ , with the property that if two nodes of  $G_1$  are adjacent, so are their images in  $G_2$ . The graph isomorphism problem is therefore the problem of determining whether two given graphs are isomorphic. In other words, it is the problem of determining if two graphs with different structures are the same. Figure 1 gives an example of isomorphic graphs, with their corresponding adjacency matrices. Notice that the entries of the matrices, where  $m_{ij} = 0$  are left blank.



**Fig. 1.** Example of Isomorphic Graphs with their corresponding adjacency matrices

For example, Graph (b) is found, by relabeling the vertices of Graph (a) according to the following permutation: (3, 5, 2, 4, 1). This means that Node 1 in Graph (a) becomes Node 3 in Graph (b), Node 5 becomes Node 1 and so on. Following in Fig.2 is an illustration of how the permutation is applied to Graph (a).



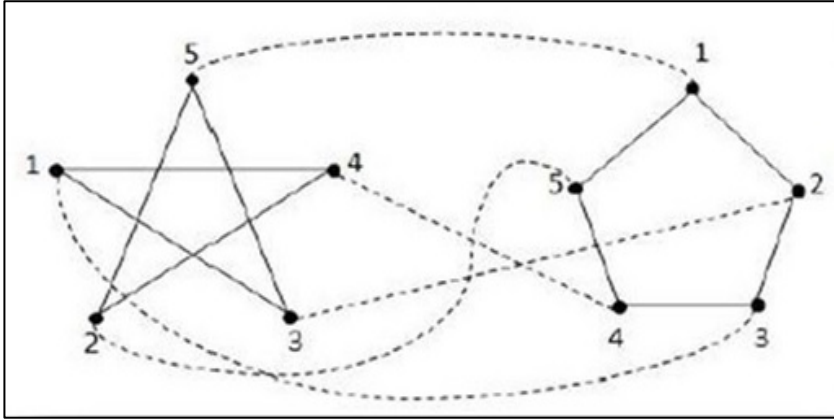


Fig. 2. Relabeling the vertices of Graph (a) using the permutation (3,5,2,4,1)

### 3 Proposed Method

#### 3.1 Graph Isomorphism Based Zero-Knowledge Proofs

The purpose of Zero-Knowledge Proof (ZKP) protocols is to enable the prover to prove an assertion to the verifier that he holds some secret knowledge, without leaking any information about the knowledge during the verification process (zero-knowledge). To demonstrate this concept, we chose to implement ZKP protocol based on Graph Isomorphism (GI). In this protocol the input to the prover and the verifier is a pair of graphs  $G_0$ ,  $G_1$ , and the goal of the prover is to convince the verifier that the graphs are isomorphic, but without revealing any information. If  $G_0$  and  $G_1$  are isomorphic, then the prover also has a permutation  $\pi$  such that  $\pi(G_0) = G_1$  (i.e.,  $\pi$  is an isomorphism).

Suppose we have two graphs  $G_1$  and  $G_2$  where  $G_2$  is generated from  $G_1$  using a secret permutation named  $\pi$ .  $G_2$  is obtained by relabeling the vertices of  $G_1$  according to secret permutation  $\pi$  by preserving the edges. The pair of graphs  $G_1$  and  $G_2$  forms the public key pair, and the permutation  $\pi$  serves as the private key. A third graph  $Q$  is either obtained from  $G_1$  or  $G_2$  using another random permutation  $\rho$ . Once the graph  $Q$  is found, the prover (represents the new node seeking entrance to the grid) sends it to the verifier who will challenge him to provide the permutation  $\sigma$  which can map  $Q$  to either  $G_1$  or  $G_2$ .

For example, if  $Q$  is found from  $G_1$  and the verifier puts a challenge to the prover to map  $Q$  to  $G_1$ , then  $\sigma = \rho^{-1}$ . In the same fashion, if  $Q$  is obtained from  $G_2$  and the verifier challenges the prover to map  $Q$  to  $G_2$ , then  $\sigma = \rho^{-1}$ . Otherwise, if  $Q$  is obtained from  $G_1$  and the verifier puts a challenge to the prover to provide the permutation that maps  $Q$  to  $G_2$ , then  $\sigma = \rho^{-1} \circ \pi$ , which is a combination of  $\rho^{-1}$  and  $\pi$ . Indeed,  $\rho^{-1}$  will be applied to  $Q$  to obtain  $G_1$  then the vertices of  $G_1$  will be modified according to the secret permutation  $\pi$  to get  $G_2$ . Eventually, if  $Q$  is obtained from  $G_2$  and the verifier challenges the prover to map  $Q$  to  $G_1$ , then  $\sigma = \rho^{-1} \circ \pi^{-1}$ .

One can observe that in the first two cases, the secret permutation  $\pi$  is not even used. Thus, a verifier could only be certain of a node's identity after many interactions. Moreover, we can also observe that during the whole interaction process, no clue was given about the secret itself this makes it strong grid authentication mechanism.

### 3.2 Pseudo Code

Given  $G1$  and  $G2$  such that  $G2 = \pi(G1)$ , the interactions constituting iterations of the graph isomorphism based ZKP protocol are demonstrated below:

- Step 1:** Prover randomly selects  $x \in \{1,2\}$
- Step 2:** Prover selects a random permutation  $\rho$ , and generates  $Q = \rho(Gx)$
- Step 3:** Prover communicates the adjacency matrix of  $Q$  to the verifier
- Step 4:** Verifier communicates  $y \in \{1,2\}$  to prover and challenges for  $\sigma$  that maps  $Q$  to  $Gy$
- Step 5:** If  $x=y$  the prover sends  $\sigma = \rho^{-1}$  to the verifier
- Step 6:** If  $x=1$  and  $y=2$  the prover sends  $\sigma = \rho^{-1} \circ \pi$  to the verifier
- Step 7:** If  $x=2$  and  $y=1$  the prover sends  $\sigma = \rho^{-1} \circ \pi^{-1}$  to the verifier
- Step 8:** Verifier checks if  $\sigma(Q) = Gy$  and access is granted to the prover accordingly

A number of iterations of these interactions are needed for the verifier to be totally convinced of the prover's identity, as the prover can be lucky and guess the value of  $y$  before sending  $Q$ .

### 3.3 Prototype

The researchers have implemented a prototype version of graph isomorphism based ZKP protocol for authenticating a grid. Graph isomorphism has been chosen because of its ease of implementation. The JAVA-implementation of the ZKP protocol in authentication of the grid is presented below. The advantage of this implementation is that it allows us to verify the adjacency matrices at each level of the simulation and to evaluate the correctness of the algorithm. The JAVA implementations are mainly based on the pseudo-code provided in Pseudo code section above. So as to demonstrate the logic used in the program, we considered a simple graph with 4(four) nodes. In reality, however, the graphs need to be very large, consisting of several hundreds of nodes (in our case the number of users requesting for grid resources), and present a GI complexity in the order of NP-complete.

There are three functions that are used in the implementation, namely, "Generate\_Isomorphic", "Permutuate", and "Compare\_Graphs" also one main function.

#### a) "Generate\_Isomorphic" function

Generate\_Isomorphic function enables us to apply a random permutation ( $\pi$  or  $\rho$ ) to a specific graph and to get another graph that is isomorphic to the first graph. For example, the declaration  $G2 = \text{Generate\_Isomorphic}(G1, \pi, n)$  is used to apply  $\pi$  to the

graph G1 so as to get the graph G2; where the variable „n” represents the number of nodes in the graph( the number of users requesting for grid resources).

b) “Permutuate” function

This function is used after receiving the challenge response from the prover. Indeed, once the verifier receives  $\sigma$  from the prover, his objective is to apply that permutation to the graph Q to check if he will get the expected response G1 or G2. This function is declared as follows, “Permutuate (Q,sigma,n)” where  $\sigma$  is being applied to a graph Q.

c) “Compare\_Graphs” function

Once the verifier receives the response  $\sigma$  and applies it to Q to get another graph that we will refer it as R for explanation purpose, the “Compare\_Graphs” function is used to check whether or not  $R = G1$  or  $R = G2$  depending on the case.

The function is declared as “Compare\_Graphs(G2,R,n).” Here, for instance, the function is used to compare both graphs G2 and R.

### 4 Simulations and Results

In this section we will depict how the implemented algorithm works through simulation. At first, the simulator requests for the value of “n” which represents the size of the graphs (In practise the number of grid users requesting for resource at one point of time). Once the size is set(the value of „n” is known), the user has the choice to either launch the interaction or to stop the simulation. The user is then prompted to provide the values of the graph G1. After all the values are entered, the user is asked for the value of the secret  $\pi$  and then for the value of the random permutation  $\rho$ . The user is then requested to choose the graph from which to create the graph Q and the graph that the verifier could ask the prover to generate when challenged. After these graphs are specified, the user is eventually asked for the value of  $\sigma$ . Once the value of  $\sigma$  entered, the simulator verifies all the inputs and declares if the prover is correct or incorrect. In other words, the simulator verifies if the prover knows the secret or not. In the real environment it is similar to proving the veracity of the grid user before allowing access to the grid. At last, the user is requested to repeat the process if he/she is willing. In order to test the correctness of the algorithm we simulate the code with some input examples and validate the results. In fact, the simulations were ran with the values of n,  $\pi$ ,  $\rho$  and the graph G1 illustrated below.

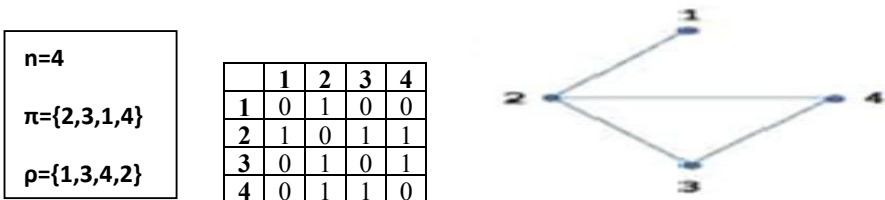


Fig. 3. Graph G1 and its Adjacency Matrix representation

If  $Q$  is obtained from  $G1$  ( $Q = G1 * \rho$ ) and the verifier challenges for  $G1$ , we will have the equation:

$$\sigma = \rho^{-1} \quad (1)$$

From equation 1 we have  $\sigma = \rho^{-1} = \{1, 4, 2, 3\}$  and to get the inverse of  $\rho = \{1, 3, 4, 2\}$ , we assign indices running from 1 to 4 to its vertices, and then process as follows. The first index "1" is at position 1, so its position remains the same. Index 2 is at the position 4; so the second vertex of  $\rho^{-1}$  is 4. Index 3 is at the second position; therefore the next vertex in  $\rho^{-1}$  is 2. Finally, index 4 is at the third position; therefore, the last element of  $\rho^{-1}$  is 3. Following this logic, the vertices of  $\rho^{-1}$  are obtained. If  $Q$  is obtained from  $G2$  ( $Q = G2 * \rho$ ) and the verifier challenges for  $G2$ , we will have  $\sigma = \rho = \{1, 4, 2, 3\}$ . If  $Q$  is obtained from  $G1$  ( $Q = G1 * \rho$ ) and the verifier challenges for  $G2$ , we will get the equation

$$\sigma = \rho^{-1} \circ \pi \quad (2)$$

From equation 2 we have  $\sigma = \rho^{-1} \circ \pi = \{1, 4, 2, 3\} \circ \{2, 3, 1, 4\} = \{2, 4, 3, 1\}$ . To equate and get the value of  $\sigma$ , we consider the vertices of  $\rho^{-1}$  as indices in  $\pi$  and observe which values they are associated with. The whole process consists of the following steps namely the first vertex of  $\pi$  is 2, the fourth vertex of  $\pi$  is 4. and the second vertex of  $\pi$  is 3 and the third vertex of  $\pi$  is 1. If  $Q$  is obtained from  $G2$  ( $Q = G2 * \rho$ ) and the verifier challenges for  $G1$ , we will have the equation:

$$\sigma = \rho^{-1} \circ \pi^{-1} \quad (3)$$

From equation 3 we have  $\sigma = \rho^{-1} \circ \pi^{-1} = \{1, 4, 2, 3\} \circ \{3, 1, 2, 4\} = \{3, 4, 1, 2\}$ . Now that all the parameters are set, we can run the actual simulation. One can note that the permutations are applied to the rows as well as the columns of the adjacency matrix in the above process. Simulation results will be depicted in the following section so as to demonstrate the three different cases (i.e.  $\sigma = \rho^{-1}$ ,  $\sigma = \rho^{-1} \circ \pi$ , and  $\sigma = \rho^{-1} \circ \pi^{-1}$ ).

```

The value of n is: 4
Start the interaction?(1:Yes,0:No):1
The value of the original graph G1 is: G1=[0 1 0 0;1 0 1 1;0 1 0 1;0 1 1 0]
The value of the secret pi is :[2 3 1 4]
The value of rho is :[1 3 4 2]
Choose the graph from which to create the graph H to give the verifier[0:G1;1:G2]:0
The verifier randomly chooses a graph for the prover to generate [1:G1;2:G2]:0
The value of sigma is :[ 1 4 2 3]
The prover has proved himself!

Does the verifier want to repeat the process? (1:Yes;0:No): 0
    
```

**Fig. 4.** Demonstration of the case in which  $Q = G1 * \rho$  and when the verifier challenges the prover to map  $Q$  to  $G1$

After the first simulations is over we obtained  $G_2 = G_1 * \pi$  and  $Q = G_1 * \rho$  as depicted in figure 3 and figure 4 below respectively.

	1	2	3	4
1	0	0	1	1
2	0	0	1	0
3	1	1	0	1
4	1	0	1	0

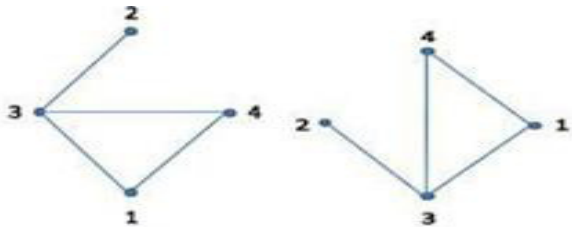


Fig. 5. Graph  $G_2$  in two distinct forms and its adjacency matrix

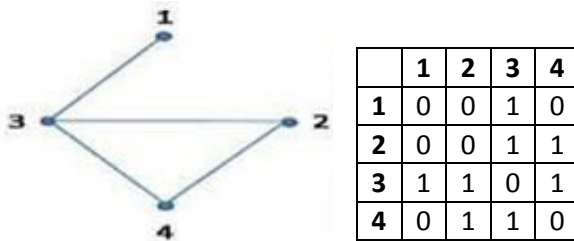


Fig. 6. Demonstration of graph  $Q = G_1 * \rho$  and its adjacency matrix

At last, let us see the situation where the prover enters incorrect responses to the challenges. The following figure depicts this case.

```

The value of n is:4
Start the interaction?(1: Yes, 0: No):1
The value of the original graph G1 is: G1=[0 1 0 0;1 0 1 1;0 1 0 1;0 1 1 0]
The value of the secret pi is:[2 3 1 4]
The value of rho is:[1 3 4 2]
Choose the graph from which to create the graph Q to give the verifier (0:G1,1:G2):0
The verifier randomly chooses a graph for the prover to generate(1:G1,2:G2):1
The value of sigma is:[1 2 4 3]
The prover is proved wrong! He does not know the secret!

Does the verifier want to repeat the process?(1:Yes,0:No):1
Choose the graph from which to create the graph Q to give the verifier (0:G1,1:G2):0
The verifier randomly chooses a graph for the prover to generate(1:G1,2:G2):0
The value of sigma is:[2 1 3 4]
The prover is proved wrong! He does not know the secret!

Does the verifier want to repeat the process?(1:Yes,0:No):1
Choose the graph from which to create the graph Q to give the verifier (0:G1,1:G2):1
The verifier randomly chooses a graph for the prover to generate(1:G1,2:G2):0
The value of sigma is:[1 4 3 2]
The prover is proved wrong! He does not know the secret!

Does the verifier want to repeat the process?(1:Yes,0:No):|
    
```

Fig. 7. Demonstration of the case where a user provides wrong answer for the challenge

## 5 Conclusion and Future Work

A new approach for authenticating grid using ZKP protocol based on Graph Isomorphism (GI) is proposed. Our simulation results show that the proposed method provides a much higher level of security for authenticating users in the grid by hiding the secret during the entire authentication process. Besides, it enables one identity to be used for various accounts. The implementation could be modified further so that the system is interactive and user friendly. Moreover, one can integrate the implementation of graph isomorphism based zero knowledge proof within a grid portal.

## References

- [1] Foster, I.: *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. Wiley, Argonne Illinois (2002)
- [2] Daniel, K., Ludek, M., Michal, P.: *Survey of Authentication Mechanisms for Grids*. In: CESNET Conference, pp. 200–204. Czech Academy of Science Press, Prague (2008)
- [3] Lum, J.: *Implementing Zero Knowledge Authentications with Zero Knowledge*. In: *The Python Papers Monograph Proceedings of PyCon Asia-Pacific*, Melbourne (2010)
- [4] Lior, M.: *A Study of Perfect Zero-Knowledge Proofs*, PhD Dissertation, University of Victoria, British Columbia, Victoria (2008)

# Recognition of Marathi Isolated Spoken Words Using Interpolation and DTW Techniques

Ganesh B. Janvale<sup>1</sup>, Vishal Waghmare<sup>2</sup>, Vijay Kale<sup>3</sup>, and Ajit Ghodke<sup>4</sup>

<sup>1</sup> Symbiosis International University, Symbiosis Centre for Information Technology,  
Pune-411 057(MS), India

<sup>2,3</sup> Dept. of CS&IT, Dr. Babasaheb Ambedkar Marathwada University,  
Aurangabad-411004, India

<sup>4</sup> Sinhgad Institute of Business Administration & Computer Application (SIBACA),  
Lonavala, India

ganesh@scit.edu, {vishal.pri12,vijaykale1685}@gmail.com,  
ajit\_1974\_in@yahoo.com

**Abstract.** This paper contains a Marathi speech database and isolated Marathi spoken words recognition system based on Mel-frequency cepstral coefficient (MFCC), optimal alignment using interpolation and dynamic time warping. Initially, Marathi speech database was designed and developed through Computerized Speech Laboratory. The database contained Marathi isolated words spoken by the 50 speakers including males and females. Mel-frequency Cepstral Coefficients were extracted and used for the recognition purpose. The 100% recognition rate for the isolated words have been achieved for both interpolation and dynamic time warping techniques.

**Keywords:** Speech Data base, CSL, MFCC, Speech Recognition and statistical method formatting.

## 1 Introduction

Currently, a lot of research is going on speech recognition and synthesis. The speech recognition understands basically what some speak to computer, asking a computer to translate speech into corresponding textual message, where as in speech synthesis, a computer generate artificial spoken dialogs. The speech is one of the natural forms of communication among the humans. The Marathi is an Indo-Aryan language, spoken in western and central India. There are 90 million of fluent speakers all over world [1][2]. The amount of work in Indian regional languages has not yet reached to a critical level to be used it as real communication tool, as already done in other languages in developed countries. Thus, this work was taken to focus on Marathi language. It is important to see that whether Speech Recognition System for Marathi can be carried out similar pathways of research as carried out in English [3]. Present work consists of the Marathi speech database and speech recognition system. The first part describes technical details of the database and second words recognition system. This paper is also split into four parts i.e. Introduction, Database, Words recognition System and Result and Conclusion.

## 2 Spoken Marathi Isolated Words Database

### 2.1 Features of Prosody in Marathi Words

Standard Marathi is based on dialects used by academic and printed media. There are 42 dialects of Marathi some of these are Ahirani, Khandeshi, Varhadi, Zadiboli, Vadvali, Samavedi and Are Marathi. The phonetics inventory of Marathi (Devnagar) along with International Phonetic Alphabets (IPA) is shown in Figure 1.a and b [4].

Devnagary	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ	औ	अं	अः
Transliterated	A	Āa	i	ī	u	ū	ṛ	e	Ai	o	au	aṁ	aḥ
IPA	/ə/	/a/	/i/		/u/		/ru/	/e/	/əi/	/o/	/əu/	/əṁ/	/əḥ/

Fig. 1a. Vowels in Marathi Language along with Transliteration and IPA

क	ka	/kə/	च	ca	/tsə/	ट	ṭa	/t̪ə/	प	pa	/pə/
ख	kha	/kʰə/	छ	cha	/tʃʰə/	ठ	ṭha	/t̪ʰə/	फ	pha	/fə/
ग	ga	/gə/	ज	ja	/zə/	ड	ḍa	/d̪ə/	ब	ba	/bə/
घ	gha	/gʰə/	झ	jha	/zʰə/	ढ	ḍha	/d̪ʰə/	भ	bha	/bʰə/
ङ	ṅa	/ŋə/	ञ	ña	/ɲə/	ण	ṇa	/ɳə/	म	ma	/mə/
त	ta	/t̪ə/	य	ya	/jə/	ष	ʃa	/ʃə/			
थ	tha	/t̪ʰə/	र	ra	/rə/	स	sa	/sə/			
द	da	/d̪ə/	ऌ	ṛa	/t̪ə/	ह	ha	/hə/			
ध	dha	/d̪ʰə/	ल	la	/lə/	ळ	ḷa	/l̪ə/			
न	na	/nə/	व	va	/wə/	क्ष	kʃa	/kʃə/			
			श	śa	/ʃə/	ज्ञ	jña	/j̪ɲə/			

Fig. 1b. Consonants in Marathi Language along with Transliteration and IPA

Marathi words are formed with the combination of vowels and consonants. e.g. 'Aai' means 'Mother'. This is pronounced as 'a i /' according to International phonetics alphabet (IPA) and composed of two vowels. The collection of utterances in digital format is called computerized speech database and is required for recognition purpose. The isolated spoken words [5] were collected in 35 different sessions from 50 speakers. Initially all the speakers were asked to pronounce of Marathi words, the speaking skill has been examined by examiner; speakers have been selected on the bases of test. After selection, they have been trained how to speak the given words.



## 2.2 Acquisition Setup

The samples were recorded in 15 X 15 X 15 feet room with sampling frequency 11 KHz in normal temperature and humidity. The microphone was kept 5 – 7 cm from speakers. Whole speech database was collected with the help of Computerized Speech Laboratory (CSL). It is an input/output recording device for a PC, which has special features for reliable acoustic measurements. CSL offers input signal-to-noise performance typically 20-30dB. Analog Inputs with 4 channels two XLR and two phono-type, 5mV-10.5V peak-to-peak, channels 3 and 4, switchable AC or DC coupling, calibrated input, adjustable gain range >38dB, 24-bit A/D, Sampling rates::<90dB F.S., 8000-200,000Hz, THD + NFrequency Response (AC coupled): 20 to 22kHz +.05dB at 44.1kHz [6][7].

## 2.3 Isolated Marathi Words Corpus

There are 12 vowels and 36 consonants in Marathi alphabets. Spoken words from the selected speakers were recorded and stored in different files and folders with respective to speakers' initial names in the 'wav' format. The database was included mostly the words starting from each vowels, some of the words with phonetics.

## 3 Words Recognition System

The recognition system was developed using Mel – Frequency Cepstral Coefficient (MFCC) [8]. The following are the step to find the MFCC features.

### 3.1 Speech Signal

The waveform of spoken word 'Ati' along with pronounces of vowels is shown in figure 2. The words of exciting signal and impulse response of vocal tract is called speech signal as shown in Equation (1).

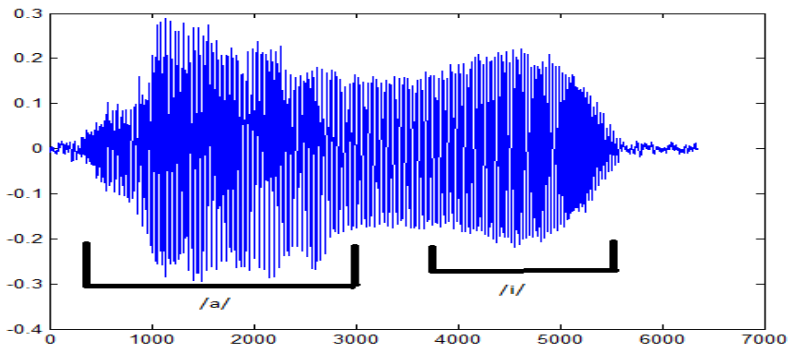


Fig. 2. Waveform of Spoken word 'Ati'

$$S[n] = e[n] * \theta[n] \quad (1)$$

Where,  $S[n]$ ,  $e[n]$  and  $\theta[n]$  are speech signal, exciting signals and impulse response of vocal tract respectively.

### 3.2 Pre-emphasizing

Speech is emphasized with filter  $1-az^{-1}$  where, “a” is between 0.9 and 1. In the time domain, the relationship between the output  $S'_n$  and input  $S_n$  of the pre-emphasis by the default value of a is 0.97 as shown in equation (2).

$$S'_n = S_n * aS_{n-1} \quad (2)$$

### 3.3 Framing and Windowing

The signal is remained stationary at 20ms. Calculation of number of frame is done by multiplying the signal, consisting of  $N$  samples, with a rectangular window function of a finite length as shown in Equations (3), (4) and (5).

$$S_{frames}[n] = S[n]W[n] \quad (3)$$

$$W[n] = \begin{cases} 1 \\ 0 \end{cases} \quad (4)$$

$1 - N \cdot r \leq n < N \cdot (r+1)$ ,  $r = 0, 1, 2, 3, \dots, M-1$   
 0- otherwise

$$N = fs \cdot tframe \quad (5)$$

Where ‘ $M$ ’ is the number of frames, ‘ $fs$ ’ is the sampling frequency and  $tframe$  length of frame measured in time. The frame is shifted 10 ms so that the overlapping between two adjacent frames is to avoid the risk of losing the information from the speech signal. Each frame that contains nearly stationary signal blocks the windowing function is applied. There are a number of different windows functions to minimize the discontinuities. The Hamming window has a very good relative side lobe amplitude and good frequency resolution compared to the rectangular window. The window function is described by Equation (6).

$$x[n] = 0.54 - 0.46 \cdot \left( \frac{2\pi \cdot n}{N-1} \right) \quad (6)$$

### 3.4 Fourier Transform

A 512 point Fast Fourier Transform was applied and calculated using equation 7, for good frequency resolution. The resolution of the frequency spectrum is primarily

decided by the main lobe, while the degree of leakage depends on the amplitude of the side lobes.

$$X(k) = \sum_{n=1}^n x_n * e^{(-j*2*\pi*(k-1)*(n-1)/N)} \quad (7)$$

Where  $1 \leq n \leq N$

### 3.5 Mel-Frequency Filter Bank

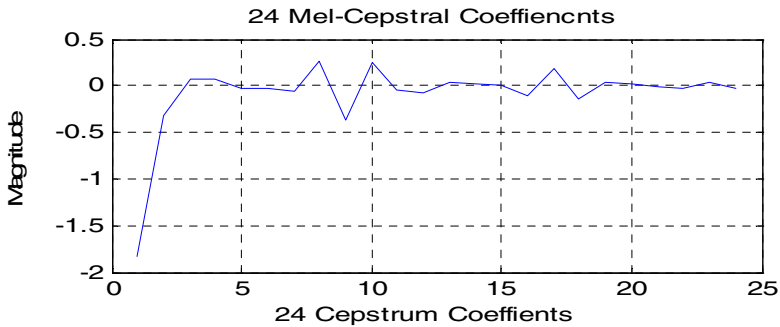
A 24 triangular shaped band-pass filter banks are created by calculating a number of peaks uniform spaced in the mel – scale and then transforming them back to the normal frequency scale. The transmission from linear frequency to mel-frequency is shown in equation (8).

$$mel = 2595 \cdot \log\left(1 + \frac{frequency}{700}\right) \quad (8)$$

The mel frequency spectrum furthermore reduces the amount of data without losing vital information in a speech signal. The resolution as a function of the frequency is logarithmic.

### 3.6 Discrete Cosine Transform

The DCT is widely used in the area of speech processing and is often used when working with cepstrum coefficients. In a frame, there are 24 mel cepstral coefficients obtained, out of 24 only 13 coefficient has been selected for the recognition system as shown in figure 3.



**Fig. 3.** 24 Mel-Cepstral Coefficients

For the word recognition system, we have selected 13 Mel-Cepstral Coefficients frame wise. Figure 4 shows the mean of the 13 Mel Cepstral coefficients of all the frames of spoken for the word.

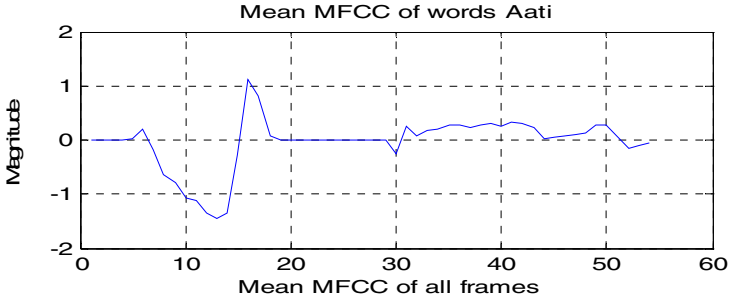


Fig. 4. Mean of MFCC of Marathi word i.e. Aati

### 3.7 Distance Measures

There are some commonly used distance measures i.e. Euclidean Distance, City Block, Weighted Euclidean Distance and Mahalanobis distance. A Euclidean Distance measure is the “standard” distance measures between the two vectors in feature space. Euclidean distance of two vectors  $x$  and  $p$  is measured using the equation (9).

$$d^2 Euclid (\vec{x} \cdot \vec{p}) = \sum_{n=1}^N (x_i - p_i)^2 \tag{9}$$

#### 3.7.1 Optimal Alignment between Two Time Series by Simple Interpolation

As the Euclidean distance measure the standard distance of two same in length vectors, but speech signals are not same in size. To make the vector same in size, for this purpose position ( $i'$  prime) of prime value is calculated by equation (10). Simple interpolation has been calculated by using the equation (11) for example two speech signals having different frame numbers i.e.  $m$  and  $n$  as shown in figure 5.

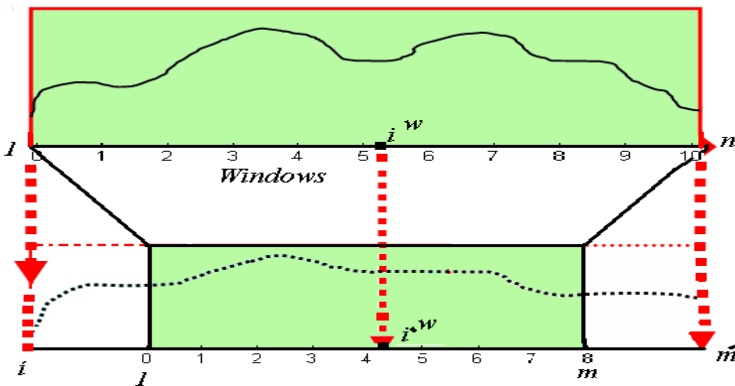


Fig. 5. Speech signals having different frames

$$I' = \left(\frac{m-1}{n-1}\right)i + \left(\frac{n-m}{n-1}\right) \quad (10)$$

$Wk'$  is values of a frame at a position  $k'$  which can be defined as  $k'$  is not integer. So that integer part is separated from the fraction.

$$I'' = \text{int}(k'), f = k' - \text{int}(k')$$

Where 'int' is integer, so that the interpolated values  $Wk'$  at the position  $k'$  is calculated as shown in equation (11)

$$Wk' = Wk''(1-f) + Wk''' \cdot f \quad (11)$$

Now, both the number of frames is same, is applied the Euclidean distance on the different signals.

### 3.7.2 Optimal Alignment between Two Time Series by DTW Techniques

DTW algorithm is used to determine time series [9] [10]. The minimum distance warp path is calculated using the equation (12).

$$Dist(w) = \sum_{k=1}^{k=K} Dist(W_{ki} - W_{kj}) \quad (12)$$

$Dist(w)$  is typically Euclidean distance of warp path 'w', and  $Dist(W_{ki}, W_{kj})$  is the distance between the two data point indexes in the 'kth' of the warp path.

**Table 1.** Euclidean distance matrix Marathi words calculated by Interpolation Method

Average distant of spoken words by all 50 subjects						
	Aabhar	Aabhas	Aadhar	Aai	Anand	Apali
Aabhar	<b>0.04</b>	0.13	0.14	0.16	0.07	0.14
Aabhas	0.13	<b>0.02</b>	0.15	0.11	0.14	0.10
Aadhar	0.20	0.23	<b>0.02</b>	0.20	0.23	0.13
Aai	0.19	0.15	0.18	<b>0.03</b>	0.13	0.21
Anand	0.10	0.21	0.23	0.15	<b>0.03</b>	0.25
Apali	0.17	0.13	0.11	0.20	0.21	<b>0.02</b>

**Table 2.** Euclidean Distance Matrix Marathi Words calculated by DWT method

Average distant of spoken words by all 50 subjects						
	Aabhar	Aabhas	Aadhar	Aai	Anand	Apali
Aabhar	83	283	391	328	475	341
Aabhas	283	24	265	409	212	137
Aadhar	391	265	27	379	316	347
Aai	328	409	379	34	471	555
Anand	475	212	316	471	47	272
Apali	341	137	347	555	272	59

## 4 Conclusion

We have computed the distance matrixes for 750 corpuses, including males and females. During the experiment, we experienced the effectiveness of MFCC in feature extraction. For this experiment, we have used a limited number of samples. Increasing the number of samples may give the complete recognition. To compare two vectors of different in size, two methods have been used i. e. interpolation and DTW. Table 1 and 2 shows the average values of spoken words from 50 speakers. It is also found the values of diagonal elements are smaller than rest of the elements shown in all the tables. It can be seen from the tables that 100 % recognition is achieved for all subjects in both comparative approaches. The recognition by DTW is more correct than simple interpolation methods. Marathi Speech recognition system based on MFCC features seems to be successful.

## References

1. Ghadge, S.A., Janvale, G.B., Deshmukh, R.R.: Speech Feature Extraction using Mel – Frequency Cepstral Coefficient (MFCC). In: International Conference on Emerging Trends in Computer Science Communication and Information, January 9-11 (2010)
2. Rbner, L., Juury, B.-H.: Fundamental of speech Recognition, 1st edn. Pearson Education, Inc., Copyright 1993 AT & T (1993) ISBN 9780130151575
3. Samudravijaya, K., Rao, P.V.S., Agrawal, S.S.: Hindi Speech database. In: Proc. Int. Conf. Spoken Language Processing, ICSLP 2000, Beijing (October 2000)
4. [http://en.wikipedia.org/wiki/Marathi\\_language](http://en.wikipedia.org/wiki/Marathi_language)
5. Lipeika, A., Lipeikience, J., Telkonys, L.: Development of Isolated word speech Recognition system. Information 13(1), 37–46 (2002)
6. KayPENTAX.: Multi-speech and Software. Software instrumentation manual, issue G, A division of PENTAX – medical company, Bridgewater line, Lincoln Park, NJ 07035 -1488 USA (February 2007)
7. Janvale, G.B., Gawali, B.W., Deshmukh, S.N., Deshmukh, R.R., Mehrotra, S.C.: Speech Analysis through CSL. on 8th -10th at Inter – University Research Festival. Sant Gadge Baba Amravati University, Amravati (MS) India

8. Sato, N., Obuchi, Y.: Emotion Recognition using Mel – Frequency Cepstral Coefficients. *Journal of Natural Language processing* 14(4), 83–96 (2007)
9. Keogh, E., Pazzani, M.: Derivative Dynamic Time Warping. In: *International Processing of First Intl SIAM International Conference on Data Mining*, Chicago, Illinois (2001)
10. Kruskal, J., Liberman, M.: The Symmetric Time Warping Problem: From continuous to Discrete. In: *Time Warps String Edits and Macromolecules. The Theory and Practice of Sequence Comparison*, pp. 125–161. Addison – Wesley Publication Co., Reading Massachusetts (1983)

# An Automatic Process to Convert Documents into Abstracts by Using Natural Language Processing Techniques

Ch. Jayaraju, Zareena Noor Basha, E. Madhavarao, and M. Kalyani

Department of Computer Science & Engineering,  
Vignan's Lara Institute of Technology and Science, Guntur, A.P., India  
glorious.ch@gmail.com

**Abstract.** Now a days each and every people using internet and collects the information. At the same time the internet is growing exponentially, huge amount of information is available online. That's why the information overload problem is faced by every end user. So Automatic Process of Document Abstracts is recognized as an important task. For this intention we used various approaches these are Anaphora resolution, mining methods and TFxIDF. However these techniques have some limitations and mainly the drawback is from the end user's perspective, the requestor may not be aware of all the knowledge that constitutes the methods. That's why in this paper we focussed on developing Abstracts, that is Summarization method based on Natural Language Processing Techniques. At the same time it is also useful to multi-documents summarization. We explore some of the metrics and evaluation strategies, features in document abstracts or summarization.

**Keywords:** Information overload, TFxIDF, Natural Language Processing Techniques, Summarization.

## 1 Introduction

The Process of Documents to Abstracts has been a well-known field of computational linguists for many decades, but only recently has it been possible to advertise, these concepts. Abstract document is basically a short version or more purely we can state that it is a division of the original set. Text summarization is one of the segments of data preprocessing [1]. Abstracts or Summarization is also referred to as characterization [2]. The major design goal of summarization is to convey the same amount of information available in the original source text in a much smaller document. The study on Automatic Process of Documents to Abstracts has been initiated 40 years before. This area is highly interdisciplinary and related with artificial intelligence, Machine learning, natural language processing, information retrieval, information extraction. The abstract summaries are broadly categorized based on purpose, input and output. There are number of summaries be present [3]. In that one of most important is Abstractive vs Extractive Summaries. Abstractive summaries or abstracts, involve identifying important parts of the source document



and reformulating it in original terms. The phrase "important parts" here refer to the parts which contain significant information content related to the source document. Extractive summaries or extracts, involve selecting sentences, words, phrases and combining them together in a cohesive and coherent manner. Abstracts are a much harder problem because the second phase in the problem of generating grammatically correct memories using a natural language generator is hard. On this much of the research has been concentrated in finding effective algorithms and building up efficient systems, such as SMART, SUMMARIST, MEAD [4], and NEATS, which provided precious experience for future research. This paper has proposed a new viable summarizer using Natural Language Processing (NLP) techniques. The aim is to design a summarizer that not only processes traditional "smooth" documents, which are primarily textual documents with no structure, but also to process complex structured documents by retaining the structure. The paper is organized as follows. Section 2 says about the related work in this area. The proposed architecture is discussed in Section 3.

## 2 Related Work

Till now the researchers considered the rate of recurrence of the word combining the mining and document frequency. Semi-supervised learning (SSL) is one of the major learning techniques in Machine learning. It makes use of both labeled and unlabeled data for training. It assumes that amount of labeled data is very little compared to unlabeled data, which makes it a very practical alternative to supervised learning which uses only labeled data, which is very expensive. Does not consider the performance of the effective documents. Based on an iterative graph based ranking algorithm, Rada Mihalcea and Paul Tarau explained an approach for language autonomous extractive summarization. They presented that in spite of the language, their algorithm works efficiently. They did so by means of appraisal applied on single document summarization task. Those tasks were in Portuguese and as well as in English [5].

Inderjit S. Dhillon [6], worked on Theoretic Feature Clustering Algorithm for text classification, it tells about feature clustering, feature selection of word clusters. But have not considered the meaningful documents and text summarization for better search.

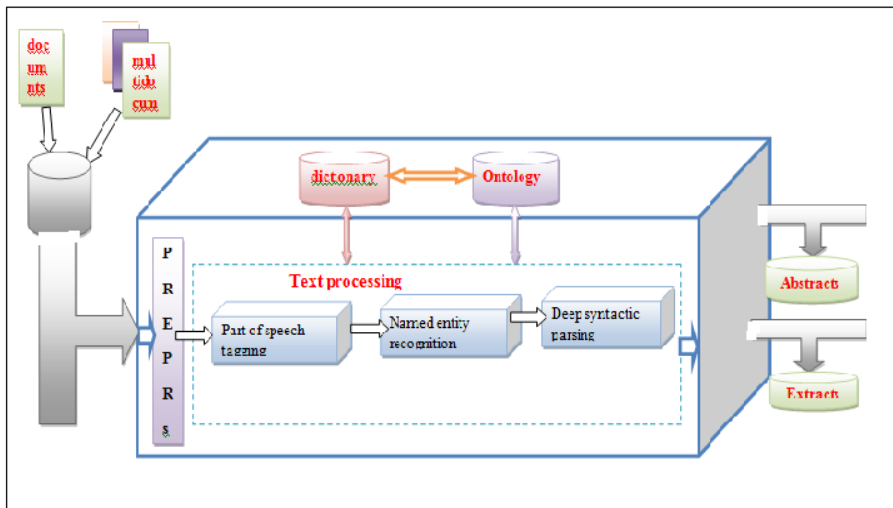
Finley Lacatusu et. al. elucidates a novel clustering based text summarization system that uses manifold sequence Alignment in order to enhance the arrangement of sentences contained by theme clusters .

Youngjoong KO, JinwooPark et al. explores improving text categorization using the importance of sentences. They discuss the importance of sentences using text summarization techniques. Four kinds of classifiers are used in this experiment. Naive bayes, Rocchio, KNN, and SVM. A supervised learning algorithm has been applied to this area using a training data set of categorized documents. The aim of this paper presented a new indexing method for text categorization using two kinds of text summarization techniques; one uses the TITLE and the other uses the IMPORTANCE of TERMS. And don't considered the how important the document is and the

comparison of documents to know which is better. To overcome these drawbacks architecture is proposed which does summarization. Section 3 gives the detailed description for this architecture.

### 3 Proposed System

The proposed Automatic Process to convert structure Documents into Abstracts summarizer has multiple steps coupled with it. The first step is a pre-processing of documents. In that mainly consist of two steps the first one is examination of documents [7], taxonomy of document .The proposed system is shown below diagram.



**Fig. 1.** Proposed framework for Automatic Process to convert Documents into Abstracts summarization

#### 3.1 Pre-processing

The process of the Automatic Process to convert Documents into Abstracts summarization the important step is pre-processing. It is based on following steps.

##### 3.1.1 Structure Examination of Documents

Document structure analysis can be regarded as a syntactic analysis problem. The order and containment relations among the physical or logical components of a document page can be described by an ordered tree structure and can be modeled by a tree grammar which describes the page at the component level in terms of regions or blocks. The principal attributes for detecting titles and section headings include font size, boldness, underline, and link properties. Once identified, heuristics are used to

classify them as titles or section headings by analyzing their relative font size. This also provides information about the overall layout and content size of each section. Content may include text, images, links and other entries. Figure 3 shows the extracted structural layout from the document of Figure 2.

Machine learning

For the journal, see [Machine Learning \(journal\)](#). See also: [Pattern recognition](#) **Machine learning**, a branch of [artificial intelligence](#), is about the construction and study of systems that can [learn](#) from data. For example, a [machine learning](#) system could be trained on email messages to learn to distinguish between spam and non-spam messages. After learning, it can then be used to classify new email messages into spam and non-spam folders. The core of machine learning deals with representation and generalization. Representation of data instances and functions evaluated on these instances are part of all machine learning systems. Generalization is the property that the system will perform well on unseen data instances; the conditions under which this can be guaranteed are a key object of study in the subfield of [computational learning theory](#). There is a wide variety of machine learning tasks and successful applications. [Optical character recognition](#), in which printed characters are recognized automatically based on previous examples, is a classic example of machine learning. In 1959, [Arthur Samuel](#) defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed".<sup>[1]</sup> [Tom M. Mitchell](#) provided a widely quoted, more formal definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".<sup>[2]</sup> This definition is notable for its defining machine learning in fundamentally [operational](#) rather than cognitive terms, thus following [Alan Turing's](#) proposal in Turing's paper "[Computing Machinery and Intelligence](#)" that the question "Can machines think?" be replaced with the question "Can machines do what we (as thinking entities) can do?" |

Fig. 2. Document structure

```
<Head>Support Vector Machines for Web Page Classification</Head> <ContentWeight>
<337></ContentWeight><ImageWeight>0</ImageWeight><LinkWeight>0</LinkWeight>
<SecHead>ABSTRACT</SecHead><ContentWeight><330></ContentWeight>
<ImageWeight>0</ImageWeight><LinkWeight>0</LinkWeight>
<SecHead>Categories and Subject Descriptors</SecHead><ContentWeight>
<5></ContentWeight><ImageWeight>0</ImageWeight><LinkWeight>0</LinkWeight>
<SubSecHead>Designs Methodology</SubSecHead><ContentWeight><77></ContentWeight>
<ImageWeight>0</ImageWeight><LinkWeight>0</LinkWeight>
<SecHead>General Terms</SecHead><ContentWeight><18></ContentWeight>
<ImageWeight>0</ImageWeight><LinkWeight>0</LinkWeight>
<SecHead>Keywords</SecHead><ContentWeight><54></ContentWeight>
<ImageWeight>0</ImageWeight><LinkWeight>0</LinkWeight>
<SecHead>INTRODUCTION</SecHead><ContentWeight><1814></ContentWeight>
<ImageWeight>0</ImageWeight><LinkWeight>0</LinkWeight>
<SecHead>SUPPORT VECTOR MACHINES</SecHead>
<ContentWeight><2892><ContentWeight><ImageWeight>0</ImageWeight>
<LinkWeight>0</LinkWeight>
```

Fig. 3. Extracted structure from Figure 2

### 3.1.2 Taxonomy of Document

Taxonomy of document is the task of approximating the unknown target function  $\Psi : D \times C \rightarrow \{T, F\}$  by means of a function  $\Phi : D \times C \rightarrow \{T, F\}$  called the classifier, such that  $\Psi$  and  $\Phi$  "coincide as much as possible". Here

- $C = \{c_1, \dots, c_{|C|}\}$  is a fixed set of pre-defined categories ;
- $D$  is a domain of documents. Depending on the application, classification may be single-label : exactly one category must be assigned to each

Document. A special case is when  $|C| = 2$ .

- multi-label: any number of categories can be assigned to each document. And it generally the attribution of documents to categories should be realized on the basis of the content of the documents, and not on the basis of metadata that may be available from an external source.

## 4 Text-Processing

The succeeding step is an text processing in that NLP Techniques have been used to create summaries of their content. Cohesion is a way of connecting different parts of text into a single theme. In other words, this is a list of semantically related words, constructed by the use of co-reference, ellipses and conjunctions. This aims to identify the relationship between words that tend to co-occur in the same lexical context.

### 4.1 Parts of Speech (Pos) Tagger

An example might be the relationship between the words "Joy" and "Loves" in the sentence: "Joy loves Truelight". For every sentence in the node (the "content"), all nouns are extracted using a Parts of Speech (POS) tagger[8]. A simple rule-based part of speech tagger which automatically acquires its rules and tags with accuracy comparable to stochastic taggers. It is highly helpful to a vast reduction in stored information or documents. The perspicuity of a small set of meaningful rules, ease of finding and implementing improvements to the tagger, and better portability from one tag set, corpus genre or language to another. All possible synonym sets are determined that each noun could be part of speech for every synonym set. It is simply shown as .It Assigns a part-of-speech tag to each word in the sentence.

"Joy	a	professor	at	True light	University,	was
NNP	DT	NN	IN	NNP	NNP	VBD
Awarded the Nobel Prize in computers on Monday".						
VBN	DT	NNP	NNP	IN	NNP	IN NNP

Part-of-speech tags are NN: Noun IN: Preposition VBD: Verb, past tense DT: Determiner. Part-of-speech tagging is not easy task. It case Parts-of-speech are often ambiguous .For this various approaches are used. For example ambiguity is shown in the following sentence.

"I have to <u>Go</u> Tenali" Here <u>Go</u> is VERB
"I had a <u>GO</u> at it" Here Go is NOUN

The above problem is ambiguous, mostly we can overcome this with the help of the concept called FBA. FBA [9] stands for forward-backward algorithm which has been invented by **Welch**. It is specially based on Maximization Estimation method. It is very effective when compared to previous methods like a **Markov model**, **Hidden Markov model**. FBA works without using manually annotated corpora. FBA guarantees to find a locally best set of values from the initial parameter values. It works well if small amount of manually tagged corpus given.

## 4.2 Named Entity Recognition

Named Entity Recognition [10] is one of the most important preprocessing tools for many natural language processing tasks, in that one of the main ones is document and text summarization, speech translation. The role of named entity based patterns is emphasized in measuring the document sentences and topic relevance for topic-focused extractive summarization. Supervised NER systems perform well when they are trained and tested on data from the same domain. The NER for summarization is based on the following concepts. These are Entity Type, In Title or Not, Document Has Entity in Title or Not, Document Frequency in the Corpus, Entity Neighbor, Entity Frequency. At the same time the task and domains define a simple example below.

### 4.2.1 Task and Domains

NER task is similar to those defined in some standard evaluations such as MUC-6, CoNLL-2003. Given a raw sentence, the goal is to identify name expressions and classify them into one of the following three types: PER (person), ORG (organization) and GPE (Geo-Political entity). We choose to work with these three types as they are the most frequent ones in our target domain. The Figure illustrates examples from both domains.

<b>Source</b>	<b>Example:</b>	<GPE>U.S.</GPE> <ORG>Defense</ORG> Secretary <PER>Donald H. Rumsfeld</PER> discussed the resolution ...
<b>Target Example1:</b>	<i>The ruler of &lt;GPE&gt;Saudi Arabia&lt;/GPE&gt; is &lt;PER&gt;Fahad bin Abdul Aziz bin Abdul Rahman Al-Saud&lt;/PER&gt;.</i>	
<b>Target Example2:</b>	<i>... where Sheikh &lt;PER&gt;Abdul Sattar al-Rishawi&lt;/PER&gt; and the &lt;ORG&gt;Anbar Salvation Front&lt;/ORG&gt; became a force for stability.</i>	

Fig. 4. Task and domain

Figure 4: Examples of NER task and domains. Our target domain documents are from publicly available reports (in English) on terrorism, such as those from the Combating Terrorism Center at West Point<sup>2</sup>. There are many domain-specific characteristics of our target domain.

## 5 Dictionaries

The Dictionaries have consisting of many strengths for summarization. The first one is

### 5.1 Wealth of Information [11]

- Semantics:** Lexical items are carefully analyzed and explained, and their various T L equivalents are set out clearly and helpfully.

- Grammar:** There is a commitment to include enough information (albeit often couched in opaque codes), to allow the foreign language expressions to be used correctly.

- Collocation:** This type of information is often drawn from corpora, and the the tendency now is towards including this wherever possible.

- Linguistic:** The summary list of the types of information painstakingly gathered, ordered, compressed and presented intelligibly gives enough evidence of this.

## 6 Ontology

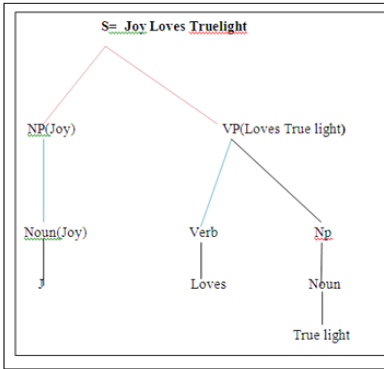
It plays a vital role in the next generation Web. The term ‘ontology’ [12] is derived from the Greek words “onto”, which means being, and “logia”, which means written or spoken discourse. Tom Gruber defines ontology as an explicit specification of a conceptualization. According to Handler ontology is “a set of knowledge terms, including the vocabulary, semantic interconnections, and some simple rules of inference and logic for some particular topic”. Existing ontology’s can be classified into the following major categories: (1) meta-ontology’s, (2) upper ontology’s, (3) domain, and (4) specialized ontology’s. The Web ontology languages (OWL) is a language to define and instantiate Web ontology’s. It was formerly called DAML+OIL language. OWL ontology may include descriptions of classes, along with their related properties and instances. OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans.

## 7 Deep Syntactic Parsing

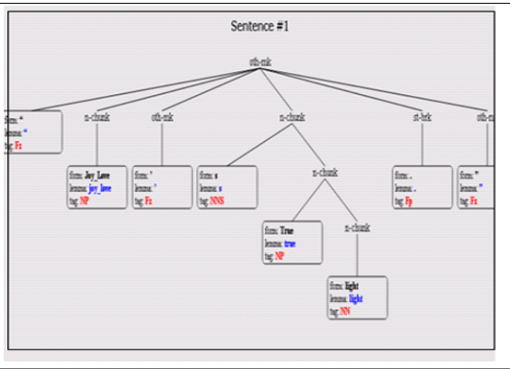
In the field of computational linguistics parsers have been developed for generating syntactic analyses [13]. Evaluating parser performance is useful for locating elements of the parser that leave room for improvement. If a parser is not applied for the purpose of a language technology application, evaluation is typically intrinsic, i.e. measuring the performance of a parser in the framework it is created in by comparing parser output to a truth at the right hand, a gold standard. This type of evaluation differs greatly from extrinsic evaluation, where the benefit of the parser to a language technological application is established. In this paper, we undertake an intrinsic evaluation of a parsing system designed for linguistic purposes – more specifically

for the linguistic annotation of text corpora – that is being employed in a document abstracts concepts. The syntactic parsing simply defined as a provides rules to put together words to form components of sentence and to put together these components to form sentences. Here parsing is a a method to perform syntactic analysis of a sentence. The example of parsing i:e Syntactic structure.

**Input: “Joy Love’s True light.” Output:**



**Fig. 5.** Syntactic structures



**Fig. 6.** Deep Syntactic parsing

## 8 Conclusion

In this paper, we proposed a method which is used in our document to abstract summarization concept. It is based on NLP Techniques. We have presented different steps. Our experimental consequence indicates that 'Generating summary using similarity score based on some text processing concepts. It is also very much useful for best outcomes of end user request. In future, we will expand this paper to acquire more enhanced domino effects by using different Text related algorithms. In addition, we will apply machine learning models for further enhanced estimation of process of document to abstract summarization.

**Acknowledgment.** We take this opportunity to acknowledge those who have been great support and inspiration through the research work.

## References

- [1] Tanasa, D.: Advanced Data Preprocessing Mining. IEEE Intelligent Systems 19(2)
- [2] Luhn, H.P.: The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development 2(2), 159–165 (1958)
- [3] Buckley, C., Cardie, C.: SMART Summarization System. In: Hand, T.F., Sundheim, B., eds. (1997)

- [4] Radev, D.R., Jing, H., Budzikowska, M.: Summarization of multiple documents: clustering, sentence extraction, and evaluation. In: ANLP/NAACL Workshop on Summarization, Seattle, WA (April 2000)
- [5] Dhillon, I.S.: A divisive information theoretic feature clustering algorithm for text classification. *Journal of Machine Learning Research* 3 (2003)
- [6] Mihalcea, R., Tarau, P.: A Language Independent Algorithm for Single and Multiple Document Summarization. University of North Texas
- [7] Alam, H., Kumar, A., Nakamura, M.: Structured and Unstructured Document Summarization: Design of a Commercial Summarizer using Lexical Chains
- [8] Santorini, B.: Part-of-Speech Tagging Guidelines for the Penn Treebank Project
- [9] [worldofcomputing.net/pos-tagging/markov-models.html](http://worldofcomputing.net/pos-tagging/markov-models.html)
- [10] A survey of named entity recognition and classification David Nadeau. Satoshi Sekine National Research Council Canada / New York University
- [11] dictionary generation for low-resourced language pairs Varga István Yamagata uNIVERSITY, Graduate School of Science and Engineering, [dyn36150@dip.yz.yamagata-u.ac.jp](mailto:dyn36150@dip.yz.yamagata-u.ac.jp)
- [12] Ontology's, Web 2.0 and Beyond. Keynote presentation at the Ontology Summit 2007 – Ontology, Taxonomy, Folksonomy: Understanding the Distinctions (March 1, 2007)
- [13] Detecting Opinions Using Deep Syntactic Analysis Caroline Brun Xerox Research Centre Europe Meylan, France



# A Novel Method for CBIR Using Texture Spectrum in Wavelet Domain

G. Rosline Nesa Kumari, M. Sudheer, and R. Tamilkodi

Godavari Institute of Engineering & Technology, Rajahmundry  
{rosemaruthu,maridisudheer}@gmail.com, tamil\_kodiin@yahoo.co.in

**Abstract.** This paper presents an effective color image retrieval scheme by contrast based texture features, which achieves higher retrieval efficiency. The proposed method Texture Spectrum in Wavelet (TSW) domain is very fast to compute and enable to speed up the wavelet computation phase for thousands of sliding windows of varying sizes in an image. Proposed TSW provides a robust contrast features for image retrieval from a lot of objects in an image can be distinguished solely by their textures without any other information. The proposed system divides the lower resolution approximation image into four sections, where each section extracts 12 contrast features and stores the features of the query image and also all images in the database and extracts the features of each image. The proposed TSW method reduces the computation of possible patterns as well as fast and retrieving accurate images. Experimental results show that the proposed method for image retrieval is more accurate, efficient and quite understandable in spite of the availability of the existing retrieving algorithms.

**Keywords:** Image retrieval, Texture, Robust Feature, Query image, DWT.

## 1 Introduction

Multimedia data including images and videos have been dramatically increased in our life due to the popularity of digital devices and personal computers. Image searching is one of the most important services that need to be supported by such systems Content-Based Image Retrieval (CBIR) systems are used to extract feature vectors that represent image properties such as color, texture, and shape [1]. CBIR is originated with the work of Kato [2] for the automatic retrieval of the images from a database, based on the color and shape. Since CBIR has widely used to describe the process of retrieving desired images from a large collection of database. The CBIR system extracts and stores the features of the query image then it go through all images in the database and extract the features of each image.

In the decade, many image retrieval systems have been successfully developed, such as the IBM QBIC system[3], developed at the IBM Almaden Research Center, the VIRAGE System [6], developed by the Virage Incorporation, the Photo book System[8], developed by the MIT Media Lab, the Visual Seek System [7], developed at Columbia University, the WBIIS System [9], developed Stanford University, and the Blob world System[5], developed at U.C.Berkeley and SIMPLIcity System[4].

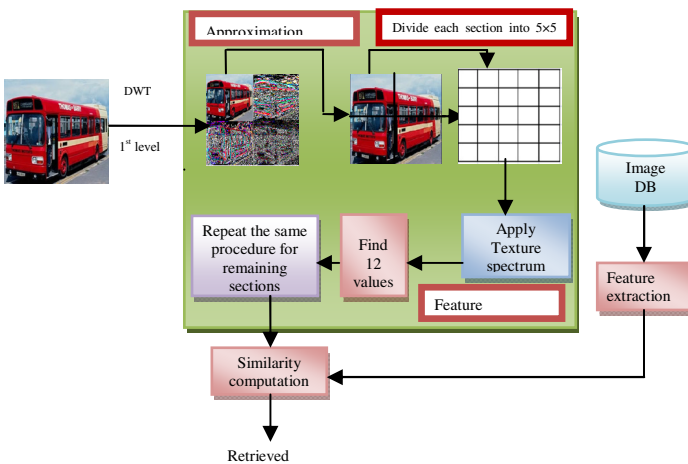
Since simply color, texture and shape features cannot sufficiently represent image semantic, semantic-based image retrieval is still an open problem. This paper proposes an image retrieval system, called Texture Spectrum in Wavelet (TSW) domain for CBIR based on Texture features.

Texture is also considered as one of the feature extraction attributes by many researchers [10-13]. Texture gives us information about the spatial arrangement of color or intensities in an image or selected region of an image. Texture may consist of some basic primitives, and may also describe the structural arrangement of a region and the relationship of the surrounding regions [19]. Many special issues of leading journals have been dedicated to this topic [14, 15, 16, 17, 18]. Content based image retrieval; we feel there is a need to survey what has been achieved in the past few years and what are the potential research directions which can lead to compelling applications. Discrete Wavelet Transformation (DWT) [20] is used to transform an image from spatial domain into frequency domain. The proposed method finds effort to extract the primitive features of the query image and compare them to those of database images. The image features under consideration are texture by using the DWT concept.

The rest of the paper is organized as follows: In section 2, a brief review of the Discrete Wavelet Transformation is presented. The proposed method is given in section 3. Section 4 describes experimental results and the performance evaluation of the proposed method. Finally, conclusion is presented in section 5 respectively.

## 2 Proposed Method

This paper proposes the algorithms for CBIR using Texture Spectrum in Wavelet Domain by using Haar Wavelet. The block diagrams of the proposed TSW method is shown in Figure 1. To demonstrate the strength of the proposed TSW method, a query image is fed into a wavelet filter bank and is decomposed into de-correlated sub bands. Each sub band captures the feature of some scale and orientation of the



**Fig. 1.** Block diagram of proposed TSW

original image. The proposed TSW decompose an image into single wavelet level, thus having 4 sub bands. The approximation image decomposition is divided into four sections and divides into 5x5 windows for extracting texture features in each section. The advantage of the proposed TSW is to reduce the 5x5 window to a 3x3 window. The proposed TSW is going to use 17 out of the 25 pixels from the 5x5 blocks of pixels, by using eight compass directions from the central pixel cp, indexed as shown in Figure 2.

The average intensity of the pixels along each direction is compared with that of the central pixel in the reduction. The reduction of 5x5 window to 3x3 window can be mathematically done by using Equation (1).

Texture spectrum is the histogram of TS (B)<sub>s</sub> with an image. Several useful features, corresponding to visually meaningful patterns may be defined on the Texture Spectrum. The proposed TSW is extracting 12 values according to TS (B)<sub>s</sub>, the first three values of the proposed TSW is surrounding contrast features that have a uniform neighborhood around the central pixel. All the pixels in the neighborhood have the same property with respect to the central pixel, they are all Brighter, Darker and equally brighter as the central pixel as shown in Figure 3.

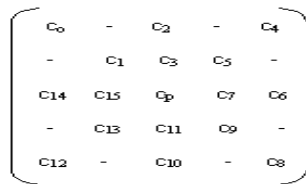


Fig. 2. Pixel intensity values for 5x5 window

$$c_i^u = \begin{cases} 2, & \text{if } c_{2i}^w + c_{2i+1}^w - 2c_p^w > 0 \\ 1, & \text{if } c_{2i}^w + c_{2i+1}^w - 2c_p^w = 0, 0 \leq i \leq 7 \\ 0, & \text{if } c_{2i}^w + c_{2i+1}^w - 2c_p^w < 0 \end{cases} \quad (1)$$

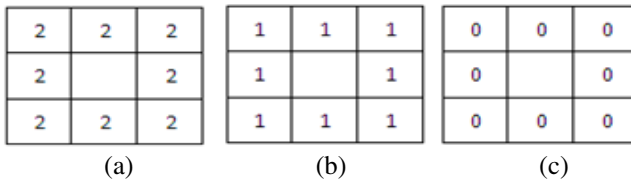
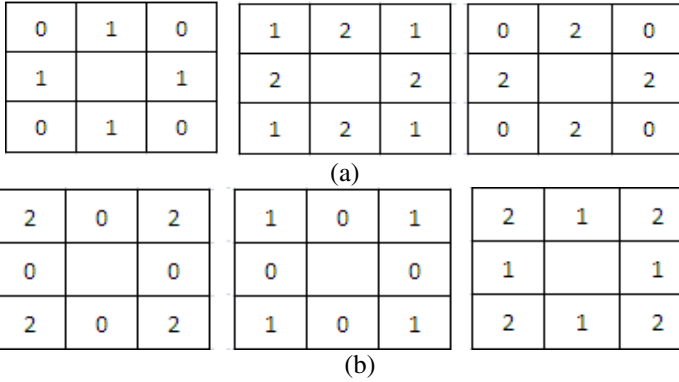


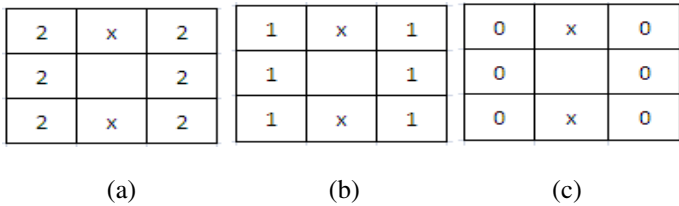
Fig. 3. Surrounding contrast pattern (a) Brighter (b) equally brighter (c) Darker

The next two values of the proposed TSW from alternating contrast features measure the frequency of occurrence of local patterns in which the brightness of the surrounding pixels alternates between being brighter and being darker than the central pixel. There are six possible configurations as shown in Figure 4.



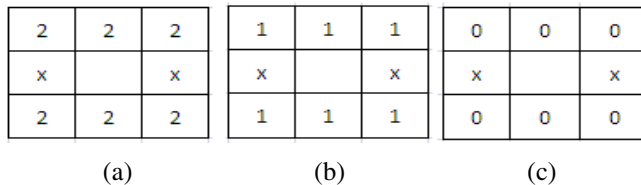
**Fig. 4.** Alternating contrast pattern (a) Brighter (b) Darker

These six possible configurations are combined into two groups. They are brighter and darker. These groups are 4-connected neighbors of collection patterns. From the vertical contrast patterns the next 3 values can be determined. This contrast pattern measures the possible values of  $x$  identical to vertical stripes in an image. The brightness of the central pixel is compared against that of the pixels in the previous and succeeding columns. These features are measure three groups of vertical Brighter, Darker and equal brighter as shown in Figure 5.



**Fig. 5.** Vertical contrast pattern (a) Brighter (b) equally brighter (c) Darker

The proposed TSW calculates 9<sup>th</sup>, 10<sup>th</sup> and 11<sup>th</sup> values using Horizontal contrasts patterns. These features are measure three groups of horizontal Brighter, Darker and equal brighter as shown in Figure 6.



**Fig. 6.** Horizontal contrast pattern (a) Brighter (b) equally brighter (c) Darker

In each category of texture spectrum, with the exception of alternating contrast there are three sub classes of patterns depending on whether the neighboring pixels

are brighter, equal in brightness or dark than the central pixel. There are only two sub classes in alternating contrast feature and that gives a total of 11 features based on texture spectrum in 5×5 windows. The last value of proposed TSW in each section is carried out contrast variation rather than intensity. The proposed TSW is measured 45<sup>0</sup> contrasts as shown in Figure 7. The contrast 45<sup>0</sup> feature is calculated based on average values of C<sub>45</sub> as given in Equation (2).

d <sub>1</sub>	d <sub>2</sub>	d <sub>3</sub>	d <sub>4</sub>	-
d <sub>5</sub>	d <sub>6</sub>	d <sub>7</sub>	-	e <sub>1</sub>
d <sub>8</sub>	d <sub>9</sub>	-	e <sub>2</sub>	e <sub>3</sub>
d <sub>10</sub>	-	e <sub>4</sub>	e <sub>5</sub>	e <sub>6</sub>
-	e <sub>7</sub>	e <sub>8</sub>	e <sub>9</sub>	e <sub>10</sub>

Fig. 7. Calculating contrast at 45<sup>0</sup>

$$C_{45} = \sum_{i=1}^{10} d_i - \sum_{i=1}^{10} e_i \tag{2}$$

When all these calculations are complete, the proposed TSW gives twelve texture spectrum features for each section; therefore the total values of proposed TSW are 12×4= 48 values. Similarly, the texture Spectrum features for all other images in the database are computed and stored.

### 3 Experimental Results

The proposed TSW method is tested with WANG Database contains 1000 images. The images can be divided into 10 categories based on their content namely Buses, Dinosaurs, Flowers, Building, Elephants, Mountains, Food, African people, Beaches and Horses with JPEG format which used in a general purpose image database for experimentation. These images are stored with size 256×256 and each image is represented with RGB color space. Sample of WANG image database is shown in Figure 8. To measure retrieval effectiveness for an image retrieval system, precision and recall values are calculated. Precision measures the ability of the system to retrieve only models that are relevant and this used the ratio of relevant retrieved images to the total number of relevant retrieved images to the total number of retrieved images.

$$\text{Precision} = \frac{\text{Number of relevant image retrieved}}{\text{Total number of images retrieved}}$$

Recall measures the ability of the system to retrieve all models that are relevant and the ratio of retrieved relevant images to the total number of relevant images to the total number of relevant images in the database.

$$\text{Recall} = \frac{\text{Number of relevant image retrieved}}{\text{Total number of relevant images}}$$

The experiment is carried out with the number of retrieved images set as 10 to compute the average precision and recall of each query image. The proposed TSW method experiments based on Sum-of-Absolute Differences (SAD) Similarity measures is shown in Equation (3).

$$SAD(f_q, f_t) = \sum_{i=0}^{n-1} |f_q[i] - f_t[i]| \quad (3)$$

Where  $f_q$ ,  $f_t$  represent query feature vector and database feature vectors and  $n$  is the number of features in each vector.



**Fig. 8.** Sample of WANG Image database

Figure 9 shows the query image Bus and retrieved images based on the query images from WANG database are shown in Figure 10. Table 1 summarizes the experiment results of proposed TSW method with ten different categories of images. Proposed TSW method retrieved Buses, Horses and Dinosaurs categories of images accurately. Building, Flower, Elephant, Food, African people categories of images are retrieved on and above average of 70%. Remaining categories of images are retrieved on an average of 40%. The overall performance of proposed TSW obtained more than 75% of retrieval images with more accurate, efficient and well- organized.

**Table 1.** Summarize the Precision and Recall results of the proposed TSW

Category of Images	Average Precision	Average Recall
Buses	0.933	0.420
Building	0.737	0.198
Flowers	0.750	0.136
Elephants	0.893	0.260
Mountains	0.524	0.232
Dinosaurs	0.972	0.305
Food	0.793	0.177
Beaches	0.655	0.207
African people	0.711	0.386
Horses	0.909	0.291
Average	0.787	0.261



Fig. 9. Query Image: Bus

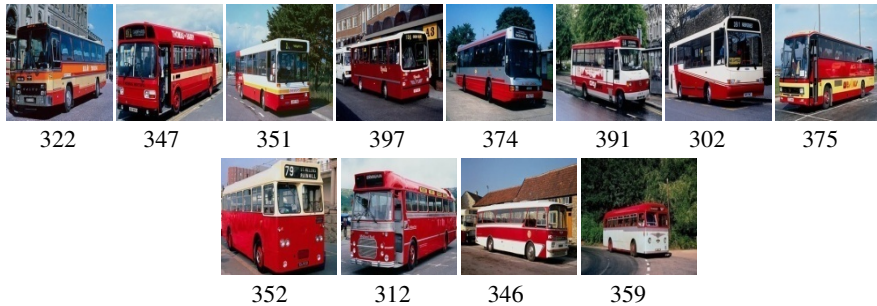


Fig. 10. Retrieved images based on query image Bus

## 4 Conclusion

This paper presents a novel method for Content Based Image Retrieval using Texture Spectrum in wavelet domain. The proposed TSW demonstrated the efficiency of DWT in texture features. Texture analysis is done using texture spectrum. The experimentation is carried out with the number of retrieved images set as 10 classes to compute the average precision and recall of each query image. The experimental result shows that the proposed method outperforms to other retrieval method gives higher average precision and Recall. More over, the computational steps are effectively reduced with the use of Wavelet transformation. The proposed TSW shows better performance compare to other wavelet based texture features methods.

## References

1. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.* 2(1), 1–19 (2006)
2. Kato, T.: Database architecture for content-based image retrieval. In: *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 1662, pp. 112–113 (1992)
3. Chang, S.-K., Yan, C.W., Dimitroff, D.C., Arndt, T.: An intelligent image database system. *IEEE Trans. Software Eng.* 14(5) (1988)
4. Narasimhalu, A.D.: Special section on content-based retrieval. *Multimedia Systems* (1995)
5. Carson, I.C., Belongie, S., Greenspan, H., Malik, J.: Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Mach. Intell.* 8(8), 1026–1038 (2002)
6. Chang, S.-K., Hsu, A.: Image information systems: Where do we go from here? *IEEE Trans. on Knowledge and Data Engineering* 4(5) (1992)

7. Gudivada, V.N., Raghavan, J.V.: Special issue on content-based image retrieval systems. *IEEE Computer Magazine* 28(9) (1995)
8. Tamura, H., Yokoya, N.: Image database systems: A survey. *Pattern Recognition* 17(1) (1984)
9. Wang, J., Wiederhold, G., Firschein, O., We, S.: Content-based Image Indexing and Searching Using Daubechies' Wavelets. *International Journal on Digital Libraries (IJODL)* 1(4), 311–328 (1998)
10. Natsev, A., Rastogi, R., Shim, K.: WALRUS: A Similarity Retrieval Algorithm for Image Databases. In: *Proceeding. ACM SIGMOD Int. Conf. Management of Data*, pp. 395–406 (1999)
11. Ardizzoni, S., Bartolini, I., Patella, M.: Windsurf: Region based Image Retrieval using Wavelets. In: *IWOSS 1999*, pp. 167–173 (1999)
12. Wouwer, G.V.D., Scheunders, P., Dyck, D.V.: Statistical texture characterization from discrete wavelet representation. *IEEE Transactions on Image Processing* 8, 592–598 (1999)
13. Livens, S., Scheunders, P., Wouwer, G.V.D., Dyck, D.V.: Wavelets for texture analysis, an overview. In: *Proceedings of Sixth International Conference on Image Processing and Its Applications*, vol. 2, pp. 581–585 (1997)
14. Gudivada, V.N., Raghavan, J.V.: Special issue on content based image retrieval systems. *IEEE Computer Magazine* 28(9) (1995)
15. Pentland, A., Picared, R.: Special issue on digital libraries. *IEEE Trans. Patt. Recog. and Mach. Intell.* (1996)
16. Narasimhalu, A.D.: Special issue on content based retrieval. *Multimedia Systems* (1995)
17. Jain, R. (Guest ed.): Special issue on Visual information management. *Comm. ACM* (December 1997)
18. Schatz, B., Chen, H.: Buiding large scale digital libraries. *Computer* (1996)
19. Vogel, J., Schiele, B.: Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition* 39(5), 897–909 (2006)
20. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: *Digital Image Processing Using MALAB*. Pearson Education (2008)



# Identification of Abdominal Aorta Aneurysm Using Ant Colony Optimization Algorithm

A. Dinesh Kumar<sup>1</sup>, R. Nidhya<sup>2</sup>, V. Hanah Ayisha<sup>3</sup>, Vigneshwar Manokar<sup>4</sup>,  
and Nandhini Vigneshwar Manokar<sup>5</sup>

<sup>1,2</sup>Department of Computer Science and Engineering,  
Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India  
{dineshngpit, nidhuraji88}@gmail.com

<sup>3</sup>Department of Computer Science and Engineering,  
MEA College of Engineering, Kerala, India

<sup>4</sup>Department of Computer Science and Engineering,  
Faculty of Engineering, Karpagam University, Coimbatore, Tamil Nadu, India  
vignesh.manohar@gmail.com

<sup>5</sup>Applied Research – Modeling and Simulation,  
Technologies Division, Bharath Corporate, India

**Abstract.** Abdominal aortic aneurysm (AAA) is a localized dilatation of the abdominal aorta. It occurs when there is a increase in the normal diameter of the blood vessels by more than 50 percent. Approximately 90 percent of abdominal aortic aneurysms occur infrarenally, but they can also occur pararenally or suprarenally. This is because of some catastrophic outcome. Due to this, the blood flow is exaggerated so the blood hemodynamic interaction forces are affected. Therefore this will tends to wall rupture. To identify the AAA, it is important to identify the blood flow interaction and the wall shear stress. The blood and wall interaction is the wall shear stress. Computational fluid dynamics (CFD) is used to get the results for the mechanical conditions within the blood vessels with and without Aneurysms. CFD contains vast computations with Navier Stroke Equations so this will be very time consuming. So to make these CFD computations very efficient, Data mining algorithms are to be used. And also DM algorithms will be a best method to predict the shear stress at the AAA. This will estimate the wall shear stress. There is in need of thousands of CFD runs in a single computer for creating machine learning data so grid computing can be used.

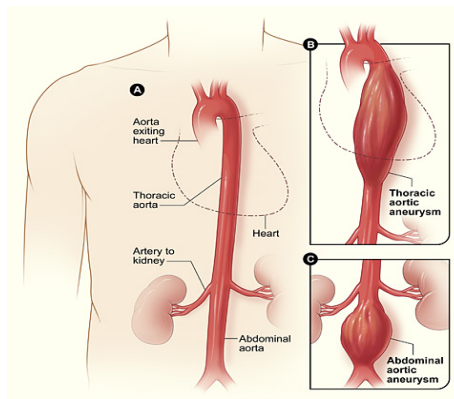
**Keywords:** Computational fluid dynamics (CFD), data mining (DM), grid computing, hemodynamic parameters, predictive modeling, Ant Colony Optimization (ACO) algorithm.

## 1 Introduction

Cardiovascular disease is a class of diseases that involve the heart or blood vessels. Cardiovascular disease refers to any disease that affects the cardiovascular system. Cardiovascular diseases remain the biggest cause of deaths worldwide. Age is

an important risk factor in developing cardiovascular diseases. It is estimated that 87 percent of people who die of coronary heart disease are 60 and older. Among these diseases the stenosis process is the most dangerous one. This is a very tedious process which will lead to stroke and aneurysm and now days this is commonly present in patients. In this paper, we are giving importance to Abdominal Aorta Aneurysm (AAA) development that may cause rupture and fatal outcome. An aneurysm is an abnormal widening or ballooning of a portion of an artery due to weakness in the wall of the blood vessel. It is a localized, blood-filled balloon-like bulge in the wall of a blood vessel. Aneurysms can commonly occur in arteries at the base of the brain and an aortic aneurysm occurs in the main artery carrying blood from the left ventricle of the heart. When the size of an aneurysm increases, there is a significant risk of rupture, resulting in severe hemorrhage, other complications or death. Aneurysms can be hereditary or caused by disease, both of which cause the wall of the blood vessel to weaken. Aneurysms may be classified by type, location, and the affected vessel. Other factors may also influence the pathology and diagnosis of aneurysms. A consensus definition of an aneurysm was established in 1991 by the Society of Vascular Surgery and the International Society for Cardiovascular Surgery as a permanent localized dilatation of an artery having at least 50% increase in diameter compared with the expected normal diameter of the artery, or of the diameter of the segment proximal to the dilatation.

To predict the AAA many patient specific studies have been done, from this it's concluded that maximum stress within the vessel wall was more appropriate criterion than maximum diameter and also the shear stress on aneurysm wall is found. Shear stress is a frictional force produced by blood flow, affects biology and structure of the wall. There are many methods to predict this.



**Fig. 1.** Detailed View of Abdominal Aorta Aneurysm

The following paragraph describes about the methods along with its disadvantages. First, the Shear stress can be obtained by the Computational Fluid dynamics (CFD). This approach has been employed to study wall-shear stress distribution on idealized models of blood vessels. CFD uses numerical methods and algorithms to solve and

analyze problems that involve fluid flows. CFD involves the Navier Stroke Equations and this CFD involves many computations. So this method is very time consuming. Since there are uncertainties in the patients it is difficult to characterize geometric variability using a small number of recorded parameters[2]. Second, Statistical assessment to identify the relationship between flow patterns and geometric attributes. The main concept is to construct the probabilistic models for the input parameters uncertainties that give a reliable output of interest very quickly, without classical CFD calculations[3-5]. An example of this idea is reported by Kolachalama who used Bayesian–Gaussian process emulator to generate a relationship between geometric parameters and maximal wall shear stress (MWSS), and to identify geometries having maximum and minimum of the wall shear stress (WSS)[7]. Third, Statistical Analysis as the Monte Carlo simulation technique. In this method computer runs for the generated random input values and the resultant data is post processed to estimate the output statistics. CPU requirement for a large sample size, this approach becomes computationally prohibitive, particularly when high-fidelity models are used[6].

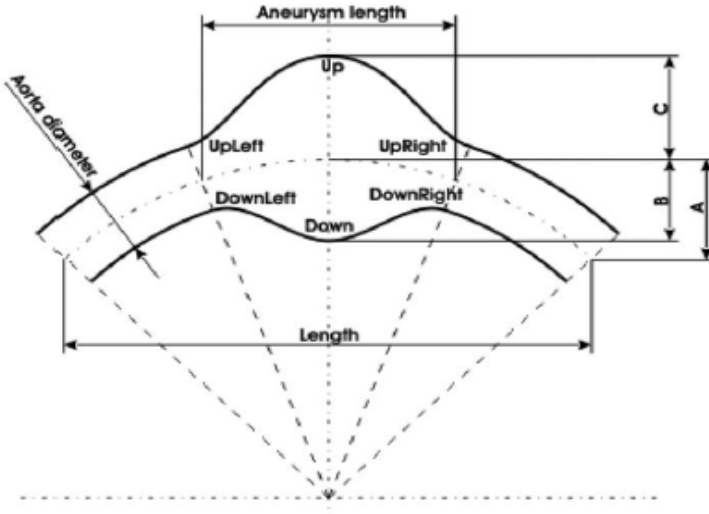
Fourth, it uses data Mining (DM) model that can be used to avoid maximum time taken by the CFD simulations. The DM model uses data mining algorithms where the input is the aneurysm shape, determined by several independent geometric parameters, while the output parameters are the MWSS over the aneurysm for peak systolic flow, and MWSS over full heart cycle. Also, the shear stress averaged over the aneurysm region and over the heart cycle can show which aneurysms are likely to rupture so that the risk can be properly quantified. The aim of this paper is to introduce Data Mining Algorithm for detecting AAA for multiple CFD runs. An Ant Colony algorithm is used for the training and testing of results.

## 2 Materials and Methods

The basic task in our approach was first to feed the DM system with necessary data acquired from the CFD simulations and then to perform machine learning. We ran a number of various aneurysm models in order to “teach” the DM system for producing the most accurate predictions. As in any CFD modeling task, there are a number of issues to be considered: geometry, boundary conditions, initial conditions, mesh generation, etc. We are using the CFD finite element (FE) models, which differ only in geometry and rely on random values of ten parameters.

Only the AAA have to be considered among various kinds of artery aneurysms. The shape of the AAA is defined by two splines with half-circle extrusions between them, as shown in Fig. 2 We divide the geometric parameters into three groups.

- 1) *aneurysm length, A, B, C*—variable parameters;
- 2) *upleft, up, upright, downleft, down, downright*—quantifying the curvature in the Bezier description fashion, variable parameters;
- 3) *length, aorta diameter*—parameters considered are constant, and these are taken from the literature for typical AAA, thus this will give ten independent parameters having a ranges as shown in Table 1.



**Fig. 2.** Parametric model of the aneurysm geometry[1]

**Table 1.** Ranges of AAA Geometrical Parameters[1]

Parameter	Range
Aneurysm Length (cm)	6 to 15
A (cm)	0 to 3
B (cm)	1.4 to 3
C (cm)	1.4 to 3
Curvature parameters (UpLeft, ...)	0.15 to 0.4

## 2.1 CFD Model of an Aneurysm

The next step in the prebuilding models for the CFD analysis. This includes the specification of boundary conditions. The identical boundary conditions are prescribed for all combinations of the variable input parameter values. We used an equivalent length at the aneurysm outlet to model the resistance to the blood flow. Other relevant quantities with fixed values are blood density  $\rho = 1.05 \text{ g/cm}^3$ , kinematic Viscosity  $\nu = 0.035 \text{ cm}^2/\text{s}$ , length = 24 cm, and aorta diameter  $D = 2 \text{ cm}$ . CFD module generates finite element models for each CFD run. All data are then transferred to Storage Element (SE) from User Interface (UI). Then, each CFD instance reaches destination worker node anywhere in the grid, determined by the Workload Management System (WMS). This procedure significantly saves time, enabling a large number of runs to be executed concurrently (Taken from Reference). The fundamental equations for flow of a viscous incompressible fluid (such as blood) are the Navier–Stokes equations[8].

$$\left( \frac{1}{\Delta t} M_v + {}^{t+\Delta t} K_{vv}^{(i-1)} + {}^{t+\Delta t} K_{\mu v}^{(i-1)} + {}^{t+\Delta t} K_{vv}^{\prime(i-1)} + {}^{t+\Delta t} J_{vv}^{(i-1)} + K_{\lambda v} \right) \Delta v^{(i)} = {}^{t+\Delta t} F_v^{\prime(i-1)}$$

These Navier–Stokes equations are used to compute the CFD. For the 3-D model of blood flow, eight-node finite element is used with linear interpolation of velocities while the pressure, taken to be constant over the element, is eliminated by a penalty parameter[9]. The incremental iterative form of the FE equilibrium equations. Where the matrices and vectors are defined in a standard FE manner. As an illustration of the CFD solution, the velocity field (left panel) and pressure distribution (right panel) for the peak systole  $t/T=0.16$  of an AAA with  $D/d=2/75$  ( $D$  is aneurysm diameter),  $d = 12.7$  mm. After initializing input parameters and creating appropriate FE mesh, a specialized CFD FE analysis module performs hemodynamic simulation. Among these quantities, the DM system only considers the wall shear stress values. A total number of models (and therefore FE meshes) constructed using random parameter values and automatic mesh generator was 6000.

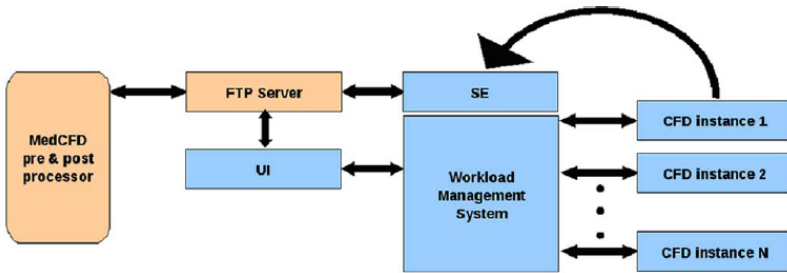


Fig. 3. Grid computing scheme[1]

Since a huge number of CFD finite element analyses had to be executed (for 6000 different geometries based on ten variable parameters) it will take more time (takes even days). So Grid Computing is used to reduce the computation time. Since each CFD run takes around 20 min on a typical personal computer, simple computational time estimation gives around 80 days to run on a single CPU. Equivalent run on a grid platform took 5 h only, while the infrastructure utilization peak during that run was around 600 CPUs at a time.

## 2.2 DM Approach

After obtaining data from a number of FE runs machine learning process should be done. For this purpose Ant Colony Optimization (ACO) Algorithm is used. The ACO algorithm looks for the minimum of the error function in weight space using the method of gradient descent. The combination of weights which minimizes the error function is considered to be a solution of the learning problem. Since this method requires computation of the gradient of the error function at each iteration step, we must guarantee the continuity and differentiability of the error function.

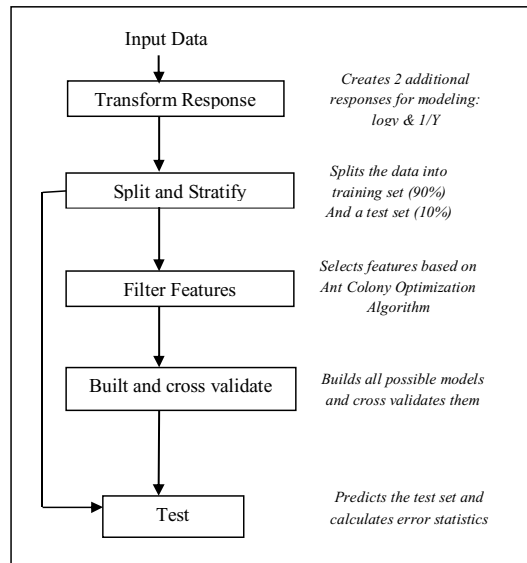
```

Training set = all training cases;
  WHILE (No. of cases in the Training set > max_uncovered_cases)
    i=0;REPEAT i=i+1;
  Anti incrementally constructs a classification rule;
  Prune the just constructed rule;
  Update the pheromone of the trail followed by Anti;
  UNTIL (i ≥ No_of_Ants) or (Anti constructed the same rule as the previous
  No_Rules_Converg-1 Ants)
    Select the best rule among all constructed rules;
    Remove the cases correctly covered by the selected rule from the
    training set;
  END WHILE

```

#### **Ant Colony Optimization (ACO) Algorithm.**

We have used the regression model for finding  $\tau$  systolic (maximum shear stress for the systolic phase) and  $\tau$  max (maximum shear stress value for the entire cycle domain). Fig. 4 shows these steps the first step is to create additional transformations to the dependent variables, in our case  $\tau$  systolic and  $\tau$  max. Sometimes, it is easier to create a regression if the dependent variable  $Y$  is first transformed as  $\log(Y)$  or  $1/Y$ . Therefore, in addition to building models for  $Y$ , the Bus automatically builds models for  $\log(Y)$  and  $1/Y$ . The second step is to sort the dependent variable in the ascending order and take every tenth out for testing. The third step is to perform variable selection. As a result, it produces up to five different subsets of features, including a solution where all variables are selected. The fourth step is to build regression models



**Fig. 4.** Overview of the main stages in building and testing

and to cross validate them. All regression models are at the end cross-validated (tenfold) and the following statistics is calculated.

- 1) Tenfold cross-validated RMSE—root mean square error after tenfold cross validation.
- 2) Tenfold cross-validated  $Q^2$ —coefficient of determination after tenfold cross validation.
- 3) Tenfold cross-validated RelSE—relative squared error after tenfold cross validation.
- 4) Size of the training dataset ( $N_{tr}$ )

The last step is to apply all regression models to the test set and following statistics are calculated.

- 1) RMSE—root means square error.

$$RMSE_{train} = \sqrt{\frac{1}{N_{train}} \sum (y_i - y'_i)^2} \quad RMSE_{test} = \sqrt{\frac{1}{N_{test}} \sum (y_i - y'_i)^2}$$

2.  $R^2$ —coefficient of determination.

$$R^2_{test} = 1 - \frac{\sum (y_i - y'_i)^2}{\sum (y_i - y'_{test})^2}$$

$y_i$  *i*th observed value in the training dataset;

$y'_i$  *i*th predicted value in the training dataset;

$y'_{test}$  *mean of the observed* values in the testing dataset;

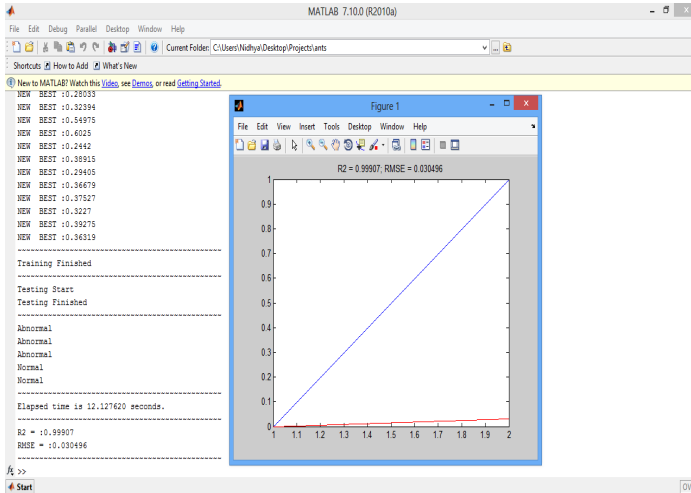
$N_{train}$  number of compounds in the training dataset;

$N_{test}$  number of compounds in the training dataset

From these values we have to produce a performance metrics in which the error rate should be a minimum. RMSE can have an error of 0.001 and  $R^2$  as 0.9.

### 3 Results and Discussion

From 85 models for property  $\tau$  systolic and the best model was found for  $\tau -1$  systolic. It is a feed forward neural network model that uses all ten input variables. From 82 models for property  $\tau_{max}$  and the best model was found for  $\tau -1_{max}$ . It is a feed forward neural network model that uses all ten input variables. The models show a very low error statistics on the test and training sets. In order to better quantify the DM method, we also analyzed an AAA model based on the patient-specific clinical data. The best model among 85 models produced by ACO algorithm (feed forward neural network that uses all ten input variables) has been chosen in order to obtain the best approximation for  $\tau$  systolic and  $\tau_{max}$ . As expected, the most realistic model from the first approach consumes the most computing time. The values of  $\tau_{max}$  and  $\tau$  systolic were taken as reference values for the parametric and DM models. It can be observed that all relative errors are below 10%, with significant computing time reduction in favor of the DM model.



**Fig. 5.** Experimented Stimulated Output in MATLAB

## 4 Conclusion

This paper provides a better performance for detecting the aneurysm. In normal methods, the computation takes more than 80 days to run in a single computer and by introducing Grid Computing it has reduced to 5 hours. The CFD calculations were performed using the FE method and the shear stress was evaluated. From these results, it can be stated that a new direction for reducing computational time for patient-specific AAA modeling is suggested. Also, this approach of coupling the computer modeling and DM method (Ant Colony Optimization Algorithm) can further facilitate the development of predictive diagnostic system for clinical practice. For future enhancement efficient data mining algorithm can be used to provide improved accuracy and performance.

## References

1. Filipovic, N., Ivanovic, M., Krstajic, D., Kojic, M.: Hemodynamic Flow Modeling Through an Abdominal Aorta Aneurysm Using Data Mining Tools, vol. 15(2) (March 2011)
2. Dobrin, P.B.: Distribution of lamellar deformation implications for properties of the arterial media. *Hypertension* 33, 806–810 (1999)
3. Taylor, T., Yamaguchi, T.: Three-dimensional simulation of blood flow in an abdominal aortic aneurysm—Steady and unsteady flow cases. *ASME J. Biomech. Eng.* 116, 89–97 (1994)
4. Finol, E., Amon, C.: Secondary flow and wall shear stress in three dimensional steady flow AAA hemodynamics. In: 2001 Adv. Bioeng., IMECE2001/BED-23013, ASME BED, vol. 51 (2001)



5. Kumar, R., Yamaguchi, T., Liu, H., Himeno, R.: Numerical simulation of 3D unsteady flow dynamics in a blood vessel with Multiple aneurysms. In: 2001 Adv. Bioeng., ASME BED, vol. 50, pp. 475–476 (2001)
6. Rubinstein, R.Y.: Simulation and the Monte Carlo Method. Wiley, Hoboken (1981)
7. Kolachalama, V.B., Bressloff, N.W., Nair, P.B.: Mining data from hemodynamic simulations via Bayesian emulation. *BioMed. Eng. Online* 6, 47 (2007) 1475-925X-6-47.J
8. Kojić, M., Filipović, N., Stojanović, B., Kojić, N.: Computer Modeling in Bioengineering. Wiley, Hoboken (2008)
9. Filipovic, N., Mijailovic, S., Tsuda, A., Kojic, M.: An implicit algorithm within the arbitrary lagrangian-eulerian formulation for solving incompressible fluid flow with large boundary motions. *Comput. Methods Appl. Mech. Eng.* 195, 6347–6361 (2006)

# Color Image Watermarking Using Wavelet Transform Based on HVS Channel

G. Rosline Nesa Kumari, Syama Sundar Jeeru, and S. Maruthuperumal

Godavari Institute of Engineering & Technology, Rajahmundry, India  
{rosemaruthu, jeerusyam, maruthumail}@gmail.com

**Abstract.** Digital image watermarking is a copyright protection technology expected at asserting intellectual property rights of digital images by embedding a copyright identifier in the contents of the image, without disturbing its quality. In this paper, a new robust and secure color watermarking scheme based on Discrete Wavelet domain is proposed. The RGB color space is converted into HSV color space and the H component is decomposition with 'Haar' which is simple, symmetric and orthogonal wavelet. The proposed embedding process, the scrambled watermark is embedded in the modification of the extraordinary value of LL band and LH band of H component to the watermarking scheme which excellent preserves the quality. The supplementary improvement of the proposed technique is taking advantage of HVS which can adaptively regulate the watermark embedding strength. Experimental results show that the method not only has better transparency, but also has good robustness such as noise, compression, filtering, cropping.

## 1 Introduction

In modern year's Copyright protection and authentication of digital data using watermarking is a powerful issue of research. The basic principle of digital watermarking technique [5] is to embed an amount of secret information in the functional part of the cover data. This data may be an image, an audio or video. Since mid 1990s this technique has attracted much attention from academic and industrial sector. After the pioneering contribution by I. J. Cox [5], digital watermarking techniques have been widely developed as an effective tool against piracy, illegal alteration of contents or improper use of image. Digital watermarking is a method that inserts some information into a multimedia object and generates a watermarked multimedia object [6, 7]. The object may be an image, audio, video or text. Watermarking has many different applications [8, 9, 10, 11, 12], such as ownership evidence, fingerprinting, authentication and integrity verification, content labeling and protection, and usage control.

Watermarking techniques can be divided into spatial [13], [14], and frequency [15], [16], [17] based methods. Watermarking algorithms that rely on spatial domains, hide the watermark by modifying the pixel values of the host image. In transform domain technique [1,2,3,4], the host image is first converted into frequency domain by transformation method such as the discrete cosine transform (DCT), discrete

Fourier transform (DFT) or discrete wavelet transform (DWT) ,etc. then, transform domain coefficients are modified by the watermark. Watermarking the process of embedding data into multimedia element can primarily for copyright protection. Because of its growing popularity, the Discrete Wavelet Transform (DWT) is commonly used in the proposed watermarking scheme increase, area increases so power consumption [18 ].

Human Visual System (HVS) plays important role in watermarking of images to maintain the perceptual similarity between original and watermarked image. HVS has been characterized with several phenomena that permit to adjust the pixel values to yield perception. These phenomena are luminance sensitivity, frequency sensitivity and texture sensitivity. Human visual model is based on the characteristics such as edges and textures, which are incorporated to determine the gain factor in watermarking. The distortion visibility is very low if the back ground contains texture. In a high texture block, energy is more distributed among the pixels. Therefore, the block having a stronger texture can have a high embedding gain factor. The paper also utilizes Torus automorphism (TA) permutation [19,20,21] to scramble the watermark before embedding and to reassemble it after extraction. This helps increase robustness to intentional attacks while preserving blindness. The paper also proposes an extended version of this technique to increase the robustness against intentional attacks even further. The algorithm is discussed in the rest of the paper.

This paper describes the efficient and robust color watermarking schemes in an attempt to improve watermark robustness to attacks. The proposed scheme embeds a scrambled watermark image by using Torus automorphism into a 2nd level wavelet color image of H components. Select a LL and LH subband having high energy and then embed scrambled the watermark. The proposed method is directly embedded into cover image using a transform before embedding. Experimental results demonstrate that there is high degree of perceptual similarity between original image and watermarked image. Further, the proposed method resists different types of attacks. The rest of the paper is organized as follows. Section 2 briefly reviews about DWT and Torus automorphism Section 3 describes the proposed color watermarking method and in section 4, the experimental results are discussed. Finally, some conclusions are drawn in section 5.

## **2 Discrete Wavelet Transformation**

The Discrete Wavelet Transform (DWT) [22] is currently used in a wide variety of signal processing applications, such as in audio and video compression, removal of noise in audio, and the simulation of wireless antenna distribution. Discrete Wavelet decomposition of image produces the multi-resolution representation of image. A multi-resolution representation provides a simple hierarchical framework for interpreting the image information. At different resolutions, the details of an image generally characterize different physical structures of the image. At a coarse resolution, these details correspond to the larger structures which provide the image

context. The following section briefly reviews about Two Dimensional Wavelet Transformation. The original image  $I$  is thus represented by set of sub images at several scales;  $\{L_d, D_{nl}\}$ ,  $n = 1, 2, \dots, d$ , which is multi-scale representation with depth  $d$  of the image  $I$ . The image is represented by two dimensional signal functions; wavelet transform decomposes the image into four frequency bands, namely, the  $LL_1$ ,  $HL_1$ ,  $LH_1$  and  $HH_1$  bands.  $H$  and  $L$  denote the highpass and lowpass filters respectively. The approximated image  $LL$  is obtained by lowpass filtering in both row and column directions. The detailed images  $LH$ ,  $HL$  and  $HH$  contain the high frequency components. To obtain the next coarse level of wavelet coefficients, the sub-band  $LL_1$  alone is further decomposed and critically sampled. Similarly  $LL_2$  will be used to obtain further decomposition. By decomposing the approximated image at each level into four sub images forms the pyramidal image tree. This results in two-level wavelet decomposition of image as shown in the Figure 1. Embedding watermarks in these regions allow us to increase the robustness of our watermark, at little to no additional impact on image quality. The fact that the DWT is a multi-scale analysis can be used to the watermarking algorithm's benefit.



**Fig. 1.** Layout of individual bands at second level of DWT decomposition

## 2.1 Torus Automorphism Permutation

To increase the robustness of the embedded watermark against intentional attacks, the proposed method use Torus automorphism (TA) permutation [19,20,21] to disarrange the watermark bits equally and randomly before embedding and reconstruct it after extraction. This scheme offers cryptographic protection against intentional attacks since the keys utilized in TA permutation (for scrambling the watermark) are also necessary in inverse TA permutation (for reconstructing the watermark after extraction). The watermark is scrambled using the following equation before it is embedded into the host image:

$$\begin{pmatrix} i^* \\ j^* \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ k & k+1 \end{pmatrix} * \begin{pmatrix} i \\ j \end{pmatrix} \text{mod } m \quad (1)$$

Equation (1) indicates that each bit of the watermark at location  $(i, j)$  will be moved to a new location  $(i^*, j^*)$ . Parameter  $m$  is obtained from the size  $m*n$  of the watermark while parameter  $k$  is arbitrarily chosen by the user. Parameters  $m$  and  $k$  are secret keys needed for both the scrambling and reconstruction of the watermark. Even with TA permutation, the proposed algorithm is still blind.

### 3 Proposed Scheme

This paper proposed a novel strategy for DWT domain robust invisible embedding and extraction through a unique approach for creation of a compound color image to serve as the effective watermark. One of the most important features that make the recognition of images possible by humans is color. Color is a property that depends on the reflection of light to the eye and the processing of that information in the brain. Usually colors are defined in three dimensional color spaces These could be RGB (Red, Green, and Blue), HSV (Hue, Saturation, and Value) or HSB (Hue, Saturation, and Brightness). The last two are dependent on the human perception of hue, saturation, and brightness [22]. Color represents the distribution of colors within the entire image. This distribution includes the amounts of each color, but not the locations of colors. The entire process of the proposed method is represented in the Figure 2.

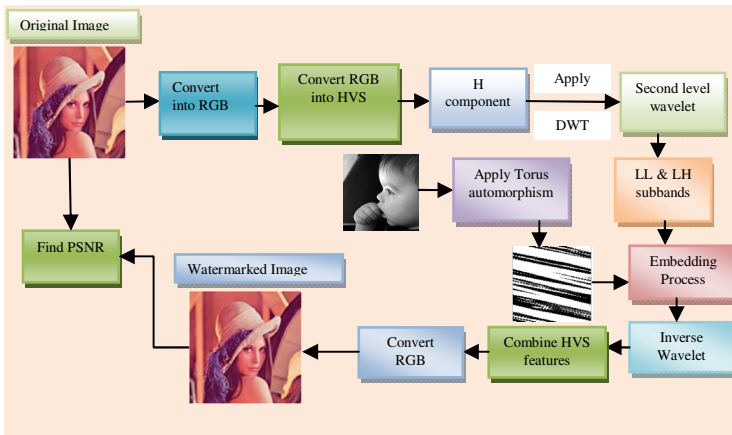


Fig. 2. Overview of Proposed Color Watermarking Scheme

#### 3.1 Embedding Watermark Procedure

The embedding algorithm uses color image as cover and gray-scale image as watermark. The color image is decomposed into Red, Green and Blue channels. RGB color model converted into HVS plane, the proposed method is considered H channel to insert the watermark image. The 2-level DWT is applied on the Hue channel of color image, which produces the frequency subband coefficients. From these  $LL_2$  and  $LH_2$  subband coefficients is selected for inserting watermark. Before embedding the watermark into selected subbands, the watermark image is scrambled by using Torus automorphism, as explained in section 2.1. The scramble watermark image is split into two shares, the first share insert into  $LH_2$  subband coefficients and second share insert into  $LL_2$  subband coefficients. Calculate the average value of  $LL_2$  and  $LH_2$ , and

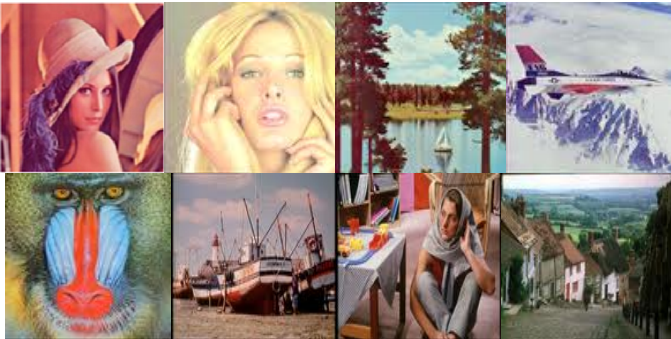
compare each and every pixel value with average value of  $LL_2$  and  $LH_2$ . If the value is less than average value, select the pixel to insert the watermark bit in the 6<sup>th</sup> LSB. Continue this procedure until the watermark bit inserted into the wavelet image. Apply inverse wavelet to obtain H component and combine HVS channels, then convert HVS into RGB to get watermarked image.

### 3.2 Extracting Watermark Procedure

Extraction procedure is a nature blind extraction which uses only watermarked color image as input. The watermarked color image is decomposed into Red, Green and Blue channels. RGB watermarked color model converted into Hue, Saturation, and Value channels. The DWT is applied on the Hue channel of watermarked color image, which produces the frequency subband coefficients. On this subband apply DWT to obtain the second level decomposition. The watermark is extracted from  $LL_2$  and  $LH_2$ , subband coefficient. The retrieved watermark can be used to determine the ownership by comparing the retrieved watermark with the assigned one. As in the definition, the goals of the reversible watermarking are to protect the copyrights and can recover the original image.

## 4 Experimental Results and Analysis

Eight 256×256 sized cover images Lena, Tiffany, Lake, F16, Baboon, Boat, Barbara, and Goodhill are considered in the proposed method, as shown in Figure 3. A binary image “Boy” of size 64×64 is used as the watermark image as shown in Figure 4. The outcome expose that there are no detectably visual degradations on the watermarked image presented in Figure 5 with a PSNR of 43. The proposed method solemnized that there are no visual degradations on the reverenced watermarked images. For all the different original test images, the watermark is successfully extracted with unit NCC.



**Fig. 3.** Cover images Lena, Tiffany, Lake, F16, Baboon, Boat, Barbara, Goodhill



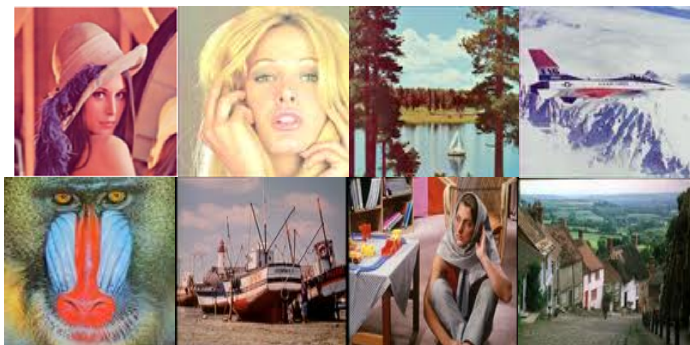
**Fig. 4.** Watermark images (a) Boy

We propose a new universal objective image quality index (IQI), which is easy to calculate and applicable to various image processing applications. Instead of using traditional error summation methods, the proposed index is designed by modeling any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Although the new index is mathematically defined and no human visual system model is explicitly employed, experiments on various image distortion types show that it exhibits surprising consistency with subjective quality measurement. It performs significantly better than the widely used distortion metric mean squared error. The structural similarity (SSIM) index is a method for measuring the similarity between two images. The SSIM index is a full reference metric, in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference. IQI and SSIM values of the proposed method is nearer to 1, this indicating the cover images and watermarked images are identical as shown in Table 1.

To provide verification of the capability of proposed method, test the proposed scheme watermarked images with Cropping (5%, 10%,15%) , (Gaussian Blur 1px, 2px, 3px), Noise(10%, 15%, 20%), Rotate (20,40,60). The Table 2 exposes the watermark recognition results of the different attack of the Barbara image. From the results, the proposed method is expert to absolutely detect the watermark in the watermarked images. Therefore, the proposed method can represent the conclusion that the proposed method can detect perfectly the watermark from the watermarked images and it is well built robustness to the geometric and non-geometric attacks.

**Table 1.** Values of various parameters for proposed method

Original images	Proposed approach			
	PSNR(dB)	IQI	SSIM	NCC
Lena	43.24	0.98	0.99	0.97
Tiffany	42.57	0.99	0.97	0.98
Lake	43.98	0.97	0.98	0.99
F16	42.45	0.98	0.99	0.98
Baboon	42.10	0.97	0.97	0.97
Boat	43.07	0.99	0.98	0.998
Barbara	42.76	0.98	0.99	0.98
Goodhill	42.72	0.99	0.97	0.97



**Fig. 5.** Watermarked images Lena, Tiffany, Lake, F16, Baboon, Boat, Barbara, Goodhill

**Table 2.** Attacks for proposed method

Attacks	PSNR	NCC
Cropping 5%	34.45	0.80
Cropping 10%	32.76	0.76
Cropping 15%	27.35	0.63
Gaussian Blur 1px	30.65	0.71
Gaussian Blur 2px	28.71	0.67
Gaussian Blur 3px	24.39	0.57
Noise 10%	31.39	0.73
Noise 15%	28.65	0.67
Noise 20%	26.10	0.61
Rotate 20	30.18	0.70
Rotate 40	26.38	0.61
Rotate 60	22.16	0.51

## 5 Conclusions

In this paper, a Color image watermarking using Wavelet Transform based on HVS Channel has been proposed. This proposed method applies DWT in the HVS channel and embed scrambled watermark into the DWT coefficient. The frequency domain technique are good for applications where exact watermark need to be extracted and channel do not consists any noise. In the proposed color digital watermarking method, the actual bits are scattered in the image in such a way that they cannot be identified by unauthorized persons and show resilience against attempts to remove the hidden data. The proposed method tested the method for several standard test images. The quantitative measure of the extracted watermark for both gray scale and color images shows the flexibility against different attacks.



## References

1. Cheng, L.M., Cheng, L.L., Chan, C.K., Ng, K.W.: Digital watermarking based on frequency random position insertion. presented at Control, Automation, Robotics and Vision Conference, vol. 2, pp. 977–982 (2004)
2. Chun-Shien, L., Shih-Kun, H., Chwen-Jye, S., Hong-Yuan Mark, L.: Cocktail watermarking for digital image protection. *IEEE Transactions on Multimedia* 2, 209–224 (2000)
3. Lu, W., Lu, H., Chung, F.-L.: Robust digital image watermarking based on subsampling. *Applied Mathematics and Computation* 181, 886–893 (2006)
4. Reddy, A.A., Chatterji, B.N.: A new wavelet based logo-watermarking scheme. *Pattern Recognition Letters* 26, 1019–1027 (2005)
5. Cox, I.J., Killian, J., Leighton, F.T., Shamoon, T.: Secure Spread Spectrum Watermarking for Multimedia. *IEEE Trans. on Image Processing* 6(12) (December 1997)
6. Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J., Su, J.: Attacks on digital watermarks: classification, estimation-based attacks and benchmarks. *IEEE Commun. Mag.* 39(9), 118–126 (2001)
7. Sequeira, A., Kundur, D.: Communications and information theory in watermarking: a survey. In: *Proceedings of SPIE Multimedia Systems and Application IV*, vol. 4518, pp. 216–227 (2001)
8. Mintzer, F., Braudaway, G., Yeung, M.: Effective and ineffective digital watermarks. In: *IEEE International Conference on Image Processing (ICIP 1997)*, vol. 3, pp. 9–12 (1997)
9. Memon, N., Wong, P.W.: Protecting digital media content. *Commun. ACM* 41(7), 35–43 (1998)
10. Barnett, R.: Digital watermarking: application, techniques, and challenges. *IEE Electron Commun. Eng. J.*, 173–183 (1999)
11. Bender, W., Butera, W., Gruhl, D., Hwang, R., Paiz, F.J., Pogreb, S.: Applications for data hiding. *IBM Syst. J.* 39(3-4), 547–568 (2000)
12. Zhao, J., Koch, E., Luo, C.: In business today and tomorrow. *Commun. ACM* 41(7), 67–72 (1998)
13. Huang, P.S., Chiang, C.S., Chang, C.P., Tu, T.M.: Robust spatial watermarking technique for colour images via direct saturation adjustment. In: *IEE Proceedings. of Vision, Image and Signal Processing*, vol. 152(5), pp. 561–574 (2005)
14. Verma, B., Jain, S., Agarwal, D.P., Phadikar, A.: A New color image watermarking scheme. *INFOCOMP Journal of Computer Science* 5(3), 37–42 (2006)
15. Xie, L., Arce, G.: Joint wavelet compression and authentication watermarking. In: *IEEE International Conference on Image Processing*, vol. 2, pp. 427–431 (1998)
16. Reddy, A.A., Chatterji, B.N.: A new wavelet based logo- watermarking scheme. *Pattern Recognition Letters* 26(7), 1019–1027 (2005)
17. Raval, M.S., Rege, P.P.: Discrete wavelet transform based multiple watermarking scheme. In: *Int. Conference on Convergent Technologies for Asia-Pacific Region*, vol. 3, pp. 935–938 (2003)
18. Anumol, T.J., Karthigaikumar, P.: DWT based Invisible Image Watermarking Algorithm for Color Images. In: *IJCA Special Issue on “Computational Science - New Dimensions & Perspectives” NCCSE*, pp. 76–79 (2011)
19. Voyatzis, G., Pitas, I.: Applications of Toral Automorphisms in Image Watermarking. In: *Proceedings of the Image Processing International Conference* (1996)

20. Chang, C., Hsiao, J., Chiang, C.: An Image Copyright Protection Scheme Based on Torus Automorphism. In: Proceedings of the 1st International Symposium on Cyber Worlds, pp. 217–224 (2002)
21. Engedy, M., Munaga, V., Saxena, A.: A Robust Wavelet Based Digital Watermarking Scheme Using Chaotic Mixing. In: Proceedings of the 1st International Conference on Digital Information Management, pp. 36–40 (2006)
22. Hsieh, S.-L., Tsai, I.-J., Huang, B.-Y., Jian, J.-J.: Protecting Copyrights of Color Images using a Watermarking Scheme Based on Secret Sharing and Wavelet Transform. *Journal of Multimedia* 3(4), 42–49 (2008)
23. Husaini, A.Z., Nizamuddin, M.: Challenges and approach for a robust image water marking algorithm. *International Journal of Electronics Engineering* 2(1), 229–233 (2010)

# Automatic Vehicle Number Plate Localization Using Symmetric Wavelets

V. HimaDeepthi, B. BalvinderSingh, and V. SrinivasaRao

Department of CSE, VR Siddhartha Engineering College, Vijayawada, India  
balvinder546@gmail.com

**Abstract.** Automatic number plate recognition (ANPR) plays a major role in real life applications and several techniques have been proposed. The localization or detection of the number plate of the vehicle images is the basis for any ANPR system. This paper proposes a robust method for localization of number plates in different conditions. There are two stages; first the preprocessing of the input image is performed and then localization is done. After preprocessing the statistical measures such as root mean square error and peak signal to noise ratio are calculated. Next the localization is done using symmetric wavelets and mathematical morphology. Experimental results show that this method gives dominant values of RMSE and PSNR. Experiments were performed on a database and also on a sample of 280 images of different countries taken from various scenes and conditions; results show that success rate of 77.14% on database and 92.14% on sample images achieved.

**Keywords:** Number plate localization (NPL), Preprocessing, Peak signal to noise ratio (PSNR), Root mean square error (RMSE), Symmetric wavelets.

## 1 Introduction

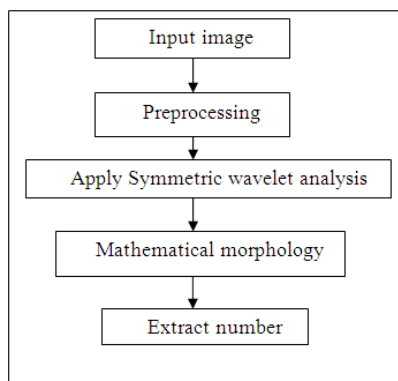
Automatic number plate recognition (ANPR) has a wide range of real-life applications such as automatic toll collection, traffic law enforcement, and road traffic monitoring [1]. Reading or locating the license number plate is the main and the first step in determining the identities of parties involved in the traffic incidents. The common aim of these applications is to reduce man power and facilitate to the automatic management. An efficient license plate localization system may become the core of fully computerized road traffic monitoring systems. Furthermore, an ANPR system can have two varieties: on-line ANPR system and off-line ANPR system. In an online system, the localization and interpretation of license plates take place instantaneously from the incoming video frames, enabling real-time tracking of moving vehicles through the surveillance camera. On the other hand, an offline system captures the vehicle images and stores them in a centralized data server for further processing, i.e. for interpretation of vehicle license plates. ANPRs are cameras mounted on stationary objects [2] (telephone poles, the underside of bridges, etc.) or on patrol cars. The cameras snap a photograph of every license plate that passes them by capturing information on up to thousands of vehicles per minute.

Some standards have been given for numbering vehicles [3]. The standard specifications of license plates are not same for all countries and may not same even for two countries because every country has their own numbering standards. Here the specifications for Indian license plates have mentioned. The numbering format [4] is as follows. AA 11 BB 1111, Where AA is the state code, 11 is the two digit district code, 1111 is the unique number and BB is the optional alphabets if the 9999 numbers are used up. Most of previous license plate detection algorithms are restricted to work under certain conditions, such as fixed backgrounds [5], known color [6], or designated ranges of the distance between camera and vehicle [7], [8]. The quality of the input image also plays major role in the overall process of recognition. Most of the researchers used median filter [9], [10], [11], histogram equalization [12], [13] techniques as preprocessing.

In this paper, we propose a novel method for preprocessing. First the input image is converted to gray version and then the process of correlation takes place using a mask. Next, the statistical measures such as RMSE and PSNR are calculated. Experimental results give dominant values such as low RMSE and high PSNR when compared to existing preprocessing methods. The localization is done by combination of symmetric wavelets (symlets) and morphology. The rest of the paper is structured as follows. Section 2 describes the proposed technique. Section 3 shows the experimental results. Finally this paper is concluded in Section 4.

## 2 Proposed Technique

The proposed technique consists of two modules. The preprocessing and number plate localization. Fig. 1 shows the flow chart of our proposed method. Input images are processed by the preprocessing module. This improves the overall success rate of the localization or detection of number plate.



**Fig. 1.** The flow chart of proposed number plate localization technique

## 2.1 Preprocessing

Initially the input image is converted to gray scale. Next zero padding is done for the image. The purpose of zero padding is to avoid wrap around error which gives rise to distortion around the edges and to get out of poor pixelizations. A kernel of 3x3 size is chosen to perform correlation with the image. Correlation involves calculating the weighted sum of neighborhood pixels. The weights are taken from kernel. Each value from the neighborhood of pixels is multiplied by the same pixel position of the kernel. The correlation is defined as,

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x + s, y + t) \quad (1)$$

where  $w(s, t)$  is the kernel of size  $m \times n$ ,  $f$  is the input image,  $g$  is the resultant image and  $a=(m-1)/2$ ,  $b=(n-1)/2$ . The kernel selection is important because correlation is performed using the kernel with the input image. The following is the kernel that we choose here. We selected this kernel after conducting several trials and is suitable for acquiring dominant values when calculating statistical measures. Next step that we done is subtracting this image from the input gray version. This gives us the preprocessed image. Fig. 2 shows the results of preprocessing. Next the assessment of quality is performed.

$$w(s, t) = \begin{bmatrix} 1 & -2 & -1 \\ -2 & 4 & -2 \\ 1 & -2 & -1 \end{bmatrix}$$

There are two types of image quality measurements [14]. One is subjective analysis and other is objective analysis. The former is concerned with how image is perceived by a viewer and designates his or her notion on it. The human eyes extract structural information from the view field, so the human visual system (HVS) is highly adapted for this purpose. Disadvantages of this type of quality assessment are time consuming and cannot be performed in real time. The latter is concerned with algorithmic evaluation. Here we performed objective analysis for measuring the quality of the preprocessed image. The most commonly used statistical measure for objective analysis is PSNR. Here we calculated PSNR and RMSE for original image and processed image. Given a input image  $f$  and test image  $g$ , both of size  $M \times N$ , the peak signal to noise ratio (PSNR) can be defined as

$$PSNR(f, g) = 10 \log_{10} (\max^2 / MSE(f, g)) \quad (2)$$

where  $\max$  is maximum pixel value of the image and MSE is mean squared error and is defined as

$$MSE(f, g) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f_{ij} - g_{ij})^2 \quad (3)$$

The root mean square error (RMSE) is the square root of the MSE. In general, low RMSE and high PSNR values indicate that the reconstructed or preprocessed image has high quality.

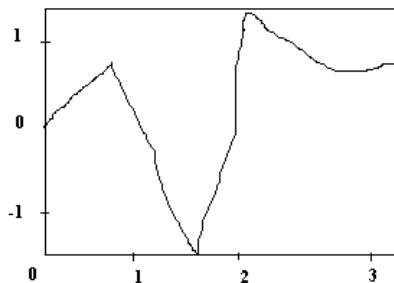


**Fig. 2.** The original input image (a) and the preprocessed image (b)

Here from Fig. 2 we observed that there is not much difference between original image and preprocessed image. The reason is that the human eye cannot distinguish original and reconstructed images when the PSNR is 40 dB or greater [15]. After calculating these objective measures the existed preprocessing filters such as median filter, wiener filter and lee filter were applied. The purpose of choosing these filters is that these reduce salt and pepper, Gaussian and speckle noises respectively. From the experimental results we observed that the proposed method reduces the random noise and gives dominant values.

## 2.2 Number Plate Localization

There are two major components constituting the localization module i.e., wavelet analysis and morphological processing. Here the type of wavelet that we choose is symlets and the morphological operations are combination of dilation and erosion. Finally, the properties of image are measured and the number plate is extracted.



**Fig. 3.** Wave function of symlet of order 2 (sym2)

**Wavelet Analysis.** Wavelet transforms have become a powerful transformation in various image processing applications since they allow both time and frequency analysis concurrently. Wavelet transforms are based on wavelets which are varying frequency and limited duration [16]. There are different types of wavelets, one among them is discrete wavelets and the transformation is discrete wavelet transform (DWT). Symlets are a type of discrete wavelets and are modified version of daubechies wavelets. These are also known as least-asymmetric wavelets [17]. There are seven types of symlets from sym2 to sym8. We used sym2 function in this experiment that gives best results.

The symlets having the characteristic of least asymmetry and highest number of vanishing moments for a given support width and the associated filters are linear phase filters. The advantage of symlets over others is least asymmetry and increased symmetry, which is useful to avoid dephasing while processing images. The procedure of applying symlets is as follows. First the preprocessed image is decomposed into four components by applying single level 2D-DWT with respect to symlet. The four components are such as approximation, horizontal, vertical and diagonal respectively. After the reconstruction takes place directly from 2D wavelet coefficients.

**Table 1.** Comparison of objective measures after preprocessing

Technique Used	RMSE	PSNR
Median filter	3.05	38.44dB
Wiener filter	3.31	37.75dB
Lee filter	4.79	34.52dB
Proposed	0.40	56.17dB

Now the edge detection is applied on these reconstructed components to get important features. Here we used sobel edge detection because it is computationally inexpensive and robust to noise. From these, after conducting several attempts, we observed that the vertical component is suitable for extracting number plate location.

**Morphological Processing.** Mathematical morphology refers to wide range of image processing operations that process images based on shapes [18]. There are several operations of morphology but we used only dilation and erosion which are basic for any morphological operations. First the edge detected vertical reconstructed image is undergone for dilation and then erosion is applied on it. This is also known as morphological closing operation. The following operations form the basis of morphology.

$$(F \oplus B)(x, y) = \max \{F(x - s, y - t) + B(s, t)\} \quad (4)$$

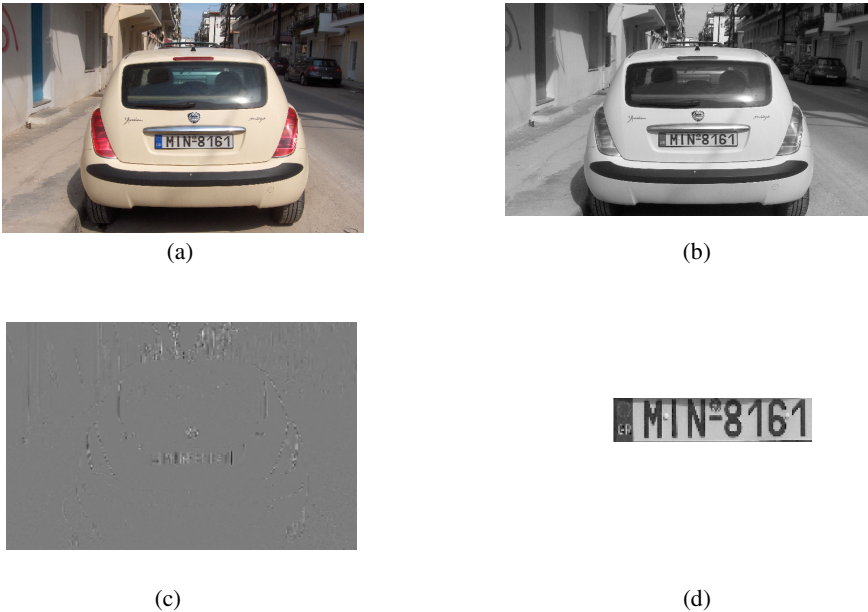
$$(F \ominus B)(x, y) = \min \{F(x + s, y + t) - B(s, t)\} \quad (5)$$

$$F \bullet B = (F \oplus B) \ominus B \quad (6)$$

where  $F(x,y)$  is image and  $B(x,y)$  is structuring element. The equations show dilation, erosion and closing respectively. The structuring element is a characteristic of certain shape and is used to carry other image processing operations [19]. The shape and size of structuring element plays a major role in morphological processing. Here we choose rectangular structural element. The dilation followed by erosion i.e., closing operation is applied using the same structuring element. Now the properties such as area and bounding box are measured. Based on these the number plate region is extracted. This results in the localization of the number plate from the input image.

### 3 Experimental Results

Two sets of images have been used for our experiments. The first set contains still images taken from Greek database [20]. This set contains images of different types such as day color, day with shadows, various angle of vision etc. The second set contains a total of 280 images of different sizes and countries such as India, England, America, Japan and Malaysia. These images are taken from various scenes such as streets, roadside, parking lots, day light, night time and dirty. Experiments were performed using Matlab 7.9 for assessing the quality of the proposed method. Fig. 4 shows the process of localization of number plate from car images.



**Fig. 4.** Process of obtaining a valid number plate region: (a) original vehicle image, (b) preprocessed image, (c) vertical detailed image, and (d) localized number plate



In Fig. 4 (a) shows the input image, (b) shows the preprocessed image, (c) shows the vertical detailed image from the decomposed version using DWT of type symlets of order 2, and (d) shows the localized number plate from the given image which is the desired region. From the first set of Greek database the success rate achieved is 78.18%. A total of 275 images were tested and 215 of these were successfully locates the number plate and the rest 60 images were failed to correctly locate the number plates. From the second set of images 22 were failed to detect the plates. The failure of detection shown in Fig. 5 is mainly due to the fact that the images were too complex in which the license plate is partially covered by shadow (a), (b) very poor lighting condition, (c) text that exists around the plate looks like a plate region, and (d) image captured from long distance and multiple cars exist.



**Fig. 5.** Examples of failures to localize number plates: (a) number plate partially covered by shadow, (b) poor lighting condition, (c) text around the plate region, and (d) long distance captured image with multiple cars

## 4 Conclusion

Compared to most of the preceded techniques that in some way restricted to localize number plates, the techniques presented in this paper are less restrictive. The proposed method consists of two modules, one for preprocessing and the other for locating number plates. The strength of the method depends on preprocessing module that gives dominant values of objective measures. The localization module using symlets and morphology produces better results after preprocessing input images. The most crucial step that we take up is using a database for assessing quality of our technique. Afterwards we used sample images of different countries for our experimental evaluation. Hence the method is not restricted to specific country and can be applicable for any country.

## References

1. Zhang, J., He, X.: A Fast Algorithm for License Plate Detection in VariouConditions. In: IEEE International Conference on Systems, Man, and Cybernetics (2006)
2. Cousineau, M.: The Global War on Terror and Automatic LicensePlateRecognition. Canadian Review of Sociology 50(1) (February 2013)
3. Singh, B.B., Deepthi, H.V.: Survey on Automatic Vehicle Number Plate Localization. International Journal of Computer Applications 67(23), 7–12 (2013)
4. Kaur, A., Jindal, S., Jindal, R.: License Plate Recognition using Support Vector Machine (SVM). International Journal of Advanced Research in Computer Science and Software Engineering 2(7) (July 2012)
5. Bai, H., Zhu, J., Liu, C.: A Fast License Plate Extraction Method on Complex Background. In: Proceedings of IEEE International Conference on Intelligent Transportation Systems, vol. 2, pp. 985–987 (2003)
6. Kim, S.K., Kim, D.W., Kim, H.J.: A Recognition of Vehicle License Plate Using a Genetic Algorithm Based Segmentation. In: Proceedings of International Conference on Image Processing, pp. 661–664 (1996)
7. Kim, S.: A Robust License-plate Extraction Method under Complex Image Conditions. In: Proceedings of 16th International Conference on Pattern Recognition, pp. 216–219 (2002)
8. Jia, W., et al.: Mean Shift for Accurate License Plate Localization. In: Proceedings of International Conference on Intelligent Transportation Systems, pp. 566–571 (2005)
9. Ssuri, P.K., Walia, E.: Vehicle Number Plate Detection using Sobel Edge Detection Technique. IJCST 1(2) (December 2010)
10. Pandya, P., Singh, M.: Morphology Based Approach To Recognize NumberPlates in India. International Journal of Soft Computing and Engineering (IJSCE) 1(3) (July 2011)
11. Jin, L., Xian, H., Bie, J., Sun, Y., Hou, H.: License Plate Recognition Algorithm for Passenger Cars in Chinese Residential Areas. Sensors (2012)
12. Sulehria, H.K., Zhang, Y., Irfan, D.: Mathematical Morphology Methodology for Extraction of Vehicle Number Plates. International Journal of Computers (2007)
13. PeterTarabek, A.: Real-Time License Plate Localization Method Based on Vertical Edge Analysis. In: Proceedings of the Federated Conference on Computer Science and Information Systems, pp. 149–154 (2012)
14. Sankara Ganesh, J., Srinivasa Rao, V., Srinivas, K.: Enhanced Noise Type Recognition Using Statistical Measures. IOSRJCE 2 (2012)
15. Kaleka, J.S., Sharma, R.: Comparative Performance Analysis of Haar, Symlets and Bior wavelets on Image compression using Discrete Wavelet Transform. International Journal of Computers & Distributed Systems 1 (August 2012)
16. Prasad, L., Iyengar, S.S.: Wavelet Analysis with Applications to Image Processing, pp. 101–115. CRC Press LLC, Boca Raton (1997)
17. Seyedzade, S.M., Mirzakuchaki, S., Tahmasbi, A.: Using Symlet Decomposition Method, Fuzzy Integral and Fisherface Algorithm for Face Recognition. In: 2nd International Conference on Computer Engineering and Applications, vol. 2, pp. 83–88 (2010)
18. Jyothirmai, M.S.V., Srinivas, K., Srinivasa Rao, V.: Enhancing shadow area using RGB color space. IOSR Journal of Computer Engineering 2 (August 2012)
19. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice Hall, Englewood Cliffs (2007)
20. Images Database,  
<http://www.medialab.ntua.gr//research/LPRdatabase.html>  
 (last date accessed: July 2013)

# Response Time Comparison in Multi Protocol Label Switching Network Using Ant Colony Optimization Algorithm

E.R. Naganathan<sup>1</sup>, S. Rajagopalan<sup>2</sup>, and S. Narayanan<sup>2</sup>

<sup>1</sup> Dept. of Computer Science,  
Hindustan University, Chennai, India  
naganathaner@hotmail.com

<sup>2</sup> Dept. of CSE, Alagappa University,  
Karaikudi, India

raja\_04@hotmail.com, narayanan\_alu@yahoo.com

**Abstract.** Multi-Protocol Label Switching (MPLS) networks transfers the data with the help of labels. MPLS creates "virtual links" between distant nodes. MPLS can encapsulate packets of various network protocols. In MPLS packet-forwarding decisions are made solely on the contents of this label, without the need to examine the packet itself. This allows to create end-to-end circuits using any protocol. Congestion Control and Congestion avoidance is the main task in Traffic Engineering. Slow Start, ECN, RED, AIMD are some of the techniques available for congestion management. This paper compares the response time in MPLS network using Ant Colony Optimization (ACO) technique to avoid congestion and gives good results in terms of response time.

**Keywords:** Ant Colony Optimization, MPLS Network, Traffic Management.

## 1 Introduction

MPLS operates at an OSI Model and is generally considered to lie between Layer 2 and Layer 3. It is often referred to as a "Layer 2.5" protocol. It provides a unified data-carrying service for both circuit-based clients and packet-switching clients. It can be used to carry many different kinds of traffic, including IP packets, as well as native ATM, SONET, and Ethernet frames. Initially MPLS is developed by IPSILON Networks. But it is defined to work over ATM. Cisco Systems, Inc., developed another related scheme and it is not restricted to ATM which is called "Tag Switching". This is given to IETF for standardization purpose. The IETF develops a protocol by including features from several other vendors' work. MPLS have the ability to support multiple service models and perform traffic management, and it also supports robust recovery framework.

The paper is organized as follows: The section 2 explains about MPLS, section 3 explains traffic engineering in MPLS. Section 4 explains Ant Colony Optimization

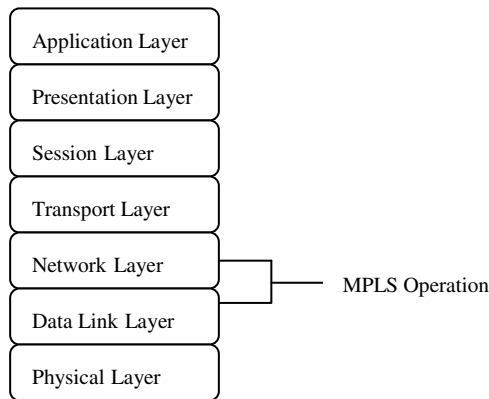
and section 5 gives the ACO algorithm. Section 6 describes the proposed work. Results are given in section 7. Section 8 concludes the paper.

## 2 MPLS

The operation of MPLS in OSI model is shown in figure 1 above. MPLS prefix the packets with MPLS header. MPLS header can have one or more "labels" which is called as label stack. Each label stack entry contains four fields:

- A 20-bit label value.
- a 3-bit Traffic Class field
- a 1-bit bottom of stack flag.
- an 8-bit TTL (time to live) field.

These labeled packets are forwarded by reading the label instead of reading the entire IP table. So the label lookup and label switching were faster than a routing table or RIB (Routing Information Base) lookup. The Label Edge Routers (LER) is placed in the periphery that is at the entry and exit points of an MPLS network. These LER push a label onto an incoming packet (Ingress LER) and pop it off (Egress LER) from the outgoing packets. The Label Switch Router (LSR) that performs the forwarding of packets according to the label. Labels are distributed using the "Label Distribution Protocol" (LDP) and the network operator can establish the Label Switch Paths (LSPs) for various purposes. These LSPs can be used to create network-based IP or the traffic can be routed through the specific paths in the network.



**Fig. 1.** Operation of MPLS in OSI Layer

## 3 Traffic Engineering in MPLS

Traffic engineering is the method of optimizing the performance of a network by analyzing, predicting and regulating the behavior of data transmitted in a network.

When the level of traffic in a network reaches or exceeds the capacity of the network then the network is said to be congested. In a congested network, one of three things can happen. For example, when a subscriber attempts to send a message or place a call:

- The user receives a busy signal
- A message is placed in a queue and is delivered according to the parameters.
- A message is rejected, returned or lost.

When message queues become inappropriately long or the frequency of busy signals becomes unacceptably high, the network is in congestion. Minimization or eliminating the congestion is the major objective of traffic engineering. Chun Tung Chou [1] proposed a virtual private network architecture, which allows granularity and load balancing. A distortion factor for heterogeneous streams in traffic engineering of MPLS backbone networks was introduced by Shekhar et al [2].

MPLS hierarchical architecture for label-switched networks is explained in [3] and can be used to address all required functions of converged/unified networks, from initial IP level authentication, configuration, security, admission control, to quality of service and policy management.

Bosco et al [4] analyses the performance of a traffic engineering (TE) strategy for MPLS based network. The implementation based on a distributed control plane is investigated and realized by means of a test bed where real signaling protocol (RSVP-TE) and routing protocols (OSPF-TE) have been implemented.

## 4 Ant Colony Optimization

The swarm intelligence – Ant colony is used for optimum congestion control. Ant colony algorithms [6], [7] have been inspired by the behavior of the real ant colony. In real ant colony, the ants search their surroundings for food where as in Ant colony algorithm; the artificial ants search the solution space to obtain the optimum solution. The strength of the pheromone deposit directs the artificial ants toward the best paths and the pheromone evaporation allows the system to forget the old information and thereby avoiding quick convergence to suboptimal solutions. ACO has been applied successfully to discrete optimization problems such as the traveling salesman problem [8], routing [9], and load balancing [10]. A number of proofs for the convergence to the optimum path of the ACO can be found in [11] and [12]. The implementation of the proposed system [13] [14] in the wired environment which provides optimum result and suggested traffic free routing.

## 5 ACO Algorithm

### 1) //Initialization Phase

For each pair (r, s), the value of  $\tau(r, s) := \tau_0$  End-for

For k := 1 to m do

Let (r, k1) be the starting node for an ant k

```

    Jk(rk1) := {1, ..., n} - rk1
    // Jk(rk1) is the set of nodes yet to be visited for //ant k in node rk1
    rk := rk1
    // rk is the node where ant k is located
    End-for
2) //Ants build their tour in this phase
    For i := 1 to n do
        If i < n Then
            For k := 1 to m do
                Choose the next node Sk
                Jk(Sk) := Jk(rk) - Sk
                Tourk(i) := (rk, Sk)
            End-for
        Else
            For k := 1 to m do
                //All ants go to initial node rk1 in this phase
                Sk := rk1
                Tourk (i) := (rk , Sk )
            End-for
        End-if
        //pheromone is updated in this phase
        For k := 1 to m do
             $T(r, s) \leftarrow (1 - \alpha) \cdot T(r, s) + \sum (1 - \alpha) \cdot T(r, s)$ 
            rk := sk // New node for ant k
        End-for
    End-for
    //In this phase global updating occurs
    For k := 1 to m do
        Compute Lk // Lk is the length of the tour done by ant k
    End-for
    Compute Lbest
    /*Update edges belonging to Lbest
    For each edge (r, s)
         $T(r, s) \leftarrow (1 - \alpha) \cdot T(r, s) + \sum (1 - \alpha) \cdot T(r, s)$ 
    End-for
3) //In this phase priority is assigned
    For k := 1 to m do
        Sort the routing table based on pheromone values
        Assigns high priority to higher pheromone density path
        Choose the best path based on priority
    End-for

```

In our previous work [15] the above algorithm is implemented to avoid congestion in MPLS network. It gives good results in terms of packet loss.

## 6 Proposed Work

The above ACO algorithm is used to calculate the response time in MPLS network. The ACO algorithm is implemented in Network Simulator (NS2). All the necessary information like nodes, links etc are written in OTcl script. The response time is calculated using the output trace file. The normal load is taken as 20% of the nodes are communicating. The Medium load is taken as 50% of the nodes are communicating and in heavy load all the nodes are communicating. Number of nodes is designed as 8, 20, 50, 75, 100 and 200. The response time in ACO is compared with OSPF and RIP.

## 7 Result

The tables 1 to 4 shows the response time in Normal load, Medium Load and Heavy load respectively. The Figures 2 to 4 shows the graphical representation of the response time in Normal load, Medium load and Heavy load respectively.

**Table 1.** Response Time in Normal load (in ms)

No of Nodes	OSPF	RIP	ACO
	54	43	41
20	67	53	51
50	82	66	62
75	102	81	77
100	125	100	95
200	155	123	118

**Table 2.** Response Time in Mediumload (in ms)

No of Nodes	OSPF	RIP	ACO
8	67	53	51
20	82	66	62
50	102	81	77
75	125	100	95
100	155	123	118
200	191	152	145

**Table 3.** Response Time in Heavy load (in ms)

No of Nodes	OSPF	RIP	ACO
8	96	82	65
20	118	102	81
50	146	125	100
75	180	155	123
100	223	191	152
200	275	236	188

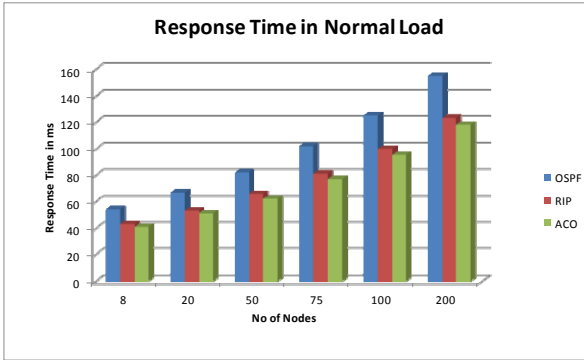


Fig. 2. Response Time in Normal Load

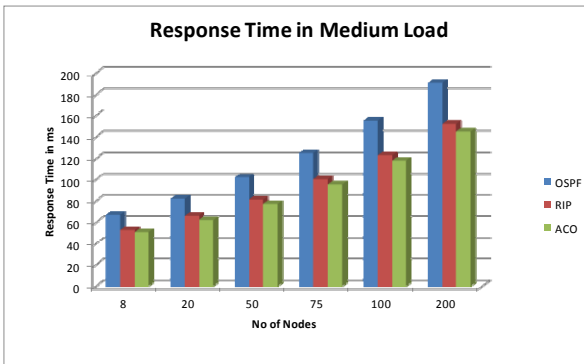


Fig. 3. Response Time in Medium Load

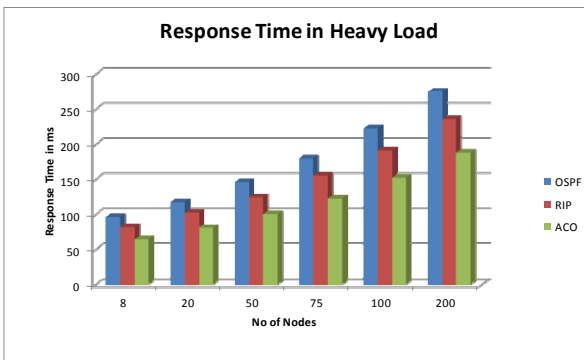


Fig. 4. Response Time in Heavy Load



## 8 Conclusion

The ACO algorithm is implemented using NS2. The number of nodes and the load conditions are varied and the response time is taken. The response time of ACO is compared with OSPF and RIP. The ACO gives good result in Normal, Medium and Heavy load conditions.

## References

1. Chou, C.T.: Traffic engineering for MPLS-based virtual private networks. *Computer Networks* 44, 319–333 (2004)
2. Srivastava, S., van de Liefvoort, A., Medhi, D.: Traffic engineering of MPLS backbone networks in the presence of heterogeneous streams. *Computer Networks* 53, 2688–2702 (2009)
3. Palmieri, F.: An MPLS-based architecture for scalable QoS and traffic engineering in converged multiservice mobile IP networks. *Computer Networks* 47, 257–269 (2005)
4. Boscoa, A., Bottab, A., Conteb, G., Iovannaa, P., Sabellaa, R., Salsanoc, S.: Internet like control for MPLS based traffic engineering: performance evaluation. *Performance Evaluation* 59, 121–136 (2005)
5. Iovanna, P., Sabella, R., Settembre, M.: Traffic engineering strategy for multi-layer networks based on the GMPLS paradigm. *IEEE Netw.* 17(2), 28–37 (2003)
6. Di Caro, G., Dorigo, M.: AntNet: A Mobile Agents Approach to Adaptive Routing. Tech. Rep. IRIDIA/97-12, Univ. Libre de Bruxelles, Brussels, Belgium (1997)
7. Schoonderwoerd, R., Holland, O., Bruten, J.: Ant like agents for load balancing in telecommunication networks. In: *Proceedings of the First Int. Conf. on Autonomous Agents*, pp. 209–216. ACM Press, New York (1997)
8. Duan, H., Yu, X.: Hybrid Ant Colony Optimization Using Memetic Algorithm for Traveling Salesman Problem. In: *Proceedings of the IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning*, pp. 92–95 (2007)
9. Subramanian, D., Druschel, P., Chen, J.: Ants and reinforcement learning: A case study in routing in dynamic networks. In: *Proceedings of the 15th Int. Joint Conf. on Artificial Intelligence*, pp. 823–838. Morgan Kaufmann, San Francisco (1997)
10. Sim, K.M., Sun, W.H.: Ant Colony Optimization for Routing and Load-Balancing: Survey and New Directions. *IEEE Transactions on Systems, Man, and Cybernetics* 33(5), 560–572 (2003)
11. Xing, L.-N., Chen, Y.-W., Wang, P., Zhao, Q.-S., Xiong, J.: A Knowledge-Based Ant Colony Optimization for Flexible Job Shop Scheduling Problems. *Applied Soft Computing* 10, 888–896 (2010)
12. Lopez-Ibanez, M., Blum, C.: Beam ACO for the traveling sales man problem with time windows. *Computers & Operations Research* 37, 1570–1583 (2010)
13. Chandra Mohan, B., Sandeep, R., Sridharan, D.: A Data Mining Approach for Predicting Reliable Path for Congestion Free Routing Using Self-motivated Neural Network. In: Lee, R. (ed.) *Soft. Eng., Arti. Intel., Net. & Para./Distri. Comp.*, vol. 149, pp. 237–246. Springer, Heidelberg (2008)

14. Chandra Mohan, B., Baskaran, R.: Redundant Link Avoidance Algorithm for improving Network Efficiency. *International Journal of Computer Science Issues* 7(3) (May 2010)
15. Rajagopalan, S., Naganathan, E.R., Herbert Raj, P.: Ant Colony Optimization Based Congestion Control Algorithm for MPLS Network. In: Mantri, A., Nandi, S., Kumar, G., Kumar, S. (eds.) *HPAGC 2011. CCIS*, vol. 169, pp. 214–223. Springer, Heidelberg (2011)

# Software Effort Estimation Using Data Mining Techniques

Tirimula Rao Benala<sup>1</sup>, Rajib Mall<sup>2</sup>, P. Srikavya<sup>3</sup>, and M. Vani HariPriya<sup>3</sup>

<sup>1</sup>Department of Information Technology,  
JNTUK, University College of Engineering, Vizianagaram-535003  
b.tirimula@gmail.com

<sup>2</sup>Department of Computer Science and Engineering,  
Indian Institute of Technology, Kharagpur  
rajib@cse.iitkgp.ernet.in

<sup>3</sup>Department of Computer Science and Engineering,  
Anil Neerukonda Institute of Technology and Sciences,  
Sangivalasa-531162, Visakhapatnam, Andhra Pradesh

**Abstract.** This paper describes an empirical study undertaken to investigate the quantitative aspects of application of data mining techniques to build models for Software effort estimation. The techniques chosen are Multi linear regression, Logistic regression and CART. Empirical evaluation using three fold cross validation procedure has been carried out using three bench marking datasets of software projects, namely, Nasa93, Cocomo81, and Bailey Basili. We observed that: (1) CART technique is suitable for Nasa93 and Nasa93\_5. (2). Multiple Linear Regression is suitable for Nasa93\_2, Cocomo81s, Cocomo81o and Basili Bailey. (3). Logistic Regression is suitable for Nasa93\_1, Cocomo81 and Cocomo81e. It is concluded that data mining techniques tend to help estimating in the best way possible as they are objective and are applicable to unlimited sets of data.

**Keywords:** Software effort estimation, CART, Logistic regression.

## 1 Introduction

Software cost estimation is a complex process where the quality of outputs heavily depends on the quality of inputs. As the internet revolution is going through high paced technological change, the shorter software project development life cycle has been introduced by IT industries to make revenues. Estimating the time and effort – which represents the output- needed to develop an IT project is one of the most challenging task faced by the project managers. Under estimate leads to cost overruns and over estimate leads to missed financial opportunities.

Data mining is the process of extracting large amounts of data from many databases and find interesting patterns among them. Many data mining techniques are available for this purpose. They play a very important role in software effort estimation. Some of the data mining techniques are Regression techniques which can

be classified as linear and non linear regression techniques respectively, ordinary least squares regression, linear regression, multiple linear regression, logistic regression, Tree based techniques like CART and M5 [3].

The techniques used in our methodology are multiple linear regression, logistic regression and CART. In multiple linear regression technique we analyze the relation between dependent and independent variables using the model equation. In this we have multiple independent variables. In logistic regression we use a probabilistic function for finding outcome of the dependent variable based on independent variables. In CART technique classification and regression analysis is done to predict the outcome. Initially a classification tree is built using the decision tree induction method and training is done and then the rules are extracted from it. These rules are applied on the testing data to obtain the result of estimation.

Fundamentals of software cost estimation, data mining techniques and data preprocessing technique are briefly reviewed in Section 2. The proposed approach is described in Section 3. In Section 4, numerical examples from Cocomo81 (Coc81), Nasa93, Bailey and Basili dataset is used to evaluate the performance. Section 5 concludes this paper.

## 2 Background

In this section we will discuss briefly fundamentals of software cost estimation and the different data mining techniques used in our study.

### 2.1 Software Cost Estimation

In the field of software cost estimation, the effort required to develop a new software project is estimated by taking into account the factors that influence the development of the new project. The specific project is then compared to a historical data set -a set of successfully executed past projects- containing the measurements of relevant metrics and associated development effort. There are many ways of estimating software effort. Some of the methods prominent are Expert judgment, analogical estimations, Functional link artificial neural network based software cost estimation (FBE), data mining techniques, algorithmic methods (Formal models) and Top down and Bottom up methods. In expert judgment the decision from many experts is taken. Each one is provided with a specification and they give their estimation. The final estimation is obtained by iterating over each estimation. This method helps in finding requirements of proposed project by using the past experience but, it's not a perfect estimation as it's just the expectation of experts. In analogical estimations we estimate using actual characteristic data. In Top down and Bottom up methods there are problems like justification of data and time consumption respectively. In algorithmic methods like COCOMO model modification becomes easy and past experience can be used. But, inputs are poor as they are applicable to unlimited data and are subjective. [3, 7, 8]

## 2.2 Data Mining Techniques

Data mining is the process of extracting data from various sources. Many data mining techniques are used for this purpose. In our study we implemented multiple linear regression, logistic regression and CART. Multiple linear regressions [6] technique helps in finding out the relation between dependent variable and the independent variable. This is the extension of linear regression as it deals with multiple independent variables. To study a given model we first built the model equation.

$$y_i = \beta_0 + \beta_1 x_i + e_i, i=1, 2, 3 \dots n \quad (1)$$

Where  $y_i$  is the new dependent value,  $\beta_0, \beta_1$  are the coefficients and  $x_i$  is the independent variable. So, in this way we initially train the data and obtain the model equation for respective data and then apply testing data to it. Error value is estimated by the difference between original  $y_i$  and new  $y_i$ . Logistic regression [1, 2] technique helps in finding out the relation between dependent variable and the independent variable. This is the extension of linear regression as it deals with multiple independent variables. To study a given model we first built the model equation,

$$y = 1/(1 + e^{-\beta x}) \quad (2)$$

Where  $y$  is the new dependent value,  $\beta$  is the coefficient and  $x$  is the independent value. In this way the data is trained and tested alternatively using the model. CART (Classification and regression trees) [5] is a tree based machine learning technique. In this technique we use the independent variables and construct a tree in which there are leaf nodes. They either represent the category to which the data belongs to or the *data values*. *Classification trees are built in case of class categorization and Regression trees are built in case of values*. In our study we trained data initially using the decision tree and then using that tree the rules are obtained. These rules are applied for testing of data.

## 2.3 Data Preprocessing Technique

In many instances data collected are raw in nature. It may have different scales. To achieve high efficiency we need to pre-process the raw dataset. In our investigated datasets the features have different scales. In order to bring these to a common scale we have used Min Max Normalization Preprocessing technique. Min Max Normalization is used to bring the data values in the range of [0, 1]. Lowest value is assigned 0 and highest value is assigned 1.

## 3 Framework

Three data mining techniques are applied on the datasets where 60% of the dataset is sent for training, 20% of the dataset kept for validation and 20% used as testing data. The model is developed using training data. The test data is applied to check the accuracy of the model. This approach has been widely used in practice and detailed by algorithmic framework depicted in Fig. 1.

Let us assume D is the dataset and T is the technique chosen. N is the min max normalization technique. Min is the minimum value and max is the maximum value. V represents the value in D. M is the model equation.

Step1: Choose D and apply N i.e.

Step 1.1: for each column in D,  $v = (v - \min) / (\max - \min)$

Step 2: Apply T to D i.e.

Step 2.1: Divide D into 3 parts D1 (Training), D2 (Validation) and D3 (Testing) with D1 has 60% in D and D2 has 20% in D and D3 having 20% in D.

Step 2.2: Apply T to D1 and obtain M and get training error

Step 2.3: Apply D2 to M and get validation error

Step 2.4: while(validation error > 0.5)

```
{
    Repeat steps 2.2 and 2.3
}
```

Step 2.5: Apply D3 to M and get testing error.

Step 3: Calculate MRE, MMRE PRED.

Step 4: Compare results

**Fig. 1.** Framework

### 3.1 Performance Metrics

To estimate accuracy of the techniques used the evaluation measures considered are Mean Magnitude of Relative Error (MMRE), Median Magnitude of Relative Error (MdmRE) and Prediction (PRED (0.25)) [7, 8].

## 4 Datasets, Experiments and Results

In this section the data sets used for evaluating the performance of our methods are Nasa93 and its variants, Cocomo81 and its variants and Basili bailey.

### 4.1 Dataset Preparation

Initially the data set is preprocessed as the values in the data set are of different scales. The data is normalized in the interval of [0, 1] using the Min-Max normalization. Then the data is divided into three parts. The first two parts are for training and validation and the rest is for testing.

### 4.2 Effort Estimation Methods

As explained earlier the techniques are applied on the datasets. To train multiple linear regression models we train the data and obtain the model equation then get the

testing results. Similarly process is followed for Logistic regression. They aim in developing a model and obtaining the result out of it. CART technique is applied by training data and building a model out of it and the rules obtained are applied to test data. In this instead of model equation a tree is built.

### 4.3 Experiment Procedure

At first data preprocessing is done using min max normalization. Then division of data into training and testing sets is done. Then 3-fold cross validation [4] is done where 2 folds represent the training data and validation data and 1 fold represents testing data. The training dataset is applied to the data mining techniques and model is build. The validation datasets are applied to the model. If the validation error is less than 0.5 then testing process is initiated otherwise training process is again performed the rotation of the folds is performed till all the data sets are tested. Then using the evaluation measures the results are compared. The analysis and results are represented in the next section.

### 4.4 Experiment Results

For our simulation we have chosen 3 data sets from promise repository to investigate the performances of 3 techniques discussed earlier, we have adopted 3-fold cross validation approach. In this approach a data set is divided into three parts of which first 2 parts are taken for the training and the last part is used for the testing purpose. Similarly, the part which is used for the testing is next used for the training. In this way each and every data set part is investigated for training and testing of our techniques. To compare the performances of the three techniques we have chosen MMRE, MDMRE, PRED measures. MMRE is the average of MRE'S over training and testing data sets. However, the other two measures are median and prediction. The lower the values of MMRE and MDMRE for an algorithm or technique the better the technique is. However the higher the PRED (0.25) the better is the technique.

We have presented 9 different tables for 3 different data sets comparing all the 3 techniques. The bold letters in the table indicates the better performance.

**Table 1.** Results of Nasa93 Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	1.4304	1.765	0.3993	0.9057	0.349	0.064
<b>LOGISTIC REGRESSION</b>	0.7442	0.9584	0.5019	1	0.2	0
<b>CART</b>	<b>0.4058</b>	<b>1.0653</b>	<b>0.1412</b>	<b>0.1412</b>	<b>0.8</b>	<b>0.52</b>

**Table 2.** Results of Nasa93\_center1 Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	0.0245	1.7199	0	0.4609	0.9583	0.4583
<b>LOGISTIC REGRESSION</b>	<b>1.23</b>	<b>2.32</b>	<b>0.08</b>	<b>0.0943</b>	<b>0.9</b>	<b>1</b>
<b>CART</b>	1.5287	0.806	0.4893	0	0.38	0.61

**Table 3.** Results of Nasa93\_center2 Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	<b>0.0355</b>	<b>0.1599</b>	<b>0</b>	<b>0.00000015</b>	<b>0.958</b>	<b>0.91</b>
<b>LOGISTIC REGRESSION</b>	1	1	1	1	0	0
<b>CART</b>	0.4704	0.7259	0.1683	0.5416	0.59	0.37

**Table 4.** Results of Nasa93\_center5 Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	0.4714	36.0487	0.0588	13.5227	0.71	0.17
<b>LOGISTIC REGRESSION</b>	0.76	1	0.79	1	0	0
<b>CART</b>	<b>0.7784</b>	<b>1.2544</b>	<b>0.279</b>	<b>0.279</b>	<b>0.41</b>	<b>0.46</b>

**Table 5.** Results of cocomo81 Dataset

<i>Methods Table VIII</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	17.6139	17.0151	0.3932	0	0.46	0.68
<b>LOGISTIC REGRESSION</b>	<b>0.513</b>	<b>0.2062</b>	<b>0.0614</b>	<b>0.1111</b>	<b>0.9102</b>	<b>0.8496</b>
<b>CART</b>	1.3505	1.9849	0.2709	0.6888	0.365	0.2698



**Table 6.** Results of cocomo81E Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	0.8644	15.29	0.1584	6.9493	0.62	0.14
<b>LOGISTIC REGRESSION</b>	<b>0.3867</b>	<b>0.1686</b>	<b>0.0426</b>	<b>0.1417</b>	<b>0.8647</b>	<b>0.758824</b>
<b>CART</b>	1.54	5.3489	0.5242	0.893	0.25	0.0051

**Table 7.** Results of cocomo81O Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	<b>0</b>	<b>3.5623</b>	<b>0</b>	<b>0.2286</b>	<b>1</b>	<b>0.55</b>
<b>LOGISTIC REGRESSION</b>	0.7536	0.9189	0.8946	1	0.08	0.083
<b>CART</b>	1.39	2.0182	0.4552	0.8074	0.2916	0.0146

**Table 8.** Results of cocomo81S Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	<b>0</b>	<b>20.6855</b>	<b>0</b>	<b>0.0274</b>	<b>1</b>	<b>0.83</b>
<b>LOGISTIC REGRESSION</b>	2.0003	4.37	0.125	0.1664	0.74	0.65
<b>CART</b>	5.9	6.6	0.6051	0	0.4545	0.6363

**Table 9.** Results of Bailey Basili Dataset

<i>Methods</i>	<i>MMRE</i>		<i>MdMRE</i>		<i>PRED(0.25)</i>	
	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
<b>MULTI LINEAR REGRESSION</b>	<b>0.0015</b>	<b>0.000828</b>	<b>0.000757</b>	<b>0.00075</b>	<b>1</b>	<b>1</b>
<b>LOGISTIC REGRESSION</b>	12.5	8.5106	0.3905	0.0627	0.4444	0.6944
<b>CART</b>	1.34	1.1836	0.6235	0.577	0.38	0.1111

## 5 Conclusion and Results

By analyzing the three data sets using the three data mining techniques namely multiple linear regression, Logistic Regression and Cart technique, we have deduced which technique is good for a particular dataset as is mentioned below. CART

technique fits for Nasa93 and Nasa93\_5. Multiple Linear Regression fits for Nasa93\_2, Cocomo81s, Cocomo81o and Basili Bailey. Logistic Regression fits for Nasa93\_1, cocomo81 and Cocomo81e. Our work is limited only to three data sets which can be expanded further. There are many promising techniques which might work better than the present ones. So, there is a great scope for extension of our work. There are many theories and techniques like rough set theory, radial basis function networks, case based reasoning, least square support vector machines which are to be analyzed for future progress.

## References

1. Bowie, D.K.: Using multi variate logistic regression analysis to predict black male student persistence at a predominately white institution. An approach investigating the relationship between engagement and persistence. PhD Dissertation. Louisiana State University and Agricultural and Mechanical College, USA (2006)
2. Saha, G.: Applying logistic regression model to the examination results data. *Journal of Reliability and Statistical Studies* 4(2), 1–13 (2011)
3. Dejaeger, K., Verbeke, W., Martens, D., Baesens, B.: Data Mining Techniques for Software Effort estimation: a Comparative study. *IEEE Transactions on Software Engineering* 1(1), 1–25 (2011)
4. Mendes, E.: *Cost Estimation Techniques for Web Projects*. IGI Publishing (2008)
5. Pickard, L., Kitchenham, B., Linkman, S.: Using simulated data sets to compare data analysis techniques used for software cost modeling. *IEE Proceeding of Software* 148(6), 165–174 (2001)
6. Brown, S.H.: Multiple linear regression analysis: a matrix approach with matlab. *Alabama Journal of Mathematics* Spring/Fall, 1–3 (2009)
7. Tirimula Rao, B., Sameet, B., Kiran Swathi, G., Vikram Gupta, K., Raviteja, C., Sumana, S.: A Novel Neural Network approach for Software Cost Estimation Using Functional Link Artificial Neural Networks. *International Journal of Computer Science and Network Security (IJCSNS)* 9(6), 126–131 (2009)
8. Tirimula Rao, B., Dehuri, S., Mall, R.: Functional Link Artificial Neural Networks for Software Cost Estimation. *International Journal of Applied Evolutionary Computation (IJAEC)* 3(2), 62–82 (2012)

# Infrared and Visible Image Fusion Using Entropy and Neuro-Fuzzy Concepts

S. Rajkumar and P.V.S.S.R. Chandra Mouli

School of Computing Science and Engineering, VIT University, Vellore 632014, India  
{rajkumarsrajkumar,mouli.chand}@gmail.com

**Abstract.** Image fusion is the process to derive the useful information from the scene captured by infrared (IR) and visible images. This derived information is used to improve the image content by enhancing the image visualization. Human identification or any living object identification in IR images is a challenging task. This paper proposes two fusion techniques namely Discrete Wavelet Transform with Neuro-Fuzzy (NF) and Entropy (EN) (DWT-NF-EN) and Integer Wavelet Transform with Neuro-Fuzzy and Entropy (IWT-NF-EN) and their results are compared and analyzed with existing fusion techniques using different quantitative measures. Subjective and objective evaluation of the results obtained is compared with other fusion techniques namely Redundancy Discrete Wavelet Transform (RDWT) and Integer Wavelet Transform and Neuro-Fuzzy (IWT-NF). The objective evaluation is done using the quantitative measures Entropy (EN), Peak Signal to Noise Ratio (PSNR) and Normalized Correlation Coefficient (NCC). From the experimental results it is observed that proposed methods provided better information (quality) using EN, PSNR and NCC measures for majority of the test images and the same is justified with the subjective results.

**Keywords:** Infrared and visible images, Integer Wavelet Transform, Discrete Wavelet Transform, Neuro-Fuzzy, RDWT, Fusion.

## 1 Introduction

Detection of objects emitting electromagnetic radiations is an important and challenging problem from the defense point of view. Surveillance of the most secured areas is required round the clock and is done automatically by setting up multiple sensors at high altitudes. The regular sensors used in the defense scenario are the infrared (IR) sensors and visible sensors. Accordingly, the images captured from these sensors are IR images and visible images respectively. The visibility of IR images in general is very dark with white color portioned boundary of object emitting the radiations. The image spans a specified region on the earth surface. The same area is also captured by the visible sensor and the visibility of the visible image is clear with the surface of the earth but the presence of the electromagnetic radiating objects is

difficult. In order to visualize the radiation emitting objects and the background surroundings clearly, both the IR image and the visible image need to be fused.

Many researchers have proposed different fusion techniques. The fusion techniques are categorized into three levels such as pixel, region and decision. Pixel level fusion is the simplest and the most popularly used technique in research field. The pixel level image fusion methods can be classified as spatial domain and transform domain. The spatial domain is the simplest method. Pixel Averaging, Pixel level Maximum and Minimum (MM) High pass filtering, Intensity-Hue-Saturation based method, Brovey method, Principle of Component Analysis (PCA) etc, fall into this category but the fused images have the problem of spatial distortion and spectral distortion. These problems are resolved in transform domain such as Contrast Pyramid [1], Laplacian Pyramid [2], Ratio Pyramid [3], Morphological Pyramid [4], Discrete Wavelet Transform (DWT) [5], Contourlet Transform (CT) [6], Curvelet Transform [7] and so on. All these methods have individual limitations in fusion process. In pyramid based methods, Contrast Pyramid method could not retain sufficient information from the source images; Laplacian Pyramid method is more sensitive to noise and hence provides wrong edges; Ratio Pyramid provides more false positive information; Morphological Pyramid creates many false edges. The majority of image fusion techniques are based on wavelet transform. In wavelet transform, regularly used DWT technique has the problem of shift invariance and it results in additive noise in the resultant fused image. The revised version of DWT is Redundancy Discrete Wavelet Transform (RDWT) [8], Integer Wavelet Transform (IWT) [9]. This method overcomes the problem of DWT, but they do have some limitations. As a whole, each method has its own limitation.

Based on the survey, it is observed that RDWT, IWT and Neuro-Fuzzy (IWT-NF) were recently used fusion techniques and are proven to be good for image fusion. To overcome these problems, two hybrid methods based on RDWT and IWT-NF are proposed.

RDWT [10] fusion technique can be performed for registered images of two different sensor images. This technique works in the three steps. In the first step, the IR and visible images are decomposed using Haar wavelet transform thus forming four subbands. In the second step, low subbands of IR and visible images are fused using average method forming a fused low sub band. The other three corresponding high subbands of IR and visible images are fused using entropy method thus forming three fused high subbands. Finally, inverse redundancy discrete wavelet transform is applied on the fused low and high subbands to form the fused image.

IWT [11] is implemented in three stages. In the first stage integer wavelet transform is performed based on the lifting scheme. In next stage output of the lifting scheme is fused using neuro-fuzzy. Finally the fused image is retrieved using inverse integer wavelet transform procedure.

The rest of the paper is organized as follows: Proposed method is described in section-2 and experimental results and performance evaluation are presented in section-3. Finally section-4 concludes the work.

## 2 Proposed Methodology

Two hybrid methods based on the two fusion techniques discussed in introduction are proposed in this section.

### 2.1 Proposed Method-I (DWT-NF-EN)

The architecture of the proposed method-1 is given in Fig. 1. This proposed system works in three steps. In first step, both the IR and visible images are decomposed by 2D Haar Wavelet Transform. In the second step, the low bands of the IR and visible image are combined using Neuro-Fuzzy method. The three high subbands of IR and visible image are fused using entropy concept. The Neuro-Fuzzified low band and the entropy applied high subbands are combined and inverse wavelet transform is applied. The resultant image is the fused image.

**Low Subband Fusion.** The IR and visible image low subbands (A, B) are fused using Neuro-Fuzzy method. The low band coefficients are converted into fuzzy domain using triangular membership function. The fuzzified coefficients are used as input to feed forward neural network for training the pattern. Using the training pattern, the fuzzy inference system is created.

**High Subband Fusion.** The high subband blocks are fused based on the entropy calculation of each block [10]. The entropy is defined as:

$$e_{jk}^i = \ln \sqrt{\left( \mu_{jk}^i - \sum_{x,y=1}^{3,3} AB_{jk}^i(x,y) / \sigma_{jk}^i \right)^2 / m^2} \quad (1)$$

where  $j=(v, d, h)$  denotes the subbands,  $k$  represents the block number,  $m=3$  is size of each block and  $i = (1,2)$  is used to differentiate the two input images A and B.  $\mu_{jk}^i$  and  $\sigma_{jk}^i$  are the mean and standard deviation of the each DWT coefficients. Using the entropy values fused image  $AB_v^F$ ,  $AB_d^F$  and  $AB_h^F$  are calculated. The fused image block is  $AB_{jk}^F$  derived from A is selected if the entropy value of the particular block of A image is greater than the particular block of B image, otherwise derived from B show in Eq. 2.

$$AB_{jk}^F = \begin{cases} AB_1^F, & \text{if } e_{jk}^1 > e_{jk}^2 \\ AB_2^F, & \text{otherwise} \end{cases} \quad (2)$$

**Reconstruction of Fusion Image.** Inverse DWT is applied on all the fused subbands to generate fused image  $AB^F$ .

$$AB^F = IDWT(AB_a^F, AB_v^F, AB_d^F, AB_h^F) \quad (3)$$

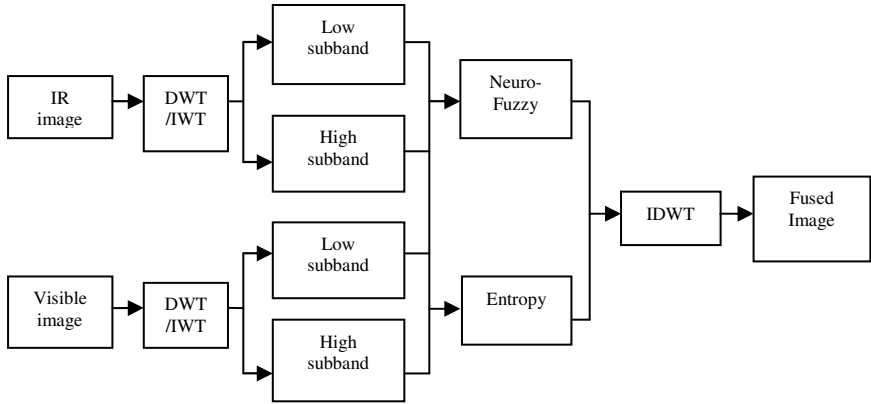


Fig. 1. Block Diagram of proposed method-1 and proposed method-2

### 2.2 Proposed Method-2 (IWT-NF-EN)

The second proposed method is given in this sub section. This method is similar to proposed method-1 except that instead of DWT, IWT is used. The block diagram for the proposed method shown in Fig. 1.

The IWT decomposition performed with lifting schemes. Lifting consists of three steps: split, predict and update. The details of the steps are as follows:

Split: Divide the source image dataset into odd subset  $S_i^0$  and even subset  $d_i^0$

Predict: Use odd subset  $S_i^{n-1}$  to predict the even subset based  $d_i^n$  on the correlation present in the source image. Find a prediction operator  $P$ , independent of the data, so that the construction of a prediction operator  $P$  is typically based on some model of the data which reflects its correlation structure.

$$d_i^n = d_i^{n-1} + \sum_k P_n(k) S_i^{n-1} \tag{4}$$

Update: Some global properties of the original data set in the subsets need to be maintained. For this to find better  $S_i^n$ . Construct an operator  $U$  and update  $S_i^n$  as

$$S_i^n = S_i^{n-1} + U_n(k) d_i^n \tag{5}$$

**Low Subband Fusion.** The decomposed IR and visible images LL band fused using neuro-fuzzy. Neuro-fuzzy means combinations of neural network and fuzzy logic. To perform that first defines membership function and fuzzy rules using these adaptive neuro-fuzzy interference system created.

**High Subband Fusion.** High subbands are fused using entropy method.

**Reconstruction of Fused Image.** Finally, inverse integer transform form is performed to get the fused image by using reverse order of lifting scheme.

### 2.3 Quantitative Analysis

Quantitative measure is helpful to measure the fused image subjective and objective information. Here, the following quantitative measures are used to analyze the fused image.

**Entropy (EN).** Entropy can reflect the amount of information in certain image. The higher value of entropy indicates the better fusion result is obtained.

**Peak Signal to Noise Ratio (PSNR).** PSNR is used to measure the quality of the fused image with respect to the source image [12]. It is defined as:

$$PSNR = 10 \log_{10} (MAX^2 / MSE) \quad (6)$$

$$MSE = \frac{1}{pq} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} [F(i, j) - I(i, j)]^2 \quad (7)$$

where MAX is the maximum value in an image. p, q are the height and weight of an image. I(i, j) is the value of input image and F(i, j) is the value of fused image.

**Normalized Correlation Coefficient (NCC).** A measure that determines the degree to which the similarity of two matrices / images possesses. Normalized Correlation Coefficient calculated as:

$$NCC = \frac{\sum_{i=1}^m \sum_{j=1}^n F(i, j) * I(i, j)}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n F(i, j)^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n I(i, j)^2}} \quad (8)$$

Where F(i,j) is the fused image value and I(i,j) is the input image value. The correlation coefficient ranges vary from -1 to +1. A -1 indicates the negative correlation and +1 indicates the positive correlation.

## 3 Experimental Results

The experimental results of fusion techniques are tested with two categories [13] (dune, trees). Each category contains five IR and visible images used in Defense. Each IR image combined with visible image consider as one set for fusion. Totally derives twenty combinations of input images. All images have the same size of 256 X 256 pixels, with 256-level gray scale. Some of the input sample images and proposed methods fused images are shown in Fig. 2.

The fused image obtained from each technique is analyzed with quantitative measures EN, PSNR and NCC. The results of all the fusion techniques, analyzed for the fused image with quantitative measures from each category are shown in Fig. 3-5.

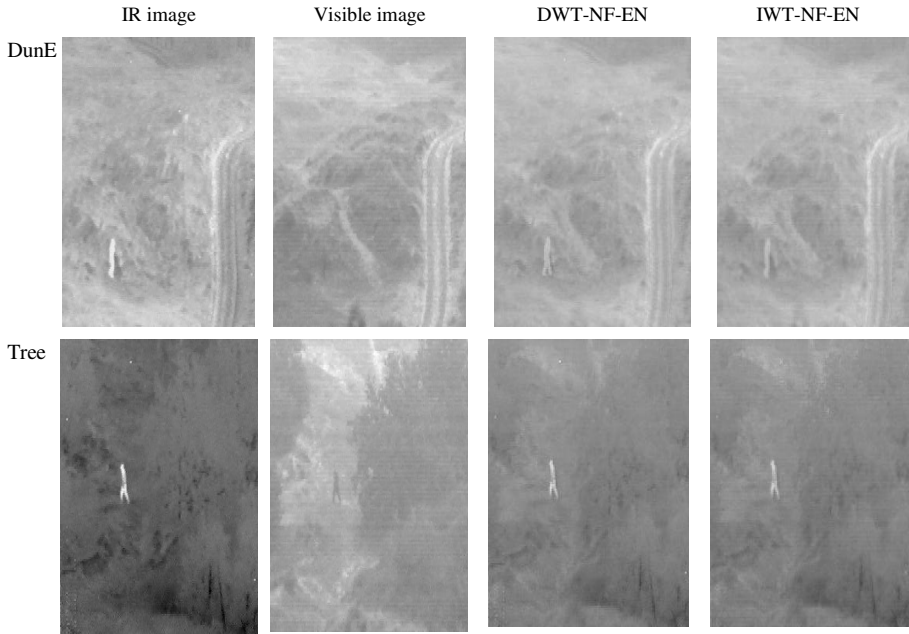


Fig. 2. Sample input and output image of each category

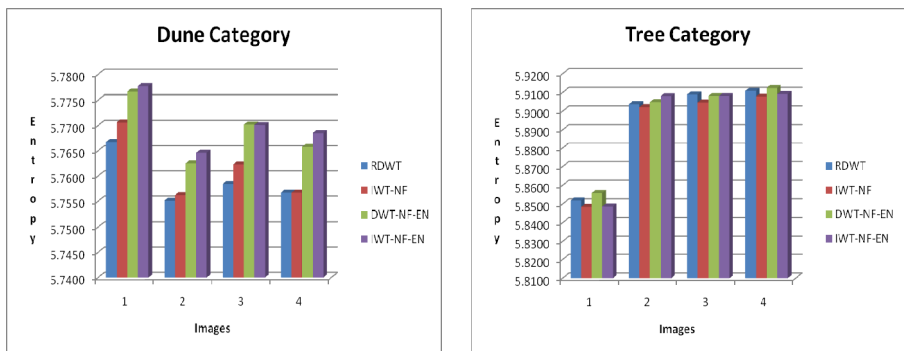


Fig. 3. Entropy comparison of same images in two categories



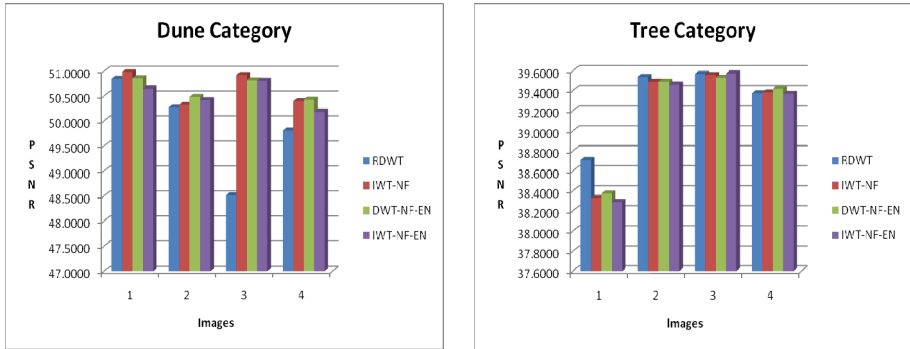


Fig. 4. Peak Signal to Noise Ratio comparison of same images in two categories

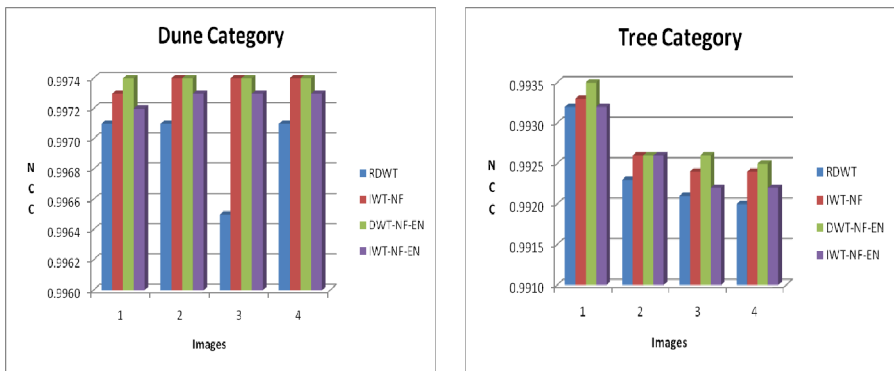


Fig. 5. Normalized Correlation Coefficient comparison of same images in two categories

## 4 Conclusion

In this paper two hybrid methods are proposed for fusion of IR and visible images. The results obtained are compared and analyzed with various quantitative measures. The results show that the proposed fusion techniques provide better results than the existing methods in both subjective and objective evaluation. This fusion technique can be widely used in different applications such as object recognition, target detection, security surveillance. In addition, it can be used as a pre-processing step for noisy images.

## References

1. Toet, A., van Ruyven, J.J., Valette, J.M.: Merging thermal and visual images by a contrast pyramid. *Optical Engineering* 28, 789–792 (1989)
2. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications* 31, 532–540 (1983)

3. Toet, A.: Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters* 9, 245–253 (1989)
4. Toet, A.: A morphological pyramidal image decomposition. *Pattern Recognition Letters* 9, 255–261 (1989)
5. Li, M., Wu, S.: A New Image Fusion Algorithm Based on Wavelet Transform. In: *Proceedings of International Conference on Computational Intelligence and Multimedia Applications*, pp. 154–159 (2003)
6. Yang, L., Guo, B.L., Ni, W.: Multimodality Medical Image Fusion Based on Multiscale Geometric Analysis of Contourlet Transform. *Neuro Computing* 72, 203–211 (2008)
7. Filippo, N., Andrea, G., Stefano, B.: Remote Sensing Image Fusion Using the Curvelet Transform. *Information Fusion* 8, 143–156 (2007)
8. Singh, R., Vastsa, M., Noore, A.: Multimodal Medical Image Fusion using Redundant Discrete Wavelet Transform. In: *Seventh International Conference on Advances in Pattern Recognition*, pp. 232–235 (2009)
9. Wang, Z., Yu, X., Zhang, L.B.: A Remote Sensing Image Fusion Algorithm Based on Integer Wavelet Transform. *Journal of Optoelectronics Laser* 19, 1542–1545 (2008)
10. Rajkumar, S., Kavitha, S.: Redundancy Discrete Wavelet Transform and Contourlet Transform for Multimodality Medical Image Fusion with Quantitative Analysis. In: *Third International Conference on Emerging Trends in Engineering and Technology*, pp. 134–139 (2010)
11. Kavitha, C.T., Chellamuthu, C.: Multimodal Medical Image Fusion Based on Integer Wavelet Transform and Neuro-Fuzzy. In: *International Conference on Signal and Image Processing*, pp. 296–300 (2010)
12. Prakash, C., Rajkumar, S., Chandra Mouli, P.V.S.S.R.: Medical Image Fusion based on Redundancy DWT and Mamdani type min sum mean-of-max techniques with Quantitative Analysis. In: *International Conference on Recent Advances in Computing and Software Systems*, pp. 54–59 (2012)
13. Saeedi, J., Faez, K.: Infrared and visible image fusion using fuzzy logic and population-based optimization. *Applied Soft Computing* 12, 1041–1054 (2011)
14. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Prentice Hall (2007)
15. Jang, J.-S.R., Sun, C.-T., Mizutani, E.: *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall (1997)

# Hybrid Non-dominated Sorting Simulated Annealing Algorithm for Flexible Job Shop Scheduling Problems

N. Shivasankaran<sup>1\*</sup>, P. Senthilkumar<sup>1</sup>, and K. Venkatesh Raja<sup>2</sup>

<sup>1</sup> Department of Mechanical Engineering, K.S.R. College of Engineering,  
Tiruchengode – 637215, Tamilnadu, India

<sup>2</sup> Department of Automobile Engineering, K.S.R. College of Engineering,  
Tiruchengode – 637215, Tamilnadu, India  
shivasankaran.n@gmail.com

**Abstract.** A new hybrid Non-dominated Sorting simulated annealing algorithm has been proposed to solve Multiobjective flexible job-shop scheduling problems (MOFJSPs). The multi objectives considered in this study are makespan, total workload of machines, workload of critical machines and total cost simultaneously. In this study, the critical or incapable machines are eliminated by non-dominated sorting of all operations and the initial solution is arrived using simulated annealing. A main feature of this proposed algorithm is its computational efficiency and simplicity in hybridization. The performance of the proposed algorithm is tested with flexible benchmark instances. The experimental results prove its performance by providing non-dominated solutions for both small and relatively larger cases in minimum computational time.

**Keywords:** Multi objective, flexible job-shop scheduling, Non-dominated Sorting, Simulated annealing algorithm.

## 1 Introduction

Scheduling in job shop involves the allocation of resources to perform a set of tasks over a period of time. These scheduling problems are known to be NP hard [1]. The complexity in computing will be further increased while considering Flexible job shop scheduling problems (FJSP). The scheduling in a FJSP consists of a routing sub-problem, that is, assigning each operation to one machine out of a set of capable machines and the scheduling sub-problem, which consists of sequencing the assigned operations on all machines in order to obtain a feasible schedule minimizing a predefined objective function. The literature of FJSP is considerably sparser than the literature of JSP.

J. Hurink et al. [2] developed Tabu-search algorithm to solve FJSP by considering an integrated approach using insertion techniques and beam search for computing initial solution. M. Mastrolilli and L. M. Gambardella [3] developed two

---

\* Corresponding author.

neighbourhood structures for Tabu search algorithm for efficient computation in FJSP. Improved GA models have been proposed for solving FJSPs [4–9]. These algorithms consider an integrated approach for chromosome representations which improves the optimal solution feasibility. W. Xia et al. [10] developed a hybrid approach using particle swarm optimization (PSO) to assign operations on machines and simulated annealing algorithm (SA) to schedule operations on each machine. J.Gao et al. [11, 12] developed a new approach for exploiting the ability of global search and local search by hybridizing genetic algorithm (GA) with bottleneck shifting for solving MOFJSP.

N. Liouane et al. [13] defined a new approach based on the combination of the ant system with tabu search algorithm for FJSP. In this method, the operations of the machine responsible for makespan  $C_{max}$  are checked for possibility of swapping between other machines for reducing the makespan. This process is done by local search method especially tabu search algorithm. M. S. Mehrabad and P. Fattahi [14] presented an effective two-phase tabu search algorithm with sequence dependent setups for FJSPs. G. Zhang et al. [15] proposed a new hybrid PSO algorithm for omitting the traditional velocity, displacement updating method using particle's effective encoding scheme. L. N. Xing et al. [16] developed an efficient search technique using two feasible move search algorithms namely random handpicked searching algorithm and full-scale general searching algorithm. They defined five modules which are listed as follows: Operation assignment; operation sequencing; feasible moves search; feasible moves evaluation and best move execution. Using this approach they are able to obtain only near-optimal solution in few cases.

J.Q. Li et al. [17] developed an effective neighborhood structure with two adaptive rules for machine assignment and a speed-up local search method with three kinds of insert and swap neighborhood structures based on public critical block theory for operation scheduling. Immune algorithms have been considered for MOFJSPs by X. Wang et al. [18] and A. Bagheri et al [19]. G. Moslehi and M. Mahnam [20] developed a hybrid algorithm using PSO and local algorithm. This algorithm provides competitive solutions with satisfactory computation time instead of providing optimal solutions. N. Shivasankaran et al. [21, 22] applied simulated annealing algorithm for scheduling repair shops and job shops with multiple objectives.

Most of the MOFJSPs constitute multiple optimal solutions. The selection of solutions is a complex job for a job planner. So in this paper, non-dominated sorting simulated annealing is used to solve the MOFJSPs. Simulated annealing helps in extensive search of solution space while non-dominated sorting reduces the time of search and local search algorithm is employed to assign the operations to machines.

The remainder of the paper is organized as follows: In Section 2, we describe the assumptions and formulation of the multi objective flexible job-shop scheduling problem in detail. In Section 3, we present our approach to solve the MOFJSP. Then, the experimental results are illustrated and analyzed in Section 4. Finally, Section 5 provides conclusion and suggestions for further study of this problem.

## 2 Problem Formulation

The flexible job shop scheduling can be formulated as follows. There are  $n$  jobs which are to be processed by  $m$  machines. Each job  $j_i$  consists of a predefined sequence of operations  $O_{ij}$ . Each operation  $O_{ij}$  is allowed to be processed by one machine out of all capable machines  $M_{ij}$ . The processing time of  $j^{\text{th}}$  operation for the job  $j_i$  on the machine  $k$  is denoted by  $P_{ijk}$ . The processing cost of  $j^{\text{th}}$  operation for the job  $j_i$  on the machine  $k$  is denoted by  $V_{ijk}$ . The task of scheduling is to assign each operation on the given machine and sequence the operations on all the machines, in order to optimize some objectives. In addition, some restrictions must be met:

- (i) Each machine can process only one job at the same time;
- (ii) Each job can be processed by a single machine at a time;
- (iii) Each operation cannot be split when being processed;
- (iii) There are no priority restrictions between the operations for different jobs;
- (iv) All jobs have equal priorities.

### Indices

- $i$  index of jobs;
- $j$  index of operation sequence;
- $k$  index of machines;

### Parameters

- $m$  total number of machines
- $n$  total number of jobs
- $n_i$  total number of operations on job  $i$
- $C_i$  completion time of the job  $i$

### Decision variable

$$X_{ijk} = \begin{cases} 1, & \text{if machine } k \text{ is selected for the operation } O_{ij} \\ 0, & \text{otherwise} \end{cases}$$

The objectives considered in this paper are to minimize the

- (1) Makespan

$$F_1 = \max ( C_i ) \tag{1}$$

- (2) Critical Machine workload

$$F_3 = \max \sum_{i=1}^n \sum_{j=1}^{n_i} X_{ijk} p_{ijk} \tag{2}$$

- (3) Total workload of the machines

$$F_2 = \sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^m X_{ijk} p_{ijk} \tag{3}$$

- (4) Total operating cost

$$F_4 = \sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^m X_{ijk} V_{ijk} \tag{4}$$

The weighted sum of the above four objective values are taken as the combined objective function:

*Minimize*

$$G = W_1 X F_1 + W_2 X F_2 + W_3 X F_3 + W_4 X F_4 \quad (5)$$

Subject to

$$W_1 \in (0, 1)$$

$$W_2 \in (0, 1)$$

$$W_3 \in (0, 1)$$

$$W_4 \in (0, 1)$$

$$W_1 + W_2 + W_3 + W_4 = 1. \quad (6)$$

Where  $W_1, W_2, W_3, W_4$  are the weight coefficient for the four objective values, respectively. The weight coefficients are fixed based on previous works of Xing et al. (2009) [16] on Kacem instances.

### 3 NSSA Algorithm

In this study we propose a hybrid algorithm known as non-dominated sorting simulated annealing (NSSA) algorithm. The non-dominated sorting introduced by Deb et al. [23] based on the principle of Pareto dominance relationship is used for the

---

**Algorithm:** *Non dominated sorting simulated annealing algorithm*

**Begin**

$r \leftarrow 1$ ;

**for**  $i=1$  **to**  $n$

*is*  $v_1(r) > v_2(r)$  // sort the order of machines based on operation time

**Swap** ( $v_1(r), v_2(r)$ );

$r \leftarrow r + 1$ ;

**End**

**Begin**

$T \leftarrow$  start temperature

Current  $\leftarrow$  generate initial solution // operation time and cost is considered

Evaluate current

Best  $\leftarrow$  current

Repeat

Count  $\leftarrow 0$

Repeat

Schedule  $\leftarrow$  generate schedule from current

Cost  $\leftarrow$  generate cost from current

Evaluate Schedule and Cost

If schedule or cost is better than best then update best

Accept schedule or cost as current with probability  $\leftarrow \min(1, e^{-(F_{\text{current}} - F_{\text{schedule}}) / T})$

Until ( $++$  count  $\leftarrow$  number of iterations at temperature  $T$ )

Decrease temperature according to cooling scheme until (stopping criteria)

**End**

---

**Fig. 1.** Pseudo code of Non-dominated sorting simulated annealing algorithm



## 5 Conclusion

In this paper, a Non-dominated sorting simulated annealing algorithm is proposed to solve the multi-objective flexible job shop scheduling problems. The performance of the presented approach is evaluated in comparison with the results obtained from literature for four representative instances. These results clearly show the decision makers, a better solution that can be implemented for such complex problems. The future research guidelines includes

- Employing the algorithm for more practical data in case of implementation of CIM concepts in manufacturing industries and scheduling manpower in repair shops;
- Developing effective theory and algorithm for this complex combinatorial optimization problem are needed;

## References

- [1] Garey, M.R., Johnson, D.S., Sethi, R.: The complexity of flowshop and job shop scheduling. *Math. Oper. Res.* 1(2), 117–129 (1976)
- [2] Hurink, J., Jurisch, B., Thole, M.: Tabu search for the job-shop scheduling problem with multi-purpose machines. *OR-Spektrum* 15(4), 205–215 (1994)
- [3] Mastrolilli, M., Gambardella, L.M.: Effective neighborhood functions for the flexible job shop problem. *J. Sched.* 3(1), 3–20 (1996)
- [4] Kacem, I., Hammadi, S., Borne, P.: Approach by localization and multiobjective evolutionary optimization for flexible job-shop scheduling problems. *IEEE Transactions on Systems, Man, and Cybern, Part C* 32(1), 1–13 (2002a), doi:10.1109/TSMCC.2002.1009117
- [5] Kacem, I., Hammadi, S., Borne, P.: Pareto-optimality approach for flexible job-shop scheduling problems: Hybridization of evolutionary algorithms and fuzzy logic. *Math. Comput. Simul.* 60(3), 245–276 (2002b)
- [6] Tay, J.C., Wibowo, D.: An effective chromosome representation for evolving flexible job shop schedules. In: Deb, K., Tari, Z. (eds.) *GECCO 2004*. LNCS, vol. 3103, pp. 210–221. Springer, Heidelberg (2004)
- [7] Ho, N.B., Tay, J.C.: GENACE: An efficient cultural algorithm for solving the flexible job-shop problem. In: *IEEE Congress on Evolutionary Computation CEC 2004*, vol. 2, pp. 1759–1766. IEEE (2004), doi:10.1109/CEC.2004.1331108
- [8] Ho, N.B., Tay, J.C.: Evolving dispatching rules for solving the flexible job-shop problem. In: *IEEE Congress on Evolutionary Computation*, vol. 3, pp. 2848–2855. IEEE (2005), doi:10.1109/CEC.2005.1555052
- [9] Zhang, H., Gen, M.: Multistage-based genetic algorithm for flexible job-shop scheduling problem. *J. Complex. Int.* 11, 223–232 (2005)
- [10] Xia, W., Wu, Z.: An effective hybrid optimization approach for multi-objective flexible job-shop scheduling problems. *Comput. Ind. Engg.* 48(2), 409–425 (2005)
- [11] Gao, J., Gen, M., Sun, L., Zhao, X.: A hybrid of genetic algorithm and bottleneck shifting for multiobjective flexible job shop scheduling problems. *Comput. Ind. Engg.* 53(1), 149–162 (2007)



- [12] Gao, J., Sun, L., Gen, M.: A hybrid genetic and variable neighborhood descent algorithm for flexible job shop scheduling problems. *Comput. Oper. Res.* 35(9), 2892–2907 (2008)
- [13] Liouane, N., Saad, I., Hammadi, S., Borne, P.: Ant systems & Local Search Optimization for flexible job shop scheduling production. *Int. J. Comput. Commun. Control* 2(2), 174–184 (2007)
- [14] Saidi-Mehrabad, M., Fattahi, P.: Flexible job shop scheduling with tabu search algorithms. *Int. J. Adv. Manuf. Technol.* 32(5-6), 563–570 (2007)
- [15] Zhang, G., Shao, X., Li, P., Gao, L.: An effective hybrid particle swarm optimization algorithm for multi-objective flexible job-shop scheduling problem. *Comput. Ind. Engg.* 56(4), 1309–1318 (2009)
- [16] Xing, L.N., Chen, Y.W., Yang, K.W.: An efficient search method for multi-objective flexible job shop scheduling problems. *J. Intell. Manuf.* 20(3), 283–293 (2009)
- [17] Li, J.Q., Pan, Q.K., Liang, Y.C.: An effective hybrid tabu search algorithm for multi-objective flexible job-shop scheduling problems. *Comput. Ind. Engg.* 59(4), 647–662 (2010)
- [18] Wang, X., Gao, L., Zhang, C., Shao, X.: A multi-objective genetic algorithm based on immune and entropy principle for flexible job-shop scheduling problem. *Int. J. Adv. Manuf. Technol.* 51(5-8), 757–767 (2010)
- [19] Bagheri, A., Zandieh, M., Mahdavi, I., Yazdani, M.: An artificial immune algorithm for the flexible job-shop scheduling problem. *Future Gener. Comput. Syst.* 26(4), 533–541 (2010)
- [20] Moslehi, G., Mahnam, M.: A Pareto approach to multi-objective flexible job-shop scheduling problem using particle swarm optimization and local search. *Int. J. Prod. Econ.* 129, 14–22 (2011)
- [21] Shivasankaran, N., Senthil Kumar, P., Nallakumarasamy, G., Venkatesh Raja, K.: A Hybrid Bubble Sorting Simulated Annealing Algorithm for Job Shop Scheduling. In: *Third International Conference on Computing Communication & Networking Technologies (ICCCNT 2012)*, pp. 1–5. IEEE (2012), doi:10.1109/ICCCNT.2012.6395981
- [22] Shivasankaran, N., Senthil Kumar, P., Nallakumarasamy, G., Venkatesh Raja, K.: Repair Shop Job Scheduling with Parallel Operators and Multiple Constraints using Simulated Annealing. *Int. J. Comput. Intell. Syst.* 6(2), 223–233 (2013)
- [23] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multi-objective genetic algorithm: NSGAI. *IEEE Transactions on Evolutionary Computation* 6(2), 182–197 (2002)
- [24] Frutos, M., Olivera, A.C., Tohmé, F.: A memetic algorithm based on a NSGAI scheme for the flexible job-shop scheduling problem. *Ann. Oper. Res.* 181(1), 745–765 (2010)

# DICOM Image Retrieval Using Geometric Moments and Fuzzy Connectedness Image Segmentation Algorithm

Amol Bhagat<sup>1</sup> and Mohammad Atique<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering, Prof Ram Meghe College of Engineering and Management Badnera, Amravati, India  
amol.bhagat84@gmail.com

<sup>2</sup> PG Department of Computer Science, SGB Amravati University, Amravati, India

**Abstract.** The Medical image database is growing day by day. Most of the medical images are stored in DICOM (Digital Imaging and Communications in Medicine) format. There are various categories of medical images such as CT scan, X- Ray, Ultrasound, Pathology, MRI, Microscopy, etc [1]. Physicians compare previous and current medical images associated patients to provide right treatment. Medical Imaging plays a leading role in modern diagnosis. Efficient image retrieval tools are needed to retrieve the intended images from large growing medical image databases. Such tools must provide more precise retrieval results with less computational complexity. This paper compares the proposed technique for DICOM medical image retrieval and shows that the proposed geometric moments and fuzzy connectedness image segmentation algorithm based image retrieval algorithm performs better as compared to other algorithms.

**Keywords:** Medical Image Retrieval, Image Enhancement, Relevance Feedback, Medical Image Processing, Soft Computing.

## 1 Introduction

Various researches are carried out in recent past years for efficient medical image indexing and retrieval [1, 2]. The overview of some of these researches is given in this paper. This paper includes the comparison of image retrieval techniques with the proposed Geometric Moments combined with Fuzzy Connectedness Image Segmentation. The comparison is carried out between proposed technique and various image retrieval techniques such as Web base Medical IR in Oracle, Pattern Similarity based Medical IR, Indexing for RF in IR, Entropy Based IR, Similarity Based Online Feature Selection, Entropy Based Feature Selection and Localized CBIR.

The implementation of web based system for medical image retrieval using Oracle Multimedia features is given in [3]. The system uses Digital Imaging and Communications in Medicine (DICOM) image file format, which contains additional information regarding image modality, acquisition device and patient identification in its header along with raw image data. The DICOM feature was introduced in Oracle Database 10g Release 2. It includes: ORDDicom object type, DICOM metadata

extraction, DICOM conformance validation and DICOM image processing and image compression. The reference of Picture archiving and communication systems (PACS) is also given in [2, 3]. PACS are used in many hospitals, which are basically computer networks for storage, distribution and retrieval of medical image data.

An efficient content based medical image retrieval scheme is proposed in [4]. It is based on PAttern for Next generation DAtabase systems (PANDA) framework for pattern representation and management. An expectation maximization [4, 5] algorithm is used for clustering feature space. The similarity between two clusters is estimated as a function of the similarity of both their structures and the measures component. Large set of radiographic images were used from the Image Retrieval in Medical Applications (IRMA) dataset [6] to carry out experiments. A fast clustering based indexing technique for exact nearest neighbor search that adapts to the Mahalanobis (generalized Euclidean) distance is described in [7, 8, 16]. A clustering-based indexing scheme is proposed, where relevant clusters are retrieved till the exact nearest neighbors are found. The combination of a statistical similarity matching technique and relevance feedback scheme in medical image database is demonstrated in [9]. The comparison is given in the [9, 10] based on Bhattacharyya and Mahalanobis with and without relevance feedback.

Online feature selection in the relevance feedback learning process to improve the retrieval performance of the region based image retrieval system is investigated in [11]. The implementation of efficient entropy based features selection for Image Retrieval is given in [12]. The system is developed using eMbedded Visual C++ 4.0 and runs in a Windows environment on a Pocket PC. ACCIO! A localized CBIR system is presented in [13]. This system does not consider that the objects are in a fixed location or have a fixed size. Improved entropy and moments based image retrieval is presented in [14, 15].

## 2 Geometric Moments with Fuzzy Connectedness Image Segmentation

Consider a 2D image composed of two regions corresponding to two objects  $O_1$  and  $O_2$  as illustrated in Fig. 1,  $O_2$  being the background.  $O_2$  itself may consist of multiple objects which are not of interest in distinguishing since object of interest is  $O_1$ . Determine an affinity relation that assigns to every pair of nearby pixels in the image a value based on the nearness of pixels in space and in intensity. To every “path” connecting every pair of pixels, such as the solid curve  $p_{c o_1}$  connecting  $c$  and  $o_1$  in Fig 1 “strength of connectedness” is assigned which is simply the smallest pair wise affinity of pixels along the path. The strength of connectedness between any two pixels such as  $o_1$  and  $c$  is simply the strength of the strongest of all paths between  $o_1$  and  $c$ . Suppose,  $p_{c o_1}$  shown in Fig 1 represents the strongest path between  $o_1$  and  $c$ . If the affinity is designed properly, then  $p_{c o_1}$  is likely to have a higher strength than the strength of any path such as the dotted curve between  $c$  and  $o_1$  that goes outside  $O_1$ .

In the original fuzzy connected method, an object such as  $O_1$  is segmented by setting a threshold on the strength of connectedness. This threshold defines a pool of

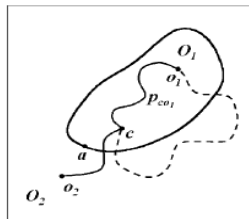
pixels such that within this pool the strength of connectedness between any two pixels is not less than the threshold but between any two pixels, one in the pool and the other not in it, the strength is less than the threshold. The basic idea in relative fuzzy connectedness is to first select reference pixels  $o_1$  and  $o_2$ , one in each object, and then to determine to which object any given pixel belongs based on its relative strength of connectedness with the reference pixels. A pixel  $c$  would belong to  $O_1$  since its strength of connectedness with respect to  $o_1$  is likely to be greater than that with  $o_2$ . This relative strength of connectedness offers a natural mechanism for partitioning pixels into regions. This mechanism eliminates the need for a threshold required in the original method and offers potentially more powerful segmentation strategies. According to the fuzzy topology theory, a field  $H = \{\eta(p)\}$  can be derived from any digital image by simply normalizing the pixel-intensity value. A fuzzy-connectedness degree can be computed for each pixel  $p$ , and this measure refers to the absolute maximum membership value. One can extract a fuzzy-connectedness measure with respect to any image pixel  $a$ , given the appropriate transform that is applied to each pixel  $p$ . For the sake of clarity, such a transformation, which gives rise to the modified field  $X^a$  (Equation 1), is given as

$$x^a(p) = 1 - |\eta(p) - \eta(a)| \tag{1}$$

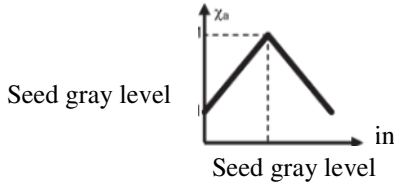
Pixel  $a$  – seed point assumes the maximum value in the modified field, as shown in Fig 2. If we define  $P(q, p)$  as connected path of points from a pixel  $q$  to a pixel  $p$  and if the seed point represents and belongs to a structure of interest, it is possible to measure the connectivity (Equation 2) associated with the structure by applying, for every  $p$ , the following:

$$C_{X^a} = c_{X^a}(p) = \text{conn}(X^a, a, p) = \max_{P(a,p)} [\min_{z \in P(a,p)} X^a(z)] \tag{2}$$

The max is applied to all paths  $P(a, p)$  from  $a$  to  $p$  and thus refers to the optimum path connecting  $p$  to the seed point, while the min is applied to all points  $z$  along the optimum path  $P(a, p)$ . The equation is named “ $\chi$ -connectivity” or “intensity connectedness,” and its application results in an image where every pixel value represents the degree of membership to the searched object. The new image produced is called the “connectivity map,” where each image element has a gray level that is dependent on the degree of connectivity with respect to seed point  $a$ .



**Fig. 1.** Illustration of the main ideas behind relative fuzzy connectedness. The membership of any pixel, such as  $c$ , in an object is determined based on the strength of connectedness of  $c$  with respect to the reference pixels  $o_1$  and  $o_2$  specified in objects  $O_1$  and  $O_2$ .  $c$  belongs to that object with respect to whose reference pixel it has the highest strength of connectedness.



**Fig. 2.** Modified  $X^a$  value as a function of the original value (in)

Geometric moments, which are also, know as Cartesian moments or regular moments are the simplest among moment functions, with the kernel function defined as a product of the pixel coordinates. For a two-dimensional density function  $p(x, y)$  the  $(p + q)^{th}$  order geometrical moments  $m_{pq}$  (Equation 3) are defined as:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q p(x, y) dx dy \tag{3}$$

If  $p(x, y)$  is a piece-wise continuous function and has non-zero values only in the finite part of the x-y plane, then moments of all orders exist for  $p(x, y)$ , and the moments sequence  $m_{pq}$  is uniquely determined by  $p(x, y)$  and vice-versa. Although originally described in continuous form, discrete formulae are commonly in use for practical reasons. If an image is considered as a discrete function  $f(x,y)$  with  $x = 0, 1, \dots, M$  and  $y = 0, 1, \dots, N$  then  $(p+q)^{th}$  order geometric moments  $m_{pq}$  (Equation 4) are defined as :

$$m_{pq} = \sum_{x=0}^M \sum_{y=0}^N x^p y^q f(x, y) \tag{4}$$

It should be noted that second equation can assume very large values, especially for high order moments (large p, q). This often leads to numerical instabilities as well as high sensitivity to noise. Furthermore, image reconstruction is not straightforward.

### 3 Proposed Fuzzy Connectedness Image Segmentation and Geometric Moments for Shape Analysis

Image Segmentation is used to find the (x, y) co-ordinates of the largest image segment and (x, y) co-ordinates of the boundary of the largest image segment. Store the information of all consecutive pixels having same discrete level. The algorithms are explained below; shape is analyzed using the combination of these algorithms.

**Algorithm for Image Segmentation:-**

**Input:** Image and Seed point.

**Output:** Segmented Image.

1. Initialization
  - 1.1 Seed point 'a'
  - 1.2  $P(q,p) \leftarrow$  path of points from a pixel q to pixel p
2. Obtain modified field by normalize the pixel intensity values.
3. Pixel 'a' selected as seed point assuming that it has maximum value in the modified field.

4. If 'a' represents and belongs to a structure of interest then
  - 4.1 Measure the connectivity associated with the structure by applying  $conn(X^a, a, p)$
5. Return the connectivity map
6. Adjust threshold if necessary
7. Select best segmented result

**Algorithm for Shape Module:-**

**Input:** Query Image

**Output:** Parameters useful in geometric moments

1. Initialization
  - 1.1 Array pixel  $\leftarrow$  composed of foreground and background pixels only
  - 1.2 mx  $\leftarrow$  width
  - 1.3 my  $\leftarrow$  height
  - 1.4 sum  $\leftarrow$  0
2. Normalize the number of pixels and update the dimension if necessary
3. For i = 1 to mx\*my do
  - 3.1 sum  $\leftarrow$  sum + (pixel[i]/foreground)
4. a  $\leftarrow$  sqrt(totalpixels/sum)
5. Find foreground background pixels, return results
6. Compute object weight of the input pixel array
7. For i = 1 to height\*width do
  - 7.1 If pixel[i] == foreground then
    - 7.1.1 sum  $\leftarrow$  sum+1
8. For i = 1 to height do
  - 8.1 xsum  $\leftarrow$  xsum + i
9. For j = 1 to width do
  - 9.1 ysum  $\leftarrow$  ysum + j
10. Compute object projection along x and y-axis
11. Result1  $\leftarrow$  xsum/sum
12. Result2  $\leftarrow$  ysum/sum

**Algorithm for Geometric Moments:-**

**Input:** Query Image

**Output:** Set of similar images to query image from the set of N images

1. Initialization
  - 1.1 pValue, qValue  $\leftarrow$  1 for 2<sup>nd</sup> order geometric moment
  - 1.2 size  $\leftarrow$  height\*width
  - 1.3 Array pixel[size]
  - 1.4 Declare vector geometricMoment
  - 1.5 moment  $\leftarrow$  0
2. For j = 1 to height do
  - 2.1 For i = 1 to width do
    - 2.1.1 If pixels[j\*width + i] == foreground then

$$2.1.1.1 \text{ moment} = \text{moment} + (i\text{-weight}[0])^{p\text{Value}} * (j\text{-weight}[1])^{q\text{Value}}$$

3. Add moment to feature vector geometricMoment

Compare this feature vector with the feature vectors of  $N$  images stored in the database using distances between them. According to the distances the system returns nearest neighbors as the query result.

## 4 Experiments Carried Out on DICOM Image Database

Images are ranked according to the similarity between image given as query and the retrieved images. At any time after firing query if user finds any image which is more relevant to him as per his query, user can change the rank of that image. This updated rank can be useful in the next retrieval. Precision is calculated as the ratio of number of relevant images retrieved to the total number of images retrieved. The performance of the proposed implemented algorithms is evaluated by analyzing the images retrieved and comparing it with the images retrieved by various other approaches. Table 1 indicates the total number images retrieved by using five different techniques. It can be observed from the table that the methods that work on the basis of color features such as Color Moments, LCH and GCH does not provide relevant results. The images that are used for evaluation are gray scale images and in practice most of the medical images are gray scale images. Therefore the image retrieval technique which uses color feature for retrieval fails. As shown in Table 1 it is found that the technique designed for gray level feature and shape feature extraction works better. Table 2 shows the precision values computed as per the relevance of the retrieved images as per the query image. The overview and comparison of the systems discussed in this paper is given in Table 3 based on the various parameters. The proposed Geometric Moment Image Segmentation techniques are also included in the table.

**Table 1.** Comparison of Color Moments, Co-occurrence, Local Color Histogram, Global Color Histogram and Geometric Moments on the basis of Number of images retrieved

Image Class	Color Moments	Co-occurrence	Local Color Histogram	Global Color Histogram	Color Moments	Geometric Moments
DICOM	100	12	100	100	100	15

**Table 2.** Precision Values

Image Class	Color Moments	Co-occurrence	LCH/ GCH	Geometric Moments	SIMPLICITY	Histogram Based
DICOM	-----	0.67	-----	0.76	0.78	0.26

## 5 Conclusion and Discussion

This paper gives the various aspects of medical image processing. This paper will motivate the researchers to utilize the methods that are available for medical image retrieval, image indexing and image enhancement to get the better result of retrieval

and to provide better treatment to the patient. While implementing any efficient medical image retrieval system the above discussed aspects should be taken into consideration. The Web based medical image retrieval system in Oracle uses the features and indexing technique provided by Oracle. It only works on the Oracle supported feature. This system doesn't support the automatic threshold selection for the image comparison.

**Table 3.** Comparison of Various Image Retrieval Systems

System	Web Based	Feature Used	Indexing	Image Category	Precision	Relevance Feedback
Web Based Medical IR in Oracle	Yes	Color, Texture and Shape feature based on Oracle supported methods	Supported	CT scan, X-ray, Ultrasound, Pathology, MRI and Microscopy	0.95	No
Pattern Similarity based Medical IR	No	Pattern Base Structure and Schema Measure Schema	: Not Supported	X-ray, MRI Radiographic Images	0.70	No
Indexing for RF in IR	No	Mahalanobis Distance	Supported	Cortina, Bio-Retina	0.80	Yes
Similarity Based Online Feature Selection	No	Region Based, Segmentation Based	Not Supported	Non Medical	0.60	Yes
Entropy Based Feature Selection	No	Entropy Based	Not Supported	Non Medical	-----	Yes
Localized CBIR	No	Expectation Maximization and Diverse Density	Not Supported	Non Medical	-----	No
Proposed Geometric Moments and Fuzzy Connected Image Segmentation	No	Segmentation Based, Shape Based	Supported	Medical/ Non Medical	0.76	Yes

So the performance and accuracy of the system can be improved by using various feature extraction methods and run time threshold selection instead of using Oracle supported methods and fixed threshold. PANDA framework doesn't support the retrieval of various kinds of medical images such as ultrasound and endoscopic images. Both these systems do not support relevance feedback. All other systems that



are discussed are not web based. The third system uses the indexing for relevance feedback. The feature extraction methods are not given. The indexing is based on distance between two images. This distance is also used as measure of similarity between images. The proposed technique gives precise results but those can be further extended to utilize Oracle supported DICOM image features. These features provide more information about the retrieved images. These features will be helpful for identifying the relevance of the images that are retrieved as result.

## References

1. Yong, R., Thomas, H.S.: Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *J. Visual Comm. and Image Representation* 10(4) (April 1999)
2. Xin, Z.Y., Tian, W.F.: Entropy- based Local Histogram Equalization for Medical Ultrasound Image Enhancement. In: *IEEE Intl. Conf. 2008* (2008)
3. Ivica, D., Pero, G., Suzana, L.: Implementation of Web-Based Medical Image Retrieval System in Oracle. In: *IEEE 2nd Intl. Conference on Adaptive Science & Technology 2009* (2009)
4. Greenspan, H., Pinhas, A.T.: Medical Image categorization and retrieval for PACS using the GMM-KL framework. *IEEE Trans. Info. Tech Biomedicine* 11(2) (March 2007)
5. Dimitris, I.K., Yannis, T.: A Pattern Similarity Scheme for Medical Image Retrieval. *IEEE Trans. Info. Tech in Biomedicine* 13(4) (July 2009)
6. Yuan, H., Zhang, X.: Statistical Modeling in the Wavelet Domain for Compact Feature Extraction and Similarity Measure of Images. *IEEE Trans. Circuits and Systems for Video Tech.* 20(3) (March 2010)
7. Lehmann, T.M., Guld, M.O., Thies, C., Plodowski, B., Keysers, D., Ott, B., Schubeert, H.: IRMA – Content based image retrieval in medical applications. In: *Proc. 14th World Congr. Med. Info (Medinfo)*, vol. 2. IOS, Amsterdam (2004)
8. Sharadh, R., Kenneth, R.: Towards Optimal Indexing for Relevance Feedback in Large Image Databases. *IEEE Trans. Image Processing* 18(12) (December 2009)
9. Rahman, M., Prabir, B., Desai, B.: A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching Techniques With Relevance Feedback. *IEEE Trans. Info Tech in Biomedicine* 11(1) (January 2007)
10. Jiang, W., Er, G., Dai, Q., Gu, J.: Similarity-Based Online Feature Selection in Content-Based Image Retrieval. *IEEE Trans. Image Processing* 15(3) (March 2006)
11. Chang, T.W., Sandes, F.E.: Efficient Entropy-based Features Selection for Image Retrieval. In: *Proc. 2009 IEEE Intl. Conference Man and Cybernetics, San Antonio, TX, USA* (October 2009)
12. Rahmani, R., Zhang, H., Cholleti, S., Fritts, J.: Localized Content-Based Image Retrieval *IEEE Trans. Pattern Analysis and Machine Intelligence* 30(11) (November 2008)
13. Junding, S.: Image Retrieval Based on Improved Entropy and Moments. In: *IEEE Proc. Intl. Conference Intelligent Information Hiding and Multimedia Signal Processing* (2006)
14. Shan, Z., Hai-tao, W.: Image Retrieval Based on Bit-plane Distribution Entropy. In: *IEEE Intl. Conference Computer Sci. and Soft. Engg.* 2008 (2008)
15. Krishnapuram, R., Choi, Y., Balasubramaniam, R.: Content-Based Image Retrieval Based on a Fuzzy Approach. *IEEE Trans. Knowledge and Data Engg.* 16(10) (October 2004)
16. Sengee, N., Bazarraghaa, B., Kim, T., Choi, H.: Weight Clustering Histogram Equalization for Medical Image Enhancement. In: *IEEE Intl. Conf. 2009* (2009)
17. Annamalai, M., Guo, D., Mavris, S., Steiner, J.: An Oracle White Paper: Oracle Database 11g DICOM Medical Image Support (September 2009)

# Soft Computing Based Partial-Retuning of Decentralised PI Controller of Nonlinear Multivariable Process

L. Sivakumar<sup>1</sup> and R. Kotteswaran<sup>2</sup>

<sup>1</sup> Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India  
lingappansivakumar@gmail.com

<sup>2</sup> Department of Instrumentation and Control Engineering,  
St. Joseph's College of Engineering, Chennai, India  
kotteswaran@gmail.com

**Abstract.** Recent developments in nature-inspired algorithms motivate the control engineers to work towards its application in industrial processes. Almost all the industrial processes are difficult to control since it involves many variables, strong interactions and inherent nonlinearities. In the present work the authors propose cuckoo search, a recent metaheuristic algorithm to fine tune the parameters of decentralised PI controller of coal gasifier which is a highly nonlinear multivariable process having strong interactions among the control loops. With the existing controller parameters the response does not able to meet the performance requirements at 0% load for sinusoidal pressure disturbance test. The PI controller for pressure loop is retuned using Cuckoo search algorithm and the best optimal values for its parameters are obtained. Performance of the system with tuned optimal controller settings is evaluated for pressure disturbance test, load change test and coal quality test.

**Keywords:** Coal gasifier, Cuckoo search algorithm, metaheuristic algorithm, multivariable process, nonlinear systems, PID Controller tuning.

## 1 Introduction

Integrated Gasification Combined Cycle (IGCC) is an efficient method of clean power and energy generation. Here Coal reacts with air (oxygen) and steam, converted into syngas (also called producer gas) under certain pressure and temperature. Purified syngas runs the gas turbine to generate power and exhaust gas from the gas turbine enters Heat Recovery Steam Generator (HRSG) to produce steam which in turn runs the steam turbine. Coal gasifier, an important and primary element in IGCC, is a highly non-linear, multivariable process, having five controllable inputs (char flow rate, air flow rate, coal flow rate, steam flow rate and limestone flow rate), few non-control inputs (boundary conditions, PSink and coal quality) and four outputs (fuel gas calorific value, bed mass, fuel gas pressure and fuel gas temperature) with a high degree of cross coupling between them. The process is a four-input, four output regulatory problem for the control design (keeping limestone at constant value). It exhibits a complex dynamic behaviour with mixed fast and slow

dynamics and it is highly difficult to control. The full model of coal gasifier has 25 states and the ultimate requirement is to find the controller constants ( $K_p$ ,  $K_i$ ) of PI controller such that all the constraints are met for all the specified loads as given in the challenge problem [1]. Control specification includes sink pressure step and sinusoidal disturbance tests (at the three different operating points), ramp change in load from 50% to 100%, and coal quality change ( $\pm 18\%$ ). Until recently a group of researchers have attempted to analyze the system, design controllers and retune the baseline controller to meet the performance objectives at all the load conditions [2-10]. Apart from the conventional techniques, soft computing approaches such as MOGA [11] and NSGA II [12] are also used to design the controller.

## 2 Cuckoo Search Based Optimization

Optimization is the process of finding a best optimal solution to meet the desired objective function. Nature-inspired metaheuristic algorithms (PSO, BFO, FA, Bee Colony, ANT Colony, BAT, etc..) are most widely used in a variety of optimization Problems including process control. One of such new algorithm is Cuckoo search algorithm developed by Xin-She Yang [13-15], which uses the breeding behaviour of certain species of cuckoos. Cuckoos lay their eggs in other birds' nests even it may remove others' eggs to increase the hatching probability of their own eggs. If a host bird discovers the eggs that are not their own, they will either throw these alien eggs away or abandon its nest and build a new nest elsewhere. Some cuckoo species are often very specialized in the mimicry in colour and pattern of the eggs of a few chosen host species so that increases their reproductivity. Cuckoo search algorithm is based on the following rules:

- Each cuckoo lays one egg at a time, and dumps it in a randomly chosen nest.
- The best nests with high quality of eggs will carry over to the next generation.
- With fixed number of available host nests, the egg laid by a cuckoo is discovered by the host bird with a probability  $p_a \in [0,1]$  discovering operate on some set of worst nests, and discovered solutions dumped from further calculations.

For a maximization problem, the fitness of a solution is directly proportional to the objective function and here each egg in a nest represents a solution, and a cuckoo egg represent a new solution, the aim is to use the new and potentially better solutions (cuckoos) to replace a not-so-good solution in the nests. This algorithm can be extended to the more complicated case where each nest has multiple eggs representing a set of solutions. For this present work, the authors use a simplest approach where each nest has only a single egg. The rules can be integrated to form the Cuckoo search algorithm as shown in figure 1. When generating new solutions  $x_i(t+1)$  for the  $i^{\text{th}}$  cuckoo, the following Levy flight is performed

$$x_i(t+1) = x_i(t) + \alpha \oplus \text{Lèvy} \quad (1)$$

Where,  $\alpha > 0$  = step sizes the step size;  $\oplus$  = entry-wise multiplications.

Step-lengths of Lèvy flight are distributed as

$$\text{Lèvy } u = t^{-\lambda}, \quad 1 < \lambda < 3 \quad (2)$$

Consecutive steps of a cuckoo form a random walk process which obeys a power-law step-length distribution with a heavy tail. In real world, if a cuckoo's egg is very similar to a host's eggs, then this cuckoo's egg is less likely to be discovered, thus the fitness should be related to the difference in solutions. Therefore, it is a good idea to do a random walk in a biased way with some random step sizes [16].

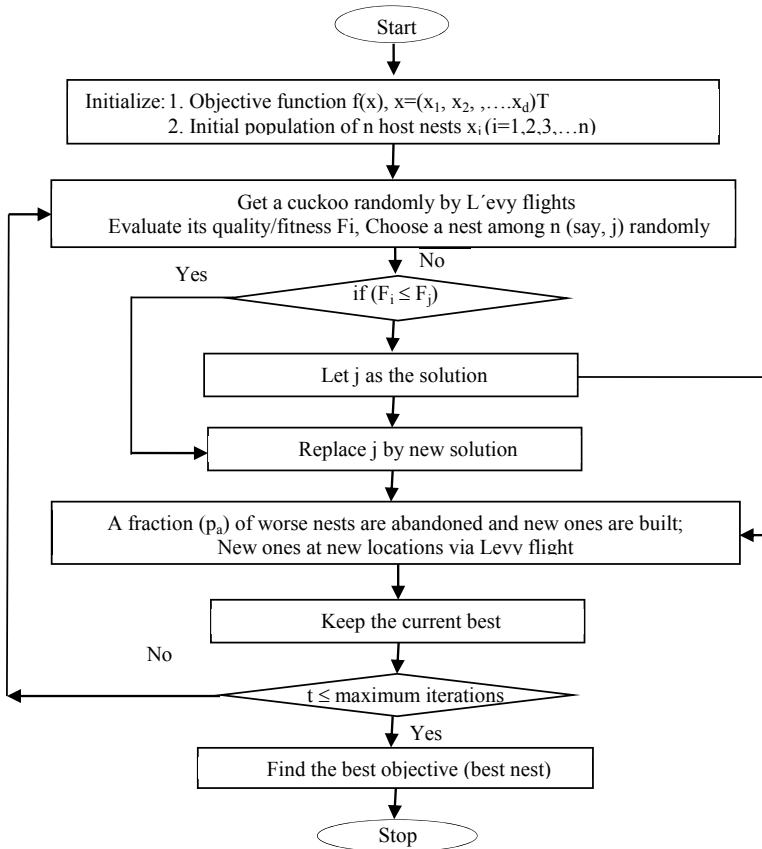


Fig. 1. Flow chart for Cuckoo search algorithm

### 3 Controller Structure

The complete transfer function model of the gasifier can be represented in the form as:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} & G_{15} \\ G_{21} & G_{22} & G_{23} & G_{24} & G_{25} \\ G_{31} & G_{32} & G_{34} & G_{34} & G_{35} \\ G_{41} & G_{42} & G_{43} & G_{44} & G_{45} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} + \begin{bmatrix} G_{d1} \\ G_{d2} \\ G_{d3} \\ G_{d4} \end{bmatrix} \times d \tag{3}$$

Where,  $G_{ij}$ =transfer function from  $i^{th}$  input to  $j^{th}$  output;  $y_1, y_2, y_3$  and  $y_4$  = Outputs;  $u_1, u_2, u_3, u_4$  and  $u_5$  = Inputs;  $d$  =sink pressure (PSink);

Limestone flow rate is set to 10% of coal flow rate and thus the process can be reduced to 4X4 MIMO process for control purpose. For a multivariable process, decentralised control schemes are usually preferred. The structure of decentralised controller used in gasifier control. It employs three PI controllers and one feedforward+feedback controller for coal flow rate.

$$G_c(s) = \begin{pmatrix} 0 & \left(K_p + \frac{1}{\tau_i s}\right) & 0 & 0 \\ K_f & 0 & K_p & 0 \\ 0 & 0 & 0 & \left(K_p + \frac{1}{\tau_i s}\right) \\ \left(K_p + \frac{1}{\tau_i s}\right) & 0 & 0 & 0 \end{pmatrix} \quad (4)$$

Gasifier process fails to satisfy the constraints [1] at 0% load operating point with the given controller structure (i.e. PGAS exceeds the limit of ±0.1bar). This major drawback can be rectified by retuning the controller parameters of the baseline controller.

### 4 Problem Formulation

Figure 2 shows the proposed scheme for Cuckoo search based retuning of pressure loop PI controller. The objective of this scheme is to meet the performance requirements at 0% load conditions.

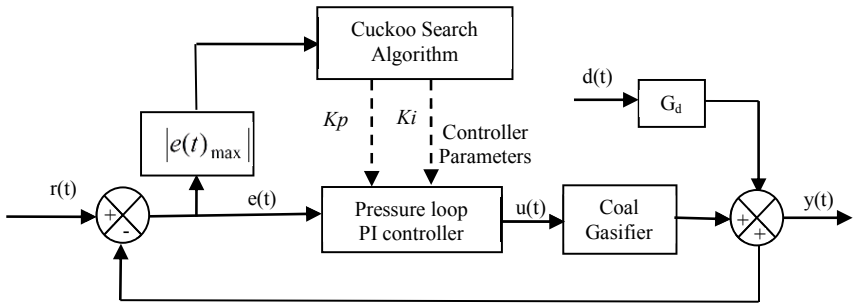


Fig. 2. Block diagram of Optimization scheme

Since control at 0% load is difficult the authors choose pressure loop PI controller to retune its parameters at 0% operating point. Proportional gain ( $K_p$ ) and integral time( $T_i$ ) of Pressure loop PI controller are taken as decision variables while the maximum Absolute Error (AE) at 0% operating point and sinusoidal pressure disturbance is considered as the objective function. Input constraints are associated with Simulink model and it is not included in the desired specifications. The controller should respond quickly enough so that the output variables do not deviate from the set point more than the specified constraints. Hence the sampling time is selected as 0.5 seconds. With the above settings Cuckoo search algorithm is executed, and maximum Absolute Error is calculated. Optimum controller settings are obtained after running the simulation for several times. The obtained proportional gain and integral time of decentralised PI controller and default controller parameters [1], provided with the challenge pack is shown in table 1.

**Table 1.** Controller parameters

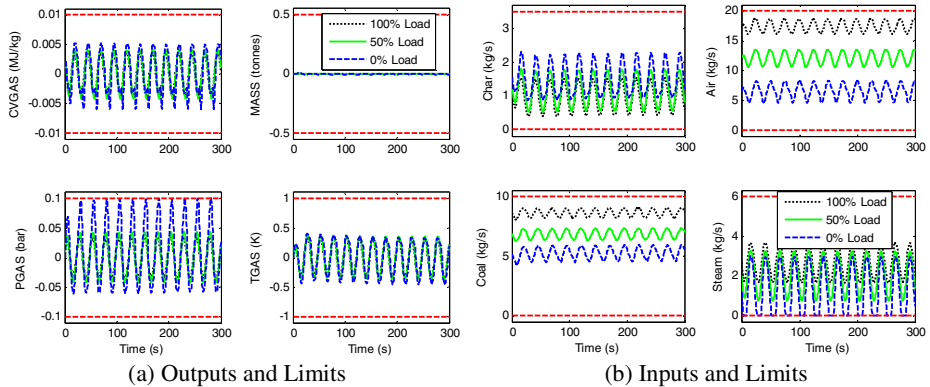
Parameter	Dixon-PI	Cuckoo search based PI controller
Pr_Kp	0.00020189	0.000354275
Pr_Ki	2.64565668e-05	7.010235151e-08

## 5 Performance Tests

Robustness of Cuckoo search based decentralised PI controller is verified by conducting performance tests (pressure disturbance, load change and coal variation). The requirement is that the response should meet the constraints [1] at 0%, 50% and 100% operating points for all performance tests.

### 5.1 Pressure Disturbance Tests

At 100% load a sinusoidal pressure disturbance (PSink) with a magnitude of 0.2bar and frequency of 0.04Hz, is applied to the gasifier. Maximum Absolute Error (AE) and Integral of Absolute Error (IAE) are calculated over a period of 300 second. This procedure is repeated for 50% and 0% operating points. Figure 3 shows the response of gasifier Cuckoo search based decentralised PI controller at 0%, 50% and 100% loads for sinusoidal pressure disturbance. All the outputs oscillate around their steadystate values and the magnitudes of these outputs are similar to [1] except for PGAS where the magnitude is reduced appreciably. The outputs meet the performance requirements comfortably, but with the existing baseline PI controller, PGAS violates the constraints at 0% load for sinusoidal pressure disturbance.



**Fig. 3.** Response to sinusoidal disturbance at 0%, 50% and 100% load

Above procedure is repeated for step disturbance with a magnitude of 0.2 bar and further analysis is carried out. Figure 4 shows the response of gasifier with Cuckoo search based decentralised PI controller at 0%, 50% and 100% loads for step pressure disturbance. The shown outputs are the deviation from the steadystate values.

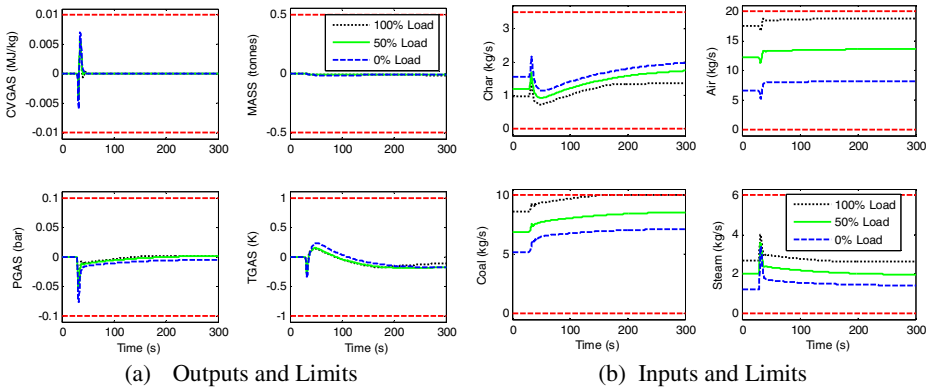


Fig. 4. Response to step disturbance at 0%, 50% and 100% load

Table 2. Summary of test output results

Test Description	Output	Maximum Absolute Error		IAE	
		Cuckoo-PI	Dixon-PI[1]	Cuckoo-PI	Dixon-PI[1]
100% Load, Step Disturbance	CVGAS	5422.38	4885.23	63878.92	60989.48
	MASS	6.94	6.94	1546.51	1597.03
	PGAS	4439.46	5018.94	173884.61	78475.47
	TGAS	0.26	0.24	62.98	65.09
50% Load, Step Disturbance	CVGAS	5864.79	5102.16	68379.15	64766.48
	MASS	8.45	8.45	887.24	840.04
	PGAS	5262.17	5790.93	226809.25	94310.73
	TGAS	0.29	0.27	73.78	77.13
0% Load, Step Disturbance	CVGAS	7048.03	5875.95	86780.39	86561.16
	MASS	11.05	11.05	1122.64	1330.92
	PGAS	7597.60	7714.53	497654.43	120167.73
	TGAS	0.36	0.32	70.52	77.05
100% Load, Sinusoidal Disturbance	CVGAS	3694.67	4101.30	1381309.32	1545471.04
	MASS	10.68	10.89	4156.20	4154.65
	PGAS	3471.73	4981.41	1297515.31	1857629.38
	TGAS	0.34	0.38	121.76	134.44
50% Load, Sinusoidal Disturbance	CVGAS	4238.37	4715.68	1569922.39	1759740.23
	MASS	12.61	12.87	5039.63	5041.36
	PGAS	4304.03	6209.91	1603092.22	2307614.42
	TGAS	0.38	0.42	134.70	149.47
0% Load, Sinusoidal Disturbance	CVGAS	5904.44	5869.69	1987706.29	2074977.65
	MASS	16.21	16.35	6178.99	6016.65
	PGAS	9995.70	11960.42	2831464.84	3845931.81
	TGAS	0.45	0.48	155.57	159.09

All the outputs meet the performance requirements comfortably. Marginal changes in other output variables are also observed. This is due to the interactions among the control loops i.e., changes in one input variable affects all the outputs. Table 2 shows the performance indices for sinusoidal and step pressure disturbance tests. All the outputs meet the performance requirement comfortably without violating the constraints.

## 5.2 Load Change Test

Stability of the gasifier and controller function across the working range of the plant is verified by load change test. For this purpose the system is started at 50% load in steady state and ramped it to 100% over a period of 600 seconds (5% per minute). The actual load, CVGAS and PGAS track their demands quickly to setpoint while Bedmass takes more time to reach its steady state, though manipulated inputs coal flow and char flow have reached their steady state immediately.

## 5.3 Coal Quality Test

The quality of syngas depends on the coal quality (carbon content and moisture content). In this test, the quality of coal increased and decreased by 18% (the maximum possible change in coal quality), and the above pressure disturbance test are conducted to verify the robustness of the controller. Input-output responses for sinusoidal and step change in PSink are verified for 300 seconds. Table 3 shows the violation of the variables under positive and negative change in coal quality. Since input constraints are inbuilt in the actuator limits, output constraints are considered to be the actual violation. TGAS and PGAS violate the limits under change in coal in coal quality for sinusoidal pressure disturbance and no output variable is found for step pressure disturbance

**Table 3.** Violation variables under coal quality change ( $\pm 18\%$ ) ( $\uparrow$  - the variable reaches its upper limit,  $\downarrow$  the variable reaches its lower limit)

Load	100%		50%		0%	
	Sine	Step	Sine	Step	Sine	Step
Coal quality increase (+18%)	Char $\downarrow$ Tgas $\uparrow$	Char $\downarrow$	Char $\downarrow$ Tgas $\uparrow$	Within limits	Char $\downarrow$ WStm $\downarrow$ Pgas $\uparrow$	Within limits
Coal quality decrease (-18%)	Coal $\uparrow$ Tgas $\downarrow$	Coal $\uparrow$	Within limits	Coal $\uparrow$	Char $\uparrow$ WStm $\downarrow$ Pgas $\uparrow$	Char $\uparrow$

## 6 Conclusion

This paper uses Cuckoo search algorithm to retune the parameters of decentralised PI controller for pressure loop of Coal gasifier. Existing controller with tuned parameters does not satisfy the performance requirements at 0% load for sinusoidal disturbance. And hence optimal tuning parameters are obtained using Cuckoo search algorithm. Pressure loop PI controller parameters are replaced by obtained controller parameters and performance tests are conducted. Pressure disturbance test shows excellent results and meets the performance requirement satisfactorily even at 0% load. Load change test and coal quality tests also provide good results. For the allowable limits of coal quality variations ( $\pm 18\%$ ), test results shows that the Cuckoo search based decentralised PI controller provides good results. Finally it can be concluded that Cuckoo search can be used to get the optimum controller parameters results and further the response can be improved by the use of Multi-Objective Cuckoo search Algorithm.



**Acknowledgement.** The authors would like to thank Dr. Roger Dixon, Director of Systems Engineering Doctorate Centre, Head of Control Systems Group, Loughborough University, UK for useful communication through email, Xin-She Yang, Senior Research Scientist, National Physical Laboratory, London, UK, Managements of St. Joseph's College of Engineering, Chennai and Sri Krishna College of Engineering & Technology, Coimbatore for their support.

## References

1. Dixon, R., Pike, A.W.: Alstom Benchmark Challenge II on Gasifier Control. IEE Proceedings - Control Theory and Applications 153(3), 254–261 (2006)
2. Chin, C.S., Munro, N.: Control of the ALSTOM gasifier benchmark problem using H2 methodology. Journal of Process Control 13(8), 759–768 (2003)
3. Al Seyab, R.K., Cao, Y., Yang, S.H.: Predictive control for the ALSTOM gasifier problem. IEE Proceedings - Control Theory and Application 153(3), 293–301 (2006)
4. Al Seyab, R.K., Cao, Y.: Nonlinear model predictive control for the ALSTOM gasifier. Journal of Process Control 16(8), 795–808 (2006)
5. Nobakhti, A., Wang, H.: A simple self-adaptive Differential Evolution algorithm with application on the ALSTOM gasifier. Applied Soft Computing 8(1), 350–370 (2008)
6. Agustriyanto, R., Zhang, J.: Control structure selection for the ALSTOM gasifier benchmark process using GRDG analysis. International Journal of Modelling, Identification and Control 6(2), 126–135 (2009)
7. Tan, W., Lou, G., Liang, L.: Partially decentralized control for ALSTOM gasifier. ISA Transactions 50(3), 397–408 (2011)
8. Sivakumar, L., Anitha Mary, X.: A Reduced Order Transfer Function Models for Alstom Gasifier using Genetic Algorithm. International Journal of Computer Applications 46(5), 31–38 (2012)
9. Kotteeswaran, R., Sivakumar, L.: Lower Order Transfer Function Identification of Nonlinear MIMO System-Alstom Gasifier. International Journal of Engineering Research and Applications 2(4), 1220–1226 (2012)
10. Huang, C., Li, D., Xue, Y.: Active disturbance rejection control for the ALSTOM gasifier benchmark problem. Control Engineering Practice 21(4), 556–564 (2013)
11. Griffin, I.A., Schroder, P., Chipperfield, A.J., Fleming, P.J.: Multi-objective optimization approach to the ALSTOM gasifier problem. Proceedings of IMechE Part I: Journal of Systems and Control Engineering 214(6), 453–469 (2000)
12. Xue, Y., Li, D., Gao, F.: 'Multi-objective optimization and selection for the PI control of ALSTOM gasifier problem. Control Engineering Practice 18(1), 67–76 (2010)
13. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms. Luniver Press (2008)
14. Yang, X.S.: Biology-derived algorithms in engineering optimization (Chapter 32). In: Olariu, Zomaya (eds.) Handbook of Bioinspired Algorithms and Applications. Chapman & Hall / CRC (2005)
15. Yang, X.-S., Deb, S.: Engineering optimization by Cuckoo Search. Int. J. Math. Model Numerical Optimisation 1(4), 330–343 (2010)
16. Valian, E., Mohanna, S., Tavakoli, S.: Improved Cuckoo Search Algorithm for Global Optimization. International Journal of Communications and Information Technology 1(1), 31–44 (2011)

# Multi Objective Particle Swarm Optimization for Software Cost Estimation

G. Sivanageswara Rao, Ch.V. Phani Krishna, and K. Rajasekhara Rao

Department of Computer Science & Engineering, KL University,  
Guntur, India

**Abstract.** Planning, monitoring-control and termination activities are classified as Software Project Management. The most vital activity in project management is planning which states the resources required to complete the project successfully. To complete the project successfully Software Cost Estimation is very important. Software Cost Estimation is the process of predicting the cost and time required. The basic input for the software cost estimation is coding size and set of cost drivers, the output is Effort in terms of Person-Months (PM's). In this paper, we have proposed a model for tuning parameters of COCOMO model Software Cost Estimation using Multi Objective (MO) Particle Swarm Optimization. The parameters of model tuned by using MOPSO considering two objectives Mean Absolute Relative Error and Prediction. The dataset COCOMO is considered for testing the model. It was observed that the model we proposed gives better results when compared with the standard COCOMO model. It is also observed, when provided with enough classification among training data may give better results.

**Keywords:** KDLOC-thousands of delivered lines of code, PM- person months, PSO- particle swarm optimization, COCOMO- constructive cost estimation, MO- Multi Objective.

## 1 Introduction

Software Engineering is a systematic approach to the development, maintenance and retirement of software. The resources include the number and skill level of the people, and the amount of computing resources [6, 8]. Cost for a project is a function of many parameters. Size is a primary cost factor in most models and can be measured using lines of code (or) thousands of delivered lines of code (KDLOC) or function points. The A number of models have been evolved to establish the relation between Size and Effort for Software Effort Estimation. There are two major types of cost estimation methods: algorithmic and non-algorithmic. Algorithmic models vary widely in mathematical sophistication. Some are based on simple Arithmetic formulas using such summary statistics as means and standard deviations. Others are based on regression models and differential equations [7, 20, and 21]. Some of the famous models are COCOMO [1], SLIM, Function Point, Price to win and Delphi model. The parameters of the algorithms are tuned using Genetic Algorithms[17], Fuzzy

models[10,19,24], Soft-Computing Techniques[9,22], Computational Intelligence Techniques, Heuristic Algorithms, Neural Networks[14,15], Radial Basis[23], MO Genetic Algorithm[3] and Regression. In this paper the parameters are tuned by using Multi Objective Particle Swarm Optimization for Software Effort Estimation.

## 2 Background

This section discusses the COCOMO model and Multi Objective PSO for fine tuning parameters in software effort estimation.

### 2.1 Constructive Cost Model (COCOMO)

[Boehm, 1981] described COCOMO as a collection of three variants, they are Basic model, Intermediate model, and Detailed model. The Basic COCOMO Model computes effort E as function of program size, and it is same as single variable method [7, 15, 20, and 21]. The Effort calculated using the following equation

$$\text{Effort} = a * (\text{size})^b \quad (1)$$

Where a and b are the set of values depending on the complexity of software (for organic projects a=2.4, b=1.05, for semi-detached a=3.0, b=1.1.2 and for embedded a=3.6, b=1.2).

An Intermediate COCOMO model effort is E is function of program size and set of cost drivers or effort multipliers. The Effort calculated using the following equation

$$\text{Effort} = a * (\text{size})^b * \text{EAF} \quad (2)$$

Where a and b are the set of values depending on the complexity of software (for organic projects a=3.2, b=1.05, for semi-detached a=3.0, b=1.1.2 and for embedded a=2.8, b=1.2) and EAF (Effort Adjustment Factor) which is calculated using 15 cost drivers. Each cost driver is rated from ordinal scale ranging from low to high. The Effort calculated using the following equation

$$\text{Effort} = a * (\text{size})^b * \text{EAF} * \text{sum (WI)} \quad (3)$$

COCOMO called as COCOMO II. It is a collection of three variants, Application composition model, early design model, and Post architecture model.

### 2.2 Multi Objective Particle Swarm Optimization

#### Particle Swarm Optimization (PSO)

Swarm Intelligence (SI) is an innovative distributed intelligent paradigm for solving optimization problems that originally took its inspiration from the biological examples by swarming, flocking and herding phenomena in vertebrates. Particle Swarm Optimization (PSO)[ Dr. Russell C. Eberhart and Dr. James Kennedy in 1995]

[11,12,13,16,18] incorporates swarming behaviors observed in flocks of birds, schools of fish, or swarms of bees, and even human social behavior, from which the idea is emerged. The modifications of the particles positions can be mathematically modeled according to the following equations:

$$V_i^{k+1} = V_i^k + c_1 * \text{rand}() * (Pbest - S_i^k) + c_2 * \text{rand}() * (Gbest - S_i^k) \tag{4}$$

$$S_i^{k+1} = S_i^k + V_i^{k+1} \tag{5}$$

Where,  $S_i^k$  is current search point,  $S_i^{k+1}$  is modified search point.,  $V_i^k$  is the current velocity ,  $V_i^{k+1}$  is the modified velocity,  $V_{pbest}$  is the velocity based on  $Pbest$  ,  $V_{gbest}$  = velocity based on  $Gbest$ ,  $c_j$  is the weighting factors.  $\text{Rand}()$  are uniformly distributed random numbers between 0 and 1.

**Multi Objective PSO**

A general single-objective optimization problem is defined as minimizing (or maximizing)  $f(x)$  subject to  $g_i(x) \leq 0, i = \{1 \dots m\}$ , and  $h_j(x) = 0, j = \{1 \dots p\} x \in \Omega$ . A solution minimizes (or maximizes) the scalar  $f(x)$  where  $x$  is an  $n$ -dimensional decision variable vector  $x = (x_1, \dots, x_n)$  from some universe  $\Omega$ . Observe that  $g_i(x) \leq 0$  and  $h_j(x) = 0$  represent constraints that must be fulfilled while optimizing (minimizing or maximizing)  $f(x)$ .  $\Omega$  contains all possible  $x$  that can be used to satisfy an evaluation of  $f(x)$  and its constraints. Many (may be most) real-world problems involve the optimisation of two or more objectives A multi objective optimization is defined as  $U = [U_1 U_2 \dots U_n]$  Where  $U$  is the control variable vector,  $n$  is the no of control variable and Objective function is  $\text{Min / Max } F = \{ f_1(U), f_2(U), \dots, f_m(U) \}$  Subject to  $G_j(U) \leq 0, j=1,2,\dots,m; L_j(U) = 0, j=1,2,\dots,p$ . In order to make single objective, each objective has some weight, combine the objectives into single weighted formula

$$W_1 * f_1(U) + W_2 * f_2(U) + \dots + W_m * f_m(U) \tag{6}$$

and normalize the weights using

$$W_1 + W_2 + \dots + W_m = 1. \tag{7}$$

**3 Proposed Methodology for Software Effort Estimation**

The following section introduces the methodology that has been used on the proposed model in order to tune the parameters. Earlier the COCOMO parameters are tuned by using regression analysis. The proposed model parameters of the COCOMO are tuned by using Particle Swarm Optimization with two objectives, which are Mean Absolute Relative Error (MARE) and Prediction (n). The parameters should minimize the MARE and Maximize the Prediction accuracy. The objectives defined as

$$\% \text{ MARE} = \text{mean} \left[ \frac{\text{abs}(\text{Measured Effort} - \text{Estimated Effort})}{(\text{measured effort})} \right] \times 100 \quad (8)$$

Prediction (n) is number of projects having less than n% error in the measured value.

### 3.1 Methodology

The Methodology/ Algorithm is used to tune the parameters are

Step 1: Start

Step 2: Initialize the m particles random position and velocity vectors  $[P_1, P_2 \dots P_m]$  and  $[V_1, V_2 \dots V_m]$  respectively for parameters to be tuned.

Step 3: Initialize all the particles as Pbest particles.

Step 4: Evaluate the two fitness functions  $f_1(U)$ ,  $f_2(U)$  using equations 7 and 8 for all the particles. The objective of  $f_1(U)$  is minimization and objective of  $f_2(U)$  is also maximization.

Step 5: This step converts Multi Objective into Single Objective by using weighted ranking method. For each two objectives assign ranks for all the particles. Add the ranks of objectives assigned to each particle. Final fitness is minimization.

Step 6: if fitness (p) better than fitness (Pbest) then Pbest = p.

Step 7: set the best of Pbest as a Gbest.

Step 8: update the particles velocities using the equations 4 and 5.

Step 9: repeat the steps 4 to 8 until particles exhaust that is no change in the objectives.

Step 10: Give the Gbest value parameters as optimal solution.

### 3.2 Proposed Model

We considered intermediate COCOMO model for tuning parameters. The proposed model is

$$\text{Effort} = a * (\text{Size})^b * \text{EAF} + c \quad (9)$$

Where a, b are cost parameters and c is bias factor. Size is coding size measured in KDLOC and Effort is in terms of Person Months (PM's) In order to tune the

parameters the above methodology multi objective particle swarm optimization is used.

#### 4 Implementation and Performance Measures

The following section describes the experimentation part of work, and in order to conduct the study and to establish the affectivity of the models two datasets of 20 projects and 21 projects from COCOMO dataset were used. We have implemented the above methodology for tuning parameters a, b and c in “C” language. The performance measures considered is equations 6 and 7. By running the “C” implementation of the above methodology we obtain the following parameters for the proposed model.

**Table 1.** Estimated Efforts of Proposed Models using MOPSO

Project No	Size	EA F	Measured Effort	COCOMO Effort	Estimated Effort MOPSO Model	COCOMO Error %	MOPSO Error %
1	46	1.17	240	212	237	13.207	1.699
2	16	0.66	33	39	37.2	15.384	11.26
4	6.9	0.4	8	9.8	10	18.367	19.67
10	24	0.85	79	108	77	26.851	2.690
14	1.9	1.78	9	10.7	9	15.887	0.111
21	2.14	1	7.3	7	6.8	4.2857	6.725
22	1.98	0.91	5.9	5.8	6.1	1.7241	3.752
28	34	0.34	47	44	49	6.8181	3.983
30	6.2	0.39	8	8.4	8.9	4.7619	10.01
31	2.5	0.96	8	8.9	7.5	10.112	6.241
32	5.3	0.25	6	4.7	6	27.659	0
33	19.5	0.63	45	46	45	2.1739	0
38	23	0.38	36	33	34.2	9.0909	5.293
42	8.2	1.9	41	55	45.1	25.454	9.171
43	5.3	1.15	14	22	17.5	36.363	20.09
44	4.4	0.93	20	14	12.2	42.857	63.93
49	21	0.87	70	68	66.8	2.9411	4.743
51	28	0.45	50	47	50.5	6.3829	1.048
52	9.1	1.15	38	42	32.1	9.5238	18.56
53	10	0.39	15	17	14	11.764	7.296

##### *Experiment 1.*

Totally 20 projects are considered. Number of iterations 100, Number of particles considered are 50. The inertia weight  $w$  is taken to be 0.5. and the weighting factors  $c1$  and  $c2$  are identically taken to be 2.0 from the literature.  $a=1.538113$ ,  $b=1.270503$  and

$c=2.800148$ . The range of  $a$  is  $[1, 10]$   $b$  is  $[-5, 5]$  and  $c$  is  $[-5, 5]$ . **Table 1** shows estimated effort of our proposed models.

#### Experiment 2:

Totally 21 projects are considered. Number of iterations 100, Number of particles considered are 50.  $a=3.960064$ ,  $b=1.103581$  and  $c=-5.420986$ . The range of  $a$  is  $[1, 10]$   $b$  is  $[-5,5]$  and  $c$  is  $[-5,5]$ . Table 2 shows estimated effort of our proposed models.

**Table 2.** Estimated Efforts of Proposed Models using MOPSO

Project No	Size	EAF	Measured Effort (Actual)	COCOMO Effort	Estimated Effort-MOPSO	COCOMO Error %	MOPSO Error %
1	46	1.17	240	212	311.44	8.990326	22.94
2	16	0.66	33	39	50.3	11.92624	34.40
3	4	2.22	43	30	35.17	36.95862	22.24
4	6.9	0.4	8	9.8	7.92	22.69974	0.887
5	22	7.62	107	869	908.96	22.66312	18.26
6	30	2.39	423	397	398.43	6.525573	6.166
7	18	2.38	321	214	223.44	47.88725	43.66
8	20	2.38	218	243	251.66	9.933995	13.37
9	37	1.12	201	238	233.11	15.87182	13.77
10	24	0.85	79	108	106.85	27.13894	26.06
11	3	5.86	73	60	72.58	17.90941	0.568
12	3.9	3.63	61	52	59.12	15.22093	3.164
13	3.7	2.81	40	38	41.72	4.793031	4.139
14	1.9	1.78	9	10.7	8.89	19.11694	1.207
15	75	0.89	539	443	407.98	23.53037	32.11
16	90	0.7	453	326	392.19	32.38168	15.50
17	38	1.95	523	430	422.29	22.02246	23.84
18	48	1.16	387	339	323.84	14.82191	19.50
19	9.4	2.04	88	89	90.35	1.106741	2.606
20	13	2.81	98	133	183.26	19.0982	46.52
21	2.14	1	7.3	7	3.74	8.003391	99.77

## 5 Results and Discussions

For the Experiment-1 we considered the small projects which size less than 50 KDLOC. The MARE and Prediction accuracy is good. For the Experiment-2 we considered the large projects, the MARE and Prediction accuracy is good in some cases which is a limitation of multi objective. The results are tabulated in Table 3. It was observed that the model may gives better results and when provided with enough classification among training data set.

**Table 3.** Performance and Comparisons

<b>Experiment-1</b>	<b>COCOMO</b>	<b>MOPSO Proposed Model</b>
MARE	16.1306	9.0143
Prediction(25%)	20	24
<b>Experiment-2</b>		
MARE	18.1548	20.9717
Prediction(25%)	17	15

## 6 Conclusion and Future Scope

The accuracy of the cost estimation model is measured in terms of its error rate. In this paper new model was proposed to estimate the software cost. In order to tune the parameters the multi objective particle swarm optimization methodology algorithm is applied. It is observed from the results that MOPSO gives better results. On testing the performance of the model in terms of the MARE and Prediction the results were found to be useful. It is also noticed the presence of non-linearity in the data items being considered during the present work for training and testing the tuning parameters and the best way to bring in some linearity among such data items is through clustering techniques. By using clustering method the data items may be divided into number of clusters and PSO be used then for parameter tuning of each cluster. These clusters and tuned parameters may be trained on Neural Networks by using efficient back propagation algorithms and results be compared for improvements as a part of future work with a view to generate even better models for software development effort estimation.

## References

1. Lawrence, H.P.: A general empirical solution to the macro software sizing and estimating problem. *IEEE Transactions on Software Engineering* SE-4(4), 345–361 (1978)
2. Joos, H.-D., et al.: A multi-objective optimisation-based software environment for control systems design. In: *2002 IEEE International Symposium Computer Added Control System Design Proceedings*, September 18-20, pp. 7–14 (2002)
3. Kavita, C.: GA based Optimization of Software Development effort estimation. *IJCST* 1(1), 38–40 (2010)
4. Rodríguez, D.: Multi Objective simulation optimization in software project management. In: *ACM 978-1-4503-0557-0/11/07, GECCO 2011*, July 12-16 (2011)
5. Laumanns, M.: Evolutionary Multi Objective optimization. *International Journal of Computational Intelligence Research* 2 (2006) ISSN 0973-1873
6. John, W.B., Victor, R.B.: A meta model for software development resource expenditures. In: *Proceedings of the Fifth International Conference on Software Engineering*, pp. 107–129 (1981), doi:CH-1627-9/81/0000/0107500.75@IEEE
7. Chris, F.K.: An Empirical Validation of Software Cost Estimation Models. *Management of Computing-Communications of ACM* 30(5), 416–429 (1987)
8. Rajiv, D.B., Chris, F.K.: Scale Economies in New Software Development. *IEEE Transactions on Software Engineering* 15(10), 1199–1205 (1989)



9. Krishna Murthy, S., Douglas, F.: Machine Learning Approaches to estimating software development effort. *IEEE Transactions on Software Engineering* 21(2), 126–137 (1995)
10. Pedrycz, W., Peters, J.F., Ramanna, S.: A Fuzzy Set Approach to Cost Estimation of Software Projects. In: *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*, May 9-12, pp. 1068–1073 (1999)
11. Wu, B., Zheng, Y., Liu, S., Shi, Z.: CSIM: A Document Clustering Algorithm Based on Swarm Intelligence, pp. 477–482 (2002), doi:DOI:0-7803-7282-4/02@IEEE
12. Wei, P., Kang-ping, W., Chun-guang, Z., Long-jiang, D.: Fuzzy Discrete Particle Swarm Optimization for Solving Traveling Salesman Problem. In: *Proceedings of the Fourth International Conference on Computer and Information Technology (CIT 2004)*, pp. 1–5 (2004), doi:0-7695-2216-5/04
13. Matthew, S.: *An Introduction to Particle Swarm Optimization*, Department of Computer Science, November 7, pp. 1–8. University of Idaho (2005)
14. Nasser, T.: Neural Network Approach for Software Cost Estimation. In: *Proceedings of the International Conference on Information Technology: Coding and Computing, ITCC 2005* (2005), doi:0-7695-2315-3/05IEEE
15. Xishi, H., Danny, H., Jing, R., Luiz, F.C.: Improving the COCOMO model using a neuro-fuzzy approach. *Elsevier-Applied Soft Computing* 7(2007), 29–40 (2005), doi:10.1016/j.asoc.2005.06.007
16. Ajith, A., He, G., Hongbo, L.: *Swarm Intelligence: Foundations, Perspectives and Applications*. In: Abraham, A., et al. (eds.) *Swarm Intelligent Systems*. SCI, pp. 3–25. Springer, Heidelberg (2006)
17. Alaa, F.S.: Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Projects. *Journal of Computer Science* 2(2), 118–123 (2006)
18. Chan, F.T.S., Tiwari, M.K.: *Swarm Intelligence: Focus on Ant and Particle Swarm Optimization*, pp. 1–548. I-TECH Education and Publishing (2007) ISBN 978-3-902613-09-7
19. Harish, M., Pradeep, B.: Optimization Criteria for Effort Estimation using Fuzzy Technique. *CLEI Electronic Journal* 10(1), 1–11 (2007)
20. Magne, J., Martin, S.: A Systematic Review of Software Development Cost Estimation Studies. *IEEE Transactions on Software Engineering* 33(1), 33–53 (2007)
21. Rahul, P., Thomas, Z.: Building Software Cost Estimation Models using Homogenous Data. In: *IEEE First International Symposium on Empirical Software Engineering and Measurement*, pp. 393–400 (2007), doi:0-7695-2886-4/07
22. Alaa, S., David, R., Aladdin, A.: Development of Software Effort and Schedule Estimation Models Using Soft Computing Techniques. *IEEE Transaction*, 978-1-4244-1823-7/08/IEEE, 1283–1289 (2008)
23. Prasad Reddy, P.V.G.D., Sudha, K.R., Rama, S.P., Ramesh, S.N.S.V.S.C.: Software Effort Estimation using Radial Basis and Generalized Regression Neural Networks. *Journal of Computing* 2(5), 87–92 (2010)
24. Prasad Reddy, P.V.G.D., Sudha, K.R., Rama, S.P., Ramesh, S.N.S.V.S.C.: Fuzzy Based Approach for Predicting Software Development Effort. *International Journal of Software Engineering* 1(1), 1–11 (2010)

# An Axiomatic Fuzzy Set Theory Based Feature Selection Methodology for Handwritten Numeral Recognition

Abhinaba Roy, Nibaran Das, Ram Sarkar, Subhadip Basu,  
Mahantapas Kundu, and Mita Nasipuri

Computer Science and Engineering Department,  
Jadavpur University,  
Kolkata-700032, India

abhinabaroy1990@gmail.com,  
{nibaran,rsarkar,subhadip,mahantapas,mnasipuri}@cse.jdvu.ac.in

**Abstract.** A new feature selection methodology on the basis of features' combined class separability power, using the framework of Axiomatic Fuzzy Set (AFS) theory has been proposed here. The AFS theory provides the rules for logic operations needed to interpret the combinations of features from the fuzzy feature set. Based on these combinational rules, class separability power of the combined features is determined and subsequently the most powerful subset of the feature set is selected. The performance of this methodology is evaluated upon for recognition of handwritten numerals of five popular Indic scripts viz. *Bangla*, *Devanagari*, *Roman*, *Telugu* and *Arabic* with SVM based classifier using gradient based directional feature set and quad-tree based longest-run feature set separately and compared with six widely used feature selection techniques. From the experimental results, it has been found that the methodology provides higher recognition accuracies with lesser or equal numbers of features selected for each dataset.

**Keywords:** Feature selection, Axiomatic Fuzzy Set Theory, Handwritten character recognition.

## 1 Introduction

Identification of appropriate feature set, a priori for a classification problem is a challenging task to researchers. Existing feature selection methods are mainly categorized into two broad groups. In the first group of methods, the merits of the features are evaluated first with respect to the data set using some learning algorithms. The second group of methods use some heuristics based techniques to judge the merit of the features before selection. The former is referred as wrappers and the latter as filters [1, 2]. Although wrappers have been used for feature selection and region selection [3], filters are more preferred. Most prominent subset selection techniques in this category concentrate on feature similarity [4], correlation [5], minimal-redundancy-maximal-relevance [6]. For problems where result of classification holds the key to performance of the entire system, such as in OCR, the redundancy of

features do not deteriorated the performance of the system as much as the features affecting in a negative way. In cases like these, the class separability can be deemed as the main criterion. But selection of features depending upon the class separability power of the features has not been addressed much. There are very few works in literature using entropy measurements such as in [7, 8], where the class separability power of the individual feature is judged by using fuzzy entropy measurements by some ranking. But, it is a known fact that good individual features do not necessarily serve as good features when combined together [9]. In this work we focus on how combinations of individual features affect upon their class separation power and then go on to select best possible feature subset among them.

Checking for the possible combinational results of the features is lengthy and difficult task due to the granularity of data for expressing the data as a whole [10]. One useful approach that has been used by researchers [10, 11] is to fuzzify the features using some fuzzy linguistic terms such as small, large etc. Each fuzzy feature corresponds to a fuzzy set, denoted by a membership function and reduced granularity, thus providing us with a better interpretation of some local information of the original feature.

Axiomatic Fuzzy Set (AFS) theory [12, 13] is a new approach related to the semantic interpretations of fuzzy attribute. In AFS theory, fuzzy sets (membership functions) and logical operations on them are algorithmically determined according to the distributions of original data and the semantics of the fuzzy sets. The AFS framework supports the studies on how to convert the information in available data into the membership functions and their fuzzy logic operations. This aspect of AFS theory has been used in mining fuzzy association rules [14], generating class based fuzzy classification rules [15], multiple criteria decision making [16]. Fuzzification of original features along with application of AFS theory on them to implement association rules on those fuzzy features provides an effective means of indirectly measuring the combinational effects of the original features. This has been applied in our present approach. Basic overview of AFS theory can be found in [12] which has been used in our present methodology discussed in the next section.

## 2 Overview of AFS Theory

### 2.1 AFS Algebra

The following example is used to illustrate AFS algebra.

**Example 1:** Let  $X = \{x_1, x_2, \dots, x_{10}\}$  be a set of 10 employees with some features which are described by real numbers, categorical/boolean values and ranks. It is shown in **Table 1**.

Any fuzzy concept on  $X$  may associate to one or more features. For instance, the fuzzy concept “rich” associates a single feature “wealth” and the fuzzy concept “old high skilled males” associates three features “age”, “skill” and “male/female”. The concepts associated with a single feature are viewed as fuzzy

(or numeric) linguistic terms of the corresponding feature. For instance, the fuzzy concepts (fuzzy linguistic terms) “old” and “about 40 years old” associate to the feature “age”.

**Table 1.** A Sample distribution of six features within a set of ten employees

Employee	Age	wealth	Male/Female	skill	Rank	Credit
X <sub>1</sub>	25	0	1	prim	4 <sup>th</sup>	0
X <sub>2</sub>	19	0	0	prim	6 <sup>th</sup>	0
X <sub>3</sub>	50	34	0	high	3 <sup>rd</sup>	1
X <sub>4</sub>	80	80	1	none	1 <sup>st</sup>	1
X <sub>5</sub>	34	2	1	med	2 <sup>nd</sup>	0
X <sub>6</sub>	37	28	0	high	7 <sup>th</sup>	1
X <sub>7</sub>	45	90	1	med	5 <sup>th</sup>	1
X <sub>8</sub>	70	45	1	none	2 <sup>nd</sup>	1
X <sub>9</sub>	60	98	0	med	4 <sup>th</sup>	1
X <sub>10</sub>	31	0	0	none	8 <sup>th</sup>	0

Let  $M = \{m_1, m_2, \dots, m_8\}$  be a set of fuzzy terms on  $X$  and each  $m \in M$  refers to a single feature. Where,  $m_1$ : “old employee”,  $m_2$ : “rich employee”,  $m_3$ : “high work skill employee”,  $m_4$ : “young employee”,  $m_5$ : “the employee about 45 years old”,  $m_6$ : “male”,  $m_7$ : “female” (i.e., not male),  $m_8$ : “employee with credit”.  $M$  can be viewed as elementary terms of the corresponding features. From these, we can form complex terms such as  $\gamma = m_1 m_6 + m_1 m_3 + m_2$  which represents the concept of “old male employees” or “high work skill old employees” or “rich employees”. The general form is  $\sum_{i \in I} (\prod_{m \in A_i} m)$ ,  $A_i \subseteq M$ .

**Definition 1:** ([13]) Let  $M$  be a non-empty set. Then the set  $EM^*$  is defined by

$$EM^* = \{ \sum_{i \in I} (\prod_{m \in A_i} m) \mid A_i \in 2^M, i \in I, \} \tag{1}$$

where,  $I$  may be any nonempty index set.

A binary relation  $R$  is defined as, for,  $\sum_{i \in I} (\prod_{m \in A_i} m)$ ,  $\sum_{j \in J} (\prod_{m \in B_j} m) \in EM^*$ ,  $[\sum_{i \in I} (\prod_{m \in A_i} m) R \sum_{j \in J} (\prod_{m \in B_j} m)] \leftrightarrow$

$$\forall A_i (i \in I) \exists B_h (h \in J) \text{ such that } A_i \supseteq B_h$$

$$\forall B_j (j \in J) \exists A_k (k \in I) \text{ such that } B_j \supseteq A_k$$

It is clear that  $R$  is an equivalence relation. The quotient set  $EM^*/R$  is denoted by  $EM$ . The  $\sum_{i \in I} (\prod_{m \in A_i} m)$  and  $\sum_{j \in J} (\prod_{m \in B_j} m)$  are equivalent under  $R$ . Thus the semantics they represent are equivalent. Each  $\sum_{i \in I} (\prod_{m \in A_i} m) \in EM$  is called a fuzzy concept.

## 2.2 AFS Structures of the Data

An AFS structure, a triple  $(M, \tau, X)$ , gives rise to various lattice representations of the membership degrees and fuzzy logic operations of the concepts in  $EM$ .

**Definition 2:** Let  $X, M$  be sets and  $2^M$  be the power set of  $M$ . Let  $\tau: X \times X \rightarrow 2^M$ .  $(M, \tau, X)$  is called an AFS structure if  $\tau$  satisfies the following axioms:

$$AX1: (x_1, x_2) \in X \times X, \tau(x_1, x_2) \subseteq \tau(x_1, x_1);$$

$$AX2: \forall (x_1, x_2), (x_2, x_3) \in X \times X, \tau(x_1, x_2) \cap \tau(x_2, x_3) \subseteq \tau(x_2, x_3);$$

Here,  $X$  is called the universe of discourse,  $M$  is called the concept set and  $\tau$  is called the structure.

In real world applications,  $\tau$  can be defined as follows.

$$\tau(x, y) = \{m | m \in M, xR_my\} \in 2^M \tag{2}$$

where  $R_m$  represents binary relation of simple concept  $m \in M$ , and  $xR_my$  means the degree of  $x$  belonging to attribute  $m$  is larger than or equal to that of  $y$ .

For a fuzzy concept  $\xi = \sum_{i \in I} (\prod_{m \in A_i} m) \in EM$  the membership function of  $\xi$  is defined as,

$$\mu_\xi(x) = \sup_{i \in I} (\prod_{\gamma \in A_i} \mathcal{M}_\gamma(A_i^\tau(x))) \tag{3}$$

where,  $A^\tau(x) = \{y \in X | \tau(x, y) \supseteq A\}$  and  $A^\tau(x)$  Is the set of all elements in  $X$  whose degree of belonging to  $\prod_{m \in A} m$  is less than or equal to that of  $x$ , and  $\mathcal{M}_\gamma(\cdot)$  is a function whose value lies between 0 to 1.

## 3 Present Work

It has already been discussed before and also in [11] that fuzzification of features make the feature selection problem simpler by providing better interpretability. In our present problem, we fuzzify the features in the following way, for every feature  $f \in F$ , where  $F$  is the set of all the features, there is a corresponding fuzzy feature  $m_f$  which is defined as “nearness of an element to its true class with respect to other classes due to feature  $f$ ”. Value of  $m_f$  for an element  $x \in X$  is denoted as  $m_f^x$  which is determined in the following way:

$$m_f^x = \frac{\text{spread}(C_x)/x_f - M_f^{c_x}}{\sum_{i \in C} (\text{spread}(i)/x_f - M_f^i)} \tag{4}$$

Where,  $x_f$  is the value of feature  $f$  of element  $x$ ,  $M_f^i$  is the mean of all the values of feature  $f$  for all the elements belonging to class  $i$ ,  $\text{spread}(i)$  is the measurement of spread of a class  $i$  (standard deviation is used in the present work), and  $c_x$  is the true class of the sample  $x$ , and  $C$  is the set of all the classes. It is to be noted that, too many fuzzy sets and too many fuzzy rules can make the fuzzy model difficult and increase the computational burden. So, fuzzy association of the original features to the

fuzzy features is a one to one mapping i.e.  $mF$  is the set of all the fuzzy features then  $|mF| = |F|$  i.e. dimensionality is not increased in the fuzzy feature space than in the original feature space.

In a system with more than two elements, overall interaction of all the elements is best indicated by the interaction between two elements and their combination with others. In a similar way here, we look for the interactions of any two fuzzy features and ultimately come up with the best possible subset of them which interact best among themselves.

Using AFS theory [12], two fuzzy features  $m_i, m_j \in mF$  are utilized here to calculate  $\mu_{m_i m_j}(x)$ . The  $\mu_{m_i m_j}(x)$  value indicates the apparent class-separation power by the combination of fuzzy features  $m_i$  and  $m_j$ . Another class separation power  $supp(m_i)$  is introduced here in the following way:

$$supp(m_i, m_j) = \frac{\sum_{x \in X} \mu_{m_i m_j}(x)}{|X|} \tag{5}$$

Now we create feature matrix  $E$  of size  $|mF| \times |mF|$ , where.  $E(i, j)$  is  $supp(m_i, m_j)$ .  $E(i, j)$  denotes elements of  $E$  in the  $i^{th}$  row and  $j^{th}$  column Since focus of our work is on the combinational effects of the (fuzzy) features, we make the value of diagonal elements, i.e.,  $E(i, i), i \in [1, |mF|]$  to zero. So, the effect of individual features (on itself) is nullified. Now, the selection of the best subset features which perform best among themselves is reduced to the problem of selecting  $k \times k$  maximum sum sub-matrix form the matrix  $E$  discussed in Algorithm 1.

**Algorithm 1**

---

```

For  $x \in X$  /* $X$  is the set of training samples */
Calculate  $m_f^x$  corresponding to  $x_f$ ; /* $f \in F$ (the set of features);  $m_f^x$  is the
fuzzy features for an element  $x^*$ /
For  $m_i, m_j \in mF$  /* $mF$  is the set of fuzzy features*/
For  $x \in X$ 
Calculate  $\mu_{m_i m_j}(x)$ ;
For  $m_i, m_j \in mF$ 
calculate  $supp(m_i, m_j)$ ;
Create matrix  $E$ ;
find maximum sum  $k \times k$  submatrix from  $|mF| \times |mF|$  matrix  $E$ ;
/*  $k$  is the required number of features to be selected; maximum sum
sub-matrix selection is constrained by the fact that when  $i^{th}$  row from
 $E$  is selected  $i^{th}$  column is also to be selected for consideration. */
    
```

---

We select the  $k \times k$  submatrix in such a way that if  $i^{th}$  row from the original matrix  $E$  is selected for consideration, then  $i^{th}$  has to be selected too (i.e.  $i^{th}$  feature is selected for consideration). Thus  $k$  fuzzy features (rows and columns of the original matrix  $E$ ) are the best selection of  $k$  fuzzy features. Since the fuzzy features are mapped to the original feature in a one-to-one manner, the instance of  $k$  fuzzy features coming out as the best selected features, signify that the corresponding

$k$  original features are performing optimally in terms of class separability under the given problem consideration.

## 4 Results and Discussions

To evaluate the strength and usefulness of our methodology we have implemented the method on numeral datasets of five popular Indic scripts *Bangla*, *Devanagari*, *Roman*, *Telugu* and *Arabic*. Except *Roman*, other datasets are freely available at [www.code.google.com/cmaterdb](http://www.code.google.com/cmaterdb). The Roman numeral database is created by randomly selecting 600 data samples from each of the 10 classes of numerals from MNIST dataset [17].

Two popularly used feature sets- 8 directional gradient (DG) features[18] and longest run (LR) features[19] are used for experimental purposes. The SVM classifiers with RBF kernel is used here with empirically chosen SVM parameters. For DG features, average recognition accuracy of  $95.15\% \pm 0.512\%$  are observed on the above mentioned five numeral databases. Using the AFS theory based feature selection (AFSFS) technique, an average improvement of  $1.57\%$  in recognition accuracy with a mean reduction in feature dimensionality by  $12.2\%$  are observed. For LR features, mean recognition of  $96.32\% \pm 1.135\%$  are observed with a  $14.6\%$  reduction in feature dimensionality. The present feature selection technique improves the mean recognition accuracy by  $1.35\%$  with  $16.6\%$  of mean reduction in feature dimensionality. Table 2 and Table 3 show the performance of the present AFSFS technique against that of six other most widely used ranking based features selection techniques viz. chi square (Ch. s.), correlation, gain ratio (Gainr.), information gain (Info. gain), symmetrical uncertainty (Symm. U.) and relief. A general comparison between the six ranking techniques can be found in [20]. These six techniques were implemented using WEKA [21], a data mining learning tool.

**Table 2.** Performance of AFSFS with DG features and comparison with related techniques

Dataset	Feature selection techniques							Feature Reduction
	Ch.s.	Correlation	Gainr.	Info. gain	Relief	Symm.U.	AFSFS	
<i>Bangla</i>	96.1	96.2	95.45	95.4	97.15	96.8	97.45	14%
<i>Devanagari</i>	95.9	96.0	96.1	96.1	96.2	96.1	96.5	17%
<i>English</i>	94.5	94.7	94.9	95.0	95.2	94.9	95.45	17%
<i>Telugu</i>	95.3	96.6	96.5	96.7	97.0	96.5	97.2	17%
<i>Arabic</i>	95.5	96.4	96.5	95.9	96.7	96.5	97.0	8%

**Table 3.** Performance of AFSFS with LR features and comparison

Dataset	Feature selection techniques							Feature Reduction
	Ch.s.	Correlation	Gainr.	Info. gain	Relief	Symm.U.	AFSFS	
<i>Bangla</i>	96.0	96.9	97.0	96.9	97.2	96.7	97.7	11%
<i>Devanagari</i>	95.8	95.9	95.9	95.9	96.0	96.1	96.9	20%
<i>English</i>	97.2	96.9	97.1	97.0	97.1	97.2	97.45	20%
<i>Telugu</i>	98.2	98.9	98.5	98.2	99.0	98.3	99.4	11%
<i>Arabic</i>	95.0	95.3	95.0	96.6	96.9	96.1	97.5	20%

## 5 Conclusion

A novel methodology for assessing of the combinational effects of more than one feature on class separability is presented here with the help of AFS theory. As discussed earlier, a feature that performs well individually might not be a good combinatory with other features. Our approach tries to come out of that problem.

In this work, we have concentrated on measuring the effectiveness of the feature selection technique and not concentrating on comparison of recognition performances with the existing benchmarks in this domain. Also, it is worth mentioning in this context that we have not explored data preprocessing extensively. So superior preprocessing techniques and classifier combinations may be incorporated in feature order to attain better recognition accuracies for respective public-domain databases.

**Acknowledgment.** Authors are grateful to the “Center for Microprocessor Application for Training Education and Research”, “Project on Storage Retrieval and Understanding of Video for Multimedia” of Computer Science & Engineering Department, Jadavpur University laboratory, for providing infrastructure facilities during progress of the work.

## References

1. Tsamardinos, I., Aliferis, C.F.: Towards principled feature selection: Relevancy, filters and wrappers. In: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics. Morgan Kaufmann Publishers, Key West (2003)
2. Hall, M.A., Smith, L.A.: Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference, p. 239. AAAI Press (1999)
3. Roy, A., Das, N., Basu, S., Sarkar, R., Kundu, M., Nasipuri, M.: Region selection in handwritten character recognition using Artificial Bee Colony Optimization. In: EAIT, pp. 189–192 (2012)
4. Mitra, P., Murthy, C., Pal, S.K.: Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 301–312 (2002)
5. Hall, M.A.: Correlation-based feature selection for machine learning. The University of Waikato (1999)
6. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1226–1238 (2005)
7. Lee, H.-M., Chen, C.-M., Chen, J.-M., Jou, Y.-L.: An efficient fuzzy classifier with feature selection based on fuzzy entropy. *Trans. Sys. Man Cyber. Part B* 31, 426–432 (2001)
8. Luukka, P.: Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications* 38, 4600–4607 (2011)
9. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *The Journal of Machine Learning Research* 3, 1157–1182 (2003)
10. Rezaee, M.R., Goedhart, B., Lelieveldt, B., Reiber, J.: Fuzzy feature selection. *Pattern Recognition* 32, 2011–2019 (1999)



11. Li, Y., Wu, Z.F.: Fuzzy feature selection based on min–max learning rule and extension matrix. *Pattern Recognition* 41, 217–226 (2008)
12. Liu, X., Pedrycz, W.: *Axiomatic Fuzzy Set Theory and Its Applications*. STUDFUZZ, vol. 244. Springer, Heidelberg (2009)
13. Xiaodong, L.: The fuzzy theory based on AFS algebras and AFS structure. *Journal of Mathematical Analysis and Applications* 217, 459–478 (1998)
14. Wang, X., Liu, X., Pedrycz, W., Zhu, X., Hu, G.: Mining axiomatic fuzzy set association rules for classification problems. *European Journal of Operational Research* 218, 202–210 (2012)
15. Ren, Y., Liu, X., Cao, J.: A parsimony fuzzy rule-based classifier using axiomatic fuzzy set theory and support vector machines. *Information Sciences* 181, 5180–5193 (2011)
16. Tao, L., Chen, Y., Liu, X., Wang, X.: An integrated multiple criteria decision making model applying axiomatic fuzzy set theory. *Applied Mathematical Modelling* (2011)
17. <http://yann.lecun.com/exdb/mnist/>
18. Roy, A., Mazumder, N., Das, N., Sarkar, R., Basu, S., Nasipuri, M.: A new quad tree based feature set for recognition of handwritten Bangla numerals. In: 2012 IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA), pp. 1–6 (2012)
19. Das, N., Reddy, J.M., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M., Basu, D.K.: A statistical-topological feature combination for recognition of handwritten numerals. *Appl. Soft Comput.* 12, 2486–2495 (2012)
20. Roy, A., Das, N., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M.: A Comparative Study of Feature Ranking Methods in Recognition of Handwritten Numerals. In: IEEE International Conference on Signal Processing, Computing and Control (ISPCC), September 26–28 (2013)
21. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11, 10–18 (2009)

# Brownian Distribution Guided Bacterial Foraging Algorithm for Controller Design Problem

N. Sri Madhava Raja and V. Rajinikanth

Department of Electronics and Instrumentation, St. Joseph's College of Engineering,  
Sholinganallur, Chennai 600119, Tamilnadu, India

**Abstract.** Bacterial Foraging Optimization (BFO) algorithm is widely adopted to solve a variety of engineering optimization tasks. In this paper, the Brownian Distribution (BD) strategy guided BFO algorithm is proposed. During the optimization exploration, BD monitors and controls the chemotaxis operation of the BFO algorithm in order to enhance the search speed and optimization accuracy. In the proposed algorithm, after undergoing a chemotaxis step, each bacterium gets mutated by a BD operator. In the proposed work, this algorithm is employed to design the PID controller for an AVR system and unstable reactor models. The success of the proposed method has been confirmed through a comparative analysis with PSO, BFO, adaptive BFO and PSO + BFO based hybrid methods existing in the literature. The result shows that, for unstable reactor models, the BD guided BFO algorithm provides better optimization accuracy compared to other algorithms considered in this study.

**Keywords:** Bacterial Foraging Algorithm, Brownian Distribution, PID controller design, AVR system, unstable reactor.

## 1 Introduction

Traditional unconstrained optimization procedures such as steepest decent method, Newton's method and Quasi - Newton's method are widely considered to find solutions for a wide range of engineering optimization problems [1]. These methods sometime fail to provide optimal solutions for complex engineering problems such as nonlinear and nondifferential problems. Hence, a number of nature inspired metaheuristic algorithms are proposed by the scientists in recent years [1, 2].

In this work, we adopted the Bacteria Foraging Optimization (BFO) algorithm initially proposed by Passino [3]. In BFO, a group of artificial bacteria work together to find the best possible solutions in 'D' dimensional search space. In literature, a variety of techniques have been proposed by the researchers to enhance the performance of conventional BFO algorithm. Chen *et al.* proposed a Cooperative BFO) algorithm [4]. Recently Rajinikanth and Latha proposed an Enhanced BFO (EBFO) algorithm [5]. Above discussed modified BFO algorithms require large search time and convergence rate compared to Particle Swarm Optimization (PSO), Genetic Algorithm (GA) and Ant Colony Optimization (ACO) [5].

The drawbacks of BFO algorithm can be minimized with hybridization techniques. In hybrid algorithm, the chemotaxis operation (swim and tumble) of the BFO algorithm is mutated using other optimization methods in order to enhance the convergence rate and accuracy. Recently, a variety of hybridization techniques are proposed and implemented for a class of engineering optimization problems [6-10].

In the present work, a novel hybrid method is developed by combining the Brownian Distribution (BD) and BFO algorithm. In this, the BD controls the tumble operator in chemotaxis process in order to enhance the optimization accuracy. After undergoing a chemotaxis step, each bacterium gets mutated by a BD operator and offers enhanced result compared to traditional BFO algorithm. This algorithm is then employed to design the PID controller for an AVR system and an unstable bioreactor process.

## 2 Bacteria Foraging Algorithm

Bacterial Foraging Optimization (BFO) is a nature inspired stochastic search technique based on mimicking the foraging behavior of *E. coli* bacteria [3]. Due to the merits such as high computational efficiency, easy implementation and stable convergence, it is widely applied to solve a range of complex engineering optimization problems. A detailed theoretical analysis of BFO could be found in the literature [3-13]. In this work, the classical BFO algorithm is considered and the initial BFO parameters are assigned based on the paper by Rajinikanth and Latha [5]

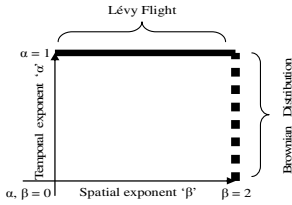
## 3 Random Walk Strategy

The optimization search process in recently developed nature inspired algorithms such as firefly algorithm and cuckoo search are guided by the random walk such as Lévy Flight (LF) strategy [2]. A detailed description of Lévy Flight (LF) and Brownian Distribution (BD) could be found in [2, 14, 15]. LF is a random walk, which consists of a sequence of arbitrary steps and is similar to the path traced by a molecule as it travels in a liquid or a gas, and the search path of a foraging animal [15]. The expression Lévy flight is used to describe specialized random walks in which the flight span, the length between two successive change of direction, are drawn from a probability distribution. The Lévy flight essentially provides a random walk while the random step length is drawn from a Lévy distribution; which has an infinite variance with an infinite mean. Brownian search is the simplest form of random walk, which can be obtained by altering the LF expression. Lévy flight is superdiffusive markovian process, whose step length is drawn from the Lévy distribution in terms of a simple power-law formula;

$$L(s) \sim |s|^{-1-\beta} \quad \text{where } 0 < \beta \leq 2. \quad (1)$$

The Brownian walk is a subdiffusive non markovian process, which obeys a Gaussian distribution with zero mean and time-dependent variance. The ratio of the exponents

$\alpha / \beta$  provides the relationship between sub and super diffusion. From Fig 1, the observation is that, when  $\beta < 2\alpha$  the undecided continuous random walk is completely superdiffusive, and for  $\beta > 2\alpha$  effectively subdiffusive. For  $\beta = 2\alpha$ , the search process exhibits the same scaling as ordinary Brownian motion [15].



**Fig. 1.** Relation between LF and BD

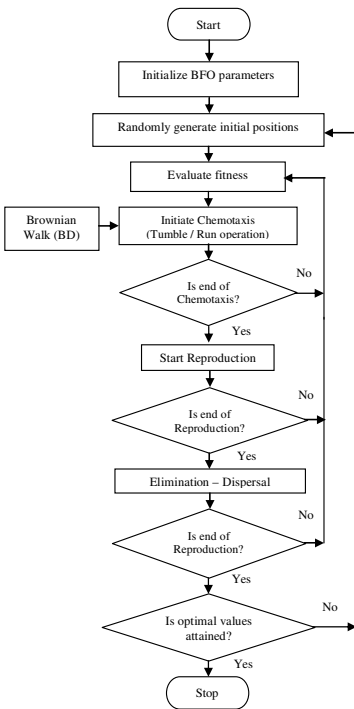
In this work, we adapted the LF expression (Eqn 2) considered by Gandomi *et al.* [16]. Eqn. 2 is modified to obtain the BD expression (Eqn. 3) based on the guidelines given in the literature [15].

$$L(s) = \beta \Gamma(\beta) \sin\left(\frac{\beta\pi}{2}\right) \frac{1}{\pi} \cdot |s|^{1/\beta} \tag{2}$$

$$B(s) = \beta \Gamma(\beta) \sin\left(\frac{\beta\pi}{2}\right) \frac{1}{\pi} \cdot |s|^{\alpha/2} \tag{3}$$

where  $\beta$  is the spatial exponent with the  $0 < \beta \leq 2$ ,  $\alpha$  is the temporal exponent with the range  $0 < \alpha \leq 1$ , and  $\Gamma(\beta)$  is a Gamma function.

### 4 BD Guided BFO Algorithm



**Fig. 2.** Flowchart of proposed algorithm

In existing hybridization methods such as GA guided BFO and PSO guided BFO, it is essential to assign the initial algorithm parameters. Choosing initial parameters such as number of agents, cost function, and total number of iterations required for both the algorithms are a challenging job. In hybrid method, the GA / PSO algorithm is used to speed up the process and the BFO is responsible to provide the optimal value. The drawbacks such as trapping in local optima and premature convergence in GA / PSO may degrade the performance of BFO. Hence, it is necessary to integrate a non-algorithmic technique, to improve the performance of classical BFO algorithm. Nurzaman *et al.* [14] reported that, the search strategy of E. coli bacteria can be representation in the form of LF and BD. The report also states that, the BD has larger target density compared to LF. Hence, in this work, we constructed a novel hybrid method by combining the BD with the BFO algorithm. The chemotaxis operation of the BFO algorithm can be expressed as in Eqn. 4 [3, 9, 11];

$$\theta^i(j+1, k, l) = \theta^i(j, k, l) + C(i) \frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \tag{4}$$

where  $\theta^i(j, k, l)$  shows  $i^{\text{th}}$  bacterium at  $j^{\text{th}}$  chemotactic,  $k^{\text{th}}$  reproductive and  $l^{\text{th}}$  elimination-dispersal step;  $C(i)$  is the step size in the random direction, and  $\Delta(i)$  is a random vector of size  $[-1, 1]$ .

In the proposed work, Eqn.4 is modified as shown below;

$$\theta^i(j+1, k, l) = \theta^i(j, k, l) + C(i) \frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \oplus B(s) \tag{5}$$

where the symbol  $\oplus$  represents the entry wise multiplication, and  $B(s)$  is the Brownian Distribution (BD) operator.

The proposed algorithm is employed to design the parallel form of PID controller. Mathematical representation of PID controller is given in Eqn 6.

$$\text{Parallel PID structure} = K_p \left( I + \frac{I}{\tau_i s} + \tau_d s \right) = \left( K_p + \frac{K_i}{s} + K_d s \right) \tag{6}$$

$$\text{where } : \tau_i = K_p / K_i, \tau_d = K_d / K_p.$$

A weighted sum of multiple objective function is considered as given in Eqn 7 with two parameters, such as  $ISE$  and  $M_p$ .

$$J(K_p, K_i, K_d) = (w_1 \cdot ISE) + (w_2 \cdot M_p) \tag{7}$$

where  $w_1 = w_2 = 1$ .

Fig 2 shows the flow chart of the proposed algorithm. In this, the Brownian strategy controls the tumble operator in chemotaxis process in order to enhance the search speed and also the optimization accuracy. After undergoing a chemotaxis step, each bacterium in this algorithm gets mutated by a BD operator and offers enhanced result compared to traditional BFO algorithm. The BD guided bacteria constantly explores the search space until the objective function is minimized. When optimal values are obtained, the algorithm terminates the search and displays the values.

## 5 Results and Discussions

In this section, the capability of the proposed technique is demonstrated with the PID Controller design problem for an Automatic Voltage Regulator (AVR) and unstable reactor models. A comparative analysis is also carried with PSO, classical BFO, and PSO + BFO based hybrid algorithm existing in the literature [9].

### Example 1: Automatic Voltage Regulator (AVR)

An Automatic Voltage Regulator (AVR) widely studied in the literature is considered [8, 17, 18]. Fig 3 shows the block diagram of a benchmark AVR system.

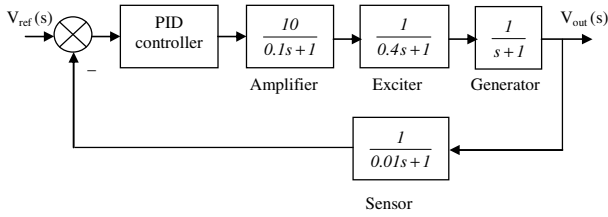


Fig. 3. Block diagram of a benchmark AVR system

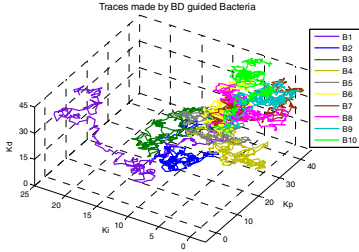


Fig. 4. Convergence of BD guided BFO algorithm in three dimensional space

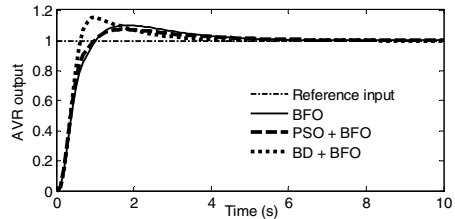


Fig. 5. Reference tracking for AVR

Table 1. Optimal values and its performance evaluation

Process	Algorithm	Controller parameter			Iteration	ISE	M <sub>p</sub>
		K <sub>p</sub>	K <sub>i</sub>	K <sub>d</sub>			
AVR	BFO	0.8518	0.5076	0.3183	152	0.331	0.102
	PSO+BFO	0.9152	0.4663	0.2977	124	0.307	0.075
	BD+PSO	0.8295	0.6627	0.1996	119	0.312	0.155
Bioreactor	PSO	-0.4066	-0.0501	-0.1197	71	11.58	0.808
	BFO	-0.5374	-0.0702	-0.0537	85	5.900	0.751
	ABFO	-0.6113	-0.0714	-0.0518	66	5.704	0.720
	PSO+BFO	-0.7388	-0.1904	-0.2519	52	3.089	0.772
	BD+PSO	-0.8115	-0.2108	-0.4174	61	2.621	0.607
CSTR	BFO	1.5935	0.0133	9.0072	224	51.05	0.888
	PSO+BFO	1.4006	0.0137	7.9550	177	55.09	0.853
	BD+PSO	1.7042	0.0098	9.1113	182	49.07	0.879

The aim is to design an optimal PID controller, which maintains the AVR output based on the reference value. The search boundary for controller parameter is assigned as: K<sub>p</sub> : min 0% to max 50%; K<sub>i</sub> : min 0% to max 25%; and K<sub>d</sub> : min 0% to max 50%. Five independent runs are performed with the BFO, PSO+BFO and BD+BFO and the best value among the search is considered as optimal value.

The algorithms continuously adjust the parameters K<sub>p</sub>, K<sub>i</sub>, and K<sub>d</sub> until the objective function J (K<sub>p</sub> K<sub>i</sub> K<sub>d</sub>) is minimised to J<sub>min</sub> (K<sub>p</sub> K<sub>i</sub> K<sub>d</sub>). Fig 4 shows the search traces made by the BD guided bacteria in a three dimensional search space. When iteration

increases, all the bacteria reach to the best possible value in the search universe. Table 1 shows the final controller values and their appraisal. From Table 1 and Fig 5, the observation is that the search time (number of iterations) taken by the BD+BFO is smaller compared to the conventional BFO and PSO + BFO based hybrid algorithm. The overshoot produced by the proposed method based PID controller is slightly larger than other methods. But, the proposed method improves the rise time and settling time of the process compared to BFO and hybrid algorithm based controller.

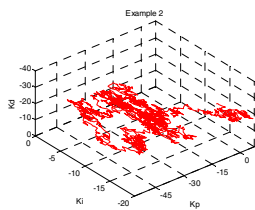
**Example 2:**

**(a) Unstable Bioreactor**

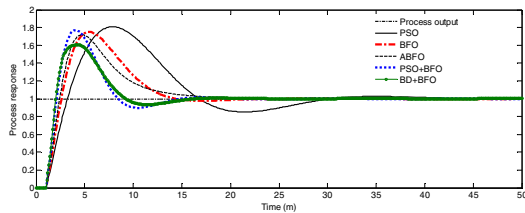
An unstable bioreactor model widely discussed in the literature is considered [5, 12]. The problem is to design a PID controller to regulate the feed concentration in order to maintain the product concentration based on the reference input. The unstable second order model of the process is provided in Eqn. 8;

$$G(s) = \frac{-0.9951s - 0.2985}{s^2 + 0.1302s - 0.0509} \cdot e^{-1s} \tag{8}$$

For the model, the following search boundary is assigned for controller parameters:  $K_p$ : min -50% to max 0%;  $K_i$ : min -12% to max 0%; and  $K_d$ : min -40% to max 0%. Five independent runs are executed with the proposed and existing algorithms and the best value among the search is presented in Table 1. A relative analysis is also presented with the existing PID values discussed in the paper by Rajinikanth and Latha [5]. Fig 6 (a) shows the traces made by a single bacterium on a bounded three dimensional search space for BD+BFO algorithm. It shows that, when iteration increases, all the bacteria move towards the optimal value.



(a) Traces by bacterium



(a) Servo response

**Fig. 6.** Algorithm performance on bioreactor model

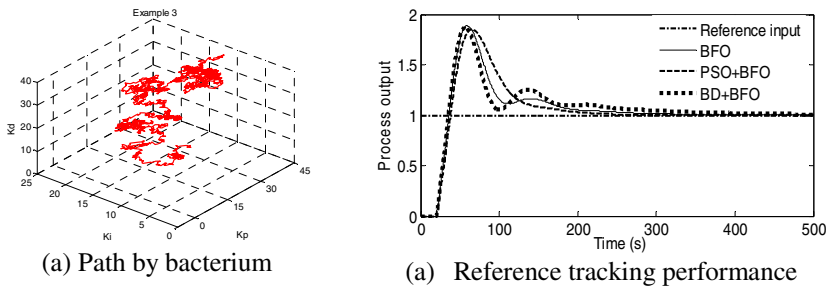
Fig 6 (b) depicts the reference tracking performance. From this figure and Table 1, the observation is that the proposed BD+BFO provides superior values for ISE, overshoot, rise time, and settling time compared to existing PSO, BFO, ABFO and PSO+BFO algorithms.

### (b) Isothermal Continuous Stirred Tank Reactor (CSTR)

The unstable process model of an isothermal Continuous Stirred Tank Reactor (CSTR) model is discussed in [5, 12, 13] is given in Eqn. 9;

$$G(s) = \frac{3.3226}{(99.69s - 1)} \exp^{-20s} \quad (9)$$

During the optimization exploration, the three dimensional search boundary for PID controller is assigned as:  $K_p$  : min 0% to max 50%;  $K_i$  : min 0% to max 25%; and  $K_d$  : min 0% to max 50% . Five runs are performed and the best value is presented in the Table 1. Fig 7 (a) depicts the traces made by a single BD guided bacterium during the optimization search.



**Fig. 7.** Algorithm performance on CSTR model

Fig 7 (b) shows the reference tracking response for the isothermal CSTR model. The PID controller tuned using the proposed method provides improvement in rise time, settling time and ISE value compared to classical BFO and PSO+BFO based hybrid algorithm.

## 6 Conclusion

In this article, a novel method has been proposed to improve the performance of conventional Bacterial Foraging Optimization (BFO) algorithm. A random walk strategy known as the Brownian Distribution (BD) is employed to guide the BFO algorithm in order to improve its convergence speed and the optimization accuracy. A comparative analysis is performed between other heuristic algorithms such as PSO, BFO, ABFO, and PSO + BFO. From this analysis, it is noted that, the PID controller designed using the proposed method provides satisfactory result on AVR system and better result on the unstable reactors compared to other optimization algorithms considered in this work.



## References

1. Liu, G.P., Yang, J.-B., Whidborne, J.F.: *Multiobjective Optimization and Control*. Prentice Hall, New Delhi (2008)
2. Yang, X.-S.: *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, UK (2008)
3. Passino, K.M.: Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Systems Magazine* 22(3), 52–67 (2002)
4. Chen, H., Zhu, Y., Hu, K.: Cooperative Bacterial Foraging Optimization. *Discrete Dynamics in Nature and Society* 2009, Article ID 815247, 17 pages (2009), doi:10.1155/2009/815247
5. Rajinikanth, V., Latha, K.: Controller Parameter Optimization for Nonlinear Systems Using Enhanced Bacteria Foraging Algorithm. *Applied Computational Intelligence and Soft Computing* 2012, Article ID 214264, 12 pages (2012), doi:10.1155/2012/214264
6. Pandi, V.R., Biswas, A., Dasgupta, S., Panigrahi, B.K.: A hybrid bacterial foraging and differential evolution algorithm for congestion management. *Euro. Trans. Electr. Power* 20(7), 862–871 (2010), doi:10.1002/etep.368
7. Ganesan, T., Vasant, P., Elamvazuthy, I.: A hybrid PSO approach for solving non-convex optimization problems. *Archives of Control Sciences* 22(1), 87–105 (2012)
8. Kim, D.H.: Hybrid GA–BF based intelligent PID controller tuning for AVR system. *Applied Soft Computing* 11(1), 11–22 (2011)
9. Korani, W.M., Dorrah, H.T., Emara, H.M.: Bacterial foraging oriented by particle swarm optimization strategy for PID tuning. In: *Proceedings of the 8th IEEE International Conference on Computational Intelligence in Robotics and Automation*, pp. 445–450 (2008)
10. Anguluri, R., Abraham, A., Snasel, V.: A Hybrid Bacterial Foraging - PSO Algorithm Based Tuning of Optimal FOPI Speed Controller. *Acta Montanistica Slovaca* 16(1), 55–65 (2011)
11. Das, S., Biswas, A., Dasgupta, S., Abraham, A.: Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications. In: Abraham, A., Hassanien, A.-E., Siarry, P., Engelbrecht, A. (eds.) *Foundations of Computational Intelligence Volume 3*. SCI, vol. 203, pp. 23–55. Springer, Heidelberg (2009)
12. Rajinikanth, V., Latha, K.: Bacterial Foraging Optimization Algorithm based PID controller tuning for Time Delayed Unstable System. *The Mediterranean Journal of Measurement and Control* 7(1), 197–203 (2011)
13. Rajinikanth, V., Latha, K.: Setpoint weighted PID controller tuning for unstable system using heuristic algorithm. *Archives of Control Sciences* 22(4), 481–505 (2013), doi:10.2478/v10170-011-0037-8
14. Nurzaman, S.G., Matsumoto, Y., Nakamura, Y., Shirai, K., Koizumi, S.: From Lévy to Brownian: A Computational Model Based on Biological Fluctuation. *PLoS ONE* 6(2), e16168 (2011), doi:10.1371/journal.pone.0016168
15. Metzler, R., Klafter, J.: The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports* 339(1), 1–77 (2000)
16. Gandomi, A.H., Yang, X.-S., Talatahari, S., Alavi, A.H.: Firefly algorithm with chaos. *Commun. Nonlinear Sci. Numer. Simulat.* 18(1), 89–98 (2013)
17. Mukherjee, V., Ghoshal, S.P.: Intelligent particle swarm optimized fuzzy PID controller for AVR system. *Electric Power Systems Research* 77(12), 1689–1698 (2007)
18. Pan, I., Das, S.: Frequency domain design of fractional order PID controller for AVR system using chaotic multi-objective optimization. *Electrical Power and Energy Systems* 51, 106–118 (2013)

# Trapezoidal Fuzzy Shortest Path (TFSP) Selection for Green Routing and Scheduling Problems

P.K. Srimani<sup>1</sup>, G. Vakula Rani<sup>2</sup>, and Suja Bennet<sup>2</sup>

<sup>1</sup> Dept. of CS & Maths, Bangalore University, Bangalore, India

pk\_srimani@gmail.com

<sup>2</sup> CMRIMS, Bangalore, India

hod.mca@ims.cmr.ac.in, sujben@yahoo.co.in

**Abstract.** The routing of vehicles represents an important component of many distribution and transportation systems. Finding the shortest path is one of the fundamental and popular problems. In real life applications, like vehicle green routing and scheduling, transportation, etc. which are related to environmental issues the arc lengths could be uncertain due to the fluctuation with traffic conditions or weather conditions. Therefore finding the exact optimal path in such networks could be challenging. In this paper, we discuss and analyze different approaches for finding the Fuzzy Shortest Path. The shortest path is computed using the ranking methods based on i) Degree of Similarity ii) Acceptable Index, where the arc lengths are expressed as trapezoidal fuzzy numbers. The Decision makers can choose the best path among the various alternatives from the list of rankings by prioritizing the scheduling which facilitates Green Routing.

**Keywords:** Fuzzy Shortest Path, Bellman's Dynamic Programming, Trapezoidal Fuzzy shortest path, Degree of Similarity, Acceptable Index.

## 1 Introduction

The vehicle routing and scheduling problems have been studied with much interest within the last four decades. During the last few years, Operations Research (OR) has extended its scope to include environmental applications. In the world, about 73% of the oil is used for transportation purposes. There is a need to design efficient plans for sustainable transportation. Advances in the transportation planning process and efficiency of transportation systems are the key components of the development of sustainable transportation. The routing of vehicles represents an important component of many distribution and transportation systems. The class of problems like routing and scheduling models that relate to environmental issues are known as Green Routing and Scheduling Problems (GRSP), which discuss different problems that relate to sustainable logistics, waste management [1] etc. Some variants of routing and scheduling problems in connection with environmental considerations are : (i) the arc routing problem, which is considered as a major component in waste management, and (ii) the time-dependent vehicle routing problem which allows one

to indirectly decrease gas emissions involved by transportation activity by avoiding congested routes. Fuzzy logic could be used successfully to model situations in a highly complex environment where a suitable mathematical model could not be provided. For modelling traffic and transportation processes characterized by subjectivity, ambiguity, uncertainty and imprecision, fuzzy logic approach happens to be a very promising mathematical approach. The use of fuzzy logic is advantageous in several situations especially in decision making processes where the description through algorithms is difficult and the associated criteria are multiplied.

The paper is organized as follows: Section 2, discusses the related work. Section 3 provides the prerequisites ie preliminary definitions and concepts required for the computation and analysis of fuzzy numbers. Various algorithms to compute the shortest path in a fuzzy network are demonstrated in Section 4. In Section 5, Experiments using these algorithms are conducted and the results are compared and presented graphically. Section 6 presents conclusions and provides an excellent platform for future work.

## 2 Related Work

The fuzzy shortest-path problem was first introduced in 1980 [2] , where the authors used Floyd's algorithm and Ford's algorithm to treat the fuzzy shortest path problem. Authors [4] proposed a dynamical programming recursion-based fuzzy algorithm. Authors [5] found the fuzzy shortest path length in a network by means of a fuzzy linear programming approach. Authors of [6] defined a new comparison index between the sums of fuzzy numbers by considering interactivity among fuzzy numbers and presented an algorithm to determine the degree of possibility for each arc on a network. Authors [7] developed two types of Fuzzy Shortest Path network problems, where the first type of Fuzzy Shortest Path Problem uses triangular fuzzy numbers and the second type uses level  $(1 - \beta, 1 - \alpha)$  interval valued numbers. The main result from their study was, the Shortest Path in the fuzzy sense correspond to the actual path in the network, and the Fuzzy Shortest Path Problem is an extension of the crisp case. [8] represented each arc length as a triangular fuzzy set and a new algorithm is proposed to deal with the FSPP. [10] studied the Fuzzy Shortest Path Problem Based on Degree of Similarity. Thus numerous papers have been published on the FSPP. In this paper, we discuss and analyze different approaches for finding the Fuzzy Shortest Path and illustrated by considering an example.

## 3 Pre-requisites

In this section, an overview of the fuzzy-set concepts and definitions are presented.

**Definition 1:** A *fuzzy number* is a quantity whose value is imprecise, rather than exact as is the case with "ordinary" (single-valued) numbers. Any fuzzy number can be thought of as a function whose domain is a specified set, usually the set of real numbers, and whose range is the span of non-negative real numbers between, and

including, 0 and 1. Each numerical value in the domain is assigned a specific "grade of membership" where 0 represents the smallest possible grade, and 1 is the largest possible grade. Let  $X = \{x\}$  be a universe, i.e. the set of all possible (feasible, relevant) elements to be considered. Then a fuzzy set (or a fuzzy subset)  $A$  in  $X$  is defined as a set of ordered pairs, where  $A = \{x, \mu_A(x), x \in X\}$  is the membership grade or degree of association of  $x$  in  $A$ , where 0 value indicates non belongingness and 1 indicates full belongingness.

**Definition 2:** A *fuzzy quantity* is defined as a fuzzy set in the real line  $R$ , that is, an  $R \rightarrow [0, 1]$  mapping  $A$ . If  $A$  is upper semi-continuous, convex, normal, and has bounded support, then it is called a fuzzy number.

## 4 Methodology

In this section, a detailed description about the method of analysis is presented which comprises the Shortest Path using i) Bellman Dynamic Programming for Crisp Arc Lengths(BDPCL) ii) Bellman Dynamic Programming for fuzzy arc lengths (BDPFL) and iii) Trapezoidal Fuzzy Shortest Path Selection using Degree of Similarity (DS) and Acceptable Index(AI).

### 4.1 Bellman Dynamic Programming for Crisp Arc Lengths (BDPCL)

Bellman ‘s Dynamic Programming for the shortest path problem can be formulated as follows: given a network with an acyclic directed graph  $G = (V, E)$  with  $n$  vertices numbered from 1 to  $n$  such that 1 is the source and  $n$  is the destination. The shortest path is given by

$$\begin{aligned}
 &f(n)=0 \\
 &f(i) = \min\{d_{ij} + f(j) \mid \langle i, j \rangle \in E\}
 \end{aligned}
 \tag{1}$$

Here  $d_{ij}$  is the weight of the directed edge  $i, j$ , and  $f(j)$  is the length of the shortest path from vertex  $i$  to vertex  $n$ .

### 4.2 Bellman Dynamic Programming for Fuzzy Arc Lengths (BDPFL)

Triangular fuzzy number  $A$  can be defined by a triplet  $(a, b, c)$  defined on  $R$  with the membership function defined as

$$\mu_A(x) = \begin{cases} (x - a)/(b - a), & a \leq x \leq b \\ (c - x)/(c - b), & b \leq x \leq c \\ 0, & \text{otherwise} \end{cases}
 \tag{2}$$

The edge weight in the network, denoted by  $d_{ij}$  and the edge weight should be expressed as triangular fuzzy number.

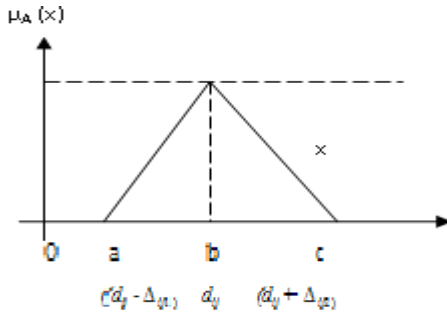


Fig. 1. Fuzzy number  $d_{ij}$

The Signed distance of (a, b) from the origin is given by  $1/2(a+b)$ . Similarly, the signed distance of (a, b, c) can be given by  $d^* = 1/4(a+2b+c) = d_{ij} + 1/4 \Delta_{ij}$  (3)  
 If  $\Delta_{ij1} = \Delta_{ij2}$ , thus the fuzzy problem becomes crisp.

### 4.3 Trapezoidal Fuzzy Shortest Path (TFSP)

The trapezoidal shape is originated from the fact that there are several points whose membership function (degree) is maximum (=1). Trapezoidal Fuzzy number A can be defined by a quadruplet (a, b, c, d) defined on R with the membership function defined as

$$\mu_A(x) = \begin{cases} (x - a)/(b - a), & a \leq x \leq b \\ 1, & b \leq x \leq c \\ (d - x)/(d - c), & c \leq x \leq d \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

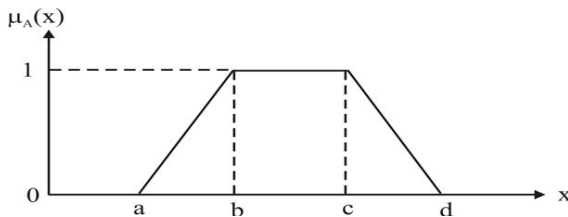


Fig. 2. Trapezoidal fuzzy number

Construct a fuzzy graph (network)  $G = (V,E)$  with pure fuzziness, where V is the set of vertices or nodes and E is the set of edges or arcs. Here G is an acyclic digraph. Let the arc length  $L_{ij}$  be the trapezoidal fuzzy numbers. Consider all possible paths  $P_i$  from the source vertex '1' to the destination vertex 'n' and find the corresponding path lengths  $L_i, i = 1, 2, \dots, n$  and set  $L_i = (a_i, b_i, c_i, d_i)$ . If  $b = c$ , the trapezoidal fuzzy number becomes the triangular fuzzy number.

**i) Degree of Similarity (DS):** Degree of Similarity is the measurement of the nearness degree of two fuzzy sets. Let  $d_{min}$  (Fuzzy Shortest Length) =  $A1 = (a, b, c, d)$  and  $L_i$  ( $i^{th}$  fuzzy path length) =  $A2 = (a_i, b_i, c_i, d_i)$  be two Triangular Fuzzy Numbers. If  $a \leq a_i, b \leq b_i, c \leq c_i$  and  $d < d_i$  then the *Degree of Similarity (DS)* between  $A1$  and  $A2$  can be calculated as follows:

$$SD(d_{min}, L_i) = \begin{cases} 0 & \text{if } L_i \cap d_{min} = \Phi \\ 1/2 \times (d - a_i)^2 / ((d - c) + (b_i - a_i)) & \text{if } L_i \cap d_{min} \neq \Phi \text{ where } c < x < d, a_i < x < b_i \\ 1/2 \times [(d - a_i) + (c - b_i)] & \text{if } L_i \cap d_{min} \neq \Phi \text{ where } b_i \leq x \leq c. \end{cases} \quad (5)$$

**ii) Acceptability Index (AI):** The *Acceptability Index (AI)* of the proposition  $L_{min} = (a, b, c, d)$  is preferred to

$L_i = (a_i, b_i, c_i, d_i)$  is given by

$$AI(L_{min} < L_i) = \frac{-(d + a_i)}{(c - d) - (b_i - a_i)} \quad (6)$$

Using this Acceptability Index we define the ranking order based on highest Acceptability Index, i.e., in Acceptability Index,  $L1 < L2$  iff  $AI(L_{min} < L1) > AI(L_{min} < L2)$ .

## 5 Experiments and Results

In this section , the experiment is carried out to predict the selection of shortest path using a)Crisp values b) Trapezoidal Fuzzy Crisp values and 3) Trapezoidal Fuzzy values. Techniques like like i) Degree of Similarity ii). Acceptable Index. are used to determine the rank of Trapezoidal Fuzzy Shortest Paths by considering the network as shown in the fig 3.

### 5.1 Bellman Dynamic Programming for Crisp Arc Lengths(BDPCL)

Consider a network Fig 3, Bellman–Ford algorithm computes the shortest paths in weighted directed graphs follows.

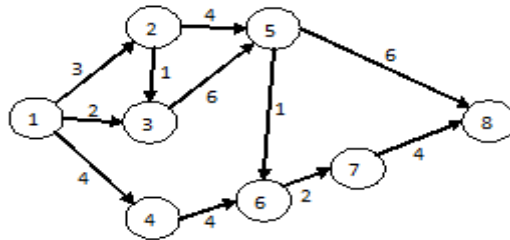


Fig. 3. Network

$$f(8) = 0 \qquad f(5) = \min \{ (d_{5j} + f(j) \mid \langle 5,j \rangle \in E \} = 6$$

$$f(7) = \min \{ (d_{7j} + f(j) \mid \langle 7,j \rangle \in E \} = d_{78} = 4 \qquad f(6) = \min \{ d_{67} + f(7) \} = 2 + 4 = 6$$

$7 <$

$$\begin{aligned}
 f(5) &= \min \{ d_{58}, d_{56} + f(6) \} = \min(6, 7) = 6 & f(4) &= \min \{ (d_{46} + f(6)) \} = 4 + 6 = 10 \\
 f(3) &= \min \{ (d_{35} + f(5)) \} = 12 & f(2) &= \min \{ d_{25} + f(5), d_{23} + f(3) \} = 10 \\
 f(1) &= \min \{ d_{12} + f(2), d_{13} + f(3), d_{14} + f(4) \} = 13
 \end{aligned}$$

Therefore the shortest path ( P1-> $d_{12} + d_{25} + d_{58}$  ) Length = 13

### 5.2 Bellman Dynamic Programming For Fuzzy Arc Lengths (BDPFL)

Considering our network in Figure 3, we get the following inequalities:

<p><b>Step1:</b> <math>d_{12} + f(2) &lt; d_{13} + f(3)</math>  <math>d_{12} + d_{25} + d_{58} &lt; d_{13} + d_{35} + d_{58}</math></p> <p><b>Step2:</b> <math>d_{12} + f(2) &lt; d_{14} + f(4)</math>  <math>d_{12} + d_{25} + d_{58} &lt; d_{14} + d_{46} + d_{67} + d_{78}</math></p>	<p><b>Step3:</b> <math>d_{25} + f(5) &lt; d_{23} + f(3)</math>  <math>d_{25} + d_{58} &lt; d_{23} + d_{35} + d_{58}</math></p> <p><b>Step4:</b> <math>d_{58} + f(8) &lt; d_{56} + f(6)</math>  <math>d_{58} &lt; d_{56} + d_{67} + d_{78}</math></p>
--	--

To express the uncertainty in edge weights we use trapezoidal fuzzy numbers. The values of parameters are chosen in such a way so as to satisfy the above inequalities and then the fuzzy arc lengths based on equation (3) are calculated: $\Delta_{12} = 0.5$   $\Delta_{23} = 0.9$   $\Delta_{13} = 1$   $\Delta_{67} = 1$   $\Delta_{56} = 1.5$   $\Delta_{46} = 1.3$   $\Delta_{58} = 0.8$   $\Delta_{78} = 0.9$   $\Delta_{35} = 1.2$   $\Delta_{14} = 0.8$   $\Delta_{25} = 1$   
 Defuzzified Arc Lengths:  $d^*_{12} = d_{12} + \Delta_{12} = 3.125$ ;  $d^*_{13} = d_{13} + \Delta_{13} = 2.25$ ;  $d^*_{14} = 4.2$ ;  $d^*_{25} = 4.25$ ;  $d^*_{23} = 1.225$ ;  $d^*_{35} = 6.3$ ;  $d^*_{46} = 4.325$ ;  $d^*_{56} = 1.375$ ;  $d^*_{58} = 6.2$ ;  $d^*_{67} = 2.25$ ;  $d^*_{78} = 4.225$ ;  $d^*_{46} = 4.325$ .

**Table 1.** Fuzzy Arc crisp Path Lengths

Paths	Path Lengths	Rank
P1: 1 – 2 – 5 – 8	$3.125 + 4.25 + 6.2 = 13.575$	1
P2: 1 – 3 – 5 – 8	$2.25 + 6.3 + 6.2 = 14.75$	2
P3: 1 – 4 – 6 – 7 – 8	$4.2 + 4.325 + 2.25 + 4.225 = 15$	3
P4: 1 – 3 – 5 – 6 – 7 – 8	$2.25 + 6.3 + 1.375 + 2.25 + 4.225 = 16.4$	5
P5: 1 – 2 – 5 – 6 – 7 – 8	$3.125 + 4.25 + 1.375 + 2.25 + 4.225 = 15.225$	4
P6: 1 – 2 – 3 – 5 – 8	$3.125 + 1.225 + 6.3 + 6.2 = 16.85$	6

### 5.3 Trapezoidal Fuzzy Shortest Path Length (TFSPL)

Consider the trapezoidal fuzzy numbers for the network are as follows: $d'_{12} = (2.8, 3, 3.7, 4.2)$ ;  $d'_{13} = (1.5, 2, 3.5, 6.2)$ ;  $d'_{14} = (3.8, 4, 5, 6.5)$ ;  $d'_{25} = (3, 4, 6, 8)$ ;  $d'_{23} = (0.7, 1, 2.2, 3)$ ;  $d'_{35} = (5.7, 6, 7.5, 10)$ ;  $d'_{46} = (3.8, 4, 5.5, 6)$ ;  $d'_{56} = (0.5, 1, 3, 5)$ ;  $d'_{58} = (5.7, 6, 7.1, 7.8)$ ;  $d'_{67} = (1, 2, 4, 6)$ ;  $d'_{78} = (3.8, 4, 5.5, 6.5)$

**Table 2.** The shortest path length procedure is based on the [4] method.

**Algorithm-1:**  
**Input:**  $L_i = (a_i', b_i', c_i', d_i')$ ,  $i=1, 2, \dots, n$  where  $L_i$  denotes the trapezoidal fuzzy length.  
**Output:**  $d_{min} = (a, b, c, d)$  where  $d_{min}$  denotes the FSL.  
**Step 1:** Form the set  $U$  by sorting  $L_i$  in ascending orders of  $b_i'$  and  $c_i'$ .  $U = \{U_1, U_2, \dots, U_n\}$  where  $U_i = (a_i', b_i', c_i', d_i')$ ,  $i=1, 2, \dots, n$   
**Step 2:** Set  $d_{min} = (a, b, c, d) = U_1 = (a_1', b_1', c_1', d_1')$   
**Step 3:** let  $i=2$   
**Step 4:** Calculate  $(a, b, c, d)$   
 $a = \min(a, a_i')$   

$$b = \begin{cases} b & b < a_i' \\ \frac{(b \times b_i') - (a \times a_i')}{(b + b_i') - (a + a_i')} & b > a_i' \end{cases}$$
  

$$c = \begin{cases} c & c < b_i' \\ \frac{(c \times c_i') - (b \times b_i')}{(c + c_i') - (b + b_i')} & c > b_i' \end{cases}$$
  
 $d = \min(d, d_i')$   
**Step 5:** Set  $d_{min} = (a, b, c, d)$   
**Step 6:** Set  $i = i + 1$   
**Step 7:**  $i < n+1$  goto Step 4  
**Step 8:** Compute the above procedure to get the Fuzzy Shortest Length (FSL)  
**Step 9:** Identify the SP with the highest Similarity Degree or highest Acceptable Index the between FSL and  $L_i$ , for  $i= 1$  to  $n$ .

All the possible paths and the corresponding trapezoidal fuzzy path lengths and the Fuzzy Shortest Length (FSL) are computed by using Algorithm-1 and presented in Table 3.

**Table 3.** Trapezoidal fuzzy Paths Lengths & FSL Computational Procedure

Paths	Path Lengths	FSL
$P_1: 1 - 2 - 5 - 8$	(11.5, 13, 16.8, 20)	$d_{min} = (11.5, 13, 16.8, 20)$
$P_2: 1 - 3 - 5 - 8$	(12.9, 14, 18.1, 24)	$d_{min} = (11.5, 12.94, 15.45, 18.1)$
$P_3: 1 - 4 - 6 - 7 - 8$	(12.4, 14, 20, 26)	$d_{min} = (11.5, 12.68, 15, 18.1)$
$P_4: 1 - 3 - 5 - 6 - 7 - 8$	(12.5, 15, 23.5, 33.7)	$d_{min} = (11.5, 12.62, 15, 18.1)$
$P_5: 1 - 2 - 5 - 6 - 7 - 8$	(11.1, 14, 22.2, 29.7)	$d_{min} = (11.1, 12.196, 14.77, 18.1)$
$P_6: 1 - 2 - 3 - 5 - 8$	(14.9, 16, 20.5, 25)	$d_{min} = (11.1, 12.196, 14.77, 18.1)$

**5.4 Results**

The shortest path selection can be done by using Degree of Similarity (eqn.5) or Acceptable Index (eqn.6). It has been identified with the highest DS or AI. The

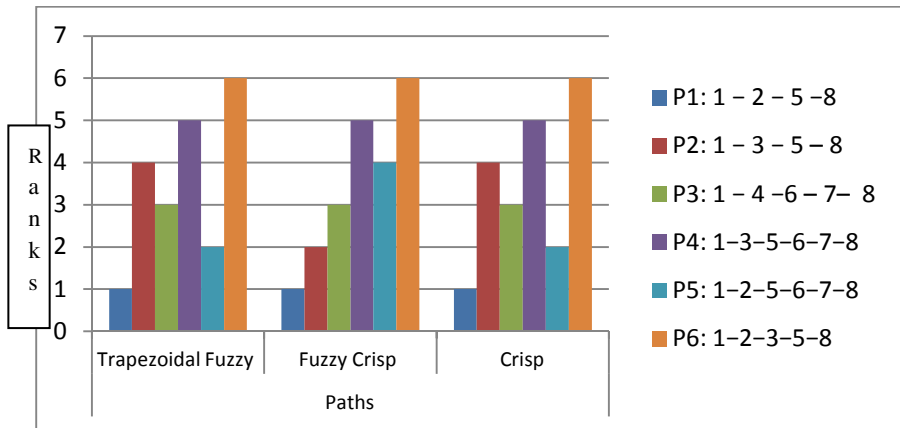


ranking given to the paths based on highest DS which helps the Decision Maker to identify the preferable path alternatives. The Degree of Similarity and the Acceptable Index between FSL and Li, for  $i = 1$  to 5 are computed and presented in the Table 4.

**Table 4.** Results of TFSP using DS and AI

Paths	DS	AI	Rank
P1: 1 - 2 - 5 - 8	4.15	1.063	1
P2: 1 - 3 - 5 - 8	2.985	1.025	4
P3: 1 - 4 - 6 - 7 - 8	3.23	1.026	3
P4: 1-3-5-6-7-8	2.689	0.993	5
P5: 1-2-5-6-7-8	3.885	1.028	2
P6: 1-2-3-5-8	1.155	0.964	6

The shortest path selection will be done on a network with fuzzy arc lengths where the shortest path is identified using the concept of ranking function with regard to the fact that the Decision Maker can choose the best path among various alternatives from the list of rankings. The Rank comparison graph of different approaches i) Crisp Arc Lengths, ii) Fuzzy Crisp Arc Lengths and iii) Trapezoidal Fuzzy Arc Lengths for Shortest Path selection is presented in figure 4.



**Fig. 4.** Ranks VS Paths

## 6 Conclusions

The study of Green Routing and Scheduling problems is gaining momentum in recent years. We have discussed shortest path algorithms by using a) Crisp values b) Trapezoidal Fuzzy Crisp values 3) Trapezoidal fuzzy values. The rank given to the paths based on these techniques helps the Decision Maker to identify the preferable path alternatives. Prioritizing the scheduling and the selection of the paths automatically results in the reduction in exhaust fumes facilitates Green routing.

Finally, it is concluded that the results of the present investigation would provide an excellent platform for making effective and efficient decisions in the case Green Routing and Scheduling Problems and also to design efficient plans for sustainable transportation.

## References

1. Touati-Moungla, N., Jost, V.: On green routing and scheduling problem, hal-00674437, version 1 - 7 (March 2012)
2. Dubois, D., Prade, H.: Theory and Applications: Fuzzy Sets and Systems. Academic Press, New York (1980)
3. Dubois, D., Prade, H.: Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences* 30, 183–224 (1983)
4. Klein, C.M.: Fuzzy Shortest Paths. *Fuzzy Sets and Systems* 39, 27–41 (1991)
5. Lin, K., Chen, M.: The Fuzzy Shortest Path Problem and its Most Vital Arcs. *Fuzzy Sets and Systems* 58, 343–353 (1994)
6. Okada, S., Soper, T.: A shortest path problem on a network with fuzzy arc lengths. *Fuzzy Sets and Systems* 109, 129–140 (2000)
7. Yao, J.S., Lin, F.T.: Fuzzy shortest-path network problems with uncertain edge weights. *Journal of Information Science and Engineering* 19, 329–351 (2003)
8. Chuang, T.N., Kung, J.Y.: The fuzzy shortest path length and the corresponding shortest path in a network. *Computers and Operations Research* 32, 1409–1428 (2005)
9. De, P.K., Bhinchar, A.: Computation of Shortest Path in a Fuzzy Network: Case Study with Rajasthan Roadways Network. *International Journal of Computer Applications* (0975 – 8887) 11(12) (December 2010)
10. Sujatha, L., Elizabeth, S.: Fuzzy Shortest Path Problem Based on Index Ranking. *Journal of Mathematics Research* 3(4) (November 2011)

# A Fused Feature Extraction Approach to OCR: MLP vs. RBF

Amit Choudhary<sup>1</sup> and Rahul Rishi<sup>2</sup>

<sup>1</sup> Maharaja Surajmal Institute, New Delhi, India  
amit.choudhary69@gmail.com

<sup>2</sup> UIET, Maharshi Dayanand University, Rohtak, India  
rahulrishi@rediffmail.com

**Abstract.** This paper is focused on evaluating the capability of MLP and RBF neural network classifier algorithms for performing handwritten character recognition task. Projection profile features for the character images are extracted and merged with the binarization features obtained after preprocessing every character image. The fused features thus obtained are used to train both the classifiers i.e. MLP and RBF Neural Networks. Simulation studies are examined extensively and the proposed fused features are found to deliver better recognition accuracy when used with RBF Network as a classifier.

**Keywords:** MLP, RBF, Hybrid Feature Extraction, Character Recognition, OCR, Neural Network.

## 1 Introduction

In the domain of off-line handwriting recognition, an artificial neural network became very popular in the late 70s and emerged as fast and most reliable classifier tool resulting in excellent recognition accuracy. The accuracy of the OCR System in recognizing the off-line handwritten character mainly depends on the selection of feature extraction technique and the classification algorithm employed. The main motive behind this work is to compare the handwritten character recognition accuracy of the Multi-layered Perceptron (MLP) and Radial Basis Function Neural Network (RBFNN) classifiers trained with Hybrid (Fusion of Binarization and Projection Profile) Features extracted from the input character image samples.

Multi-layered Perceptron (MLP) and Radial Basis Function (RBF) networks are basically non-linear layered feed forward networks. These networks work as universal approximators. RBF Neural Network is also a type of MLP but with Radial Basis as the activation function in the hidden layer neurons. It is therefore not surprising to find that there always exists an RBF network capable of accurately mimicking a specified MLP or vice versa.

## 2 Related Work

The area of OCR is becoming an integral part of document scanners, and is used in many applications such as postal processing, script recognition, banking, security (i.e. passport authentication) and language identification. The research in this area has been ongoing for over half a century and the outcomes have been astounding with successful recognition rates for printed characters exceeding 99%, with significant improvements in performance for handwritten cursive character recognition where recognition rates have exceeded the 90% mark [1].

The purpose of feature extraction is to achieve most relevant and discriminative features to identify a symbol uniquely [2]. Many feature extraction techniques are proposed and investigated in the literature that may be used for numeral and character recognition. Consequently, recent techniques show very promising results for separated handwritten numerals recognition [3], however the same accuracy has not been attained for cursive character classification [2]. It is mainly due to ambiguity of the character without context of the entire word [4]. Second problem is the illegibility of some characters due to nature of cursive handwriting, distorted and broken characters [5]. Finally, the segmentation process may cause some irregularities depending on the approach adopted [6, 7].

## 3 Comparison of MLP and RBF Neural Networks

MLP may require a smaller number of parameters than the RBF network for the same degree of accuracy, but RBF Network is capable of implementing arbitrary non-linear transformations of the input space as illustrated by the XOR problem, which cannot be solved by any linear perceptron but can be solved by the RBF Network. Hence, the choice of the classifier lies in the problem domain. For different types of pattern classification activities, one neural network as a classifier may be proved to be superior to the other. In this work, the superiority of the two classifiers is judged for the task of handwritten character recognition.

## 4 Sample Preparation Using Fused Features Extraction

Projection profile features in the vertical direction is computed by tracing the character image column wise along the y-axis and counting the number of black pixels in each column. As the entire digit images are resized into 15×12 pixels, there are 12 columns and 15 rows. Hence, the vertical projection profile will contain 12 values, each value representing the sum of number of all black pixels present in that particular column. Similarly, for horizontal projection profile, character image is traced horizontally along the x-axis. The row wise sum of number of black pixels present in each row will constitute the 15 values of horizontal projection profile. Left diagonal projection profile is computed by traversing the character image along the left diagonal. The black pixels are counted in left diagonal direction and the sum of the number of black pixels for each left diagonal line of traversing generates 26 values representing the left diagonal projection profile. In the same way, 26 values

representing the right diagonal projection profile are obtained by adding the number of black pixels in each right diagonal line of traversing.

Projection profile values obtained in all the four orientations are combined to form a single feature vector. The length of this feature vector is 79 (12+15+26+26=79). A feature vector representing the character 'c' is obtained by concatenation of these four set of features and can be written as:

```
{4 8 5 5 4 4 4 4 5 1 2 2
4 5 3 1 2 2 2 1 2 2 1 2 4 1 1 6
0 0 0 0 0 3 6 5 5 3 3 1 1 1 2 1 1 2 2 2 2 2 3 2 1 0
0 0 0 1 2 3 2 1 1 1 1 1 2 3 2 3 2 3 3 4 3 5 4 1 1 0}
```

This projection profile feature vector of character 'c' can be represented in the form of a column matrix as shown in Fig.1(a).

After preprocessing of the handwritten character images, the features are extracted by using binarization technique for each character image from the local database of handwritten characters prepared. In the proposed experiment, 1300 (50×26=1300) handwritten characters have been involved. As all the handwritten character images has been resized to 15 × 12 (in order to make all the character images of uniform size), the feature vector of a single character is a column vector of size 180 × 1. The feature vector of character 'c' is shown in Fig.1(b).

The feature vectors of each individual character using binarization and projection profile are combined. The process of combining involves the concatenation of the two feature vectors in such a way that the binarization feature vector is appended at the end of the projection feature vector. The combined feature vector is termed as "Hybrid Feature Vector". The process is repeated for all the characters in the character database. The hybrid feature vector for character 'c' obtained by fusion of projection and binarization feature vectors is shown in Fig.1(c).

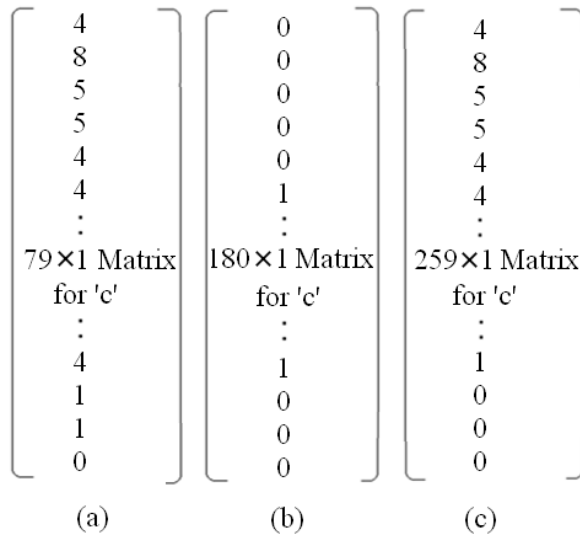


Fig. 1. (a) The Projection Feature Vector, (b) Binarization Feature Vector, (c) Hybrid Feature Vector of Character 'c'

It can be observed that the length of the hybrid feature vector is 259 as the binarization feature vector of length 180 is appended at the end of the projection feature vector of length 79.

The hybrid feature vector of all the 26 characters (a-z) are created in the form of column matrix of size 259×1 each. All these 26 feature vectors are combined to form a sample which is a binary matrix of size 259×26 as shown in Fig.2.

$$\begin{pmatrix}
 2 & 4 & 4 & 4 \\
 3 & 12 & 8 & 6 \\
 7 & 9 & 5 & 7 \\
 6 & 4 & 5 & 7 \\
 5 & 4 & 4 & 7 \\
 5 & 4 & 4 & 6 \\
 \vdots & \vdots & \vdots & \vdots \\
 259 \times 1 \text{ Matrix} & 259 \times 1 \text{ Matrix} & 259 \times 1 \text{ Matrix} & \dots & 259 \times 1 \text{ Matrix} \\
 \text{for 'a' } & \text{for 'b' } & \text{for 'c' } & & \text{for 'z' } \\
 \vdots & \vdots & \vdots & & \vdots \\
 1 & 1 & 1 & & 0 \\
 1 & 0 & 0 & & 0 \\
 1 & 0 & 0 & & 0 \\
 0 & 0 & 0 & & 0
 \end{pmatrix}$$

Fig. 2. Input Sample Created by Fusion of Projection and Binarization Features

### 5 Implementation and Discussion of Results

The number of neurons in the input and output layers has been fixed at 259 and 26 respectively. The 259 input neurons are equivalent to the input character’s size because the feature vector of length 180 obtained through binarization process is appended at the end of the projection feature vector of length 79. The number of neurons in the output layer is 26 because there are 26 English alphabets. The number of neurons for the MLP classifier has been kept 80, by trial and error method, for optimal result.

Both the neural network classifiers have been trained with 50 sets of each character i.e. 1300 (50×26=1300) character image samples from the local database has been involved in the training process. Each character at the input will put a ‘1’ at that neuron in the output layer in which the maximum trust is shown and rest neuron’s result into ‘0’ status. The output of both the network classifiers is a binary matrix of size 26×26 each because there are 26 characters and every character has 26×1 output vector. The first 26×1 column stores the first character's recognition output; the following column will be for next character and so on for the whole sample of 26 characters.

The recognition results obtained when MLP and RBF network classifiers are used to recognize the handwritten characters using Hybrid Features (features extracted

from fusion of binarization and projection features) are shown the form of confusion matrices as shown in Fig.3 and Fig.5 respectively. The confusion matrix shows the confusion among the recognized characters while testing the neural network’s recognition accuracy. The equivalent 3-D plots drawn in the ‘MATLAB Environment’ are also shown in Fig.4 and Fig.6 respectively.

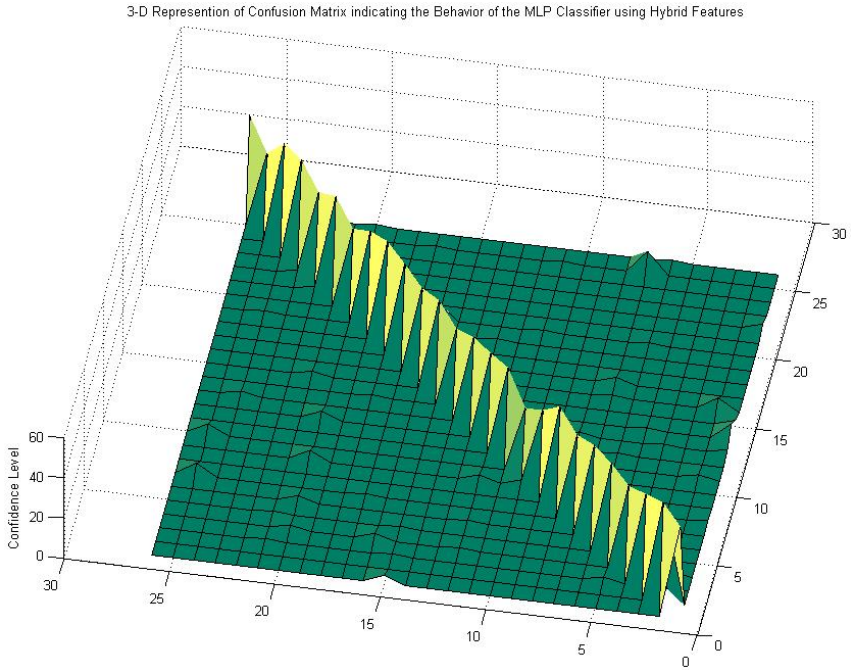
During recognition using MLP, the confusion among the different characters is explained in Fig.3. Character ‘a’ is presented 50 times to the neural network and is classified 46 times correctly. It is miss-classified four times as ‘o’. Character ‘b’ is classified 50 times correctly. Character ‘c’ is misclassified as ‘e’ two times and one time each as ‘o’ and ‘u’ and is classified correctly 46 times. The average recognition accuracy of 88.76% is quiet good for this handwritten character recognition experiment using MLP as a classifier.

The RBF neural network has also been trained with 50 sets of each character i.e 1300 (50× 26=1300) character image samples from the database. Character ‘a’ is presented 50 times to the neural network and is classified 49 times correctly as shown in Fig.5. It is miss-classified one times as ‘o’. Character ‘j’ is classified 42 times correctly and misidentified as character ‘i’ four times and two times each as character ‘g’ and ‘y’ out of a total of fifty trials.

Recognition accuracy for each character (a-z) as well as overall recognition accuracy is displayed in Fig.5. The average recognition accuracy of 93.46 % is excellent for this handwritten character recognition experiment using RBF as a classifier.

Alphabets	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	Success (%)
a	46	0	0	0	0	0	0	1	0	0	0	0	2	8	3	0	0	0	0	0	0	3	0	0	0	0	92
b	0	50	0	2	0	0	0	0	0	0	2	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	100
c	0	0	46	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92
d	0	0	0	42	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	84
e	0	0	2	0	45	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	90
f	0	0	0	0	0	46	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	92
g	0	0	0	0	0	0	43	0	0	7	0	0	0	0	0	1	3	0	0	0	0	0	0	0	11	0	86
h	0	0	0	0	0	0	0	49	0	0	4	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	98
i	0	0	0	0	0	0	0	0	39	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78
j	0	0	0	0	0	0	0	0	3	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	64
k	0	0	0	0	0	0	0	0	0	0	44	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	88
l	0	0	0	1	0	1	0	3	0	0	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86
m	0	0	0	0	0	0	0	0	0	0	0	0	43	1	0	0	0	0	1	0	0	0	0	0	0	0	86
n	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0	0	0	0	0	0	0	0	0	0	78
o	4	0	1	0	1	0	0	0	0	0	0	0	0	0	45	0	0	0	1	0	0	0	0	0	0	0	90
p	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	43	0	0	0	8	0	0	0	0	0	0	86
q	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	1	0	0	94
r	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	60	0	0	0	0	0	0	0	0	0	100
s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	0	94
t	0	0	0	2	0	3	0	0	5	0	0	4	0	0	0	0	0	0	0	40	0	0	0	0	1	0	80
u	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	48	3	0	0	0	0	0	96
v	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	42	0	0	0	0	0	84
w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	50	0	0	0	0	0	100
x	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	50	0	0	0	100
y	0	0	0	0	0	0	6	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	37	0	0	74
z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0	96
Overall Character Recognition Accuracy =																									88.76 %		

Fig. 3. Confusion Matrix representing the MLP Classifier Performance using Hybrid Features



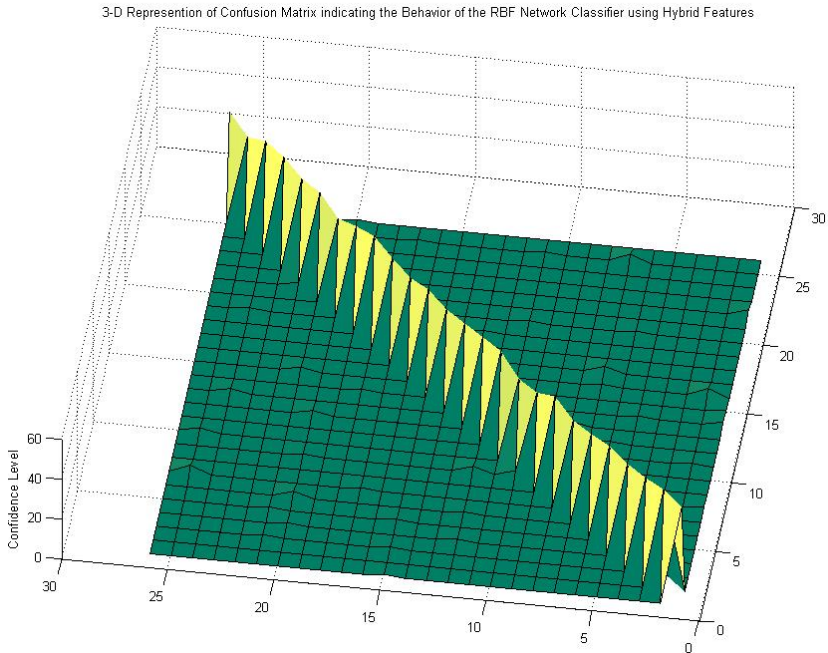
**Fig. 4.** Three-Dimensional Plot of Confusion Matrix representing the MLP Behavior

Alphabets	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	Success (%)
a	49	0	0	0	0	0	0	1	0	0	0	0	2	1	0	0	0	0	0	0	0	3	0	0	0	0	98
b	0	50	0	1	0	0	0	0	0	0	1	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	100
c	0	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96
d	0	0	0	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94
e	0	0	0	0	48	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	96
f	0	0	0	0	0	46	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	92
g	0	0	0	0	0	0	45	0	2	0	0	0	0	0	1	3	0	0	0	0	0	0	0	4	0	0	90
h	0	0	0	0	0	0	0	49	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	98
i	0	0	0	0	0	0	0	0	43	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86
j	0	0	0	0	0	0	0	0	3	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84
k	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	96
l	0	0	0	1	0	1	0	0	3	0	0	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94
m	0	0	0	0	0	0	0	0	0	0	0	0	46	1	0	0	0	1	0	0	0	0	0	0	0	0	92
n	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0	0	0	90
o	1	0	1	0	1	0	0	0	0	0	0	0	0	0	47	0	0	0	1	0	0	0	0	0	0	0	94
p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	0	0	5	0	0	0	0	0	0	90
q	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	1	0	0	94
r	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	60	0	0	0	0	0	0	0	0	100
s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	94
t	0	0	0	1	0	3	0	0	1	0	2	0	0	0	0	0	0	0	0	43	0	0	0	0	0	1	86
u	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	48	0	0	0	0	0	96
v	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	47	0	0	0	0	0	84
w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	100
x	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	50	0	0	0	100
y	0	0	0	0	0	4	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	0	0	88
z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	98

Overall Character Recognition Accuracy = 93.46 %

**Fig. 5.** Confusion Matrix representing the RBF Network Performance using Hybrid Features





**Fig. 6.** Three-Dimensional Plot of Confusion Matrix representing the RBF Network Behavior

## 6 Conclusion

The recognition accuracy of a classifier system mainly depends on the quality of training samples, the techniques employed to extract features and the type of the classifier involved. The proposed method of fusion of character features obtained from binarization and projection profile techniques showed the remarkable enhancement in the performance for the handwritten character recognition using the MLP and RBF network classifiers. By using Hybrid features, the excellent character recognition accuracy of 88.76% and 93.46% is achieved for MLP and RBF classifiers respectively.

## References

1. Alginahi, Y.: Preprocessing Techniques in Character Recognition. In: Mori, M. (ed.) Character Recognition, pp. 1–20. InTechopen Publishers (2010) ISBN: 978-953-307-105-3
2. Blumenstein, M., Liu, X.Y., Verma, B.: An investigation of the modified direction feature for cursive character recognition. *Pattern Recognition* 40, 376–388 (2007)
3. Wang, X., Ding, X., Liu, C.: Gabor filters-based feature extraction for character recognition. *Pattern Recognition* 38(3), 369–379 (2005)
4. Cavalin, P.R., Britto, A.S., Bortolozzi, F., Sabourin, R., Oliveira, L.S.: An implicit segmentation based method for recognition of handwritten strings of characters. In: *Proceedings of ACM Symposium on Applied Computing*, pp. 836–840 (2006)

5. Blumenstein, M., Verma, B., Basli, H.: A Novel Feature Extraction Technique for the Recognition of Segmented Handwritten Characters. In: Proceedings of the 7th International Conference on Document Analysis and Recognition, pp. 137–141. IEEE Computer Society Press, Edinburgh (2003)
6. Blumenstein, M., Verma, B.: Analysis of Segmentation Performance on the CEDAR Benchmark Database. In: Proceedings of the 6th International Conference on Document Analysis and Recognition, pp. 1142–1146. IEEE Computer Society Press, Seattle (2001)
7. Choudhary, A., Rishi, R., Ahlawat, S.: A New Character Segmentation Approach for Off-Line Cursive Handwritten Words. Elsevier Procedia Computer Science 17, 434–440 (2013)

# Fusion of Dual-Tree Complex Wavelets and Local Binary Patterns for Iris Recognition

N.L. Manasa, A. Govardhan, and Ch. Satyanarayana

Jawaharlal Nehru Technological University, Andhra Pradesh, India  
Nadipally.m@gmail.com

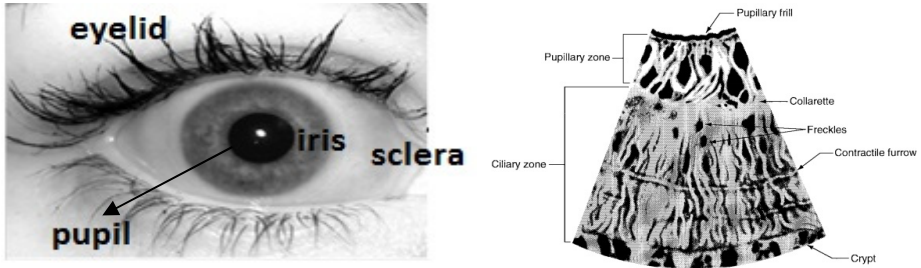
**Abstract.** Iris, the most exclusive biometric trait, is a significant begetter of research since late 1980s. In this paper, we propose new feature fusion methodology based on Canonical Correlation Analysis to combine DTCW and LBP. Complex Wavelet Transform is used as an abstract level texture descriptor that gives a global scale invariant representation, while Local Binary Pattern (LBP) lay emphasis on local structures of the iris. In the proposed framework, CCA maximizes the information from the above two feature vectors which yield a more robust and compact representation for iris recognition. Experimental results demonstrate that fusion of *Wavelet* and *LBP* features using *CCA* attains 98.2% recognition accuracy and an EER of 1.8% on publicly available CASIA IrisV3-LAMP dataset [19].

**Keywords:** Biometrics, Iris Recognition, Dual Tree Complex Wavelet Transform, Local Binary Pattern, Canonical Correlation Analysis, Cosine similarity measure.

## 1 Introduction

Secure human authentication has long been an inviting goal. The last decade has witnessed a significant surge in biometric based user authentication systems for security-critical applications [1]. Among the wide variety of biometrics available, iris is contemplated as the most data rich biometric structure on the human body. Besides providing a reliable authentication system in corporate or enterprise environment, it is also at the core of banking, e-commerce, forensics etc. Recently, Indian government's initiative for unique identification scheme AADHAAR has adopted iris recognition for 1.2 billion individuals. This paper concerns the mathematical analysis of the ocular random patterns within the iris of an eye for uniquely identifying an individual.

Iris is a thin annular structure around the pupil of the human eye. Its complex pattern is constructed of many idiosyncratic features such as fibres, freckles, furrows, arching ligaments, ridges, serpentine vasculature, rings, rifts, corona, some of which may be seen in Fig 1. All these establish a distinctive signature for human authentication. Patterns in human iris have abundance of invariance. Iris patterns emerge during the eighth month of the fetal term and remain stable throughout the life time of an individual. On the lines of precision, [2] report successful authentication across millions of cases without a single failed test. Given its non-invasive nature and affordable hardware solutions [3], iris based authentication systems have become an indispensable tool for many high-security applications.

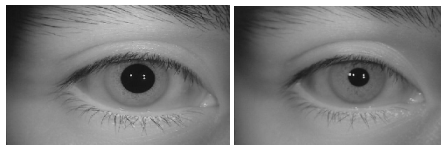


**Fig. 1.** (left)Components of the eye (right)Structure of patterns in the cross section of the coronal view of iris

Typically, iris recognition starts with segmenting the annular disk in the eye by detecting the inner and outer boundary of the pupil and the sclera embedded between the upper and lower eyelids. Most of the existing methods [2] [24] use polar transformations in the second step for rotation invariance that arises due to tilt of the head during acquisition. The annular disk is thus converted into a strip [23] and effects of rotations are converted into just circular shifts. In the third step, relevant features are extracted from the segmented annular regions which give a compact representation of the complex pattern present in the iris. Finally, the extracted feature template is then matched with the existing database to uniquely identify the person.

Challenges in iris recognition are as follows:

- (i) Accurate segmentation of the annular region is imperative for using polar transformation. Occlusion due to the eye lids and presence of eye lashes both effect the segmentation and polar transformation. The assumption of limbic and pupillary boundaries having a common center is often not true, thus leading to errors in segmentation.
- (ii) Cooperative datasets [19] based methods [2] assume centered gaze and constant illumination conditions, elastic deformation of iris texture due to pupil expansion and contraction is one of the most common and challenging issues in iris recognition[26], see Fig 2. These methods cannot be extended to non cooperative datasets [25].
- (iii) [2] assumes that the iris is planar and estimates affine transformation to correct for angular gaze but recently [23] showed the existing of some curvature component.
- (iv) Most of the methods are limited by the distance between the camera and the iris during acquisition; a more robust scale invariant method is required.



**Fig. 2.** Example iris images in CASIA-Iris-Lamp[26]

[24] classified existing methods on iris recognition into 4 categories: texture [4], phase [5], zeros crossing [6] and keypoint descriptors [13]. Though texture, phase and zeros crossing methods have shown to work well on constrained iris datasets with good resolution input, limited occlusion and minimal illumination artifacts, they perform poorly on non-cooperative datasets. Keypoint descriptors on the other hand have gained a lot of interest in the computer vision community for object recognition tasks. Interest points are detected in images and a compact representation termed *descriptor* is constructed to characterize a local neighborhood around the interest points. SIFT, SURF, GLOH have shown to be robust to changes in illumination, noise, and minor changes in the viewpoint. However these methods cannot be directly extended for iris recognition as the iris patterns are very similar locally and correlation between neighboring patterns must be considered. [25] proposed a region based sift method which computes the descriptor after dividing the iris into left right and central regions. Another approach based on SURF [24] coined keytexel feature descriptor has been proposed which computes the features after normalization and enhancement. [25] on one hand performs well for non-cooperative datasets but it is limited in accuracy on constrained datasets [24]. [24] on the other hand relies on accurate segmentation of the annular region which limits its extension to the non-cooperative datasets.

In this paper we propose a novel approach to overcome the above mentioned challenges and drawbacks. A feature descriptor fusion method based on canonical correlation analysis is used to combine two features, a global feature set and a local feature descriptor: Dual Tree Complex Wavelets (DTCW)[18] and local binary patterns (LBP)[17] respectively. Canonical Correlation Analysis (CCA) is one of the important statistical multi-data processing methods which deals with the mutual relationships between two random vectors.

As patterns in human iris have abundance of invariance, the inter-class and intra-class variability of iris features makes it difficult for just one set of features to capture this variability. The Global features are insensitive to affine transformations, noise, and captures large between-class variance and small within-class variance while Local features capture significant within-class variance. DTCW captures the global information ensuring invariance over a wide range of scales which helps in discriminating between locally similar regions [25]. DTCW features compensate the error in localizing the iris region as they are invariant to the rotation and the inexact iris localization. The LBP on the other hand captures local fine textures effectively, they are also sensitive to position and orientation of iris image. A chance of recognition rate being compromised in the case of very large databases is high with different iris images having similar global features. Hence it is important to use local texture features in combination with global texture features for accurate recognition. Sun et al [7, 8] propose a two-stage cascaded classifier method in which the first stage is a traditional Daugman like classifier and the second stage is a global classifier which are areas enclosed by zero crossing boundaries. Later Sun et al tries to improve the cascaded classifier using LBP to extract local texture features [9]. This fusion method significantly improves the recognition performance of Daugman's approach [9]. Similarly, Vatsa et al. [11], Park and Lee [12], Zhang et al. [10] also show an

improvement in performance by using more than one feature to capture the distinct iris patterns. They fuse the features at the image level where as the proposed method maximizes the correlation between the feature sets at feature level. Experiments depict that the combination of the two features yield better performance than either alone [12].

The proposed method relies on coarse segmentation and extracts both the feature descriptors. Canonical correlation analysis is used to combine the features at the descriptor level which ensures that the information captured from both the features are maximally correlated and eliminate the redundant information giving a more compact representation. See Fig. 3.

## 2 Fusion Based Iris Recognition Approach

The input to the current iris recognition system is a monochromatic image with varying illumination, gaze and distance between the camera and the iris.

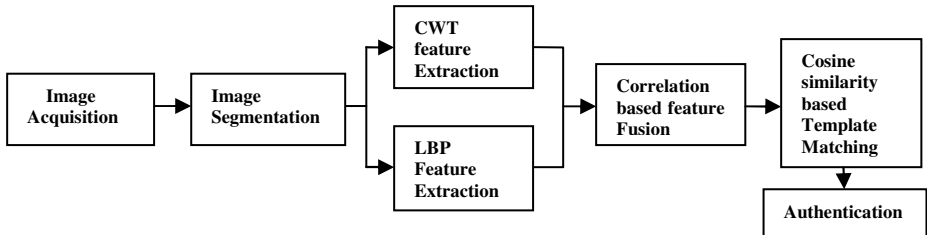


Fig. 3. Iris authentication pipeline of the proposed method

### 2.1 Iris Segmentation

Localizing the two landmark spherical regions inside the input eye image corresponding to iris and pupil is the motive of the segmentation step. The obstructing factors will be occlusions from eyelids, eyelashes, specular reflections caused by imaging under natural light. For isolating these artifacts, simple thresholding was used. Then we employ Canny edge detector to extract gradients in all directions which gave us the edge map. Consequently, circular Hough Transform was employed to detect iris and pupil boundaries as stated in [21].

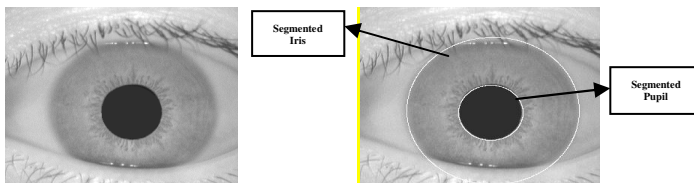


Fig. 4. Original Iris Image obtained from a monochromatic camera (left) and Segmented Iris Image (right)

## 2.2 Feature Extraction

The framework used in this approach integrates two complementary features of the coarsely segmented iris template, viz., Complex Wavelet features and Local Binary pattern features to take advantage of different characteristics of eye patterns.

### Dual-Tree Complex Wavelet Transform Based Feature Extraction

Discrete Wavelet Transforms based methods have been successfully applied to a variety of problems like denoising, edge detection, registration, fusion etc., Discrete wavelet transforms have 4 basic problems namely, oscillation, shift variance, aliasing and lack of directionality. By using complex valued basis functions instead of real basis functions these four problems can be minimized. Complex wavelet transform (CWT) [18] is represented in the form of complex valued scaling functions and complex valued wavelet functions,

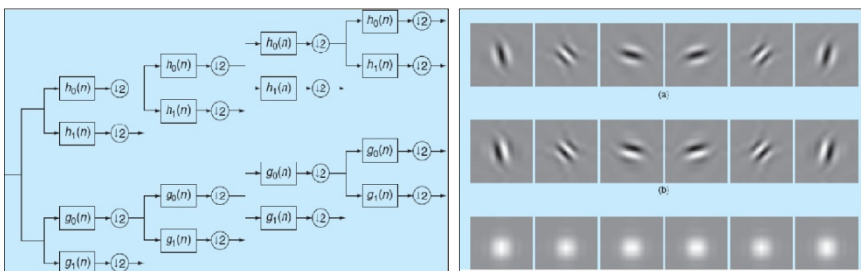
$$\Psi_c(t) = \Psi_r(t) + j\Psi_i(t) \tag{1}$$

where  $\Psi_r(t)$  are real and even and  $j\Psi_i(t)$  are imaginary and odd.

There is no one unique extension of the standard DWT into the complex plane, [27] proposed *Dual tree Complex Wavelet Transform* to realize CWT. The dual tree CWT employs two real discrete wavelet transforms (DWT); the initial DWT gives the real part of the transform and the subsequent DWT gives the imaginary part. The analysis filter bank (FB) is used to implement the dual-tree CWT as illustrated in Fig 5. To design an overall transform, two sets of filters are used. Each set of filter represents real wavelet transform. This overall transform is approximately analytic. In dual-tree CWT, consider the 2-D wavelet  $\psi(x,y) = \psi(x)\psi(y)$  associated with the row-column implementation of the wavelet transform, where  $\psi(x)$  is a complex wavelet given by  $\psi(x) = \psi_h(x) + j\psi_g(x)$ . Obtain  $\psi(x,y)$  for the expression,

$$\psi(x,y) = [\psi_h(x) + j\psi_g(x)][\psi_h(y) + j\psi_g(y)] = \underbrace{\psi_h(x)\psi_h(y) - \psi_g(x)\psi_g(y)}_{\text{Real part}} + j\underbrace{[\psi_g(x)\psi_h(y) + \psi_h(x)\psi_g(y)]}_{\text{Imaginary part}} \tag{2}$$

The implementation of the Dual-tree CWT is straight forward as given in [28]. The feature vector obtained from the DTCWT is 250 in length.



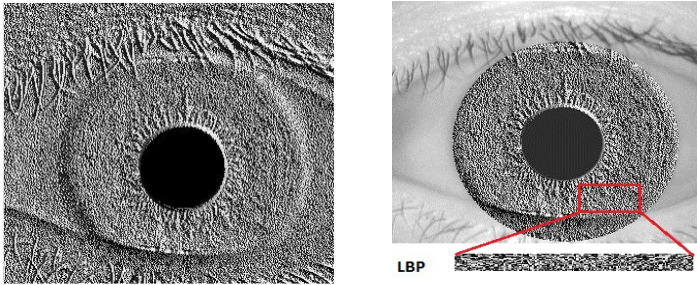
**Fig. 5.** (left) DCWT filter bank (right) Typical wavelets associated with the 2-D dual-tree CWT (a) illustrates the real part of each complex wavelet; (b) illustrates the imaginary part; (c) illustrates the magnitude [18]

### Local Binary Pattern (LBP) Based Feature Extraction

LBP captures the local level texture variations. Local binary patterns introduced by Ojala et al. [17] use local texture descriptor. In its simplest form, an LBP description of a pixel is created by thresholding the values of the 3x3 neighborhood of the pixel against the central pixel and explicating the result as a binary number. The Local binary pattern (LBP) operator was originally designed as a texture descriptor. The LBP operator attributes a label to every pixel of an image by thresholding the 3x3 neighborhood of each and every pixel value with the center pixel value and assigns binary value (0,1) based on the following equation,

$$I(x,y)=\begin{cases} 0 & \text{if } N(x;y)<I(x;y) \\ 1 & \text{else } N(x;y)\geq I(x;y) \end{cases} \quad (3)$$

where  $I(x, y)$  is the center pixel value and  $N(x, y)$  is neighborhood pixel value. After thresholding, central pixel value is represented by a binary number (or decimal number) called label. Histogram of these labels is used as texture descriptor.



**Fig. 6.** (left) LBP features on the whole eye image (right) Computed LBP features on segmented Iris region

LBP operator deals with textures at different scales using neighborhoods of different sizes. Local neighborhood can be defined using circular neighborhood. Circular neighborhood is a set of evenly spaced sampling points on a circle, whose center is the pixel to be labeled. Radius of circle controls the spatial resolution of operator and number of sampling points controls angular space quantization. Interpolation is used when a sampling point does not fall in the mediate of a pixel. Notation (P; R) will be used for pixel neighborhoods which contemplate P sampling points on a circle of radius R. Fig 6 shows the extracted LBP features of a segmented iris image. To achieve the gray level invariance we subtract the center pixel value from all circular neighborhood pixel values and assume that this difference is independent of center pixel value.

### 2.3 Feature Fusion

Canonical correlation analysis can be defined as the complication of finding two sets of basis vectors, one for  $\mathbf{x}$  and one for  $\mathbf{y}$  in a way that the correlations between the



projections of the variables onto these basis vectors are mutually maximized. Linear combinations  $x = \mathbf{x}^T \hat{\mathbf{w}}_x$  and  $y = \mathbf{y}^T \hat{\mathbf{w}}_y$  of the two variables is maximized as follows,

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E[\hat{\mathbf{w}}_x^T \mathbf{x} \mathbf{y}^T \hat{\mathbf{w}}_y]}{\sqrt{E[\hat{\mathbf{w}}_x^T \mathbf{x} \mathbf{x}^T \hat{\mathbf{w}}_x] E[\hat{\mathbf{w}}_y^T \mathbf{y} \mathbf{y}^T \hat{\mathbf{w}}_y]}} = \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}} \quad (4)$$

The maximum of  $\rho$  with respect to  $\mathbf{w}_x$  and  $\mathbf{w}_y$  is the maximum canonical correlation. The projections onto  $\mathbf{w}_x$  and  $\mathbf{w}_y$ , i.e.  $x$  and  $y$ , are called canonical variates. The canonical correlations between  $\mathbf{x}$  and  $\mathbf{y}$  can be established by solving the Eigen-value equations,

$$\begin{cases} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \hat{\mathbf{w}}_x = \rho^2 \hat{\mathbf{w}}_x \\ \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \hat{\mathbf{w}}_y = \rho^2 \hat{\mathbf{w}}_y \end{cases} \quad (5)$$

where the eigen-values  $\rho^2$  are the squared canonical correlations and the eigen-vectors  $\hat{\mathbf{w}}_x$  and  $\hat{\mathbf{w}}_y$  are the normalized canonical correlation basis vectors.  $\mathbf{C}_{xx}$  and  $\mathbf{C}_{yy}$  are the within-sets covariance matrices of  $\mathbf{x}$  and  $\mathbf{y}$  respectively and  $\mathbf{C}_{xy} = \mathbf{C}_{yx}^T$  is the between-sets covariance matrix. The number of non-zero solutions to these equations are bounded to the smallest dimensionality of  $\mathbf{x}$  and  $\mathbf{y}$ .

Just one eigen-value equations needs to be solved since the solutions are related by,

$$\begin{cases} \mathbf{C}_{xy} \hat{\mathbf{w}}_y = \rho \lambda_x \mathbf{C}_{xx} \hat{\mathbf{w}}_x \\ \mathbf{C}_{yx} \hat{\mathbf{w}}_x = \rho \lambda_y \mathbf{C}_{yy} \hat{\mathbf{w}}_y, \end{cases} \quad (6)$$

In this work we apply method proposed by [20] based on the formulation presented. Let the two feature extractors be trained by  $L$  training images. Let  $A = [a_1, a_2, \dots, a_L]$  and  $B = [b_1, b_2, \dots, b_L]$  be the corresponding outputs of the two feature extractors, and  $n_1$  and  $n_2$  be the dimensions of the two outputs, where  $n_1, n_2 \leq L$ . The covariance matrices for  $\mathbf{A}$  and  $\mathbf{B}$  are given as  $\mathbf{C}_{aa}$  and  $\mathbf{C}_{bb}$  respectively.  $\mathbf{C}_{ab}$  is the between-set covariance matrix.  $\hat{\mathbf{w}}_x$  and  $\hat{\mathbf{w}}_y$  are canonical basis vectors of feature vectors  $\mathbf{A}$  and  $\mathbf{B}$ . Let  $a_i$  and  $b_i$  are two feature vectors of input image  $i$ . Then the fusion of these two feature vectors is define as

$$\mathbf{F}_i = \begin{bmatrix} \hat{\mathbf{w}}_x^T a_i \\ \hat{\mathbf{w}}_y^T b_i \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{w}}_x & 0 \\ 0 & \hat{\mathbf{w}}_y \end{bmatrix}^T \begin{bmatrix} a_i \\ b_i \end{bmatrix} \quad (7)$$

## 2.4 Template Matching

For the matching purpose we use cosine similarity measure. Cosine similarity measure is defined as cosine angle between test image fused feature vector and training image fused feature vector as follows,

$$\arg \max_{j \in [1, 2, \dots, L]} \left( \frac{\mathbf{F}_i^T \mathbf{F}_j}{\|\mathbf{F}_i\| \|\mathbf{F}_j\|} \right) \quad (8)$$

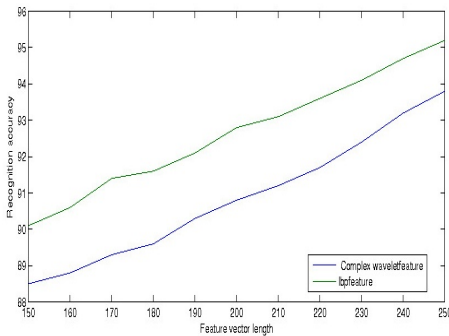
The maximum value according to Equ (8) is estimated as an authenticated iris match.

### 3 Experiments and Results

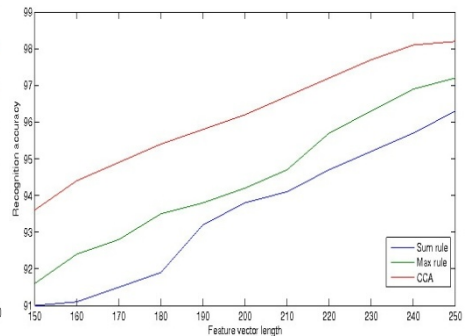
We test our algorithm on publicly available CASIA IrisV3-LAMP dataset [19], CASIA-Iris-Lamp was collected using a hand-held iris sensor. A lamp was turned on/off close to the subject to introduce more intra-class variations during acquisition. As stated in Section 1.1, elastic deformation of iris texture (Fig.2) due to pupil expansion and contraction under different illumination conditions is one of the most common and challenging issues in iris recognition [26]. So CASIA-Iris-Lamp is good for studying problems of non-linear iris normalization and robust iris feature representation. Statistics of the dataset are as follows, no. of subjects 411, no. of classes 819, no. of images 16,212, resolution 640\*480. 80% of the images were used as training samples and remaining are used for testing purpose.

In the first experiment, we check for the proposed method's recognition accuracy using Complex Wavelet feature and LBP feature separately. Highest recognition rates of DTCW and LBP are 93.8% and 95.2% with 256 and 250 feature vector lengths respectively and the same is plotted in Fig. 7 thus showing the performance gain of LBP over DTCW for shorter length of the feature vector for iris recognition.

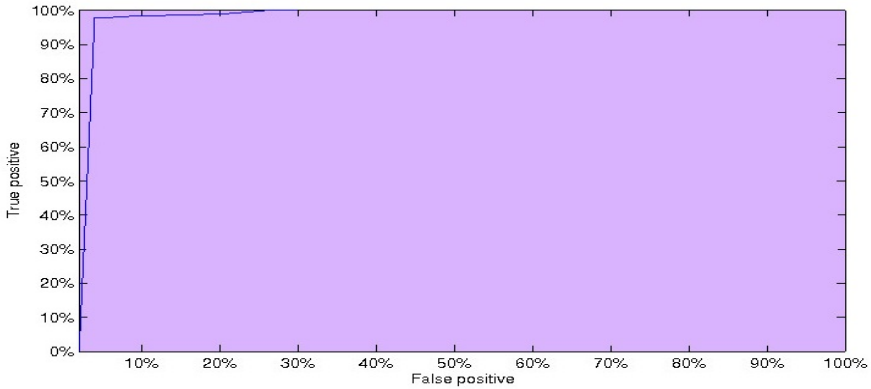
We perform the second experiment to illustrate the superior performance of the Correlation based fusion framework. We also compare the performance of the proposed framework with two other widely used fusion techniques, namely, Sum Rule and Max Rule. From Fig 8 it is evident that the recognition accuracy of the proposed method surpasses the rest by 1.2% for a feature vector length 250, see Table 1.



**Fig. 7.** Recognition accuracy vs variable feature vector length



**Fig. 8.** Fusion method vs variable feature vector length



**Fig. 9.** Receiver Operating Characteristic curve

Widely used standard for estimating biometric system’s performance, Receiver Operating Characteristics Curve (ROC) drawn against True positive rate and True Negative rate for the CASIA IrisV3 dataset. From the curve we can understand that the CCA based fusion method for iris based biometric human authentication suggested in this paper possesses an Equal Error Rate of **1.8%** which is considerably finer in comparison with the state-of-the-art iris recognition systems mentioned in this paper. For certainty, we also validate the proposed method with the state-of-the-art techniques for iris recognition, see Table 2.

**Table 1.** Recognition accuracy of three fusion methods

Sl.no.	Fusion method	Recognition Accuracy(%)
1	Sum rule	96.3
2	Max rule	97.2
3	CCA	98.2

**Table 2.** Comparative analysis of different Iris Recognition approaches [22]

Different approach	Recognition Accuracy (%)
Amir azizi	96.5
K.Masoud	95.9
A.T Zaim	95
Li Ma	94.9
Vladan	94.7
Hanho Sung	94.54
Robert W. Ives	93
Boles	92.64
Hao Meng	87.4
Agus Harjoko	84.25
Najafi and Ghofrani	98
<b>Proposed</b>	<b>98.2</b>
<b>Fusion approach</b>	

## 4 Conclusion

A diverse iris based biometric recognition system based on fusion framework is proposed in this paper. We infer that Canonical correlation maximizes the information from two feature vectors of the same pattern by combining them at the descriptor level which ensures that the information captured from both the features are maximally correlated and eliminates the redundant information giving a more compact representation.

## References

1. Global biometrics market revenue to reach \$20 billion by 2018, <http://www.biometricupdate.com>
2. Daugman, J.: How Iris Recognition Works. *IEEE Trans. on Circuits & Systems for Video Technology* 14(1) (January 2004)
3. Iris identification solutions, <http://www.neurotechnology.com/verieye-technology.html>
4. Kim, J., Cho, S., Choi, J., Marks II, R.: Iris recognition using wavelet features. *Journal of VLSI Signal Processing Systems* 38(2) (2004)
5. Miyazawa, K., Ito, K., Aoki, T., Kobayashi, K.: An efficient iris recognition algorithm using phase-based image matching. In: *Proc. of IEEE Int. Conf. on Image Processing*, vol. 2 (2005)
6. Boles, W., Boashash, B.: A human identification technique using images of the iris and wavelet transform. *IEEE Transactions on Signal Processing* 46(4) (1998)
7. Sun, Z., Wang, Y., Tan, T., Cui, J.: Cascading statistical and structural classifiers for iris recognition. In: *International Conference on Image Processing* (2004)
8. Sun, Z., Wang, Y., Tan, T., Cui, J.: Improving iris recognition accuracy via cascaded classifiers. *IEEE Trans. Syst. Man Cyber.* 35(3), 435–441 (2005)
9. Sun, Z., Tan, T., Qiu, X.: Graph matching iris image blocks with local binary pattern. In: Zhang, D., Jain, A.K. (eds.) *ICB 2005. LNCS*, vol. 3832, pp. 366–372. Springer, Heidelberg (2005)
10. Zhang, P.-F., Li, D.-S., Wang, Q.: A novel iris recognition method based on feature fusion. In: *International Conference on Machine Learning and Cybernetics*, pp. 3661–3665 (2004)
11. Vatsa, M., Singh, R., Noore, A.: Reducing the false rejection rate of iris recognition using textural and topological features. *Int. Journal of Signal Processing* 2(2), 66–72 (2005)
12. Park, C.-H., Lee, J.-J.: Extracting and combining multimodal directional iris features. In: Zhang, D., Jain, A.K. (eds.) *ICB 2005. LNCS*, vol. 3832, pp. 389–396. Springer, Heidelberg (2005)
13. Mehrotra, H., Majhi, B., Gupta, P.: Annular iris recognition using SURF. In: Chaudhury, S., Mitra, S., Murthy, C.A., Sastry, P.S., Pal, S.K. (eds.) *PREMI 2009. LNCS*, vol. 5909, pp. 464–469. Springer, Heidelberg (2009)
14. CASIA IrisV3 Description document, <http://www.idealtest.org/findTotalDbByMode.do?mode=Iris>
15. Selesnick, I.W.: Hilbert Transform Pairs of Wavelet Bases. *IEEE Sig. Proc. Letters* (2001)
16. DTCW, <http://eeweb.poly.edu/iselesni/WaveletSoftware/dt2D.html>

17. Ojala, T., Pietikinen, M., Menp, T.: Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. *IEEE Trans. on PAMI* 24(7), 971–987 (2002)
18. Selesnick, I.W., Baraniuk, R.G., Kingsbury, N.C.: The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine* 2(2), 123–151 (2005)
19. CASIA Image Iris database,  
<http://www.idealtest.org/findTotalDbByMode.do?mode=Iris>
20. Sun, Q.-S., Liu, Z.-D., Heng, P.-A., Xia, D.-S.: A theorem on the generalized canonical projective vectors. *Pattern Recognition* 38, 449–452 (2005)
21. National Recognition of Human Iris Patterns for Biometric Identification,  
<http://people.csse.uwa.edu.au/pk/studentprojects/libor/LiborMasekThesis.pdf>
22. Najafi, M., Ghofrani, S.: Iris Recognition Based on Using Ridgelet and Curvelet Transform. *International Journal of Signal Processing, Image Processing and Pattern Recognition* 4(2) (June 2011)
23. Daugman, J.: New methods in iris recognition. *IEEE Trans. Systems, Man, Cybernetics B* 37(5), 1167–1175 (2007)
24. Mehrotra, H., Pankaj, K., Majhi, B.: Fast segmentation and adaptive SURF descriptor for iris recognition. *Journal of Mathematical and Computer Modelling* 58, 132–146 (2013)
25. Belcher, C., Du, Y.: Region-based SIFT approach to iris recognition. *Optics and Lasers in Engineering* 47, 139–147 (2009)

# Evaluation of Feature Selection Method for Classification of Data Using Support Vector Machine Algorithm

A. Veeraswamy<sup>1</sup>, S. Appavu Alias Balamurugan<sup>2</sup>, and E. Kannan<sup>1</sup>

<sup>1</sup> VELTECH Dr. RR & DR.SR Technical University, Chennai, Tamil Nadu, India  
{ammisetty.veeraswamy, ek081966}@gmail.com

<sup>2</sup> Department of Information Technology,  
KLN College of Information Technology, Madurai, Tamil Nadu, India  
app\_s@yahoo.com

**Abstract.** One may claim that the exponential growth in the amount of data provides great opportunities for data mining. In many real world applications, the number of sources over which this information is fragmented grows at an even faster rate, resulting in barriers to widespread application of data mining. This paper proposes feature selection by using Support Vector Machine is to reduce the computational complexity and increase the classification accuracy of the selected feature subsets and also paper evaluates the approach by comparing it with existing feature selection algorithms over 6 datasets from University of California, Irvine (UCI) machine learning databases. The proposed method shows better results in terms of number of selected features, classification accuracy, and running time than most existing algorithms.

**Keywords:** Feature Selection, Classification, Data Mining J48, K-Star, SVM.

## 1 Introduction

Data mining is a form of knowledge discovery essential for solving problems in a specific domain. Classification is a technique used for discovering classes of unknown data. As the world grows in complexity, overwhelming us with the data it generates, data mining becomes the only hope for elucidating the patterns that underlie it [1]. The manual process of data analysis becomes tedious as size of data grows and the number of dimensions increases, so the process of data analysis needs to be computerized.

The term Knowledge Discovery from data (KDD) refers to the automated process of knowledge discovery from databases. The process of KDD is comprised of many steps namely data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation [1]. A "feature" or "attribute" or "variable" refers to an aspect of the data. Usually before collecting data, features are specified or chosen. Features can be discrete, continuous, or nominal. Generally,

features are characterized as: Relevant: These are features which have an influence on the output and their role cannot be assumed by the rest [1]. Irrelevant: Irrelevant features are defined as those features not having any influence on the output, and whose values are generated at random for each example. Redundant: A redundancy exists whenever a feature can take the role of another. The Problem of selecting some subset of a learning algorithms input variables upon which it should focus attention, while ignoring the rest.

Feature selection is the process of selecting the some features may be redundant or irrelevant thus not contributing to the learning process [1]. Feature selection, a process of choosing a subset of features from the original ones, is frequently used as a preprocessing technique in data mining. It has proven effective in reducing dimensionality, improving mining efficiency, increasing mining accuracy, and enhancing result comprehensibility. Feature selection methods can broadly fall into the wrapper model and the filter model [1]. The quality of the data is one such factor. if information is irrelevant or redundant, or the data is noisy and unreliable, then knowledge discovery during training is more difficult. Feature subset Selection is the process of identifying and removing as much of the irrelevant and redundant information as possible. Machine learning algorithms differ in the amount of emphasis they place on feature selection [2].

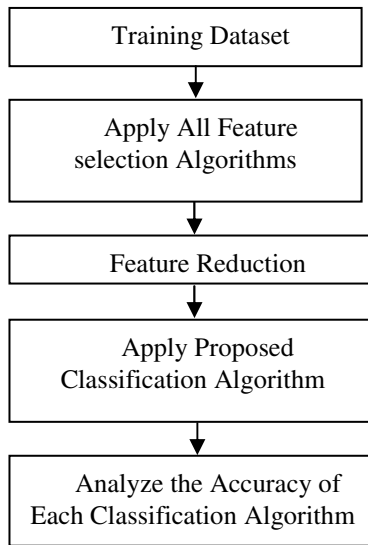
## 2 Proposed Work

The SVM is the first algorithm produced by Vladimir Vapnik's statistical learning theory frameworks [4]. Later, the SVM is extended by Cortes and Vapnik to cover binary classification problems with misclassifications [5]. The most significant discovery in terms of enabling the using SVMs in embedded systems [3] is probably attributed to John Platt [6]. Using his sequential minimal optimization method we are able to train SVMs using an insignificant memory footprint. Later, the SVM is extended to most of the other classic machine learning frameworks such as regression, novelty detection, ranking, clustering etc. [7] [8] [9].

In this paper, we introduce a novel approach for feature selection in high dimensional data using Support Vector Machine. Support Vector Machines [10] are basically binary classification algorithms. Support Vector Machines (SVM) are a classification system derived from statistical learning theory. It has been applied successfully in fields such as text categorization, hand-written character recognition, image classification, bio sequences analysis, etc. The SVM separates the classes with a decision surface that maximizes the margin between the classes. The surface is often called the optimal hyper plane, and the data points closest to the hyper plane are called support vectors. The support vectors are the critical elements of the training set. The mechanism that defines the mapping process is called the kernel function [10].

The SVM can be adapted to become a nonlinear classifier through the use of nonlinear kernels. SVM can function as a multiclass classifier by combining several

binary SVM classifiers. The output of SVM classification is the decision values of each attribute for each class, which are used for probability estimates. The probability values represent "true" probability in the sense that each probability falls in the range of Zero to One, and the sum of these values for each pixel equals one. Classification is then performed by selecting the highest probability. SVM includes a penalty parameter that allows a certain degree of misclassification, which is particularly important for no separable training sets. The penalty parameter controls the trade-off between allowing training errors and forcing rigid margins. It creates a soft margin that permits some misclassifications, such as it allows some training points on the wrong side of the hyper plane. Increasing the value of the penalty parameter increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well [11]. The paper also evaluates the approach by comparing it with existing feature selection algorithms over 6 datasets from University of California, Irvine (UCI) machine learning databases [6]. The proposed method shows better results in terms of number of selected features, classification accuracy, and running time than most existing algorithms.



**Fig. 1.** Data Flow Analysis diagram for Feature selection with Classification

### 3 System Implementation of Proposed Algorithm

The proposed algorithm is implemented using Java. The stepwise approach is as follows. The input to the system is given as an attribute-relation file format (ARFF) file. A table is created in Oracle using the name specified in @relation". The attributes specified under @attribute" and instances specified under @data" are retrieved from the ARFF File and then they are added to the created table. This procedure is



followed for providing the training set as well as test set. The created table acts as the dataset and is given as the input to the proposed algorithm.

The combination of attribute value should occur at least once in the dataset, because while finding the dependency between attribute values if a combination of attribute value did not occur once, then it will lead to alternate zeros resulting in zero probability and dependency that cannot be found. Thus, the above condition is checked before a combination of attribute value is given to the proposed method. The probabilities are calculated for the given input. Based on the probabilities, the dependent attributes are identified.

## 4 Experimental Results

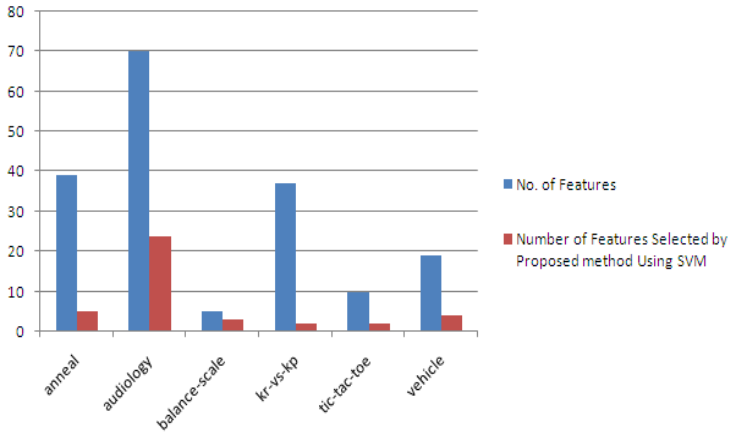
The feature selection using SVM is applied to many datasets, and the performance evaluation is done. We presented the performance evaluation on 6 datasets. All together 8 datasets are selected from the UCI machine learning repository and the UCI knowledge discovery in databases (KDD) archive [12].

A summary of datasets is presented in Table 1. For each dataset, we run all seven feature-selection algorithms, SVM, consistency subset eval, Info Gain attribute eval, Gain Ratio attribute eval, OneR attribute eval, Chi Squared attribute eval, principal components, classifier subset eval, respectively, and record the number of selected features for each algorithm. We then apply SVM, Naïve Bayes, decision tree (J 48), K-Star ,Random Tree, OneR on the original dataset as well as each newly obtained dataset containing only the selected features from each algorithm and recorded the overall accuracy by 10 fold cross validation.

From Table 2, it is found that for all the datasets, some feature selection algorithms will select the attributes that are fewer than the number of attributes selected by the proposed method. The main idea in the proposed method is finding the dependency between the attributes in deciding the class attributes value, and also the probabilities will decide the dependency between a set of attributes. Therefore, the proposed method removes the dependent attributes and identifies the perfect attributes which are sufficient for the classification of the datasets and also improve the classification accuracy.

**Table 1.** Details description of dataset used in the experiment

S.NO	Name of the Dataset	No. of Attributes	No. Of Instances	Selected Attributes on Proposed Algorithm
1	anneal	39	898	5
2	audiology	70	226	24
3	balance-scale	5	625	3
4	kr-vs-kp	37	3196	2
5	tic-tac-toe	10	958	2
6	vehicle	19	846	4



**Fig. 2.** Number of features selected by the proposed method

**Table 2.** Number of selected features for each feature selection algorithm

S.NO	Name of the Dataset	Cfs	Chi	Gain Ratio	Info Gain	One Attribute	PCA	Constituent Subset Value	Proposed Algorithm
1	anneal	7	38	38	38	38	37	8	5
2	audiology	6	69	69	69	69	63	13	24
3	balance-scale	1	4	4	4	4	4	4	3
4	kr-vs-kp	3	36	36	36	36	31	6	2
5	tic-tac-toe	1	9	9	9	9	16	8	2
6	vehicle	11	18	18	18	18	7	18	4

**Table 3.** Accuracy of each Classification algorithm applying for each Dataset

S.NO	Name of the Dataset	K-Star	ONE-R	J48	Proposed Algorithm	Naïve bayes	Random Tree
1	anneal	95.66	83.63	97.22	97.33	85.97	96.43
2	audiology	70.80	46.46	69.47	81.86	68.14	59.29
3	balance-scale	63.52	56.32	63.52	87.52	63.52	77.76
4	kr-vs-kp	90.43	66.45	90.43	95.56	90.43	88.45
5	tic-tac-toe	69.94	69.93	69.94	98.32	69.94	73.48
6	vehicle	66.43	51.53	68.32	74.34	48.46	64.30
Average Accuracy		<b>76.13</b>	<b>62.39</b>	<b>76.48</b>	<b>89.15</b>	<b>71.08</b>	<b>76.62</b>

Table 3 Specifies the Accuracy of all Classification Algorithms like K-Star, OneR, J48, Naïve Bayes and Random Tree. These Accuracy are Compared by Proposed Algorithm.

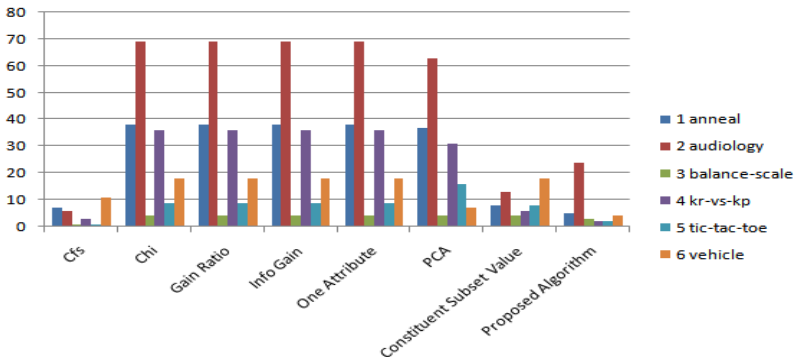


Fig. 3. No. Of Features selected by each Feature selection Algorithm

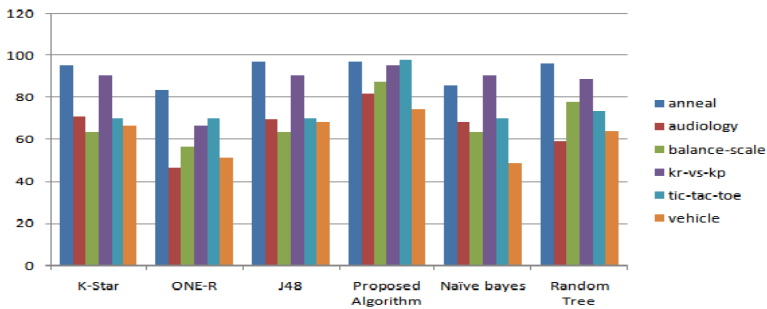


Fig. 4. Accuracy of each Classification algorithm applying for each Dataset

Table 4. Applying Each Feature Selection Algorithm with K-Star and Compared With Proposed Algorithm

S.NO	Name of the Dataset	Cfs	Chi	Gain Ratio	Info Gain	One Attribute	PCA	Constituent Subset Value	K-Star	Proposed Algorithm
1	anneal	88.08	98.99	95.76	98.66	98.66	96.1	88.08	95.66	97.33
2	audiology	53.53	99.55	99.55	99.55	99.55	99.11	50.01	70.80	81.86
3	balance-scale	74.72	45.76	45.76	45.76	88.48	45.76	45.76	63.52	87.52
4	kr-vs-kp	70.91	72.62	72.62	72.62	72.62	92.45	70.52	90.43	95.56
5	tic-tac-toe	82.46	17.84	17.84	17.84	17.84	96.86	39.03	69.94	98.32
6	vehicle	66.78	25.65	25.65	25.65	25.65	68.55	25.65	66.43	74.34
Average		72.75	60.07	59.53	60.01	67.13	83.14	53.18	76.13	89.15

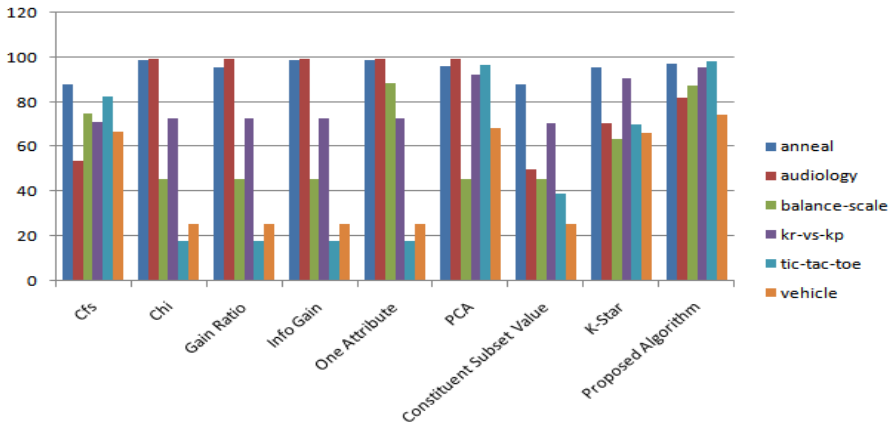


Fig. 5. Performance analysis Graph by using the Proposed Algorithm

## 5 Conclusion

This paper proposes a feature selection algorithm based on SVM. The algorithm can remove redundancy from the original dataset. The main idea provided is to find the dependent attributes and remove the redundant ones among them. The technology to obtain the dependency needed is based on SVM. A new attribute reduction algorithm of using SVM is implemented and evaluated through extensive experiments via comparison with related attribute reduction algorithms. In this paper, we consider the task of feature selection and investigate the performance of nine feature selection algorithms.

Our findings can be summarized as follows:

1) In feature selection approach, we have shown that SVM is a promising approach for automatic feature selection. It outperforms most existing algorithms in terms of number of selected features, classification accuracy. Well-established algorithms, such as Info Gain attribute eval, Gain Ratio attribute eval, OneR attribute eval, Chi Squared attribute eval, principal components are also more complex than SVM feature Selection, SVM based feature selection runs very efficiently on large datasets, which makes it very attractive for feature selection in high dimensional data.

2) We have implemented a new feature selector using SVM and found that it performs better than the popular and computationally expensive traditional algorithms.

3) We compared the performance of a number of algorithms on the UCI machine learning repository datasets.

## References

- [1] Classification and Feature Selection Techniques in Data Mining, Sunita Beniwal. Jitender Arora International Journal of Engineering Research & Technology (IJERT) 1(6) (August 2012) ISSN: 2278-0181
- [2] Hall, M.A.: Feature Selection for Discrete and Numeric Class Machine Learning. University of Waikato, Hamilton
- [3] Rasmus Ulslev Pedersen. Using Support Vector Machines for Distributed Machine Learning PhD thesis, University of Copenhagen (2005)
- [4] Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, NY (1995)
- [5] Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* 20(3), 273–297 (1995)
- [6] Platt, J.: Fast training of support vector machines using sequential minimal optimization. In: *Advances in Kernel Methods — Support Vector Learning*. MIT Press (1999)
- [7] Scholkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J.: Support vector method for novelty detection
- [8] Scholkopf, B., Bartlett, P.L., Smola, A., Williamson, R.: Shrinking the tube: a new support vector regression algorithm. In: Kearns, M.S., Solla, S.A., Cohn, D.A. (eds.) *Advances in Neural Information Processing Systems 11*, pp. 330–336. MIT Press, Cambridge (1999)
- [9] Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support vector clustering. *Journal of Machine Learning Research* 2, 125–137 (2001)
- [10] Vapnik, V.N.: *Statistical Learning Theory*. Wiley, New York (1998)
- [11] Hsu, C.W., Chang, C.C., Lin, C.J.: *A practical guide to support vector classification* (2003),  
<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- [12] Blake, C.L., Merz, C.J.: *UCI Repository of Machine Learning Databases*, Department of Information and Computer Science. University of California, Irvine (1998),  
<http://www.ics.uci.edu/mllearn>

# Electrocardiogram Beat Classification Using Support Vector Machine and Extreme Learning Machine

C.V. Banupriya<sup>1</sup> and S. Karpagavalli<sup>2</sup>

<sup>1</sup> Department of Computer Science, PSGR Krishnammal College for Women,  
Coimbatore, India

banupriya.venkat@gmail.com

<sup>2</sup> GR Govindarajulu School of Applied Computer Technology,  
Coimbatore, India

karpagam@grgsact.com

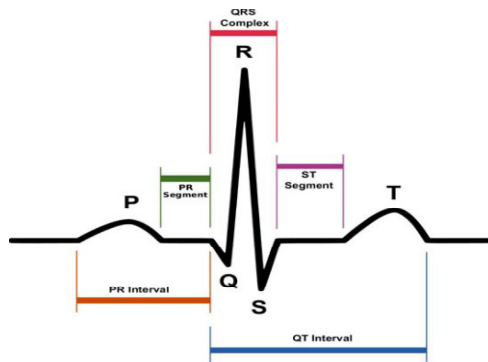
**Abstract.** The Electrocardiogram (ECG) is of significant importance in assessing patients with abnormal activity in their heart. ECG Recordings of the patient taken for analyzing the abnormality and classify what type of disorder present in the heart functionality. There are several classes of heart disorders including Premature Ventricular Contraction (PVC), Atrial Premature beat (APB), Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Paced Beat (PB), and Atrial Escape Beat (AEB). To analyze ECG various feature extraction methods and classification algorithms are used. The proposed work employed discrete wavelet transform (DWT) in feature extraction on ECG signals obtained from MIT-BIH Arrhythmia Database. The Machine Learning Techniques, Support Vector Machine (SVM) and Extreme Learning Machine (ELM) have been used to classify four types of heart beats that include PVC, LBBB, RBBB and Normal. The Performance of the classifiers are analyzed and observed that ELM-Radial Basis Function Kernel taken less time to build model and out performs SVM in predictive accuracy.

**Keywords:** Electrocardiogram, Wavelet, Support Vector Machine, Extreme Learning Machine.

## 1 Introduction

Electrocardiography represents the electrical activity of the heart. Electrocardiogram is a record of the origin and propagation of electrical potential through cardiac muscles obtained using sensors at limb extremities of the human. It provides useful and important information to cardiologists about the rhythm and functioning of the heart so as to perform diagnostic analysis with maximum accuracy. The classification of the ECG beats is an important task in the coronary care unit, and essential tool for diagnosis of heart diseases. The figure 1 shows the typical structure of the ECG. The P wave, the QRS complex, and the T wave are three characteristic features of the waveform which are easily identified. These waves are associated with the activation of the atria, activation of the ventricles, and re-polarization of the ventricles. A patient

may have different ECG waveforms and in a single ECG different types of beats may present that are unlike each other [1]. The beats may be normal beat, Premature Ventricular Contraction (PVC) beat, Left Bundle Branch Block beat (LBBB), Right Bundle Branch Block beat (RBBB), Atrial Premature beat (APB) and Paced beat (PB).



**Fig. 1.** Structure of the ECG signal

Researchers have developed different techniques and algorithms for automated processing of ECG signals for various medical applications. Some of them used time domain and some use frequency domain for depiction.

Thaweesak, et al. performed the classification of ECG using SVM to classify the 3 classes, premature ventricular contraction (PVC), Normal and Atrial Premature Contraction heart diseases [2]. Wisnu Jatmiko, et al. employed Back-Propagation Neural Networks and Fuzzy Neuro Learning Vector Quantization (FLVQ) as classifier in ECG classification [3]. In their work, they used only the MLII lead as source data. The classes that are considered are Left Bundle Branch Block beat (LBBB), Normal beat (NORMAL), Right Bundle Branch Block beat (RBBB), Premature Ventricular Contraction (PVC). Maedeh Kiani Sarkaleh, et al. [4], proposed a Neural Network (NN) based algorithm for classification of Paced Beat (PB), Atrial Premature Beat (APB) arrhythmias as well as the normal beat signal. They used the features obtained by the Discrete Wavelet Transform (DWT) along with timing interval features to train the Neural Network.

In proposed work, efficient models are built using Support Vector Machine and Extreme Learning Machine to classify the four different heart beats, Normal, Premature Ventricular Contraction (PVC), Left Bundle Branch Block (LBBB) and Right Bundle Branch Block (RBBB) in ECG. The ECG signals are obtained from MIT-BIH Arrhythmia Database. This work involves the various tasks that include preprocessing, feature extraction, building models through training and testing the models.

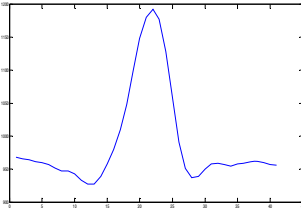
## 2 Feature Extraction

### 2.1 Preprocessing

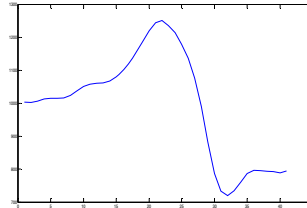
Preprocessing of ECG signals need to be performed for effective feature extraction. In preprocessing signal extension, de-noising and decomposition operations are performed. The number of samples in ECG record has been extended to 50,000 samples. In denoising, the signals are decomposed using Daubechies Wavelet with decomposition level 4 (db4).

### 2.2 Feature Extraction

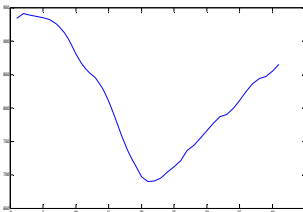
In this proposed work suitable combination of morphological features and temporal features has been extracted from the ECG beats. Actual feature extraction carried out in two phases. The first phase involves the cutting of the Normal, PVC, LBBB and RBBB beats by making use of the annotation files which exist in MIT-BIH arrhythmia database. The annotation file provides beat type and sample number (R), from that R sample R+20 and R-20 a total of 41 samples is cut from the continuous ECG signal for each type of beat. The 41 samples of Normal beat, PVC beat, LBBB beat, RBBB beat are visualized in figure 2, 3, 4, and 5 respectively. The second phase involves identification of the peaks and locations in 41 samples of each type of beat. The morphological features that describe the basic shape of the beats are: amplitude of P-peak (Pamp), amplitude of Q-valley (Qamp), amplitude of R-peak (Ramp), amplitude of S-valley (Samp) and amplitude of T-peak (Tamp).



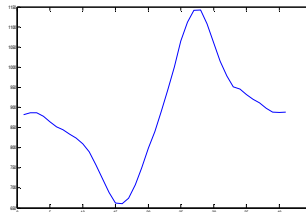
**Fig. 2.** Normal Beat



**Fig. 3.** PVC



**Fig. 4.** LBBB



**Fig. 5.** RBBB



Features that describe the position of waves in the window of beat are: position of P-peak (Ploc), position of Q-valley (Qloc), position of R-peak (Rloc), position of S-valley (Sloc) and position of T-peak (Tloc). Along with the 10 morphological features, the temporal feature QRS duration has been calculated using starting position of the Q wave and end of the S wave.

### 3 Support Vector Machine

Support Vector Machine a new approach to supervised pattern classification which has been successfully applied to a wide range of pattern recognition problems. Support vector machine is a training algorithm for learning classification and regression rules from data. SVM is most suitable for working accurately and efficiently with high dimensionality feature spaces [5-7].

The machine is presented with a set of training examples,  $(x_i, y_i)$  where  $x_i$  is the real world data instances and  $y_i$  is the labels indicating which class the instance belongs to. For the two class pattern recognition problem,  $y_i = +1$  or  $y_i = -1$ . A training example  $(x_i, y_i)$  is called positive if  $y_i = +1$  and negative otherwise. The basic idea of SVM is that it projects data points from a given two class training set in a higher dimensional space and finds an optimal hyper-plane. The optimal one is the one that separates the data with the maximal margin. SVM identify the data points near the optimal separating hyper-plane which are called support vectors. The distance between the separating hyper-plane and the nearest of the positive and negative data points is called the margin of the SVM classifier. The separating hyper-plane is defined as

$$D(x) = (w \cdot x) + b$$

Where  $x$  is a vector of the dataset mapped to a high dimensional space, and  $w$  and  $b$  are parameters of the hyper-plane that the SVM will estimate. The nearest data points to the maximum margin hyper-plane lie on the planes

$$(w \cdot x) + b = +1 \text{ for } y = +1$$

$$(w \cdot x) + b = -1 \text{ for } y = -1$$

The width of the Margin is given by  $m = \frac{2}{\|w\|}$

In multiclass SVM, the formulation of direct method known as Crammer and Singer Method [8] is

$$\text{Minimize } \frac{1}{2} \sum_{k=1}^N (w_k^T) w_k + C \sum_{i=1}^n \xi_i$$

Subject to the constraints

$$w_{w_i}^T \phi(x_i) - w_k^T \phi(x_i) \geq e_k^t - \xi_i, \forall K \neq K_i$$

Where  $k_i$  is the class to which the training data  $x_i$  belong,

$$e_k^i = 1 - C_k^i$$

$$C_k^i = \begin{cases} 1 & \text{if } K_i = K \\ 0 & \text{if } K_i \neq K \end{cases}$$

The decision function for a new input data  $x_i$  is given by

$$\hat{D}_j = \arg \max_K \{f_k(x_j)\}$$

Where  $f_k(x_j) = w_k^T \phi(x_j) - \gamma_k$

## 4 Extreme Learning Machine

Extreme Learning Machine (ELM) is a new learning algorithm for Single-hidden Layer Feed forward neural Networks (SLFNs) with supervised batch learning [9] which provides good generalization performance for both classification and regression problems at highly fast learning speed. The output function of the generalized SLFN is given by,  $F(x) = \sum_{i=1}^L \beta_i h_i(x)$

Where  $h_i(x)$  is the output of the  $i^{\text{th}}$  hidden node. The steps in ELM algorithm are, Given a training set  $N = \{(X_k t_k) | X_k \in R^n, t_k \in R^m, k = 1, \dots, N\}$ , an activation function  $g(x)$  and the number of hidden neurons  $\tilde{N}$ ,

- Randomly assign input weights  $w_i$  and biases  $b_i$  according to some continuous probability density function.
- Calculate hidden layer of output matrix H.
- Calculate the output weights

In kernel based ELM, If the hidden layer feature mapping  $h(x)$  is unknown to users, users can be described a kernel function for ELM. ELM Kernel function is given by,  $KELM(x_i, x_j) = 1/H f(x_i) \cdot f(x_j)$ . That is, the data has feed through the ELM hidden layer to obtain the feature space vectors, and their co-variance is then calculated and scaled by the number of hidden units. The main difference is that where ELM explicitly generates the feature space vectors, but in SVM or in other kernel methods only similarities between feature space vectors are used.

## 5 Experiment and Results

The experiment has been carried out using the ECG signals from the MIT-BIH arrhythmia database [10]. The database comprise 48 recordings, each one of 30 minutes duration and includes two leads, the modified limb lead II and one of the modified leads V1, V2, V4 or V5. The 48 recordings of ECG signals cover all the 16 different heart beat types including normal beat. In this experiment, a total of six ECG records with numbers: 100, 106, 119, 208, 109 and 118 are used from the database to

collect different type of beats. A total of 400 beats are extracted where 100 beats for each type i.e., Normal, PVC, LBBB, RBBB. To collect Normal and PVC beats, MLII lead is used and for LBBB and RBBB beats, V1 lead is used.

In the dataset, for training 80% of the data used and 20% used for testing. First the training set data used to build the model in SVM using the tool *SVM<sup>Light</sup>*. The dataset is trained with linear, polynomial and RBF kernel with dissimilar parameter settings for C-regularization parameter. In holder of polynomial and RBF kernels, the default settings for d (Degree of polynomial) and gamma are used in SVM. Extreme Learning Machine learning algorithm has been implemented in MATLAB. The SLFN network is trained with the training dataset and the accuracy and error rate of the model is observed. The performance of the classifiers is evaluated using test dataset and summarized in Table 1.

**Table 1.** Performance of SVM and ELM classifiers

Classifiers	Kernels	Learning	Prediction
		Time (secs)	Accuracy (%)
SVM	Linear	0.02	92.12
	Polynomial	0.03	94.28
	RBF	0.02	95.34
ELM	Linear	0.0083	98.16
	Polynomial	0.0267	98.65
	RBF	0.0181	99.48

From the results it has been observed that the predictive accuracy of the classifier ELM with three kernels is better than SVM. The training time taken to build the model using ELM kernel is significantly low, compared to SVM kernels.

## 6 Conclusion

An electrocardiogram (ECG) recording of a patient is important clinical information for the medical experts to diagnose the heart functionality of the patient or assess the patient before any surgery. The proposed work automated the ECG beat classification task using powerful supervised classification techniques namely Extreme Learning Machine and Support Vector Machine. The outcome of the experiments indicates that the Kernel based ELM classifier is more accurate and faster than the Kernel based SVM classifier model. Automated systems to classify the ECG will enable the doctors in their decision-making process and to take effective decisions for the patients with heart problems.

## References

1. Luthra, A.: ECG Made Easy. Japee Brothers Publishers (2007)
2. Yingthawornsuk, T.: Classification of Cardiac Arrhythmia via SVM. In: 2nd International Conference on Biomedical Engineering and Technology, IPCBEE, vol. 34. IACSIT Press, Singapore (2012)
3. Jatmiko, W., Nulad, W.P., EllyMatul, I., Mursanto, P.: Heart Beat Classification Using Wavelet Feature Based on Neural Network. Wseas Transactions on Systems 10(1) (January 2011) ISSN: 1109-2777
4. KianiSarkaleh, M., Shahbahrami, A.: Classification of ECG Arrhythmias Using Discrete Wavelet Transform and Neural Networks. International Journal of Computer Science, Engineering and Applications (IJCSSEA) 2(1) (February 2012)
5. Joachims, T., Schölkopf, B., Burges, C., Smola, A.: Making large-Scale SVM Learning Practical. In: Advances in Kernel Methods -Support Vector Learning. MIT Press, Cambridge (1999)
6. Shawe-Taylor, J., Cristianini, N.: Support Vector Machines and other kernel-based learning methods. Cambridge University Press, UK (2000)
7. Vapnik, V.N.: Statistical Learning Theory. J. Wiley & Sons, Inc., New York (1998)
8. Koby, C., Singer, Y.: On the Algorithmic Implementation of Multi-class Kernel-based Vector Machines. Journal of Machine Learning Research 2, 265–292 (2001)
9. Siew, C.K., Huang, G.B.: Extreme Learning Machine with Randomly Assigned RBF Kernels. International Journal of Information Technology 11(1) (2005)
10. Mark, R., Moody, G.: MIT-BIH Arrhythmia Database (1997),  
<http://ecg.mit.edu/dbinfo.html>

# Genetic Algorithm Based Approaches to Install Different Types of Facilities

Soumen Atta\* and Priya Ranjan Sinha Mahapatra

Dept. of Computer Science & Engineering, University of Kalyani, Kalyani-741 235, India  
{soumen.atta,priya}@klyuniv.ac.in

**Abstract.** Given a set  $P$  of  $n$ -points (customers) on the plane and a positive integer  $k$  ( $1 \leq k \leq n$ ), the objective is to find a placement of  $k$  circles (facilities) such that the union of  $k$  circles contains all the points of  $P$  and the sum of the radii of the circles is minimized. We have proposed a Genetic Algorithm (GA) to solve this problem. In this context, we have also proposed two different algorithms for  $k=1$  and 2. Finally, we have proposed a GA to solve another optimization problem to compute a placement of fixed number of facilities where the facilities are hazardous in nature and the range of each such facility is circular.

**Keywords:** Facility Location, Enclosing Problem, Optimization Problem, Genetic Algorithm.

## 1 Introduction

The facility location problem often computes a placement of fixed number of user friendly facilities to serve a set of customers. Examples of such user friendly facilities are post offices, ATM counters, fire stations, ambulances, mobile towers, sensors etc. A customer gets service if and only if s/he is within the *range* (or *service zone*) of at least one facility. The place where the facility is installed is called the *center* of the facility. Notice that the *service cost* (or *cost*) to facilitate a customer is often proportional to the distance between the customer and the *center* of the facility. The distance between the *center* of a facility and the farthest customer served by this facility is called the *range* of this facility. One frequently used objective in *facility location problem* is to minimize the sum of the *service cost* for all customers. This problem is known as the *min-sum* problem in facility location [1].

In facility location model, customers are often considered as points on the plane [1]. Note that the *range* (or *coverage area*) of a facility is the area within (including the boundary) of a geometric object and the nature of the object depends on the facility to be installed. We thus define the *min-sum* problem as follows: given a set  $P$  of  $n$  input points (customers) on the plane, and some number  $k \in \mathbb{Z}^+$ , compute a placement of  $m$  geometric objects  $\{C_1, C_2, \dots, C_m\}$  (facilities) of given type such that

---

\* Corresponding author.

the union of  $k$  objects encloses (contains) all points of  $P$  and  $\sum_{i=1}^k \text{range}(C_i)$ ,  $1 \leq i \leq k$ , is minimized. This problem is known to be NP-hard [10]. Moreover this problem is NP-hard even if the  $C_i$  is a simple object such as circle [10], square [2], rectangle [3], etc.

In this work, we consider a problem in network design [4] where the task is to assign transmission range to each antenna (facility) such that each customer gets service. Note that the cost associated to an antenna is proportional to the *range* assigned to it and thus, the total cost of providing service is proportional to the sum of the transmission ranges of all antennas. We thus need to solve the following optimization problem for an optimal transmission range assignment to all antennas: given a set  $P$  of  $n$  input points on the plane and a positive integer  $k$  ( $1 \leq k \leq n$ ), the objective is to find a placement of  $k$  circles such that the union of  $k$  circles encloses (contains) all points of  $P$  and the sum of radii of the circles is minimized. We call this problem as the *k-Minimum Enclosing Circles Problem (k-MECP)*. Each of the  $k$  circles generating a solution of *k-MECP* is known as *Minimum Enclosing Circle (MEC)*. Alt et al. [3] obtained approximate result on an instance of the *k-MECP* to minimize the total cost for assigning transmission range to each antenna. We have proposed a genetic algorithm (GA) to solve this problem. In this context, we have also proposed two different GAs for  $k=1$  and 2.

In some cases, the facilities may have harmful effect(s) to the customers. Examples of such obnoxious (hazardous) facilities are soccer stadiums, garbage dumps, water purification plants, dangerous chemical factories, nuclear power plants, etc. The problem of computing a placement of such obnoxious facilities is known as *obnoxious facility location problem* [11]. Here objective is to install a facility such that the customers are as far as possible from the facility. In this work, we now consider an *obnoxious facility location problem* where the *coverage area* of each facility is circular. The problem considered is defined as follows: given a set  $P$  of  $n$  input points on the plane and positive integer  $k$  ( $1 \leq k \leq n$ ), compute a placement of  $k$  circles such that union of  $k$  circles does not enclose (contain) any point of  $P$  and the sum of radii of the circles is maximized. This problem is NP-hard [1] and we call this problem as the *k-Largest Empty Circles Problem (k-LECP)*. Each of the  $k$  circles generating a solution of the *k-LECP* is known as *Largest Empty Circle (LEC)*. Another GA is proposed to solve this problem.

Genetic Algorithms (GAs) [5][6] are randomized search and optimization techniques guided by the principles of evolution and natural genetics, and have a large amount of implicit parallelism. GAs provide near optimal solution for an objective function. Here solution is encoded as chromosomes. Biologically inspired operators like selection, crossover and mutation are used over a number of generations for generating potentially better chromosomes.

The rest of the article is organized as follows: In Section 2 we have first proposed a GA based approach to solve the *k-Minimum Enclosing Circles Problem (k-MECP)* for  $k=1$  and 2. Then another GA is proposed to solve the generalized *k-MECP*. In Section 3 another GA is proposed to solve the generalized *k-LECP*. In Section 4 results are shown. Finally Section 5 concludes this article.

## 2 Proposed GA-Based Techniques for the $k$ -MECP

In the following subsections we have proposed GAs to solve the  $k$ -MECP for  $k=1$  and 2 and then we have proposed the generalized  $k$ -MECP.

### 2.1 GA-Based Techniques to Solve the $k$ -MECP for $k=1$

In this subsection we have proposed a GA for the following problem: given a set  $P$  of  $n$  input points on the plane, the objective is to find a placement of a circle such that the circle encloses (contains) all the points of  $P$  and the radius of the circle is minimized. This is the problem definition of the  $k$ -MECP for  $k=1$ .

#### 2.1.1 Initial Population

At first initial population is to be created by generating binary chromosomes of a given length, say  $n_{bit}$ . The size of the initial population is defined by the *population size* which is a user given parameter.

#### 2.1.2 Representation and Encoding Scheme

Here each chromosome represents a binary string of 1's and 0's. We divide the binary string into two parts arbitrarily and suppose the length of each part be  $n_1$  and  $n_2$  respectively such that  $n_1+n_2 = n_{bit}$ . We consider a square bounding box enclosing all points of  $P$  such that the size of the square (length of a side, say  $\beta$ ) is minimized. Now we generate random point within the range  $[\beta \times \beta]$  from the chromosome. For the rest of the paper we have computed the range similarly. We use a mapping function to generate real numbers in  $[v_{new\_min}, v_{new\_max}]$  where  $v_{new\_min} = 0$  and  $v_{new\_max} = \beta$ . Each part of the chromosome represents a value in  $[v_{min}, v_{max}]$  where  $v_{min} = 0$  and  $v_{max} = 2^n - 1$ ,  $n$  is the number bits representing the part of the chromosome under consideration. So each chromosome denotes a point in the range  $[\beta \times \beta]$  and this point is basically the center of the probable minimum enclosing circle. For a value  $v \in [v_{min}, v_{max}]$  we calculate  $v_{new}$  as follows:  $v_{new} = v_{new\_min} + ((v_{new\_max} - v_{new\_min}) / (v_{max} - v_{min})) \times v$ . Consider the following example, suppose the chromosome is [0 1 1 0 1 0 0 0 1 0 1 1 0 0 1 0] of  $n_{bit} = 16$  and let  $n_1 = n_2 = 8$ . So the first part of the chromosome is [0 1 1 0 1 0 0 0] and the second part of the chromosome is [1 0 1 1 0 0 1 0]. Given  $\beta = 100$ , the point represented by this chromosome would be (8.6257, 30.1961), using the above mentioned mapping function.

#### 2.1.3 Fitness of Chromosome Calculation

Here each chromosome denotes the center  $(p_{cx}, p_{cy})$  of a probable minimum enclosing circle. Observe that we want to enclose all the point of  $P$  by the circle. Therefore, we assign the fitness value of a chromosome as the distance between the center of the circle and a farthest input point from the center of this circle. This leads to consider the fitness value of a chromosome as  $\max \{ \sqrt{(p_{cx} - p_i)^2 + (p_{cy} - p_j)^2} \}$  for all  $(p_i, p_j) \in P$ . Our objective is to minimize the fitness value as we want to find out a circle with minimum radius that encloses (contains) all the points of  $P$ .

### 2.1.4 Selection

Here we have used binary tournament selection to select one chromosome out of two randomly chosen chromosomes to create the mating pool. Both the randomly chosen chromosome corresponds to two probable minimum enclosing circles. Obviously the circle with minimum radius is better solution than the other one. This leads to consider the selection criterion as follows: we select a chromosome between randomly selected two chromosomes if it has a less fitness value than the other. Here the chromosome which has less fitness value is considered to be better one than the other which has more fitness value associated with it because this is a minimization problem.

### 2.1.5 Crossover

Here we have used random cross over. Uniform cross over can also be used. In cross over process two chromosomes exchange genetic materials. Crossover takes place depending on the *crossover probability* value given as a user input. In general, the *crossover probability* value is kept near 1.

### 2.1.6 Mutation

Mutation is done by randomly changing the chromosome value. Mutation is happened based on *mutation probability* which is usually kept near 0 and *mutation probability* is also given as a user input.

We have also maintained the *elitism* property of genetic algorithm that is we have kept the best chromosome of parent generation to the next generation.

## 2.2 GA-Based Techniques to Solve the $k$ -MECP for $k=2$

In this subsection we have proposed a GA to solve the  $k$ -MECP for  $k=2$ . The  $k$ -MECP for  $k=2$  can be defined as follows: given a set  $P$  of  $n$  input points on the plane, the objective is to find a placement of two circles such that the union of two circles enclose (or contain) all the points of  $P$  and the sum of radii of the two circles is minimized.

### 2.2.1 Initial Population

We have created initial population by generating binary chromosomes of a given length, say  $n_{bit}$ . The size of the initial population depends on the parameter known as *population size*.

### 2.2.2 Representation and Encoding Scheme

Here each chromosome corresponds to a line  $L$  that partitions the given point set  $P$  of  $n$ -points into two subsets  $P_1$  and  $P_2$ . Now for each subset  $P_1$  and  $P_2$ , we find a *minimum enclosing circle* using the technique mentioned in Section 2.1 or using any other deterministic algorithm [1].

Each chromosome is a binary string of 1's and 0's. We divide the binary string into four parts randomly and suppose the length of each part be  $n_1$ ,  $n_2$ ,  $n_3$  and  $n_4$  respectively such that  $n_1 + n_2 + n_3 + n_4 = n_{bit}$ . Now we generate random point within the



range  $[\beta \times \beta]$  (as defined in the Subsection 2.1.2) from the chromosome. Each part of the chromosome is mapped to a real number using the mapping function as described in Subsection 2.1.2 and they represent the coordinate values of the two points. The line  $L$  is passing through these two points.

### 2.2.3 Fitness of Chromosome Calculation

A chromosome partitions the given point set  $P$  into two subsets  $P_1$  and  $P_2$ . Now having the *minimum enclosing circles* for  $P_1$  and  $P_2$ , we take the sum of their radii as the fitness value for the corresponding chromosome. Our objective is to minimize the fitness function as this is a minimization problem.

The *Selection*, *Cross-over* and *Mutation* are applied in the same way as described in the Sections 2.1.4 to 2.1.6. Here we have also maintained the *elitism* property of GA.

## 2.3 GA-Based Techniques to Solve the Generalized $k$ -MECP

Here we have proposed a GA to solve the generalized  $k$ -MCEP for any positive integer values of  $k$ .

### 2.3.1 Initial Population

Initial population is created by generating binary chromosomes of length, say  $n_{bit}$ . The size of the initial population is given as the user input.

### 2.3.2 Representation and Encoding Scheme

Here each chromosome represents string of 0's and 1's. Given a positive integer number  $k$  representing the number of circles, we generate  $k$  random points on the same plane as the given point set.

We divide the chromosome into  $2 \times k$  parts and the same technique described in the Section 2.1.2 has been used here to find  $2 \times k$  real numbers within the range  $[\beta \times \beta]$  (as defined in the Subsection 2.1.2). So basically each chromosome here corresponds to  $k$  points and these  $k$ -points are the centers of probable  $k$  *minimum enclosing circles*.

### 2.3.3 Fitness of Chromosome Calculation

The main task of this GA-based approach is to find out the fitness value for each chromosome. The fitness value for each chromosome is calculated as follows. A chromosome represents  $k$  random points. We take the first random point and then find out a cluster by finding its *floor*( $n/k$ ) nearest neighbour points from the given point set  $P$ , where  $n$  is the number of input points. Now we discard the input points which are included in the cluster already. Then we take the second random point and we again find out a cluster in the same way as earlier and so on. The last cluster is created by considering the rest of the points left in  $P$ . In this way we generate  $k$  clusters. Then we find a minimum enclosing circle for each of these clusters. Now we have proposed a *compaction* method and apply it to each of the  $k$  cluster (*minimum enclosing circle*). The objective of the *compaction method* is to reduce the overlapping area(s) among the  $k$  *minimum enclosing circles*. The *compaction* method will work as follows: for a

minimum enclosing circle, say  $C_i$  ( $1 \leq i \leq k$ ), we compute a farthest input point, say  $p_f$  from the center of  $C_i$ . Then we find a nearest neighbour of this farthest point  $p_f$  among all the points of  $P$ . If the nearest neighbour of  $p_f$  for the circle  $C_i$  belongs to another circle  $C_j$  ( $i \neq j$ ) then we assign the point  $p_f$  to  $C_j$ ; otherwise do nothing. We apply the *compaction* method to all  $C_i$  ( $1 \leq i \leq k$ ). So after compaction method the structure of clusters may change. Now we again find out a *minimum enclosing circle* for each of the clusters and take the sum of the radii of all these  $k$  circles as the fitness value for the chromosome under consideration.

The *Selection*, *Crossover* and *Mutation* are applied in the same way as described in the Sections 2.1.4 to 2.1.6. *Elitism* is also maintained here to keep the best chromosome over the generations.

### 3 Proposed GA-Based Techniques for the Generalized $k$ -LECP

In this Section we have proposed a GA to solve the generalized  $k$ -LECP for any positive integer values of  $k$ .

The *initial population* creation and *representation and encoding scheme* of chromosome are same as described in the Section 2.3.1 and Section 2.3.2 respectively.

#### 3.1 Fitness of Chromosome Calculation

Here each chromosome represents  $k$  random points which basically denote the probable centers of the  $k$  *largest empty circles*. For each such center we find its closest input point and the distance between such a center and the closest point of it is taken as the radius of that corresponding *largest empty circle*. Then we assign the sum of the radii of all such circles as the fitness value of the chromosome under consideration. Our objective is to maximize the fitness value.

#### 3.2 Selection

We have used binary tournament selection to select one chromosome out of the two chromosomes to create the mating pool. Both the randomly chosen chromosomes correspond to the two different sets of  $k$  probable *largest empty circles*. We select the chromosome with more fitness value than the other because this is a maximization problem.

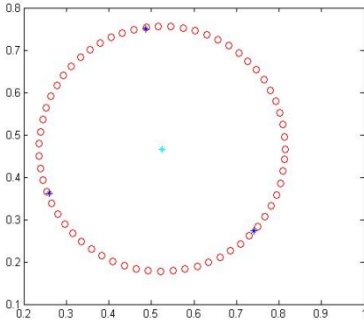
*Crossover* and *mutation* are applied in the same way as described in the Section 2.1.5 and 2.1.6 respectively. We have also mentioned *elitism* property here to keep the best chromosome of parent generation to the next generation.

## 4 Results

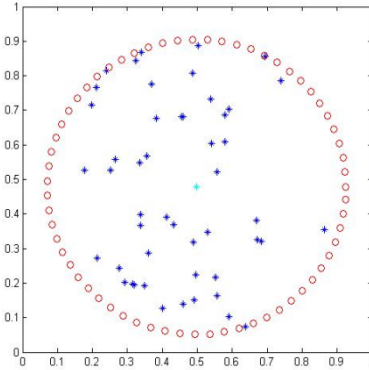
In this section experimental results are discussed.

**4.1 GA-Based Technique to Solve the  $k$ -MECP for  $k=1$**

We have generated random point set. The values of different parameters of genetic algorithm such as *population size*, *number of generations*, *cross over probability* and *mutation probability* are taken as 50, 50, 0.98 and 0.12 respectively. Our proposed GA gives the following results shown in Fig.1.(a), (b) for two points, three points and fifty points respectively.



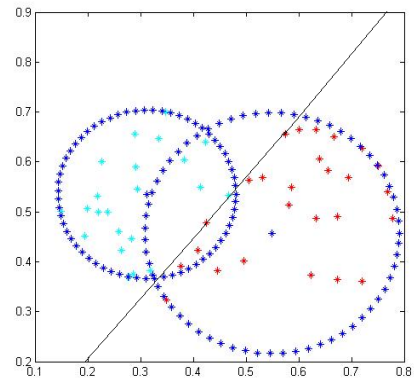
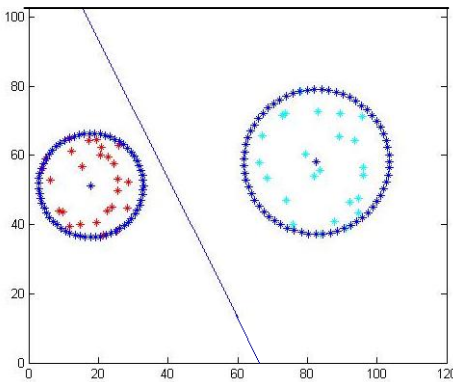
**Fig. 1.** (a)  $k$ -MECP for  $k=1$  of three points



**Fig. 1.** (b)  $k$ -MECP for  $k=1$  of fifty points

**4.2 GA-Based Technique to Solve the  $k$ -MECP for  $k=2$**

For two different set of input points of size fifty and the same parameter values of GA as mentioned in the Subsection 4.1, our proposed GA gives the following results shown in the Fig.2.(a), (b).

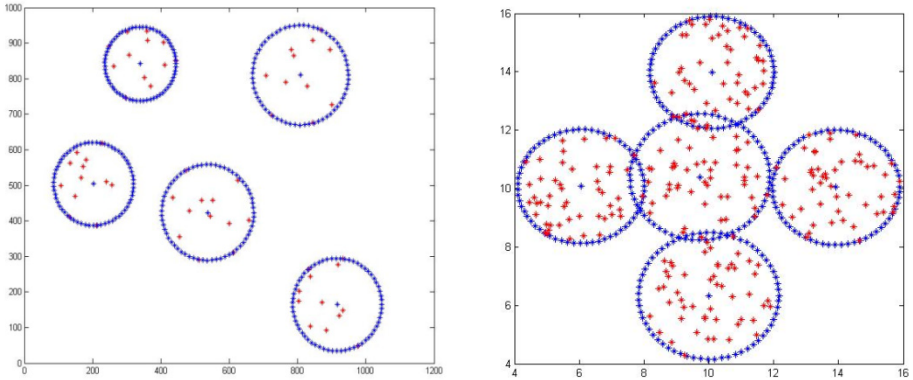


**Fig. 2.** (a), (b)  $k$ -MECP for  $k=2$  and of two different set each of fifty points

### 4.3 GA-Based Technique to Solve the Generalized $k$ -MECP

For any given value of  $k$ , the results obtained from our proposed GA are shown here. We have used the same parameter values of GA as mentioned in the Subsection 4.1. For a given point set of fifty points, Fig.3.(a) is obtained for  $k=5$  from our proposed GA.

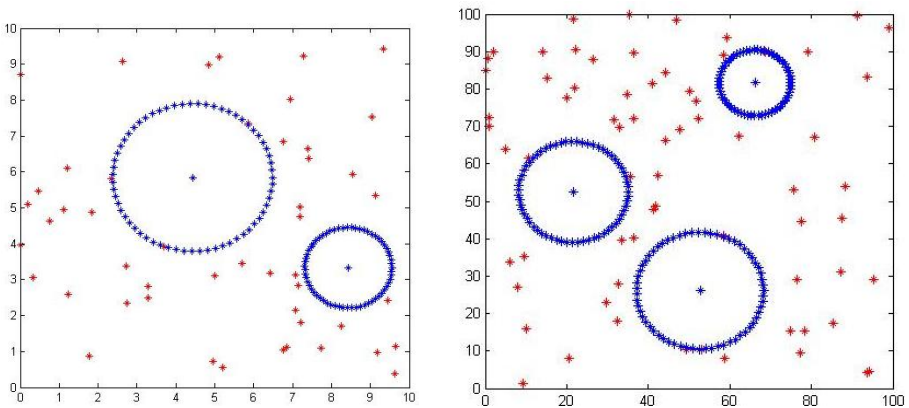
We have also used a standard data set of 250 points [7][8][9]. For the same values of GA parameter as mentioned in the Subsection 4.1 and  $k=5$ , we have obtained the result shown in the Fig.3.(b).



**Fig. 3.** (a) Generalized  $k$ -MECP for  $k=5$  and of fifty points; (b) Generalized  $k$ -MECP for  $k=5$  and of 250 points

### 4.4 GA-Based Technique to Solve the Generalized $k$ -LECP

For any given value of  $k$ , the results obtained from our proposed algorithm are shown here. We have used the same parameter values of GA as mentioned in the Subsection 4.1. For randomly generated fifty points and  $k=2$ , we have obtained the result from our proposed algorithm shown in the Fig.4.(a). For  $k=3$  and randomly generated seventy points, we have obtained the result shown in the Fig.4.(b).



**Fig. 4(a)**  $k$ -LECP for  $k=2$  of 50 points

**Fig. 4(b)**  $k$ -LECP for  $k=3$  of 70 points

## 5 Conclusion

In this article we have first proposed GA-based techniques to solve a problem ( $k$ -MECP) in network design for assigning (near) optimal transmission range to each antenna (facility) such that each customer gets service. A similar problem ( $k$ -LECP) has been studied when the facilities are hazardous in nature. The generalized  $k$ -MECP and the generalized  $k$ -LECP are known to be NP-hard problems. The proposed GA's are experimented on various synthetic and real world data sets. For the NP-hard problems, we get satisfactory results from our proposed GAs within reasonable generations.

## References

1. Drezner, Z., Hamache, H.W.: Facility location: Applications and Theory. Springer, Heidelberg (2001) ISBN 3-540-21345-7
2. Sharir, M., Welz, E.: Rectilinear and Polygonal  $p$ -Piercing and  $p$ -Center Problems. In: Proceedings of the 12th Annual Symposium on Computational Geometry, pp. 122–132 (1999)
3. Mukherjee, M., Chakraborty, K.: A polynomial-time optimization algorithm for a rectilinear partitioning problem with applications in VLSI design automation. Information Processing Letters 83, 41–48 (2002)
4. Kennington, J., Olinick, E., Rajan, D.: Wireless Network Design: Optimization Models and Solution Procedures. Springer, New York (2011)
5. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, New York (1989)
6. Davis, L. (ed.): Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York (1991)
7. Bandyopadhyay, S., Maulik, U.: Nonparametric genetic clustering: Comparison validity indices. IEEE Transactions on Systems, Man and Cybernetics, Part C 31, 120–125 (2001)
8. Bandyopadhyay, S., Maulik, U.: Genetic Clustering for Automatic Evolution of Clusters and Application to Image Classification. Pattern Recognition 35, 1197–1208 (2002)
9. Bandyopadhyay, S., Pal, S.K.: Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence. Springer, Heidelberg (2007)
10. Alt, H., Arkin, E.M., Bronnimann, H., Erickson, J., Fekete, S.P., Knauer, C., Lenchner, J., Mitchell, J.S.B., Whittlesey: Minimum-cost coverage of point sets by disks. In: Proceedings of the 22th Annual Symposium on Computational Geometry (SCG 2006), Sedona, Arizona, USA, pp. 449–458 (2006)
11. Abravaya, S., Segal, M.: Maximizing the number of obnoxious facilities to locate within a bounded region. Computers & Operations Research 37, 163–171 (2010)

# A Trilevel Programming Approach to Solve Reactive Power Dispatch Problem Using Genetic Algorithm Based Fuzzy Goal Programming

Papun Biswas<sup>1</sup> and Bijay Baran Pal<sup>2,\*</sup>

<sup>1</sup> Department of Electrical Engineering, JIS College of Engineering, Kalyani, Nadia, W.B., India

<sup>2</sup> Department of Mathematics, University of Kalyani, Nadia, W.B., India  
papunbiswas@yahoo.com, bbpal18@hotmail.com

**Abstract.** This article demonstrates how trilevel programming (TLP) in a hierarchical decision structure can be efficiently used for modeling and solving reactive power dispatch (RPD) problems of electrical power system by using genetic algorithms (GAs) in the framework of fuzzy goal programming (FGP) in uncertain environment. In the proposed approach, various objectives associated with a RPD problem are considered at three hierarchical levels in a planning horizon. In the solution process, a GA scheme is employed to obtain the individual values of objectives and thereby to evaluate the developed FGP model to reach a solution for optimal RPD decision. The proposed approach is tested on the standard IEEE 6-Generator 30-Bus System.

**Keywords:** Fuzzy goal programming, Genetic algorithm,  $L$ -index, Reactive power dispatch, Trilevel programming, Voltage profile.

## 1 Introduction

Optimization of RPD problem is one of the major issues in modern energy management system. Here, the main purpose is to minimize the real power losses of the system and improvement of voltage profile by satisfying load demand and operational constraints. Actually, an RPD problem is a multiobjective decision making (MODM) problem in power plant operational system.

The general mathematical programming model for optimum control of reactive power flow was introduced in [1]. Thereafter, the field was explored by the various active researchers in [2], [3] in the past. Now, to solve the RPD problem, a number of conventional optimization techniques in [4], [5], [6] have been discussed in the literature. To overcome the computational difficulty with nonlinear and competitive in nature objectives, global optimization techniques have been applied to solve the RPD problems in [7], [8], [9] in the recent past.

Now, it is to be observed that model parameters of most of the practical problems are often found inexact in nature. The two types of approaches for solving such

---

\* Corresponding author.

problems are stochastic programming (SP) [10], which deals with probabilistically uncertain data and fuzzy programming (FP) [11], which is based on fuzzy set theory (FST) [12] to deal with imprecisely described data. The SP approaches to RPD problem was studied in [13] in the past. Although, the FP approach to RPD problems has been discussed in [14], [15] previously, the deep study in this area is at an early stage.

In the context of solving RPD problems, although conventional MODM methods have been widely discussed in the past, it is to be noted that the conventional approaches often lead to poor performance regarding simultaneous optimization of objectives in a single framework of a traditional approach. Because, the objectives of such a problem are inherently incommensurable and effective decision cannot always be achieved owing to conflict interests of optimizing them in a decision situation. To overcome the difficulty, formulation of hierarchical decision problem in a decentralized decision system can reasonably be considered to make an appropriate decision in the decision environment.

In the above situation, trilevel programming problem (TLPP) [16] formulation of the problem as a special cases of multilevel programming problems (MLPPs) [17] in hierarchical decision system might be an effective one for solving RPD problems. Although, RPD problem in a hierarchical decision structure has been studied in [18] in the past, the deep study in this area is at an early stage. Further, the TLP approach to optimize RPD problem by employing GA based fuzzy FGP method is yet to appear in the literature.

In this paper, a GA-based FGP approach is proposed to solve the RPD problem. The proposed approach is tested on the standard IEEE 6-Generator 30-bus test system. The model solution is also compared with the approach studied in [9] previously to expound the potential use of the approach.

Now, an RPD problem is discussed in the following Section 2.

## 2 RPD Problem Description

Let there be  $N$  generators,  $G_i$ , ( $i= 1,2,\dots,N$ ),  $M$  tap-transformers,  $T_t$  ( $t= 1,2,\dots,M$ ), and  $C$  switchable *volt-ampere reactive* (VAR) sources,  $Q_c$  ( $c= 1,2,\dots,C$ ), in the power generation system. Then, let  $V_g$  be the decision variable of generator voltage of  $g$ -th generator,  $T_t$  be the decision variable of transformer tap setting of  $t$ -th transformer,  $Q_c$  be the decision variable of switchable VAR sources of  $c$ -th VAR source and  $P_D$  is the total demand associated with the system.

The objective functions and system constraints of the problem are discussed as follows.

### 2.1 Description of Objective Functions

**Power-loss Minimization Function.** The real power-loss ( $P_L$ ) in megawatt (MW) can be defined as [9]:

$$P_L = \sum_{r=1}^R g_r [V_i^2 + V_j^2 - 2V_i V_j \cos(\delta_i - \delta_j)] , \quad (1)$$

where ‘ $R$ ’ is the number of transmission lines,  $g_r$  is the conductance of the  $r$ -th line,  $V_i$  and  $V_j$  are the voltage magnitude, and  $\delta_i$  and  $\delta_j$  are the voltage phase angle at the end buses  $i$  and  $j$  of the  $r$ -th line, respectively, and ‘ $\cos$ ’ stands for *cosine* function.

**Voltage Profile Improvement Function.** Voltage profile improvement (VPI) of a system can be made by minimizing the bus voltage deviation ( $V_D$ ) from 1.0 per unit (p.u) in the system.

The objective function can be expressed as [9]:

$$V_D = \sum_{i=1}^I |V_i - 1.0| , \quad (2)$$

where ‘ $I$ ’ is the number of load buses.

**Voltage Stability Enhancement Function.** Maintenance of voltage stability and enhancement of it are the major issues concerning optimization of a RPD problem. Here, voltage stability enhancement (VSE) is achieved through minimization of voltage stability indicator,  $L$ -index in [9], values at every load bus of the system. The indicator value varies in the range between 0 (no load case) and 1 (collapse case).

The  $L$ -index at the load bus  $i$  can be express as:

$$L_i = \left| 1 + \frac{V_{oi}}{V_i} \right| ,$$

where  $V_{oi}$  and  $V_i$  indicate ‘no load voltage’ and ‘load voltage’, respectively, for bus  $i$ .

Voltage stability indicates that the condition  $L_i < 1$  should be satisfied and must not be violated on a continuous basis. Hence a global system indicator describing the stability of the complete system is  $L_{max} = \max \{L_i\}$ , where  $\{L_i\}$  contains  $L$ -indices of all load buses.

The objective function of VSE can be defined as:

$$L_{max} = \max\{L_i; i=1,2,\dots, I\} \quad (3)$$

## 2.2 Description of System Constraints

The system constraints which are inherently involved with the problem are defined as follows.

**Power Balance Constraints.** The real and reactive power balance equations are the typical load flow equations and they are represented as:

$$P_{Gi} - P_{Di} - V_i \sum_{j=1}^H V_j [G_{ij} \cos(\delta_i - \delta_j) + B_{ij} \sin(\delta_i - \delta_j)] = 0 , \quad (4)$$

$$Q_{Gi} - Q_{Di} - V_i \sum_{j=1}^H V_j [G_{ij} \sin(\delta_i - \delta_j) + B_{ij} \cos(\delta_i - \delta_j)] = 0 , \quad (5)$$

where  $j = 1, 2, \dots, H$ ; and where  $H$  is the number of all buses,  $P_{Gi}$  and  $Q_{Gi}$  are real and reactive power of the generator connected to the  $i$ -th bus, respectively;  $P_{Di}$  and  $Q_{Di}$  are the real and reactive power of the load connected to the  $i$ -th bus, respectively, and  $G_{ij}$



and  $B_{ij}$  are the transfer conductance and susceptance between bus  $i$  and bus  $j$ , respectively.

**Generation Capacity Constraints.** The upper and lower limits on generators' voltages and reactive power outputs are presented as:

$$\begin{aligned} V_{G_i}^{\min} &\leq V_{G_i} \leq V_{G_i}^{\max}, i = 1, 2, \dots, N \\ Q_{G_i}^{\min} &\leq Q_{G_i} \leq Q_{G_i}^{\max}, i = 1, 2, \dots, N \end{aligned} \quad (6)$$

where *min* and *max* stand for minimum and maximum, respectively.

**Transformer Tap-setting Constraints.** The upper and lower limits of transformer tap-settings [9] in the system are defined as follows:

$$T_t^{\min} \leq T_t \leq T_t^{\max}, t = 1, 2, \dots, M \quad (7)$$

where ' $M$ ' is the number of transformers.

**Switchable VAR Sources Constraints.** The switchable VAR sources are bounded by their limits as:

$$Q_c^{\min} \leq Q_c \leq Q_c^{\max}, c = 1, 2, \dots, C \quad (8)$$

where ' $C$ ' is the number of switchable VAR sources.

**Security Constraints.** The security constraint of voltages at load buses is defined as:

$$V_{L_i}^{\min} \leq V_{L_i} \leq V_{L_i}^{\max}, i = 1, 2, \dots, I \quad (9)$$

where ' $L$ ' stands for load.

Now, TLPP formulation of the proposed RPD problem is described in the follows Section 3.

### 3 TLPP Formulation of RPD Problem

The TLPP is actually a special case of multilevel programming (MLP) for multiobjective decision analysis of a problem with multiple decision makers (DMs) located at different decision levels in a hierarchical organizational system. In TLPP, three DMs (top- middle and bottom-level DMs) are located individually at the three different hierarchical levels and each one independently controls a vector of decision variables for optimizing his / her own objective function in the decision situation.

In TLP formulation of the proposed RPD problem, minimization of power-loss as the most important one is considered top-level problem. Minimization of voltage deviation and minimizing the voltage stability indicator are considered the middle and bottom-level problems, respectively. In a hierarchical decision process, execution of decision powers is sequential from the top-level to bottom-level in the decision environment.

Now, TLP model formulation of RPD problem is presented in the Section 3.1.

### 3.1 TLP Model Formulation

In a RPD problem, let the three vectors  $V$ ,  $T$  and  $Q$  be associated with the decision variables  $V_i$ ,  $T_i$  and  $Q_c$ , respectively.

Let the vectors  $V$ ,  $T$  and  $Q$  be controlled by the top-level, middle-level and bottom-level DMs, respectively, in a power supply decision system.

Then, TLP model of the problem can be presented as follows.

Find  $(V, T, Q)$  so as to:

$$\underset{V}{\text{Minimize}} P_L(V, T, Q) = \sum_{r=1}^R g_r [V_i^2 + V_j^2 - 2V_i V_j \cos(\delta_i - \delta_j)] \quad (\text{Top-level Problem})$$

and, for given  $V; T, Q$  solves

$$\underset{T}{\text{Minimize}} V_D(V, T, Q) = \sum_{i=1}^I |V_i - 1.0| \quad (\text{Middle-level Problem})$$

and, for given  $V$  and  $T; Q$  solves

$$\underset{Q_c}{\text{Minimize}} L_{max}(V, T, Q) = \max\{L_i\} \quad (\text{Bottom-level Problem})$$

subject to the system constraints in (4) - (9),

(10)

where  $L_{max}$  is given by (3),  $V \cap T \cap Q = \phi$ , and where  $\cap$  stands for the mathematical operation ‘intersection’.

Now, the GA scheme employed for modelling and solving the problem in (10) in the framework of FGP approach is presented in the following Section 4.

## 4 GA Scheme for RPD Problem

In the literature of GAs, a population based global random search approach, there are a number of schemes [19] for generation of new populations with the use of different operators: selection, crossover and mutation. Here, the real coded representation of a candidate solution called chromosome is considered to perform genetic operations in the solution search process. The tournament selection scheme, single-point crossover and uniform mutation operations are adopted to generate offspring in a new population in the search domain defined in the decision making environment.

The fitness value of each chromosome is determined by evaluating an objective function. The fitness function is defined as:

$$eval(E_v) = (Z_l)_v, l = 1, 2, 3; v = 1, 2, \dots, \text{pop\_size}, \quad (11)$$

where  $Z_l$  represents the objective function of  $l$ -th level DM, and where the subscript  $v$  designates the fitness value of the  $v$ -th chromosome,  $v = 1, 2, \dots, \text{pop\_size}$ .

The best chromosome with largest fitness value at each generation is determined as:

$$E^* = \max\{eval(E_v) | v = 1, 2, \dots, \text{pop\_size}\} \quad \text{or} \quad E^* = \min\{eval(E_v) | v = 1, 2, \dots, \text{pop\_size}\},$$

depends on searching of the maximum or minimum value of an objective function.

Now, formulation of FGP model of the trilevel RPD (TLRPD) problem in (10) is presented in the following Section.

## 5 FGP Model Formulation of TLRPD Problem

In the context of formulating TLRPD problem, it is assumed that all the DMs are motivated to cooperative to each other and each optimizes his / her benefit by paying an attention to the benefit of other one. Here, since top-level DM is in the leading position to make own decision, relaxation on decision of top-level DM is essentially needed to make reasonable decision by middle and bottom-level DMs individually. Therefore, relaxation of individual optimal values of objectives as well as decision vector  $\mathbf{V}$  controlled by top-level DM up to a certain tolerance level need be considered to make a reasonable balance of executing decision powers of DMs.

To cope with the situation, the fuzzy version of the problem in (10) is presented in the following Section.

### 5.1 Description of Fuzzy Goals

In a fuzzy decision situation, the objective functions are transformed into fuzzy goals.

In the sequel of making decision, since minimum values of individual objectives are always acceptable by each DM, first the independent best solutions of the three objectives are determined by employing GA scheme as:

$$(\mathbf{V}^{tb}, \mathbf{T}^{tb}, \mathbf{Q}^{tb}; P_L^{tb}), (\mathbf{V}^{mb}, \mathbf{T}^{mb}, \mathbf{Q}^{mb}; V_D^{mb}) \text{ and } (\mathbf{V}^{bb}, \mathbf{T}^{bb}, \mathbf{Q}^{bb}; L_{max}^{bb}), \text{ respectively,}$$

where  $P_L^{tb} = \text{Min}_{(\mathbf{V}, \mathbf{T}, \mathbf{Q}) \in S} P_L(\mathbf{V}, \mathbf{T}, \mathbf{Q}), V_D^{mb} = \text{Min}_{(\mathbf{V}, \mathbf{T}, \mathbf{Q}) \in S} V_D(\mathbf{V}, \mathbf{T}, \mathbf{Q})$  and

$$L_{max}^{bb} = \text{Min}_{(\mathbf{V}, \mathbf{T}, \mathbf{Q}) \in S} L_{max}(\mathbf{V}, \mathbf{T}, \mathbf{Q}), \text{ and where } tb, mb \text{ and } bb \text{ stand for top-level best, middle-level best and bottom-level best, respectively.}$$

Then, the fuzzy goals of the three levels can be successively defined as:

$$P_L \lesssim P_L^{tb}, \tag{12}$$

$$V_D \lesssim V_D^{mb}, \tag{13}$$

$$\text{and } L_{max} \lesssim L_{max}^{bb}, \tag{14}$$

where ‘ $\lesssim$ ’, refers to the fuzziness of an aspiration level.

Again, since maximum values of objectives when calculated in isolation would be the most dissatisfactory ones, the worst solutions of the objectives can be obtained by using same GA scheme as:

$$(\mathbf{V}^{tw}, \mathbf{T}^{tw}, \mathbf{Q}^{tw}; P_L^{tw}), (\mathbf{V}^{mw}, \mathbf{T}^{mw}, \mathbf{Q}^{mw}; V_D^{mw}) \text{ and } (\mathbf{V}^{bw}, \mathbf{T}^{bw}, \mathbf{Q}^{bw}; L_{max}^{bw}),$$

respectively, where  $P_L^{tw} = \text{Max}_{(V,T,Q) \in S} P_L(V,T,Q)$ ,  $V_D^{mw} = \text{Max}_{(V,T,Q) \in S} V_D(V,T,Q)$  and  $L_{max}^{bw} = \text{Max}_{(V,T,Q) \in S} L_{max}(V,T,Q)$ , and where *tw*, *mw* and *bw* stand for top-level worst, middle-level worst and bottom-level worst, respectively.

Again, the vector of fuzzy goals associated with control vector *V* concerning minimization of power-loss be defined as:

$$V \lesseqgtr V^{tb}, \tag{15}$$

Let  $V^e, (V^{tb} < V^e < V^{tw})$ , be the vector of upper-tolerance limits of achieving the goal levels of the vector of fuzzy goals defined in (15), and where *e* means tolerance.

Now, the fuzzy goals are to be characterized by the respective membership functions for measuring their degree of achievements in a fuzzy environment.

### 5.2 Characterization of Membership Function

The membership function of the fuzzy objective goal of top-level DM takes the form:

$$\mu_{P_L} [P_L(V, T, Q)] = \begin{cases} 1, & \text{if } P_L(V, T, Q) \leq P_L^{tb} \\ \frac{P_L^{tw} - P_L(V, T, Q)}{P_L^{tw} - P_L^{tb}}, & \text{if } P_L^{tb} < P_L(V, T, Q) \leq P_L^{tw} \\ 0, & \text{if } P_L(V, T, Q) > P_L^{tw} \end{cases} \tag{16}$$

where  $(P_L^{tw} - P_L^{tb})$  is the tolerance range for achievement of fuzzy goal in (12). The membership function of fuzzy objective goal of middle-level DM appears as:

$$\mu_{V_D} [V_D(V, T, Q)] = \begin{cases} 1, & \text{if } V_D(V, T, Q) \leq V_D^{mb} \\ \frac{V_D^{mw} - V_D(V, T, Q)}{V_D^{mw} - V_D^{mb}}, & \text{if } V_D^{mb} < V_D(V, T, Q) \leq V_D^{mw} \\ 0, & \text{if } V_D(V, T, Q) > V_D^{mw} \end{cases} \tag{17}$$

where  $(V_D^{mw} - V_D^{mb})$  is the tolerance range for achievement of fuzzy goal in (13).

Similarly, membership function of fuzzy objective goal of bottom-level DM takes the form:

$$\mu_{L_{max}} [L_{max}(V, T, Q)] = \begin{cases} 1, & \text{if } L_{max}(V, T, Q) \leq L_{max}^{bb} \\ \frac{L_{max}^{bw} - L_{max}(V, T, Q)}{L_{max}^{bw} - L_{max}^{bb}}, & \text{if } L_{max}^{bb} < L_{max}(V, T, Q) \leq L_{max}^{bw} \\ 0, & \text{if } L_{max}(V, T, Q) > L_{max}^{bw} \end{cases} \tag{18}$$

where  $(L_{max}^{bw} - L_{max}^{bb})$  is the tolerance range for achievement of fuzzy goal in (14).

The membership function of fuzzy decision vector *V* of leader appears as:

$$\mu_V [V] = \begin{cases} 1, & \text{if } V \leq V_i^{tb} \\ \frac{V^e - V}{V^e - V^{tb}}, & \text{if } V^{tb} < V \leq V^e \\ 0, & \text{if } V > V^e \end{cases} \tag{19}$$

where  $(V^e - V^{tb})$  is the vector of tolerance ranges for achievement of decision variables associated with *V* defined in (15).

**Note 1:**  $\mu [.]$  represents membership function.

Now, *minsum* FGP formulation of the TLRPD problem is presented in the following Section.

### 5.3 Minsum FGP Model Formulation

In the process of formulating *minsum* FGP [20] model, the defined membership functions are transformed into membership goals by assigning the highest membership value (unity) as the aspiration level and introducing under- and over-deviational variables to each of them. Then, minimization of the sum of weighted under-deviational variables of membership goals in the goal achievement function is considered.

The *minsum* FGP model can be presented as follows.

Find  $X = (V, T, Q)$  so as to:

Minimize:  $Z = \sum_{k=1}^3 w_k^- d_k^- + w_4^- d_4^-$

and satisfy

$$\begin{aligned} \mu_{P_L} &: \frac{P_L^{tw} - P_L(V, T, Q)}{P_L^{tw} - P_L^{tb}} + d_1^- - d_1^+ = 1, & \mu_{V_D} &: \frac{V_D^{mw} - V_D(V, T, Q)}{V_D^{mw} - V_D^{mb}} + d_2^- - d_2^+ = 1, \\ \mu_{L_{max}} &: \frac{L_{max}^{bw} - L_{max}(V, T, Q)}{L_{max}^{bw} - L_{max}^{bb}} + d_3^- - d_3^+ = 1, & \mu_V &: \frac{V^e - V}{V^e - V^{tb}} + d_4^- - d_4^+ = I, \end{aligned}$$

subject to the set of constraints defined in (4) - (9),

(20)

where  $d_k^-, d_k^+ \geq 0$ , ( $k = 1, 2, 3$ ) represent under- and over-deviational variables, respectively, associated with the respective membership goals.  $d_4^-, d_4^+ \geq 0$  represent the vector of under- and over-deviational variables, respectively, associated with membership goals defined for the vector of decision variables in  $V$ , and where  $I$  is a column vector with all elements equal to 1 and the dimension of it depends on dimension of  $V$ .  $Z$  represents goal achievement function,  $w_k^- > 0$ ,  $k = 1, 2, 3$ , denote the relative numerical weights of importance of achieving the aspired goal levels, and  $w_4^- > 0$  is the vector of numerical weights associated with  $d_4^-$ , and they are

determined as:

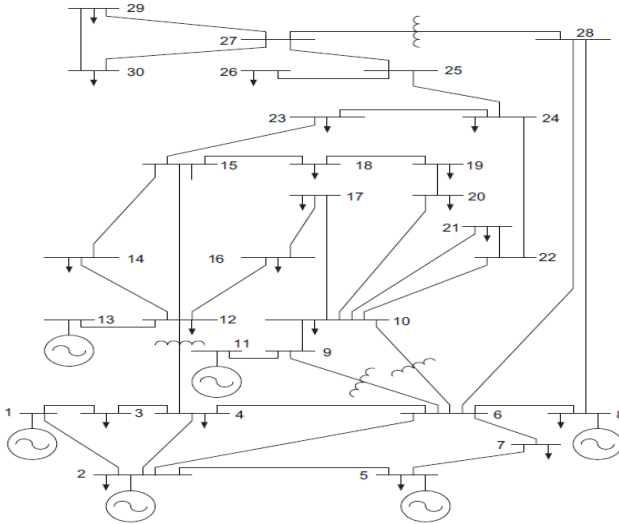
$$w_k^- = \begin{cases} \frac{1}{P_L^{tw} - P_L^{tb}}, & \text{for the defined fuzzy goal in (12)} \\ \frac{1}{V_D^{mw} - V_D^{mb}}, & \text{for the defined fuzzy goal in (13)} \\ \frac{1}{L_{max}^{bw} - L_{max}^{bb}}, & \text{for the defined fuzzy goal in (14)} \end{cases}$$

and  $w_4^- = \frac{1}{V^e - V^{tb}}$ , for the defined fuzzy goal in (15)

Now, the effective use of the *minsum* FGP model in (20) is demonstrated via a case example presented in the Section 6.

## 6 A Demonstrative Case Example

The effective use of the proposed GA based FGP approach is tested on the standard IEEE 6-generator 30-bus test system [9], which is diagrammatically shown in Figure 1.



**Fig. 1.** Single-line diagram of IEEE 30-bus test system

The system has 6 generators, 41 lines and 21 load-buses with 19-control variable (6- generator voltage magnitude, 4-tap transformer setting and 9-switchable VAR). The total system demand is 2.834 p.u.

The GA is implemented by employing GA-Toolbox under MATLAB Optimization Tool (MATLAB-Ver. R2010a) at different stages for evaluation of the problem. The execution is made in Intel Pentium IV with 2.66 GHz. Clock-pulse and 4 GB RAM.

Now, the following parameter values are introduced to solve the problem.

- Probability of crossover  $P_c = 0.8$ , probability of mutation  $P_m = 0.06$ .
- Ppopulation size = 50 and generation number = 100.

The goal achievement function  $Z$  in (20) appears as the evaluation function in the GA search process to solve the problem.

The evaluation function for determination of fitness of a chromosome appears as:

$$Eval(E_v) = (Z)_v = \left( \sum_{k=1}^3 w_k^- d_k^- + \sum_{k=4}^9 w_k^- d_k^- \right)_v, \quad v = 1, 2, \dots, \text{pop\_size} \tag{21}$$

where  $(Z)_v$  is used to represent the achievement function  $(Z)$  in (20) for measuring the fitness value of  $v$ -th chromosome in the decision process.

The best objective value  $(Z^*)$  for the fittest chromosome at a generation in the solution search process is determined as:

$$Z^* = \min \{eval(E_v) \mid v = 1, 2, \dots, \text{pop\_size}\} \tag{22}$$

Now, to formulate the TLRPD problem, the standard data presented in [9] are considered.

Following the procedure, the executable FGP model can be obtained by (20).

The achieved objective values are  $(P_L, V_D, L_{max}) = (5.169, 0.2401, 0.1579)$ .

The resultant membership values are  $(\mu_{P_L}, \mu_{V_D}, \mu_{L_{max}}) = (0.95, 0.91, 0.89)$ .

The resulting decisions associated with the given system are obtained as follows:

$$(V_1, V_2, V_5, V_8, V_{11}, V_{13}) = (1.055, 1.042, 1.035, 1.036, 1.085, 1.064),$$

$$(T_{11}, T_{12}, T_{15}, T_{36}) = (0.9536, 0.9067, 0.9990, 0.9662), \text{ and}$$

$$(Q_{10}, Q_{12}, Q_{15}, Q_{17}, Q_{20}, Q_{21}, Q_{23}, Q_{24}, Q_{29}) = (3.871, 4.151, 4.812, 3.735, 4.617, 4.828, 3.781, 4.512, 2.690).$$

The result reflects that the solution is quite satisfactory from the view point of executing the decision powers of DMs on the basis of hierarchical order of optimizing the objectives of the RPD problem.

Now, to show the potential use of the approach, the model solution is compared with the solution obtained by using the differential evolution algorithm in [9]. The achievement of objective values for individual optimization of them under the previous approach is presented in Table 1.

**Table 1.** Achieved objective values under differential evolution algorithm

Objectives	Differential evolution algorithm		
	Case1: Minimization of power-loss	Case2: Voltage profile improvement	Case3: Voltage stability enhancement
Power-loss (MW)	4.550	6.4755	7.0733
Voltage deviation (p.u)	1.9589	0.0911	1.4191
$L_{max}$ (p.u)	0.5513	0.5734	0.1246

The graphical representations of the results obtained by using the proposed method and the earlier method are displayed in Figures 2-4.

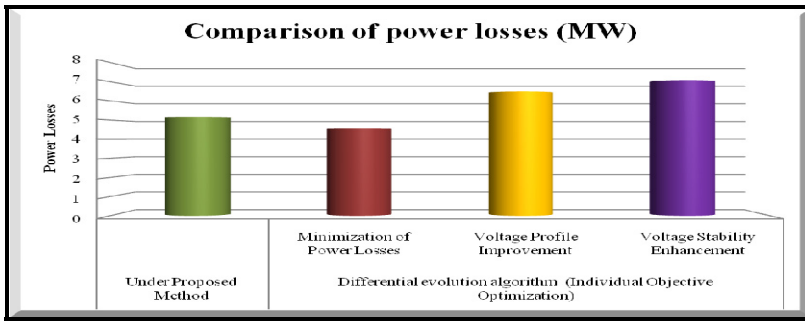


Fig. 2. Comparison of power losses (MW) achieved under different approaches

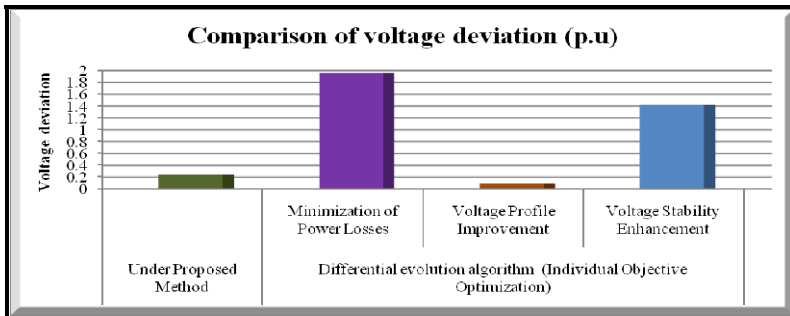


Fig. 3. Comparison of voltage deviations (p.u) achieved under different approaches

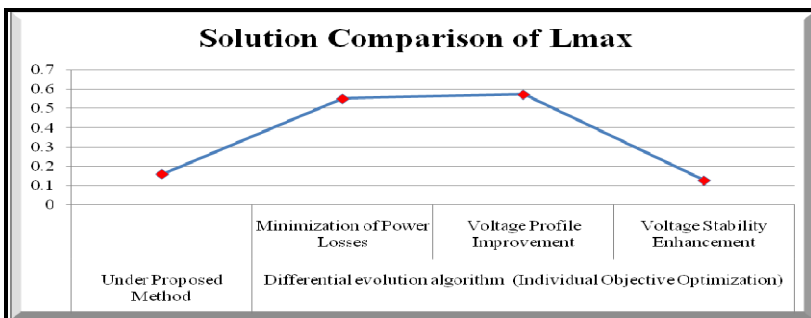


Fig. 4. Comparison of  $L_{max}$  (p.u) values achieved under different approaches

A comparison shows that the solution obtained by using the proposed GA based FGP approach is superior over the other one from the view point of achieving a compromise solution for optimizing the objectives of the RPD problem.

## 7 Conclusions

The effective use of GA based FGP approach for modeling and solving an RPD problem in a hierarchical decision structure is presented in this paper. The main



advantage of TLP formulation of the problem is that the decisions regarding optimization of objectives on the basis of hierarchy of important can be obtained here in the decision situation. However, it is hoped that the solution approach presented here may lead to future research for optimization of RPD planning problems in the current complex arena of thermal power generation and dispatch decision environment.

**Acknowledgments.** The authors are thankful to the anonymous Reviewers and Special-Session Chair-Person, Prof. J.K. Mandal, of CSI-2013 for their valuable comments and suggestions to improve the quality of presentation of the paper.

## References

1. Peschon, J., Piercy, D.S., Tinney, W.F., Tveit, O.J., Cuenod, M.: Optimum Control of Reactive Power Flow. *IEEE Transactions on Power Apparatus and Systems* PAS-87(1), 40–48 (1968)
2. Fernandes, R.A., Happ, H.H., Wirgau, K.A.: Optimal Reactive Power Flow for Improved System Operations. *International Journal of Electrical Power & Energy Systems* 2(3), 133–139 (1980)
3. Mamandur, K.R.C., Chenoweth, R.D.: Optimal Control of Reactive Power Flow for Improvements in Voltage Profiles and for Real Power Loss Minimization. *IEEE Transactions on Power Apparatus and Systems* PAS-100(7), 3185–3194 (1981)
4. Lee, K.Y., Park, Y.M., Ortiz, J.L.: An United Approach to Optimal Real and Reactive Power Dispatch. *IEEE Transactions on Power Apparatus and Systems* PAS-104(5), 1147–1153 (1985)
5. Deeb, N., Shahidehpur, S.M.: Linear Reactive Power Optimization in a Large Power Network using the Decomposition Approach. *IEEE Transactions on Power Systems* 5(2), 428–435 (1990)
6. Granville, S.: Optimal Reactive Power Dispatch through Interior Point Methods. *IEEE Transactions on Power Systems* 9(1), 98–105 (1994)
7. Yan, W., Liu, F., Chung, C., Wong, K.: A Hybrid Genetic Algorithm-interior Point Method for Optimal Reactive Power Flow. *IEEE Transactions on Power Systems* 21(3), 1163–1169 (2006)
8. Mahadevan, K., Kannan, P.S.: Comprehensive Learning Particle Swarm Optimization for Reactive Power Dispatch. *International Journal of Applied Soft Computing* 10(2), 641–652 (2010)
9. El Ela, A.A.A., Abido, M.A., Spea, S.R.: Differential Evolution Algorithm for Optimal Reactive Power Dispatch. *Electric Power Systems Research* 81, 458–468 (2011)
10. Kall, P., Wallace, S.W.: *Stochastic programming*. John Wiley & Sons, Chichester (1994)
11. Zimmermann, H.-J.: *Fuzzy Programming and Linear Programming with Several Objective Functions*. *Fuzzy Sets and Systems* 1(1), 45–55 (1978)
12. Zadeh, L.A.: *Fuzzy Sets*. *Information and Control* 8(3), 338–353 (1965)
13. Kargarian, A., Raoofat, M., Mohammadi, M.: Probabilistic Reactive Power Procurement in Hybrid Electricity Markets with Uncertain Loads. *Electric Power Systems Research* 82(1), 68–80 (2012)
14. Tomovic, K.: A Fuzzy Linear Programming Approach to the Reactive Power/voltage Control Problem. *IEEE Transactions on Power Systems* 7(1), 287–293 (1992)

15. Abdul-Rahman, K.H., Shahidehpour, S.M.: A Fuzzy-based Optimal Reactive Power Control. *IEEE Transactions on Power Systems* 8(2), 662–670 (1993)
16. Zhang, G., Lu, J., Montero, J., Zeng, Y.: Model, Solution Concept, and *K*th-best Algorithm for Linear Trilevel Programming. *Information Sciences* 180(4), 481–492 (2010)
17. Sakawa, M., Nishizaki, I.: *Cooperative and Non-cooperative Multi-Level Programming*. Springer, New York (2009)
18. Almeida, K.C., Senna, F.S.: Optimal Active-reactive Power Dispatch under Competition via Bilevel Programming. *IEEE Transactions on Power Systems* 26(4), 2345–2354 (2011)
19. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Longman Publishing Co. Inc., Boston (1989)
20. Pal, B.B., Moitra, B.N., Maulik, U.: A Goal Programming Procedure for Fuzzy Multiobjective Linear Fractional Programming Problem. *Fuzzy Sets and Systems* 139(2), 395–405 (2003)

# An Iterative Fuzzy Goal Programming Method to Solve Fuzzified Multiobjective Fractional Programming Problems

Mousumi Kumar<sup>1</sup>, Shyamal Sen<sup>2</sup>, and Bijay Baran Pal<sup>3,\*</sup>

<sup>1</sup> Department of Mathematics, Alipurduar College,  
Alipurduar Court-736122, West Bengal, India  
mousumi886@gmail.com

<sup>2</sup> Department of Mathematics, B.K.C. College, Kolkata-700108, West Bengal, India  
ssenk@yahoo.co.in

<sup>3</sup> Department of Mathematics, University of Kalyani, Kalyani-741235, West Bengal, India  
bbpal18@hotmail.com

**Abstract.** This paper presents a fuzzy goal programming (FGP) approach for modeling and solving multiobjective fractional programming problems (MOFPPs) with fuzzy numbers parameter sets. In the proposed approach, first the notion of  $\alpha$ -cut in fuzzy sets (FSs) is used to transform a problem into conventional MOFPP by using the tolerance membership functions in FSs. In model formulation, membership functions are converted into fuzzy goals for measuring the degree of satisfaction of decision maker (DM) with the solution for achievement of fuzzily described objectives of the problem. In the solution process, an iterative parametric method is addressed within the framework of *minsum* FGP model to reach the highest membership value (unity) to the extent possible in the decision making environment. The efficiency of the proposed approach is illustrated by a numerical example. The model solution is also compared with the solution obtained by using other approaches studied previously.

**Keywords:** Fractional programming, Fuzzy Goal programming, Iterative parametric method, Membership functions, Triangular fuzzy number.

## 1 Introduction

The mathematical framework of a fractional programming problem (FPP) and constructive solution procedure under the conventional linear programming was first proposed by Charnes and Cooper [1]. Thereafter, the field fractional programming was studied extensively [2, 3], from the point of view of its potential application to different real-life problems [4].

Now, since most of the real-world decision problems are multiobjective in nature, FPPs with multiplicity of objectives were studied by Steuer [5] in 1986. The goal

---

\* Corresponding author.

programming (GP) [6] approach, a promising tool for multiobjective decision analysis, have been studied by Kornbluth and Steuer [7, 8] for decision analysis with fractional objectives in crisp decision environment. But, in the real-world multiobjective decision making (MODM) situations, DMs are frequently faced with the problem of setting exact aspiration levels for each goal. To cope with the situation, fuzzy programming [9] has appeared as a robust tool for solving MODM problems in imprecise decision environment. The fuzzy programming approach to FPPs was first studied by Luhandjula [10] in 1984. From the mid-1980s to late 1990s, the methodological aspects of fuzzy set theoretic approaches [11] to MOFPPs have been studied by the active researchers in the field.

To overcome the difficulty with incommensurability and conflict in nature of objectives in a MODM environment, FGP [12, 13] as an extension of conventional GP in the area of fuzzy programming acts a prominent role to make flexible decisions of MOFPPs. FGP with the use of variable change method for solving MOFPPs has been suggested by Pal et al. [14] in recent past. Further, to overcome the computational load for linearization of objectives with the previous approach, a parametric programming based approach [15] has also been used for solving MOFPPs.

In this paper, a fully fuzzified version of MOFPP is considered where all the parameters associated with the objectives as well as system constraints are described fuzzily. In the proposed approach, the notion of  $\alpha$ -cut of fuzzy numbers [16] is introduced for achieving the solution of the problem to a desired degree of satisfaction. In fuzzy programming formulation, objective functions and constraints are decomposed into deterministic equivalents by using the concept of  $\alpha$ -cut interval [17]. Then, the objective functions are transformed into fuzzy goals by means of assigning imprecise aspiration level to each of them. In the sequel of FGP formulation, first the tolerance membership functions for measuring the degree of optimality of each of the fuzzy goals are defined, and the membership functions are then transformed into flexible goals as in conventional GP by introducing highest membership value (unity) as aspiration level and introducing under- and over-deviational variables to each of them. In the solution process, the *iterative parametric approach* is addressed to solve the problem by employing *minsum* FGP [13] methodology.

Now, the general fuzzy MOFPP formulation is presented in section 2.

## 2 Model Formulation

The general format of a fuzzily described linear MOFPP can be stated as follows.

Find  $X(x_1, x_2, \dots, x_p)$  so as to:

$$\text{Maximize } Z_{i1}(X) = \frac{\tilde{G}_{i1}X + \tilde{a}_{i1}}{\tilde{H}_{i1}X + \tilde{b}_{i1}}, \quad i=1, 2, \dots, m$$

$$\text{Minimize } Z_{j2}(X) = \frac{\tilde{G}_{j2}X + \tilde{a}_{j2}}{\tilde{H}_{j2}X + \tilde{b}_{j2}}, \quad j=1, 2, \dots, n$$

$$\text{subject to } X \in S\{X^T \in \mathfrak{R}^p \mid \tilde{C}X \begin{matrix} (\leq) \\ (\geq) \\ (=) \end{matrix} \tilde{c}, X \geq 0\}, \tag{1}$$

where  $\tilde{G}_{i1}, \tilde{G}_{j2}, \tilde{H}_{i1}, \tilde{H}_{j2}$  and  $\tilde{c}$  are vectors of triangular fuzzy numbers (TFNs),  $\tilde{a}_{i1}, \tilde{a}_{j2}, \tilde{b}_{i1}, \tilde{b}_{j2}$  are constant TFNs where  $i=1, 2, \dots, m; j=1, 2, \dots, n$  and  $\tilde{C}$  is a matrix of TFNs of order  $r \times p$  and where  $\mathfrak{R}$  is the set of real numbers,  $T$  means transposition.

Also, it is customary to assume that  $\tilde{H}_{i1}X + \tilde{b}_{i1} > \tilde{0}; \tilde{H}_{j1}X + \tilde{b}_{j2} > \tilde{0}$  to avoid any undefined situation and the feasible region  $S (\neq \emptyset)$  is bounded.

The description of TFNs and some useful operations on TFNs [17] associated with the problem is discussed in the following section.

### 2.1 Preliminaries of TFNs

- Definition of TFN

The TFN  $\tilde{A}$  can be represented as an ordered triple  $\tilde{A} = (a_1, a_2, a_3), a_t \in \mathfrak{R}; t = 1, 2, 3$ . The diagrammatic presentation of a TFN is depicted in Figure 1.

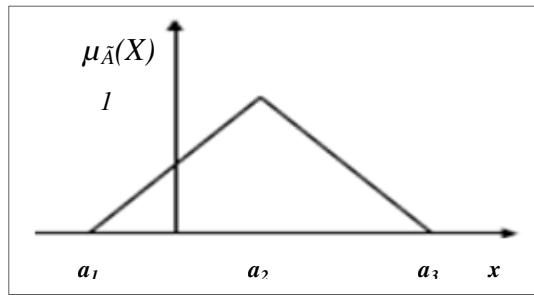


Fig. 1. Triangular fuzzy number  $\tilde{A} = (a_1, a_2, a_3)$

- Definition of Membership function of a TFN

The membership function associated with degree of achievement of  $\tilde{A}$  can be defined as [17]:

$$\mu_{\tilde{A}}(X) = \left. \begin{matrix} 0, & x < a_1 \\ \frac{x - a_1}{a_2 - a_1}, & a_1 \leq x \leq a_2 \\ \frac{a_3 - x}{a_3 - a_2}, & a_2 \leq x \leq a_3 \\ 0, & x > a_3 \end{matrix} \right\} \tag{2}$$

- Definition of  $\alpha$ - cut

The  $\alpha$ - cut of  $\tilde{A}$  is a set consisting of those elements whose membership values exceed the level  $\alpha$ , i.e,  $A_\alpha = \{x; \mu_{\tilde{A}}(X) \geq \alpha\}$ .

- Definition of Interval of a TFN

Using the notion of  $\alpha$ -cut [17],  $\tilde{A}$  can be defined by an interval  $A_\alpha = [a_\alpha^L, a_\alpha^R]$ , where  $\frac{a_\alpha^L - a_1}{a_2 - a_1} = \alpha$  and  $\frac{a_3 - a_\alpha^R}{a_3 - a_2} = \alpha$ , and where  $\alpha \in [0, 1]$ , and  $\alpha$  indicates the degree of satisfaction of achieving a value of a fuzzy number, where L and R indicate Left-Right (L-R) type representation of  $\tilde{A}$  by  $\mu_{\tilde{A}}(X)$  [16].

Then,  $A_\alpha$  can be expressed as:

$$A_\alpha = [(a_2 - a_1)\alpha + a_1, -(a_3 - a_2)\alpha + a_3]. \tag{3}$$

- Arithmetic operations on TFNs

Let  $\tilde{A} = (a_1, a_2, a_3)$  and  $\tilde{B} = (b_1, b_2, b_3)$  be two TFNs with membership functions  $\mu_{\tilde{A}}(X)$  and  $\mu_{\tilde{B}}(X)$ , respectively.

Let  $A_\alpha$  and  $B_\alpha$  be the  $\alpha$ -cuts of the fuzzy numbers  $\tilde{A}$  and  $\tilde{B}$ , respectively, Now different arithmetic operations on TFNs can be obtained as follows [17]:

- Addition

$$A_\alpha + B_\alpha = [(a_2 - a_1 + b_2 - b_1)\alpha + a_1 + b_1, -(a_3 - a_2 + b_3 - b_2)\alpha + a_3 + b_3].$$

- Subtraction

$$A_\alpha - B_\alpha = [(a_2 - a_1 - b_2 + b_1)\alpha + a_1 - b_1, -(a_3 - a_2 - b_3 + b_2)\alpha + a_3 - b_3].$$

- Multiplication

$$A_\alpha * B_\alpha = [(a_2 - a_1)(b_2 - b_1)\alpha^2 + (b_1(a_2 - a_1) + a_1(b_2 - b_1))\alpha + a_1b_1, (a_3 - a_2)(b_3 - b_2)\alpha^2 - (b_2(a_3 - a_2) + a_2(b_3 - b_2))\alpha + a_3b_3].$$

- Division

$$A_\alpha / B_\alpha = \left[ \frac{(a_2 - a_1)\alpha + a_1}{-(b_3 - b_2)\alpha + b_3}, \frac{-(a_3 - a_2)\alpha + a_3}{(b_2 - b_1)\alpha + b_1} \right],$$

provided  $a_t, b_t \geq 0, t = 1, 2, 3$  with  $-(b_3 - b_2)\alpha + b_3 > 0, (b_2 - b_1)\alpha + b_1 > 0$ .

Now, the general format of a MOFPP can be presented as follows.

## 2.2 General MOFPP Formulation

Using the notion of  $\alpha$ -cut, the MOFPP in (1) can be presented as follows:

Find  $X(x_1, x_2, \dots, x_p)$  so as to:

$$\begin{aligned}
 &\text{Maximize } Z_{i1\alpha}^L(X) = \frac{G_{i1}^L X + a_{i1}^L}{H_{i1}^R X + b_{i1}^R}, \quad \text{Maximize } Z_{i1\alpha}^R(X) = \frac{G_{i1}^R X + a_{i1}^R}{H_{i1}^L X + b_{i1}^L}, \quad i=1,2,\dots,m \\
 &\text{Minimize } Z_{j2\alpha}^L(X) = \frac{G_{j2}^L X + a_{j2}^L}{H_{j2}^R X + b_{j2}^R}, \quad \text{Minimize } Z_{j2\alpha}^R(X) = \frac{G_{j2}^R X + a_{j2}^R}{H_{j2}^L X + b_{j2}^L}, \quad j=1,2,\dots,n \\
 &\text{subject to } C_{\alpha}^L X \begin{pmatrix} \leq \\ \geq \\ = \end{pmatrix} c_{\alpha}^L \quad \text{and} \quad C_{\alpha}^R X \begin{pmatrix} \leq \\ \geq \\ = \end{pmatrix} c_{\alpha}^R, \quad X \geq 0 \tag{4}
 \end{aligned}$$

Then, Fuzzy programming formulation of the problem is presented in the following section 3.

### 3 Fuzzy Programming Problem Formulation

In the field of FP, an imprecise aspiration level is assigned to each of the objectives and tolerance limit for achievement of each of them is taken into consideration.

Let  $Z_{i1\alpha}^{bL}, Z_{i1\alpha}^{bR}, Z_{j2\alpha}^{bL}, Z_{j2\alpha}^{bR}$  be the individual best values of the defined objectives  $Z_{i1\alpha}^L, Z_{i1\alpha}^R, Z_{j1\alpha}^L, Z_{j2\alpha}^R$ , respectively, subject to the constraints in (4), where  $Z_{i1\alpha}^{bL} = \text{Max } Z_{i1\alpha}^L; Z_{i1\alpha}^{bR} = \text{Max } Z_{i1\alpha}^R, \quad i=1, 2, \dots, m, \quad \text{and} \quad Z_{j2\alpha}^{bL} = \text{Min } Z_{k\alpha}^L; Z_{j2\alpha}^{bR} = \text{Min } Z_{j2\alpha}^R, \quad j=1, 2, \dots, n$ , where  $b$  stands for the best, and  $\text{Max}$  and  $\text{Min}$  stand for maximum and minimum, respectively.

Then, the fuzzy goals of the problem can be presented as:

$$\begin{aligned}
 &Z_{i1\alpha}^L(X) \gtrsim Z_{i1\alpha}^{bL} \quad \text{and} \quad Z_{i1\alpha}^R(X) \gtrsim Z_{i1\alpha}^{bR}, \quad i=1, 2, \dots, m \\
 &Z_{j2\alpha}^L(X) \lesssim Z_{j2\alpha}^{bL} \quad \text{and} \quad Z_{j2\alpha}^R(X) \lesssim Z_{j2\alpha}^{bR}, \quad j=1, 2, \dots, n \tag{5}
 \end{aligned}$$

The lower- and upper- tolerance limits for goal achievement of the successive fuzzy goals are defined as  $Z_{i1\alpha}^{wL}, Z_{i1\alpha}^{wR}, Z_{j2\alpha}^{wL}, Z_{j2\alpha}^{wR}$ , where  $Z_{i1\alpha}^{wL} = \text{Min } Z_{i1\alpha}^L; Z_{i1\alpha}^{wR} = \text{Min } Z_{i1\alpha}^R, \quad i=1, 2, \dots, m$ , and  $Z_{j2\alpha}^{wL} = \text{Max } Z_{k\alpha}^L; Z_{j2\alpha}^{wR} = \text{Max } Z_{j2\alpha}^R, \quad j=1,2,\dots, n$ , and  $w$  stands for worst. Here  $\gtrsim$  and  $\lesssim$  indicate the fuzzy versions of the symbols  $\geq$  and  $\leq$ , respectively.

Now, to formulate the model of the problem, fuzzy goals are to be characterized by their associated membership functions.

### 3.1 Characterization of Membership Function

For  $\succ$  type of restriction, the membership function  $\mu_{Z_{i1\alpha}^{L}}(X)$  takes the form [18]:

$$\mu_{Z_{i1\alpha}^{L}}(X) = \begin{cases} 1, & \text{if } Z_{i1\alpha}^L(X) \geq Z_{i1\alpha}^{bL} \\ \frac{Z_{i1\alpha}^L(X) - Z_{i1\alpha}^{wL}}{Z_{i1\alpha}^{bL} - Z_{i1\alpha}^{wL}}, & \text{if } Z_{i1\alpha}^{wL} < Z_k(X) < Z_{i1\alpha}^{bL} \\ 0, & \text{if } Z_k(X) \leq Z_{i1\alpha}^{wL}, \end{cases} \quad (6)$$

where  $(Z_{i1\alpha}^{bL} - Z_{i1\alpha}^{wL})$  is the tolerance range for achievement of the  $i$ -th fuzzy goal,  $i=1,2,\dots,m$ .

Similarly, for  $\lesssim$  type of restriction, the membership function  $\mu_{Z_{j2\alpha}^{L}}(X)$  takes the form:

$$\mu_{Z_{j2\alpha}^{L}}(X) = \begin{cases} 1, & \text{if } Z_{j2\alpha}^L(X) \leq Z_{j2\alpha}^{bL} \\ \frac{Z_{j2\alpha}^{wL} - Z_{j2\alpha}^L(X)}{(Z_{j2\alpha}^{wL} - Z_{j2\alpha}^{bL})}, & \text{if } Z_{j2\alpha}^{bL} < Z_{j2\alpha}^L(X) < Z_{j2\alpha}^{wL} \\ 0, & \text{if } Z_{j2\alpha}^L(X) \geq Z_{j2\alpha}^{wL} \end{cases} \quad (7)$$

where  $(Z_{j2\alpha}^{wL} - Z_{j2\alpha}^{bL})$  is the tolerance range for achievement of the  $j$ -th fuzzy goal,  $j=1,2,\dots,n$ .

The FGP model formulation of the problem is described in section 4.

## 4 FGP Model Formulation

In FGP model formulation of the problem, the defined membership functions are converted into membership goals by introducing under- and over-deviational variables and assigning the highest membership value (unity) as the aspiration level to each of them. In the ‘goal achievement function’, minimization of the under-deviational variables on the basis of weights of importance of achieving the goals are taken into consideration. Such a version of FGP is the simplest and widely used approach which is called *minsum* FGP [13] for solving MODM problems.

The executable *minsum* FGP model appears as follows.

Find  $X(x_1, x_2, x_3)$  so as to:

$$\text{Minimize } Z = \sum_{i=1}^m (w_{i1}^{L-} d_{i1}^{L-} + w_{i1}^{R-} d_{i1}^{R-}) + \sum_{j=1}^n (w_{j2}^{L-} d_{j2}^{L-} + w_{j2}^{R-} d_{j2}^{R-})$$

and satisfy

$$\begin{aligned} \mu_{Z_{i1\alpha}^L} : \frac{Z_{i1\alpha}^L(X) - Z_{i1\alpha}^{wL}}{Z_{i1\alpha}^{bL} - Z_{i1\alpha}^{wL}} + d_{i1}^{L-} - d_{i1}^{L+} = 1, \quad \mu_{Z_{i1\alpha}^R} : \frac{Z_{i1\alpha}^R(X) - Z_{i1\alpha}^{wR}}{Z_{i1\alpha}^{bR} - Z_{i1\alpha}^{wR}} + d_{i1}^{R-} - d_{i1}^{R+} = 1, \quad i = 1, 2, \dots, m \\ \mu_{Z_{j2\alpha}^L} : \frac{Z_{j2\alpha}^{wL} - Z_{j2\alpha}^L(X)}{(Z_{j2\alpha}^{wL} - Z_{j2\alpha}^{bL})} + d_{j2}^{L-} - d_{j2}^{L+} = 1, \quad \mu_{Z_{j2\alpha}^R} : \frac{Z_{j2\alpha}^{wR} - Z_{j2\alpha}^R(X)}{(Z_{j2\alpha}^{wR} - Z_{j2\alpha}^{bR})} + d_{j2}^{R-} - d_{j2}^{R+} = 1, \quad j = 1, 2, \dots, n \end{aligned}$$

subject to the system constraints in (4), (8)



where  $Z$  represents the fuzzy goal achievement function,  $(d_{i1}^{L-}, d_{i1}^{L+})$  and  $(d_{i1}^{R-}, d_{i1}^{R+}) \geq 0$  with  $d_{j2}^{L-} \cdot d_{j2}^{L+} = 0$  and  $d_{j2}^{R-} \cdot d_{j2}^{R+} = 0$  represent the under- and over-deviational variables, respectively, associated with the respective membership goal,  $w_{i1}^{L-}, w_{i1}^{R-}, w_{j2}^{L-}, w_{j2}^{R-} (> 0), i = 1, 2, \dots, m; j = 1, 2, \dots, n$  represent the numerical weights of importance relative to others and associated with the respective under-deviational variables, and they are determined as [13]:

$$w_{i1}^{L-} = \frac{1}{Z_{i1\alpha}^{bL} - Z_{i1\alpha}^{wL}}, \quad w_{i1}^{R-} = \frac{1}{Z_{i1\alpha}^{bR} - Z_{i1\alpha}^{wR}}, \quad i = 1, 2, \dots, m$$

$$w_{j2}^{L-} = \frac{1}{Z_{j2\alpha}^{bL} - Z_{j2\alpha}^{wL}}, \quad w_{j2}^{R-} = \frac{1}{Z_{j2\alpha}^{bR} - Z_{j2\alpha}^{wR}}, \quad j = 1, 2, \dots, n$$

It is to be observed that the goals in (8) are actually fractional in form. To avoid the computational load with the traditional variable transformation approach [14] for linearization of the fractional goals in (8), iterative parametric linear programming approach, initially introduced by Dinklebach [15], to FGP formulation studied by Kumar and Pal [19] is introduced for achievement of membership goals in the decision situation.

The parametric linear GP (PLGP) method for solving the problem is presented in section 4.1.

### 4.1 PLGP Formulation

The fractional form of  $Z_{i1\alpha}^L(X)$  in the goal expression in (8) can be presented as follows.

$$\text{let, } Z_{i1\alpha}^L(X) = \frac{N_{i1\alpha}^L(X)}{D_{i1\alpha}^L(X)} \tag{9}$$

where  $N_{i1\alpha}^L(X) = (G_{i1}^L X + \alpha_{i1}^L) - Z_{i1\alpha}^{wL}(H_{i1}^R X + \beta_{i1}^R)$  and

$D_{i1\alpha}^L(X) = (Z_{i1\alpha}^{bL} - Z_{i1\alpha}^{wL})(H_{i1}^R X + \beta_{i1}^R)$  are linear functions of  $x_l, l = 1, 2, \dots, p$ .

Then, the goal expressions can be presented as

$$\frac{N_{i1\alpha}^L(X)}{D_{i1\alpha}^L(X)} + d_{i1}^{L-} - d_{i1}^{L+} = 1, \quad i = 1, 2, \dots, n. \tag{10}$$

Following the notion of parametric approach [19] to fractional program, the goals in (10) the equivalent linear parametric form of the goals can be presented as:

$$N_{i1\alpha}^L(X) - \lambda_{i1}^L D_{i1\alpha}^L(X) + d_{i1}^{L-} - d_{i1}^{L+} = 1, \quad i = 1, 2, \dots, m. \tag{11}$$

where  $\lambda_{i1}^L (i=1, 2, \dots, m)$  is a real parameter.

In an analogous way, linear parametric form of the other goals in (8) can be presented as:

$$N_{i1\alpha}^R(X) - \lambda_{i1}^R D_{i1\alpha}^R(X) + d_{i1}^{R-} - d_{i1}^{R+} = 1, \quad i = 1, 2, \dots, m. \tag{12}$$

$$N_{j2\alpha}^{L}(X) - \lambda_{j2}^L D_{j2\alpha}^L(X) + d_{j2}^{L-} - d_{j2}^{L+} = 1, \quad j = 1, 2, \dots, n \tag{13}$$

$$N_{j2\alpha}^R(X) - \lambda_{j2}^R D_{j2\alpha}^R(X) + d_{j2}^{R-} - d_{j2}^{R+} = 1, \quad j = 1, 2, \dots, n \tag{14}$$

where  $\lambda_{i1}^R (i=1, 2, \dots, m)$  and  $\lambda_{j2}^R, \lambda_{j2}^L (j=1, 2, \dots, n)$  are real parameters.

Then, the problem in (8) can be converted to an executable linear *minsum* FGP model as:

Find  $X(x_1, x_2, x_3)$  so as to:

$$\text{Minimize } Z = \sum_{i=1}^m (w_{i1}^{L-} d_{i1}^{L-} + w_{i1}^{R-} d_{i1}^{R-}) + \sum_{j=1}^n (w_{j2}^{L-} d_{j2}^{L-} + w_{j2}^{R-} d_{j2}^{R-})$$

and satisfy the goal expressions in (11)-(14) subject to the set of constraints in (4). (15)

The PLGP scheme adopted in the solution search process is introduced in the following section.

#### 4.1.1 PLGP Scheme

The goal expressions in (11)-(14) can be expressed in a compact form as:

$$N_k(X) - \lambda_k D_k(X) + d_k - d_k = 1, \quad k = 1, 2, \dots, 2m + 2n. \tag{16}$$

In an MODM situation, since achievement of highest membership value of a goal is not always possible in a practical decision situation, a parameter  $\epsilon$ , a sufficiently small number, is defined parametrically in such a way that  $0 < \epsilon < 1$ , ( $\epsilon < \lambda_k$ ), which is used to measure the closeness of an achieved value of a goal to its aspired level in the solution search process. Then, the execution process can be described as follows.

$$\text{Let, } F(\lambda_k, X) = [N_k(X) - \lambda_k D_k(X)] \tag{17}$$

In the execution process,  $\lambda_k$  can be defined as:

$$\lambda_k = \frac{N_k(X)}{D_k(X)}, \quad k = 1, 2, \dots, 2m + 2n, \tag{18}$$

In which  $\lambda_k$  represents the membership value for measuring goal achievement of  $k$ -th goal.

At the first stage of the solution process, renaming  $\lambda_k$  by  $(\lambda_k)_1$  and taking  $(\lambda_k)_1 = 0, \forall k$ , the problem in (15) is solved.

Let  $X_1$  be the solution achieved at the first stage.

Then, the values of  $\lambda_k (k = 1, 2, \dots, 2m + 2n)$  are determined as:

$$(\lambda_k)_2 = \frac{N_k(X_1)}{D_k(X_1)}, \quad k = 1, 2, \dots, 2m + 2n. \tag{19}$$

Using the value of  $(\lambda_k)_2$ , the problem in (15) is again solved to determine the next solution  $X_2$ .

If  $|F(\lambda_k, X_2)| < \varepsilon$  is found, then execution process terminates, and  $X_2$  is identified as the optimal solution. Otherwise,  $\lambda_k$  is updated by renaming it  $(\lambda_k)_3$  and incorporating the solution  $X_2$  in (17). Then, evaluation of the problem (15) is made to generate the next solution  $X_3$ . The process is continued until the convergence criterion is satisfied.

The convergence criterion to reach optimality is:

$$X_p^* = \begin{cases} X^*, & \text{if } |F(\lambda_{kp}, X_p)| = 0, \forall k \\ X_M^*, & \text{if } |F(\lambda_{kp}, X_p)| < \varepsilon, \forall k \end{cases}$$

where  $X^*$  represents the ideal solution for achievement of all the membership goals to the highest membership value, and  $X_M^*$  is used to represent the most satisfactory solution as a best one in the decision environment.

**Note 1:** Regarding convergence of the proposed algorithm, it is to be noted that only linear programs are involved with execution of the FGP model. Here, since the solution space is bounded ( $S \neq \emptyset$ ) and only a set of linear programs are involved there; the solution process terminates after a finite number of iterations.

To illustrative the effectiveness of the proposed approach, a numerical example is solved.

## 5 Numerical Example

A fuzzily described MOFPP with two objectives and three constraints is considered as follows.

Find  $X(x_1, x_2, x_3)$  so as to:

$$\text{Maximize } Z_1(X) = \frac{(-0.75, 0, 2)x_1 + (1.5, 2, 2.5)x_2 + (2, 4, 5.5)x_3 + (2.5, 4, 6)}{(-0.5, 1, 2)x_1 + (1, 1.5, 2)x_2 + (1, 3, 3.5)x_3 + (2, 3, 4)},$$

$$\text{Minimize } Z_2 = \frac{(6, 7, 8.5)x_1 + (2, 3.5, 4.5)x_2 + (-1, 0, 2)x_3 + (4, 5, 6)}{(4, 5, 6)x_1 + (1, 2, 3)x_2 + (1, 1.5, 2)x_3 + (3, 4, 5)},$$

subject to

$$(2, 4, 5.5)x_1 + (1, 2, 3.5)x_2 + (2, 3, 4)x_3 \leq (6, 10, 15),$$

$$(1, 2, 5)x_1 + (-1, 1, 4)x_2 + (-2, 3, 7)x_3 \geq (1, 4, 9), \quad (0, 1.5, 3.5)x_3 \leq (2.5, 5, 8),$$

$$x_i \geq 0.$$

(20)

Using the expression (4),  $\alpha \in [0, 1]$ , the MOFPP can be described as:

$$\text{Maximize } Z_1^L(X) = \frac{(0.75\alpha - 0.75)x_1 + (0.5\alpha + 1.5)x_2 + (2\alpha + 2)x_3 + (1.5\alpha + 2.5)}{(-\alpha + 2)x_1 + (-0.5\alpha + 2)x_2 + (-0.5\alpha + 3.5)x_3 + (-\alpha + 4)}$$

$$\text{Maximize } Z_1^R(X) = \frac{(-2\alpha + 2)x_1 + (-0.5\alpha + 2.5)x_2 + (-1.5\alpha + 5.5)x_3 + (-2\alpha + 6)}{(1.5\alpha - 0.5)x_1 + (0.5\alpha + 1)x_2 + (2\alpha + 1)x_3 + (\alpha + 2)}$$

$$\begin{aligned} \text{Minimize } Z_2^L(X) &= \frac{(\alpha + 6)x_1 + (\alpha + 2)x_2 + (\alpha - 1)x_3 + (\alpha + 4)}{(-\alpha + 6)x_1 + (-\alpha + 3)x_2 + (-0.5\alpha + 2)x_3 + (-\alpha + 5)} \\ \text{Minimize } Z_2^R(X) &= \frac{(-1.5\alpha + 8.5)x_1 + (-\alpha + 4.5)x_2 + (-2\alpha + 2)x_3 + (-\alpha + 6)}{(\alpha + 4)x_1 + (\alpha + 1)x_2 + (0.5\alpha + 1)x_3 + (\alpha + 3)} \end{aligned}$$

subject to

$$\begin{aligned} (2\alpha + 2)x_1 + (\alpha + 1)x_2 + (\alpha + 2)x_3 &\leq (4\alpha + 6), \\ (-1.5\alpha + 5.5)x_1 + (-1.5\alpha + 3.5)x_2 + (-\alpha + 7)x_3 &\leq (-5\alpha + 15), \\ (\alpha + 1)x_1 + (2\alpha - 1)x_2 + (5\alpha - 2)x_3 &\geq (3\alpha + 1), \\ (-3\alpha + 5)x_1 + (-3\alpha + 4)x_2 + (-4\alpha + 7)x_3 &\geq (-5\alpha + 9), \\ 1.5\alpha x_3 &\leq 2.5\alpha + 2.5, \quad (-2\alpha + 3.5)x_3 \leq -3\alpha + 8, \\ x_l &\geq 0, \quad l = 1, 2, 3 \end{aligned} \tag{21}$$

For simplicity and without loss of generality, considering  $\alpha = 0.5$ , the above MOFPP can be stated as:

Find  $X(x_1, x_2, x_3)$  so as to:

$$\text{Maximize } Z_1^L(X) = \frac{0.25x_1 + 1.75x_2 + 3x_3 + 3.25}{1.5x_1 + 1.75x_2 + 3.25x_3 + 3.5}$$

$$\text{Maximize } Z_1^R(X) = \frac{x_1 + 2.25x_2 + 4.75x_3 + 5}{0.25x_1 + 1.25x_2 + 2x_3 + 2.5}$$

$$\text{Minimize } Z_2^L(X) = \frac{6.5x_1 + 2.5x_2 - 0.5x_3 + 4.5}{5.5x_1 + 2.5x_2 + 1.75x_3 + 4.5}$$

$$\text{Minimize } Z_2^R(X) = \frac{7.75x_1 + 4x_2 + x_3 + 5}{4.5x_1 + 1.5x_2 + 1.25x_3 + 3.5}$$

subject to

$$\begin{aligned} 3x_1 + 1.5x_2 + 2.5x_3 &\leq 8, \quad 4.75x_1 + 2.75x_2 + 6.5x_3 \leq 12.5, \quad 1.5x_1 + 0.5x_3 \geq 2.5, \\ 3.5x_1 - 2.5x_2 + 5x_3 &\geq 7.5, \quad 0.75x_3 \leq 3.75, \quad 2.5x_3 \leq 6.5, \quad x_l \geq 0, \quad l = 1, 2, 3. \end{aligned} \tag{22}$$

Now following the procedure, individual best and worst values of the objective functions are obtained as

$$(Z_1^{bL}, Z_1^{wL}) = (0.703, 0.260), (Z_1^{bR}, Z_1^{wR}) = (2.461, 2.239), (Z_2^{bL}, Z_2^{wL}) = (0.945, 1.139),$$

and  $(Z_2^{bR}, Z_2^{wR}) = (1.527, 1.689)$ , respectively.

Then, the fuzzy goals of the problem become:

$$Z_1^L \gtrsim 0.703, \quad Z_1^R \gtrsim 2.461 \quad \text{and} \quad Z_2^L \lesssim 0.945, \quad Z_2^R \lesssim 1.527. \tag{23}$$

The tolerance limits of the successive goals are (0.26, 2.239, 1.139, 1.689).

Then, the membership goals can be constructed as:

$$2.257(Z_1^L(X) - 0.26) + d_{11}^{L-} - d_{11}^{L+} = 1, \quad 4.504(Z_1^R(X) - 2.239) + d_{11}^{R-} - d_{11}^{R+} = 1,$$

$$5.155(1.139 - Z_2^L(X)) + d_{12}^{L-} - d_{12}^{L+} = 1, \quad 6.173(1.689 - Z_2^R(X)) + d_{12}^{R-} - d_{12}^{R+} = 1, \tag{24}$$

Now, the executable *minsum* FGP model of the problem appears as follows.

Find  $X(x_1, x_2, x_3)$  so as to:

$$\text{Minimize } Z = 2.257d_{11}^{L-} + 4.504d_{11}^{R-} + 5.155d_{12}^{L-} + 6.173d_{12}^{R-}$$

and satisfy

$$\begin{aligned} (-0.51x_1 + 2.92x_2 + 4.86x_3 + 5.28) - \lambda_1^L(1.5x_1 + 1.75x_2 + 3.25x_3 + 3.5) + d_{11}^{L-} - d_{11}^{L+} &= 1, \\ (3.18x_1 - 0.47x_2 + 1.23x_3 - 1.69) - \lambda_1^R(0.25x_1 + 1.25x_2 + 2x_3 + 2.5) + d_{11}^{R-} - d_{11}^{R+} &= 1, \\ (2.21x_1 + 6.79x_2 + 12.85x_3 + 3.22) - \lambda_2^L(5.5x_1 + 2.5x_2 + 1.75x_3 + 4.5) + d_{12}^{L-} - d_{12}^{L+} &= 1, \\ (-0.92x_1 - 1.05x_2 + 6.86x_3 + 5.63) - \lambda_2^R(4.5x_1 + 1.5x_2 + 1.25x_3 + 3.5) + d_{12}^{R-} - d_{12}^{R+} &= 1, \end{aligned} \tag{25}$$

subject to the system constraints defined in (21).

Taking  $\epsilon = 0.02$  (a small positive number) and using the proposed algorithm the problem is solved by using the *Software* Lingo (Ver. 12). The optimal solution is achieved at the fourth iteration.

The resultant decision is  $(x_1, x_2, x_3) = (1.522, 0, 0.435)$ .

The resultant objective values are  $[Z_1^L, Z_1^R] = [0.73, 2.29]$  and  $[Z_2^L, Z_2^R] = [0.91, 1.24]$ .

The achieved membership values for the given  $\alpha = 0.5$  are obtained as

$$(\mu_{Z_{1\alpha}^L}, \mu_{Z_{1\alpha}^R}, \mu_{Z_{2\alpha}^L}, \mu_{Z_{2\alpha}^R}) = (0.92, 0.98, 0.89, 0.704).$$

The result shows that the most satisfactory decision is achieved here in the decision making environment.

**Note 2:** To explore the more effectiveness of the approach, the problem (22) is directly solved with fractional criteria in the same decision environment. Here, the *minsum* FGP model can be presented as follows.

Find  $X(x_1, x_2, x_3)$  so as to:

$$\text{Minimize } Z = 2.49d_{11}^{L-} + 2.33d_{11}^{R-} + 4.41d_{12}^{L-} + 5.35d_{12}^{R-}$$

satisfy the goal expressions in (23) subject to the given constraints in (22).

The obtained solution is  $(x_1, x_2, x_3) = (1.48, 0, 0.839)$ .

The objective values are  $[Z_1^L, Z_1^R] = [0.71, 0.89]$  and  $[Z_2^L, Z_2^R] = [0.93, 1.22]$ .

The achieved membership values for the given  $\alpha = 0.5$  are obtained as:

$$(\mu_{Z_{1\alpha}^L}, \mu_{Z_{1\alpha}^R}, \mu_{Z_{2\alpha}^L}, \mu_{Z_{2\alpha}^R}) = (1.01, 0.89, 1.22, 0.63).$$

**Note 3:** If the Zimmermann's *max-min* FP approach [18] is used to solve the problem (22) in the same decision making environment, where objective function is 'maximize  $\lambda$ ' subject to for all the defined membership functions 'less than equal to'  $\lambda$  with  $0 \leq \lambda$

$\leq 1$ , then the solution of the problem in the same decision environment is found as:  $(x_1, x_2, x_3) = (0.84, 0, 0.77)$ .

The objective values are  $[Z_1^L, Z_1^R] = [0.34, 0.45]$  and  $[Z_2^L, Z_2^R] = [1.26, 1.43]$ .

The achieved membership values for the given  $\alpha = 0.5$  are obtained as

$$(\mu_{Z_{1\alpha}^L}, \mu_{Z_{1\alpha}^R}, \mu_{Z_{2\alpha}^L}, \mu_{Z_{2\alpha}^R}) = (0.43, 0.77, 0.99, 0.55).$$

The resultant decisions for the proposed approach and other two approaches are displayed schematically in Fig. 2.

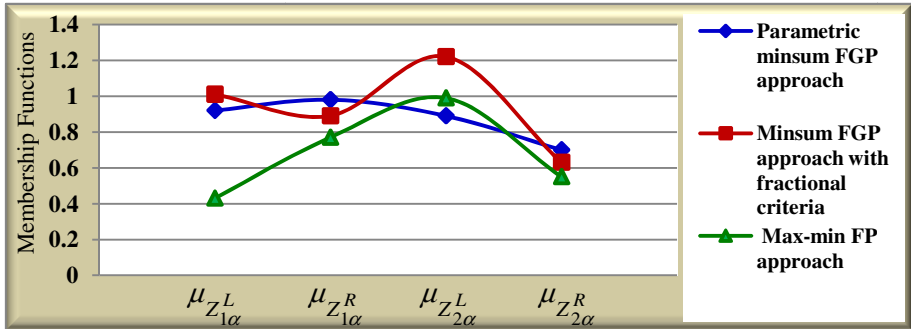


Fig. 2. Result Comparison of the proposed approach with the other approaches

A comparison shows that the decision under the proposed approach is more acceptable than the approaches studied previously in a fuzzy decision environment.

## 6 Conclusion

The main advantage of the paper is that changing the value of  $\alpha$ , the problem can be solved according to the degree of satisfaction of the DM. Again, the proposed approach is free from the computational complexities inherent to use of a traditional linearization method. Further, the use of proposed PLGP method offers the approximate solution as a best one on the basis of the accuracy desired by the DM in the decision making situation. The solution procedure presented here can be extended to solve bi-level programming [20] as well as multilevel programming [21] formulation in a highly conflicting MODM situation. The incorporation of *Type-2* fuzzy sets [22, 23] within the framework of proposed model, where fuzziness of fuzzy parameters further involves in a practical decision situation, may be considered as a prominent research problem in future study.

**Acknowledgement.** The authors are thankful to the anonymous reviewers and Prof. J. K. Mandal, the Special Session Chair of CSI-2013, for their valuable comments and suggestions which have led to improve the quality and clarity of presentation of the paper.

## References

1. Charnes, A., Cooper, W.W.: Programming with Linear Fractional Functions. *Naval Res. Logis. Quarterly* 9, 181–186 (1962)
2. Bajalinov, E.B.: *Linear-Fractional Programming Theory, Methods, Applications and Software*. Kluwer Academic Publishers, United Kingdom (2003)
3. Craven, B.D.: *Fractional Programming*. Heldermann Verlag, Berlin (1988)
4. Schaible, S.: Fractional Programming, Applications and Algorithms. *Euro. J. Oper. Res.* 7, 111–120 (1981)
5. Steuer, R.E.: *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley & Sons, New York (1986)
6. Ignizio, J.P.: *Goal Programming and Extensions*, D. C. Heath, Lexington, Massachusetts (1976)
7. Kornbluth, J.S.H., Steuer, R.E.: Goal Programming with Linear Fractional Criteria. *Euro. J. Oper. Res.* 8, 58–65 (1981a)
8. Kornbluth, J.S.H., Steuer, R.E.: Multiple Objective Linear Fractional Programming. *Manag. Sci.* 27, 1024–1039 (1981b)
9. Bellman, R.E., Zadeh, L.A.: Decision-Making in a Fuzzy Environment. *Manag. Sci.* 17, B141 – B164 (1970)
10. Ludhandjula, M.K.: Fuzzy Approaches for Multiple Objectives Linear Fractional Optimization. *Fuzzy Sets and Sys.* 13, 11 – 23 (1984)
11. Rommelfanger, H., Hanuschek, R., Wolf, J.: Linear Programming with Fuzzy Objectives. *Fuzzy Sets and Sys.* 29, 31–48 (1989)
12. Dutta, D., Rao, J.R., Tiwari, R.N.: Sensitivity Analysis in Fuzzy Linear Fractional Programming Problem. *Fuzzy Sets and Sys.* 48, 211–216 (1992)
13. Pal, B.B., Moitra, B.N., Maulik, U.: A Goal Programming Procedure for Fuzzy Multiobjective Linear Fractional Programming Problem. *Fuzzy Sets and Sys.* 139, 395–405 (2003)
14. Pal, B.B., Moitra, B.N., Sen, S.: A Linear Goal Programming Approach to Multiobjective Fractional Programming with Interval Parameter Sets. *Int. J. Math. Oper. Res.* 3, 697–714 (2011)
15. Dinkelbach, W.: On Nonlinear Fractional Programming. *Manag. Sci.* 13, 492–498 (1967)
16. Dubois, D., Prade, H.: *Fuzzy Sets and Systems: Theory and Applications*, New York, London (1980)
17. Yeh, C.T.: On improving trapezoidal and triangular approximations of fuzzy numbers. *Internat. J. Approx. Reason.* 48, 297–313 (2008)
18. Zimmermann, H.-J.: *Fuzzy Sets, Decision Making and Expert Systems*. Kluwer-Nijhoff Publishing, Dordrecht (1987)
19. Kumar, M., Pal, B.B.: Dinkelbach Approach for solving Interval-Valued Multiobjective Fractional Programming Problems using Goal Programming. *Int. J. Comp. Appl.* 57, 12–17 (2012)
20. Pal, B.B., Chakraborti, D.: Using Genetic Algorithm for solving Quadratic Bilevel Programming Problems via Fuzzy Goal Programming. *Int. J. Appl. Manag. Sci.* 5, 172–195 (2013)
21. Pal, B.B., Kumar, M., Sen, S.: Priority based Fuzzy Goal Programming Approach for Fractional Multilevel Programming Problems. *Int. Rev. Fuzzy Math.* 6, 1–14 (2011)
22. Castillo, O., Melin, P.: *Type-2 Fuzzy Logic Theory and Applications*. STUDEFUZZ, vol. 223. Springer, Heidelberg (2008)
23. Mendel, J.M., Bob, R.I.: Type-2 Fuzzy Sets Made Simple. *IEEE Tran. Fuzzy Sys.* 10, 117–127 (2002)

# An Efficient Ant Colony Based Routing Algorithm for Better Quality of Services in MANET

Abdur Rahaman Sardar<sup>1</sup>, Moutushi Singh<sup>2</sup>, Rashi Ranjan Sahoo<sup>4</sup>, Koushik Majumder<sup>3</sup>,  
Jamuna Kanta Sing<sup>4</sup>, and Subir Kumar Sarkar<sup>4</sup>

<sup>1</sup> Neotia Institute of Technology, Management and Science, West Bengal, India  
abdur.sardar@gmail.com

<sup>2</sup> Institute of Engineering & Management, Salt Lake, West Bengal, India

<sup>3</sup> Dept. of Computer Science & Engineering, West Bengal University of Technology  
koushik@ieee.org

<sup>4</sup> Jadavpur University, Kolkata, West Bengal, India  
su\_sarkar@hotmail.com

**Abstract.** The mobile ad-hoc networks (MANET) are infrastructure less and decentralized group of mobile nodes where the routers and mobile hosts are connected over wireless radio links. With the easy availability and widespread use of the highly mobile and portable devices such as laptop, sensor devices, PDAs etc, the mobile ad hoc networks are receiving increasing attention as there is a growing need for connectivity between these devices in real time, without the help of any fixed infrastructure. In MANET, the nodes communicate with each other using multi-hop communication system. The nodes and routers are free to move within the network and it is a difficult task to maintain the network topology as it changes randomly and unpredictably. Therefore, in mobile domain, establishing suitable and effective paths between the communicating end points while satisfying the time varying QOS requirements for wide variety of applications, has become a big challenge now. In this paper, we have presented a new on demand QOS routing algorithm for mobile ad hoc network with the concept of Ant Colony Optimization. This algorithm is based on swarm intelligence. It is inspired by the behaviour of the biological ants of finding optimal route to the food source in a collective but decentralized manner. With the introduction of new quality of service parameters like link stability time and available buffer size in the context of ant colony system, this algorithm provides effective routes for end to end communication.

**Keywords:** Ad-hoc network, QOS Routing, Ant Colony Optimization, Swarm Intelligence, Ant net.

## 1 Introduction

In mobile ad hoc network the nodes and routers are free to move randomly in any direction. Ad hoc is a Latin word and it means "for this purpose". MANETs are wireless networks that usually have a routable networking environment on top of a



Link Layer. Routing in MANET is considered as a challenging task as its mobility causes frequent changes to the routes. Moreover, due to the limited transmission range of the nodes, a packet needs to traverse multiple links to reach its destination. A MANET can be operated in a standalone fashion or it may be connected to the internet.

There are several applications of MANET. In military environment it is used for soldiers, tanks and planes. In case of civilian environment it is used in taxi cab network, meeting rooms, sports stadiums, boats and in small aircrafts. It is also used for personal area networking for cell phone, laptops, earphones and wrist watches. For emergency operations it is used for search and rescue, policing, firefighting, seismic activities and in medical applications. Nowadays with IEEE 802.11a and Bluetooth technology multimedia applications are also used in MANETs [1].

Routing is a very important thing in communication system. In case of MANET the nodes are not directly connected. The routing system is actually solving the problem of direct communication between source and destination. But due to increasing complexity in the modern network, routing algorithms are also facing different challenges. Traditional routing algorithms which are proposed for MANET such as AODV [2], DSR [3], and OLSR [4] route data without considering different QoS parameters. In these protocols mainly the number of hops is considered for communication. For mobile applications the new requirements are the consideration of different end to end QoS parameters with sufficient available resources [5].

In this paper we present a new approach for an on demand ad hoc routing algorithm based on ant-colony optimization. We have also accumulated some of the quality of service parameters for finding the probability of the best quality path for end to end communication in mobile ad hoc network.

The remainder of this paper is organized as follows. In section II we present the literature survey. In section III we present the proposed ant colony optimization meta heuristic based model. In section IV the proposed algorithm is discussed. Finally conclusion is given in section V.

## 2 Literature Survey

Some works related to ACO are found in the literature. A hybrid routing algorithm for MANETs based on ACO and zone routing framework of border casting is described in [9]. A new QoS routing protocol combined with the flow control mechanism has been done in [10]. This proposed routing solution is modeled by ant systems. This proposed routing protocol uses a new metric to find the route with higher transmission rate, less latency and better stability. P.Deepalakshmi. et.al [12] proposed a new on demand QoS routing algorithm based on ant colony meta heuristic. An algorithm of ant colony optimization for mobile ad hoc networks has been described in [13]. But the QoS issues such as end-to-end delay, available bandwidth, cost, loss probability, and error rate are not considered in this paper. A hybrid QoS routing algorithm has been proposed in [14]. In [14], the authors used ant's pheromone update process technique for improving QoS. But the authors described only bandwidth. Other QoS issues are not considered in this paper. In [16] the authors used the Network Time

Protocol (NTP) to predict the mobility of nodes and the amount of time the transmitter & receiver remains connected to each other. They suggest a method to find out Link Expiration Time (LET) and various ways to enhance unicast & multicast protocols by using mobility prediction. The authors of [17] suggested the relationship between the OLSR protocol and ACO for flooding. The authors of [18] stated a new congestion controlled routing LCRACO based on ant colony optimization. This algorithm uses more than one colony of ants to search for the optimal path. The next hop is chosen based on the highest concentration of pheromone. Connection is established on a multipath basis to balance the load in the network. The same authors also suggested another new algorithm MALBACO for multi agent load balanced ant colony optimization for MANETS in [19]. In [20] the author has proposed a new routing algorithm called AntHocNet. This network is based on a hybrid architecture where a reactive route setup process is used in the beginning of each new communication session in order to obtain an initial path for data forwarding, and a proactive route maintenance process is run throughout the duration of the session with the notion to keep information about the paths up-to-date and to explore new and possibly better paths, continuously adapting to the changing network environment. In [21] the author has suggested a new ant colony based routing algorithm (ARA) which is based on swarm intelligence. Ants actually communicate via stigmergy which means the indirect communication of the individuals through modifying their environment. S. Prasad, Y.P.Singh, and C.S.Rai [22] presented a novel proactive algorithm to routing called Probabilistic Ant Routing, in mobile ad hoc networks, which is inspired by Ant Colony Optimization framework and uses “ants” for route discovery, maintenance and improvement. The algorithm is based on a modification of the state transition rule of ACO routing algorithm that results in maintaining higher degree of exploration along with congestion awareness in the search space. This reduces end-to-end delay and also lowers the overhead at high node density. The comparative experimental results of the proposed algorithm with the state-of-the-art AODV reactive routing algorithm of the MANET are provided keeping mobility and density of nodes as the main consideration. The proposed algorithm is tested for different network sizes and node mobility. This proposed algorithm exhibits superior performance with respect to reactive AODV routing algorithm in terms of end-to-end delay. Hamideh Shokrani and Sam Jabbehdari [23] presented that the Mobile ad-hoc networks are infrastructure-less networks consisting of wireless, possibly mobile nodes organized in peer-to-peer fashion. The dynamic topology, limited bandwidth availability and energy constraints make the routing problem a challenging one for the wired networks. Recently a new family of algorithms inspired by Swarm Intelligence provides a novel approach to distributed optimization problems. Here ant colony finds the optimal route between the elements. The authors of [24] proposed extension to normal AODV to perform QoS routing based on bandwidth requirement and link stability constraints. Cross layer multi constraint QoS routing is proposed in [25]. In this paper Fan proposes multi constraint routing based on MAC delay metric, link reliability and throughput constraints. Fan stated that the multi constraint QoS routing problem is one NP- complete problem where a combination of additive and multiplicative metrics is considered. He also proposed a solution for reducing this NP-complete problem to one that can be solved in polynomial time. Advantage of this

approach lies in the simultaneous consideration of several important QoS metrics in path selection. But the QoS state for all paths must be discovered and kept fresh. The mechanism required for this is not discussed in his work.

### 3 Ant Colony Optimization Metaheuristic Based Model

Ant systems appeared towards the end of the nineties in Dorigo’s and Colormi’s work presented in [26]. Here we have tried to use that model for QoS based optimized route selection.

For the convenience of mathematical analysis the ad-hoc network is represented by a graph  $G(V,E)$  where  $V$  is the set of mobile nodes in the network and  $E$  is the set of links. A link is formed by two nodes  $v_i$  and  $v_j$ , where  $(v_i, v_j) \in V$ . In this paper we have considered some QoS parameters to find out the optimized route. Every link  $(v_i, v_j)$  is associated with the Bandwidth  $B(v_i, v_j)$ , Delay  $D(v_i, v_j)$ , Link stability  $T(v_i, v_j)$  and available buffer  $BUF(v_i, v_j)$ .

To find a route our algorithm uses the pheromone accumulation based on these QoS parameters. The same amount of pheromone will be deposited on each link  $(v_i, v_j)$  along the selected route  $R$ . The deposited amount of pheromone on each link  $(v_i, v_j)$  is expressed as follows:

$$\Delta\tau(v_i, v_j) = \frac{B(R)^{\beta_B} + T(R)^{\beta_T} + BUF(R)^{\beta_{BUF}}}{D(R)^{\beta_D}} \tag{5}$$

Where :

- $B(R)=\min(B(v_i, v_{i+1}), B(v_{i+1}, v_{i+2}), \dots, B(v_{k-1}, v_k))$  is the smallest link bandwidth in the route  $R$ . Here  $v_i$  is the source node and  $v_k$  is the destination node.
- $T(R)=\min(T(v_i, v_{i+1}), T(v_{i+1}, v_{i+2}), \dots, T(v_{k-1}, v_k))$ . The smallest link stability time that will be calculated by the methods mention in [27].
- $D(R)= D(v_i, v_{i+1})+D(v_{i+1}, v_{i+2}), \dots, +D(v_{k-1}, v_k)$ . The sum of delays on all links on a route  $R$ .
- $BUF(R)=\min(BUF(v_i, v_{i+1}), BUF(v_{i+1}, v_{i+2}), \dots, BUF(v_{k-1}, v_k))$ . The minimum available buffer space along the route  $R$ .
- $\beta_B, \beta_D, \beta_T, \beta_{BUF}$  indicates the relative significance of each QoS parameter on a route  $R$  These are link weight factor which varies from 0 to 1 .
- The local quality of a link represents the heuristic factor or the visibility of the ant. It is represented by the following equation:

$$\eta_{i,j} = \frac{B(v_i, v_j)^{\alpha_B} + T(v_i, v_j)^{\alpha_T} + BUF(v_i, v_j)^{\alpha_{BUF}}}{D(v_i, v_j)^{\alpha_D}} \tag{6}$$

where  $\alpha_B, \alpha_T, \alpha_{BUF}, \alpha_D$ , represent the relative importance of each QoS parameter during the link selection, which ranges from 0 to 1.

When an ANT searches for a route it chooses a node probabilistically as the next hop among the neighbors. A node  $v_i$  chooses a link  $(v_i, v_j)$  with a routing probability given as follows:

$$P_{i,j} = \frac{[\tau(v_i, v_j)]^\alpha [\eta_{i,j}]^\beta}{\sum_{k \in M} [\tau(v_i, v_k)]^\alpha [\eta_{i,k}]^\beta} \tag{7}$$

Where

- $\tau(v_i, v_j)$  is the amount of pheromone on link  $(v_i, v_j)$
- $\eta_{i,j}$  is the visibility of the link  $(v_i, v_j)$ . (an ant’s assumption about the solution at this path).
- $\alpha, \beta$  represent the relative significance of the pheromone and visibility.
- $M$  is the set of all possible neighbor nodes  $v_k$ , not yet visited by the ant.

The amount of pheromone on a link  $(v_i, v_j)$  is updated as follows:

$$\tau(v_i, v_j) = \rho * \tau(v_i, v_j) + \Delta\tau(v_i, v_j)$$

Where

- $\tau(v_i, v_j)$  is the amount of pheromone on the link  $(v_i, v_j)$ .
- $\rho$  is the evaporation factor ( $0 < \rho < 1$ ).
- $\Delta\tau(v_i, v_j)$  is the added amount of pheromone on link  $(v_i, v_j)$ .

To find out the best QoS route we consider a route with highest transmission rate, less delay, highest link stability time and maximum buffer available.

The route with maximum available bandwidth offer highest transmission rate, thus offer less transmission delay and also less congestion. The bandwidth of a route is calculated as the minimum bandwidth of all links within that route. The choice of a route with minimum delay offer faster transmission. The choice of highest stability avoids communication interruptions, packet losses and packet retransmission. The route with highest available buffer is chosen to avoid congestion and packet drops.

## 4 Proposed Algorithm

The proposed algorithm is an on demand QoS routing algorithm. Each node  $v_i$  maintains a Pheromone table “PheroTable” containing the available pheromone value on each link  $(v_i, v_j)$  which is initialized to a constant  $C$ . Each node also maintains a probability table “ProbTable” containing the transition probability to select a neighboring node  $v_j$ .

The algorithm is described as follows:

Step 1:

Each node sends a “Hello” message after a fixed interval of time with the TTL value set to 1. It also contains the instant of its creation which is used to calculate the link delay.

Step 2:

When a source node S want to set up a route to D, it send N RREQ\_ANT packets through its neighbors.

Step 3:

On each node visibility  $\eta$  is calculated by using equation (6).

Step 4:

While travelling, a RREQ\_ANT will collect minimum bandwidth, total delay, minimum link stability time and minimum available buffer on that path.

Step 5:

Calculate deposited pheromone value on the route according to equation (5).

Step 6:

When a RREQ\_ANT reaches the destination node it will send back a RREP\_ANT on the reverse path from the destination to source.

Step 7:

Update ProbTable according to equation (7)

Step 8:

On the return path update the PheroTable on each node as

$$\tau(v_i, v_j) = \rho \tau(v_i, v_j) + \Delta \tau(v_i, v_j)$$

Step 9:

Calculate total pheromone value on the path.

Step 10:

Choose a path which have the highest path preference probability.

Our proposed algorithm uses “Hello” packets to determine its neighbors. Here we have used two types of ANT packets. The RREQ\_ANT in the forward path for route discovery and RREP\_ANT is used to traverse in the reverse path from destination to source when the destination is found by the RREQ\_ANT. During route discovery RREQ\_ANT maintain a list of visited nodes. On the forward path it will also determine minimum rate, minimum link stability, minimum available buffer and total delay. The visibility of each link is also computed using equation (6) and that will be stored in the visibility list. When RREQ\_ANT reaches the destination it will immediately send a RREP\_ANT in the reverse path from destination to source by taking the node id from visited list. During this traversal it will deposit the equal amount of pheromone  $\Delta \tau(v_i, v_j)$  on each node in the reverse path. Update the “ProbTable” and “PheroTable” on each node. Finally choose a path with highest path preference probability.

Route maintenance can also be done using this algorithm. As multiple ant agents propagating through the network, multiple paths are found from the source to the destination during the route discovery phase. Therefore, in case of a link failure in the current path, an alternate path having the next highest path preference probability will be selected. If no such alternate path exists, then a new route discover will be initiated.

## 5 Conclusions

In this work, we have proposed a new on demand QOS routing algorithm for mobile ad hoc networks using the idea of Ant Colony Optimization in order to find routing

paths with better quality of service support. This algorithm is based on swarm intelligence and it follows the behaviour of the biological ants to find the optimal route to the food source in a collective but decentralized manner. With the introduction of new quality of service parameters like link stability time and available buffer size in the context of ant colony system, we are expecting that this algorithm shall provide effective routes for better end to end communication. In our future work we simulate this algorithm for comparison with other related algorithms.

## References

1. Goyal, P., Parmar, V., Rishi, R.: MANET: Vulnerabilities, Challenges, Attacks, Application. *IJCEM International Journal of Computational Engineering & Management* 11 (2011) ISSN (Online): 2230-7893
2. Perkins, C.E., Belding-Royer, E.M., Das, S.: Ad Hoc On demand Distance Vector (AODV) Routing. IETF RFC 356
3. Johnson, D., et al.: Dynamic Source Routing in Ad Hoc Wireless Networks. IETF Internet Draft (March 2003)
4. Clausen, T., Jacquet, P.: Optimized Link State Routing Protocol. IETF Internet Draft (July 2003)
5. Chakrabarti, S., Mishra, A.: QoS issues in Ad Hoc Wireless Network. *IEEE Communications Magazine* 39(2), 142–148 (2001)
6. RFC 793, Transmission Control Protocol (September 1981)
7. Chiu, D., Jain, R.: Analysis of the Increase/Decrease Algorithms for Congestion Avoidance in Computer Networks. *Journal of Computer Networks and ISDN* 17(1), 1–14 (1989)
8. Belkadi, M., Lalam, M., M'Zoughi, A., Aoudjit, R., Daoui, M., Ait Ali, K.: Am'elioration des Performances de TCP dans les MANETs par la Technique Xon/Xoff. In: 9<sup>eme</sup> Conference Magre'bine sur les Technologies de l'information, MCSEAI 2006 Proceedings, pp. 86–90 (2006)
9. Wang, J., Osagie, E., Thulasiraman, P., Thulasiram, R.K.: HOPNET: A hybrid ant colony optimization routing algorithm for mobile ad hoc network, Department of Computer Science. University of Manitoba, Winnipeg
10. Belkadi, M., Lalam, M., M'zoughi, A., Tamani, N., Daoui, M., Aoudjit, R.: Intelligent Routing and Flow Control in MANETS. *Journal of Computing and Information Technology - CIT*, 233–243 (March 18, 2010)
11. Banik, S., Roy, B., Saha, B., Chaki, N.: Design of QoS Routing Framework based on OLSR Protocol. In: ARTCOM 2010, pp. 171–173. IEEE Explorer, Kochin (2010)
12. Deepalakshmi, P., Radhakrishnan, S.: Ant Colony Based QoS Routing Algorithm For Mobile Ad Hoc Networks. *International Journal of Recent Trends in Engineering* 1(1) (May 2009)
13. Jawahar, A.C.S.: Ant Colony Optimization for Mobile Ad-hoc Networks. Department of Electrical Engineering. Rutgers
14. Singh, R., Singh, D.K., Kumar, L.: Ants Pheromone for Quality of Service Provisioning In Mobile Adhoc Networks. *International Journal of Electronic Engineering Research* 2(1), 101–109 (2010) ISSN 0975 – 6450
15. Kamali, S., Opatrny, J.: A Position Based Ant Colony Routing Algorithm for Mobile Ad-hoc Networks. *Journal of Networks* 3(4) (April 2008)
16. Su, W., Lee, S.-J., Gerla, M.: Mobility Prediction in Wireless Networks, Department of Computer Science. University of California, Los Angeles

17. Banik, S., Roy, B., Dey, P., Chaki, N., Sanyal, S.: QoS Routing using OLSR with Optimization for Flooding. *International Journal of Information and Communication Technology Research* 1(4), 164–168 (2011) ISSN-2223-4985
18. Chaki, R., Sinha, D.: LCRACO – A New Load & Congestion Controlled Routing Based on Ant Colony Optimization. In: *Proceedings of International Workshop on Internet and Distributed Computing Systems (IDCS 2008)*, Khulna, Bangladesh, December 24 (2008)
19. Chaki, R., Sinha, D.: MALBACO – A New Multi Agent Load Balanced Ant Colony Optimization for MANETS. In: *World Congress on Nature & Biologically Inspired Computing (NaBIC 2009)* (2009)
20. Ducatelle, F.: *Adaptive Routing in Ad Hoc Wireless Multi-hop Networks*. Faculty of Informatics. Università della Svizzera Italiana
21. Günes, M., Sorges, U., Bouazizi, I.: ARA – The Ant-Colony Based Routing Algorithm for MANETS, Department of Computer Science, Informatik 4. Aachen University of Technology, Aachen, Germany
22. Prasad, S., Singh, Y.P., Rai, C.S.: Swarm Based Intelligent routing for MANET. *International Journal of Recent trends in Engineering* 1(1) (May 2009)
23. Shokrani, H., Jabbehdari, S.: A Survey of Ant Based routing Algorithm for Mobile Ad-Hoc Networks. In: *International Conference on Signal Processing Systems* (2009)
24. Thenmozhi, D.S., Lakshmi pathi, R.: Highly Ensured QoS Routing in Mobile Ad Hoc Networks Based on Multiple Constraints. *International Journal of Computer Applications* 8(4) (October 2010) ISSN 0975-8887
25. Fan, Z.: QoS routing using lower layer information in adhoc networks. In: *Proc. Personal, Indoor and Mobile Radio Communications Conf.*, pp. 135–139 (September 2004)
26. Dorigo, M., Maniezzo, V., Colomi, A.: The Ant System: Optimization by a colony of cooperating agents. *IEEE Transactions on systems, Man, and Cybernetics–Part B* 26(1), 1–13 (1996)
27. Belkadi, M., et al.: Intelligent Routing and Flow Control in MANETS. *Proc. Journal of Computing and Information Technology, CIT* 18(3), 233–243 (2010)

# An Enhanced MapReduce Framework for Solving Protein Folding Problem Using a Parallel Genetic Algorithm

A.G. Hari Narayanan, U. Krishnakumar, and M.V. Judy

Department of Computer Science and IT, Amrita School of Arts and Sciences,  
Amrita Vishwa Vidyapeetham, Kochi  
{hariag2002, asaskochi, judy.nair}@gmail.com

**Abstract.** Parallel Genetic algorithms have proved to be a successful method for solving the protein folding problem. In this paper we propose a simple genetic algorithm with optimum population size, mutation rate and selection strategy which is parallelized with MapReduce architecture for finding the optimal conformation of a protein using the two dimensional square HP model. We have used an enhanced framework for map Reduce which increased the performance of the private clouds in distributed environment. The proposed Genetic Algorithm was tested several bench mark of synthetic sequences. The result shows that GA converges to the optimum state faster than the traditional

**Keywords:** Protein Folding Problem, Hadoop, MapReduce, Parallel Genetic Algorithm.

## 1 Introduction

Proteins fold rapidly to their functional state (native state). The native state is the global energy minimum or a low-lying Meta stable conformer. The primary structure of a protein is the amino acid sequence of its polypeptide chain, while the secondary structure is the local arrangement of a polypeptide's backbone atoms without regard to the conformations of its side chains. Under certain physiological conditions, the primary structure of a protein spontaneously folds into a precise three-dimensional form called its tertiary structure or native state that determines its functional properties. Finding energetically low lying conformations given a sequence of amino acids is termed as "The Protein Folding Problem"(PFP)[1]. Currently, the primary structures of approximately 91939 proteins are known. Only a small percentage of these have known native states. Efforts aimed at solving the Protein Folding Problem have involved the optimization of a potential energy function that approximates the thermodynamic state of a protein macromolecule. Since an algorithm using such a potential function does not give insight into how a protein folds, these approaches are instead known as Protein Structure Prediction. In this study protein conformations and intra molecular interactions are modeled using the simplified HP Bead Model. Parallel Genetic algorithms have proved to be a successful methodology for searching



the hyper-surface because the population based characteristic of Genetic Algorithms (GA) allows the fitness function of each individual to be computed in parallel.

The cost of parallel execution architectures and infrastructure are much reduced with the emerging use of cloud computing and general purpose computing. Cloud Computing solutions offer a parallel distributed computational environment together with an on demand resource handling and allocation to easily scale up. HadoopMapReduce is a framework for developing applications that rapidly process vast amounts of data in parallel on large clusters of computing nodes. This choice was motivated by the fact that it is becoming the de-facto standard MapReduce implementation and it has been used also in industry. Moreover, it is well supported to work not only on clusters, but also on the cloud [17] and on graphic cards [18], thus being an ideal candidate for high scalable parallelization of GAs.

In this paper we demonstrate the how a normal genetic algorithm can be converted into map and reduce primitives. We implement the MapReduce program and demonstrate its scalability to protein folding problem. We use the Non-conventional Map Reduce which removes the barriers of the traditional Map Reduce.

## 2 Genetic Algorithm for PFP

### 2.1 The HP Model

The hydrophobic-hydrophilic model (HP model) by Dill [3] is a simple abstraction that captures the essence of the important concepts of Protein Structure Prediction. In the HP model, amino acids are divided into two categories: hydrophobic (H) and hydrophilic (P). The primary sequence of a protein is therefore  $S \sum \{H, P\}^+$ . Using this simplification, optimization models can be developed that seek to maximize interactions between adjacent pairs of hydrophobic amino acids (or hydrophobes). Adjacency is considered only in the cardinal directions of a lattice upon which the sequence is embedded. In an HP lattice, vertices represent amino acids and edges represent connecting bonds. Black squares at the vertices indicate hydrophobes, while white squares indicate hydrophilic amino acids. A lattice can be two or three dimensional, and either square, cubic or triangular. The hydrophobic-hydrophobic (HH) contacts are the basis for the evaluation function. Every pair of hydrophobes that are adjacent on the lattice and not consecutive in the primary sequence is awarded a value  $\epsilon$  (usually  $-1$ ). The sum of all such values gives the energy of the conformation. The amino acid sequence is "folded" on a two-Dimensional square lattice on which at each point, the chain can turn 90° left or right, or continue ahead. Figure 1 shows a 20-length sequence embedded on a square lattice, HH contacts indicated by dotted arrows.

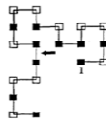


Fig. 1. HP sequence of length 20 on a square lattice with energy

### 2.2 Traditional Genetic Algorithm Model for PFP using HP Model

The genetic algorithm implementation is described in Unger and Moulton [8]. A new strategy MKBR was introduced in [13], as an intermediate selection strategy to improve the efficiency of the algorithm. The solutions are encoded as conformations themselves, which are treated directly in the spirit of genetic operators. The process starts with  $V$  extended structures. In each generation each structure is subject to a number of mutation steps with rate ranging from 0.01 to 0.20. Each mutation is the same as a single Monte Carlo (MC) step [8] and is subject to similar acceptance criteria as in a MC process. At the end of this MC stage [10], the crossover operation is performed. The chance  $p(S_i)$  of a structure being selected for crossover is proportional to its energy value  $E_i$ , That is

$$p(S_i) = \frac{E_i}{\sum_{j=1}^N E_j} \tag{1}$$

Thus, the lower energy conformations have a higher chance of being selected. For a pair of selected structures a random point is chosen along the sequence and the X-terminal portion of the first structure is connected to the C-terminal portion of the second structure (see Fig. 2). As there are three ways to join the parts together (connecting the chains with angles of 0°, 90° or 270°), these possibilities are tested in a random order to find one that is valid (That is, where no residue from one structure occupies a lattice point used by a residue from the other). If none of the three ways led to a self-avoiding structure, then another pair of structures is selected. Once a valid structure  $S_k$  is created, its energy  $E_k$ , is evaluated and compared to the averaged energy  $E_{ij} = (E_i + E_j)/2$  of its "parents" [9,10]. The structure is accepted if  $E_k \leq E_{ij}$ , or if the energy will be increased based on the decision:

$$R_{nd} < \exp\left[\frac{(E_{ij} - E_k)}{C_k}\right] \tag{2}$$

This crossover operation is repeated until  $N - 1$  newly accepted hybrid structures have been constructed to constitute the population of the next generation. We allow a higher acceptance rate for bad moves that increase the energy for mutation steps than for crossovers. This strategy maintains the diversity of the population and prevents premature convergence to a few low energy conformations

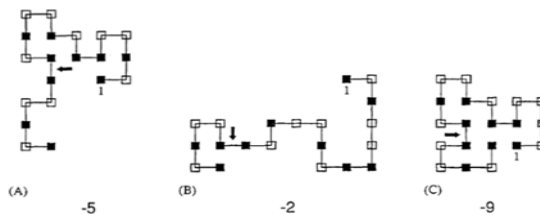


Fig. 2. HP The crossover operation

In the crossover stage pairsof structures are randomly (based on their energies) cut and pasted. In this example the cut point was randomly chosen to be after residue 14. Joining the first 14 residues of (A) with the last 6 residues of (B) and applying a randomly chosen  $270^\circ$  rotation at the joint achieves the compact structure in (C). In this case, the energy value of the hybrid (C) is -9, lower than the energies -5 and -2 of its "parents". The hybrid is always accepted if its energy is lower than the averaged energies of its parents or non-deterministically accepted according to its energy increase. In the crossover stage pairs of structures are randomly (based on their energies) cut and pasted. In this example the cut point was randomly chosen to be after residue 14. Joining the first 14 residues of (A) with the last 6 residues of (B) and applying a randomly chosen  $270^\circ$  rotation at the joint achieves the compact structure in (C). In this case, the energy value of the hybrid (C) is -9, lower than the energies -5 and -2 of its "parents". The hybrid is always accepted if its energy is lower than the averaged energies of its parents or non-deterministically accepted according to its energy increase.

### 3 Implementation Using Parallel Genetic Algorithm with Non-conventional MapReduce

The strategies proposed in the literature to parallelize Genetic Algorithms are discussed from [11]. Different methodologies exist in parallelizing genetic algorithms. The fitness evaluation level (i.e., global parallelization model), population level (i.e., coarse-grained parallelization or island model) and the individual level (i.e., fine-grained parallelization or grid model) are the most popular ones. In the global parallelization model, a node acting as a master, manages the population (i.e., applying genetic and selection operators) and distributes the individuals among slave nodes which compute only the fitness values of the individuals. The main advantage of using such a model is that it does not require any change to the design of traditional GA since the individual fitness evaluation is independent from the rest of the population. In the island model the population is subdivided in several subpopulations of relatively large size which are located in several islands (i.e., nodes). Thus, a Genetic Algorithm is executed on each subpopulation and such subpopulations exchange information by allowing some individuals to migrate from one island to another according to a given temporal criteria. The main expected advantages of this model are: (i) different subpopulations could explore different portions of the search-space; (ii) migrating individuals injects diversity into the converging population. Finally, in the grid model each individual is placed on a grid (i.e., each individual is assigned to a node) and all GA operations are performed in parallel evaluating simultaneously the fitness and applying locally selection and genetic operations to a small neighboring. A drawback of this approach is the overhead due to the frequent communications between grid nodes.

#### 3.1 Enhanced Mapreduce Framework

MapReduce is an elegant and flexible paradigm which enables to develop large-scale distributed applications [18]. It is expressed in terms of two distinct functions, namely

Map and Reduce, which are combined together in a divide-and-conquer way where the Map function is responsible to handle the parallelization while the Reduce collects and merges the results. In particular, a master node splits the initial input in several pieces; each one identified by a unique key, and distributes them via the Map function to several slave nodes (i.e., Mappers) which work in parallel and independently from each other performing the same task on a different piece of input. As soon as each Mapper finishes its own job the output is identified and collected via the Reducer function. In particular, each Mapper produces a set of intermediate key/value pairs which are exploited by one or more Reducers to group together all the intermediate values associated to the same key and to compute the list of output results. It is worth mentioning that the different intermediate keys emitted by the Mapper functions affect the way the model distributes the computation of each Reducer on different machines. Thus, the program automatically invokes and allocates a number of distinct Reducers that correspond to the number of distinct intermediate keys. Several different implementations of MapReduce have been proposed. The most famous one is the AppEngineMapReduce [18], built on the top of the distributed Google File System, and HadoopMapReduce [14]. We have implemented the algorithm using HadoopMapReduce. HadoopMapReduce is an open-source project of the Apache Software Foundation aiming at supporting developers in realizing applications that rapidly process vast amounts of data in parallel on large clusters of computing nodes.

In the Non-conventional MapReduce[16] we devise a technique by which we bypass the sorting mechanism and modify the invocation of reduce function so that it can be called with a single record. Reduce function works at one record at a time since each value itself is a protein conformation. Reducers no longer needs to wait for remotely read from Mappers and then to be grouped. Due to this performance gets improved as now Reducers need not to wait till the Mappers complete their whole work and shuffling gets completed. We don't store the intermediate results as a whole for each and every key. A separate thread is maintained for each and every Mapper. A single Buffer retrieves all records. A separate thread executes the Reduce function and is passed with one record from buffer at a time in a first in first out manner.

### **3.2 The Proposed Parallel Genetic Algorithm Based on Mapreduce**

We use a hybrid model by combining the first and second parallelization strategies using the enhanced Map Reduce. The distributed model designed [10] based on the polytypic concept of a species being represented by several types that are capable of mating and producing viable offspring is used. Breeding and evaluation are typically carried out in isolation on each island. To be consistent with the biological motivations, it was noted that migration should occur after a period of stasis [10]. However, due to the difficulty of defining stasis or equilibrium, migration occurred after G generations.

In the sequential version everything is done in a single processor, while in the parallel version, the processing load is divided into several processors mappers (slaves), under the coordination of a master processor. Encapsulate each iteration of the GA as a separate MapReduce job and parallelize the chromosome fitness

evaluation task to several *Mappers*, while *Reducers* are responsible to collect the results pertaining to respective islands based on the key value and to perform the genetic operations (i.e., parents selection, crossover and mutation, survival selection and migration process) needed to produce a new generation following a global parallelization model. Migration is done after  $g$  generations during which the best individuals in each island is sent to each neighbour, replacing the worst individuals.

Figure 3 shows the proposed architecture based on HadoopMapReduce and composed by the following main components: a *Parallel Genetic Algorithm*, a *Master*, a number of *Mappers* and *Reducers*(considered as an island), together with two other units, namely *InputStream* and *OutputStream*, which are responsible to split the data for the *Mappers* and to store the *Reducer* output into the Hadoop Distributed File System (HDFS) respectively.

These components communicate with each other exploiting the HDFS distributed file system provided by Hadoop, while the communications within the Hadoop framework (i.e., those between the master and slave nodes) are carried out via socket using SSH (Secure SHell).The *Parallel Genetic Algorithm* module takes a sequence of amino acids as input. Once the GA is terminated it returns a predicted structure as output.

**Split Phase:**In this phase the *InputFormat* module gets the current population (i.e., the sequence of amino acids composing the current population) from the HDFS and processes it in order to split it in crunch of data (i.e., input split) to be distributed among the *Mapper* modules. The number of input splits is dynamically computed on the basis of the number of available *Mappers*.

The *Master* module is responsible to coordinate and supervise the assignment of resources and the computations taking care also of the load balancing aspects. In more details, once the *InputFormat* begins to emit the  $\langle key, value \rangle$  pairs - exploiting the *RecordReader* component of Hadoop - the underlying Hadoop framework is automatically notified and the *Master* component is invoked to assign the input split produced by the *InputFormat* to the available *Mappers*

**Map Phase:** In this phase each *Mapper* carries out its task on the received input split in a parallel and independent way. In particular, each *Mapper* is responsible to perform the genetic operations, Once such evaluation is completed, each *Mapper* generates a new pair  $\langle key, value \rangle$ , where value is a pair  $\langle chromosome, fitness value \rangle$ . The key is generated by the *Master* module based on the number of islands and is assigned to *Reducers*.

**Reduce phase:** As soon as a *Mapper* evaluates a chromosome the corresponding data (i.e., key (island number), chromosome, and fitness value) is sent to the *Reducer*. Once the entire population corresponding to the island has been available to the *Reducer* it can perform the survival selection and apply on the new generation the crossover and mutation operators to produce a new offspring to be evaluated in the next *MapReduceJob*. Note that to obtain the entire population the *Reducer* should wait until all *Mappers* have replied. But it can start performing genetic operations as and when it receives a pair of chromosomes from the single buffer. However selection

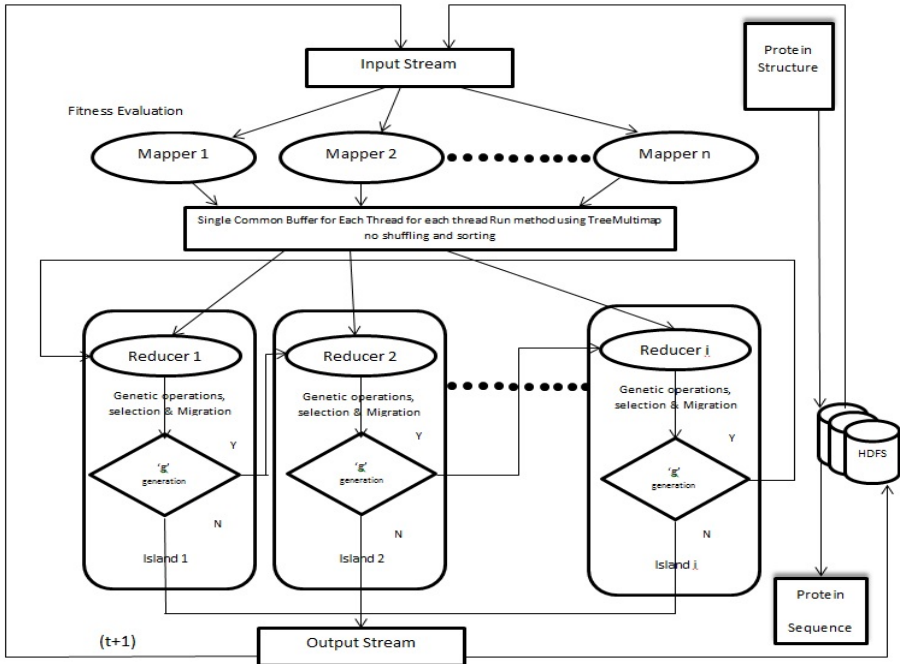


Fig. 3. Architecture of the proposed GA with enhanced MapReduce framework

Table 1. Pseudocode for Parallel GA using Table 2. Pseudocode for MapReduce Mapper()

```

map (key, value)
//key: subpopulation  $s_i$  or Island Number,
//value: (chromosome, fitness value)
for each chromosome in InputStream
EmitIntermediate (key, value)
    
```

```

Create mapReduce job for the entire population
MapReduce Mapper(); do
Concurrently for each  $i=1$  to  $N$  subpopulations
MapReduce Reducer();
While(!stopping criteria()) Endwhile
    
```

Table 3. Pseudocode for MapReduce Reducer()

```

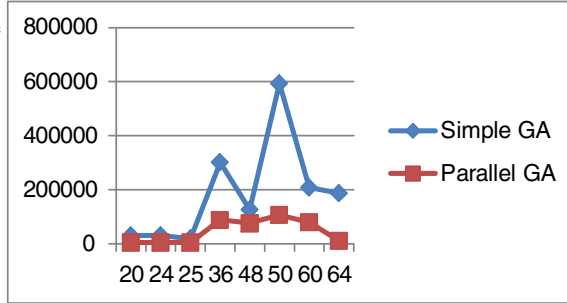
reduce (key, value) // key: subpopulation  $s_i$  or Island Number //value: (chromosome,
fitness value)
result =null list
for each  $v$  in value
select 2 offsprings for recombination;
apply crossover operator to them;apply mutation operator to them;add best to result
Insert(key,value) in TreeMultiMap {If  $g$  generations are reached then perform migration
with the
neighboring subpopulation
    
```



(60)PPHHHRHHHHHHHHHHPPPHHHHHHHHHHHHRPHPPPHHHHHHHHHHHHHPPPP  
 HHHHHHRPHRHP,  
 (64)HHHHHHHHHHHHHRHRPHRPHRPPHHRPPHHRPPHHRPPHHRPPHHRPPHHRPPH  
 HRHHHHHHHHHHHHH

**Table 4.** Comparison of the number energy evaluations

Length	Optimal Energy	Simple GA	Parallel GA
20	-9	30492	5040
24	-9	30491	5045
25	-8	20400	5047
36	-14	301339	88319
48	-22	126547	76541
50	-21	592887	107451
60	-34	208781	79980
64	-42	187393	11234



**Fig. 5.** The x-axis shows the sequence number while the y-axis shows the number of energy evaluations

The chart shows the variation in performance of the compared methods (Figure 5). We use higher mutation rates in combination with crossover. The x-axis shows the length of sequence while the y-axis shows the number of energy evaluations. The population size was 100. With PGA we were able to speed up the convergence of the GA.

### 5 Conclusion and Further Discussions

Parallel genetic Algorithm outperforms the traditional GA significantly on protein folding problems, especially as the problem size increases in terms of time and optimality. Results also showed that a random migration selection was more effective and that an aggressive selection strategy caused the population to get stuck in local optima. Memory management is the main problem which could arise for large sets of data as we are storing partial results obtained after the Map Stage in memory only. This could result in overflow. We suggested a solution to this problem by moving the contents which is least recently used into files. A Hash table could be used to keep the track of files which have been moved on to the file and for faster access. We can try to implement Non-Conventional MapReduce model on GPUs that is our compute intensive tasks Map on GPUs and Reduce on CPUs. Various other hybrid models of parallelization can be tested on benchmark problems like protein folding.

### References

1. Johnson, C.M., Katikireddy, A.: A Genetic Algorithm with Backtracking for Protein structure Prediction. In: GECCO 2006, Seattle, Washington, USA, July 8-12 (2006)
2. De Jong, K.A.: Analysis of the Behavior of a Class of Genetic Adaptive Systems. Ph.D. Dissertation, The University of Michigan, Ann Arbor, MI (1975)



3. Dill, K.A.: Theory for the folding and stability of globular proteins. *Biochemistry* 24(6), 1501–1509 (1985)
4. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading (1989)
5. Grefenstette, J., Gopal, R., Rosmaita, B., van Gucht, D.: Genetic Algorithms for the Traveling Salesman Problem. In: *Proceedings of the 1st ICGA*, pp. 160–168 (1985)
6. Haupt, R.L., Haupt, S.E.: *Practical Genetic Algorithms*. John Wiley & Sons, New York (1998)
7. Metropolis, X., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087–1092 (1953)
8. Unger, R., Moulton, J.: A genetic algorithm for three dimensional protein folding simulations. In: *Proceedings of the Fifth International Conference on Genetic Algorithms (ICGA 1993)*, Urbana-Champaign, IL, July 17-21, pp. 581–588. Morgan Kaufmann, San Francisco (1993)
9. Whitley, D.: The GENITOR Algorithm and Selection Pressure: Why Rank-Based Allocation of Reproductive Trials is Best. In: *Proceedings of the 3rd International Conference on Genetic Algorithms ZCGA, 1989*, pp. 116–121 (1989)
10. Martin, W.N., Lienig, J., Cohoon, J.P.: Population Structures Island (migration) models: evolutionary algorithms based on punctuated equilibria. In: *Handbook of Evolutionary Computation (May 97 Release)*
11. Di Geronimo, L., Ferrucci, F., Murolo, A., Sarro, F.A.: Parallel Genetic Algorithm Based on Hadoop MapReduce for the Automatic Generation of JUnit Test. In: *2012 IEEE Fifth International Conference on Proceedings of Software Testing, Verification and Validation (ICST) Issue Date: April 17-21, 2012*
12. Wiese, K., Goodwin, S.D.: Parallel Genetic Algorithms for Constrained Ordering Problems. In: *Proceedings of the 11th International Florida Artificial Intelligence Research Symposium, FLAIRS 1998*, pp. 101–105 (1998)
13. Judy, M.V., Ravichandran, K.S.: A solution to protein folding problem using a genetic algorithm with modified keep best reproduction strategy. In: *IEEE Congress on Evolutionary Computation (2007)*, Library of Congress: 2007928155, pp. 4776–4780. IEEE Xplore (2007) ISBN: 1-4244-1340-0
14. Private Cloud setup, <http://www.akashsharma.me/private-cloud-setup-using-eucalyptus-and-xen/>
15. Apache Hadoop Map Reduce, <http://hadoop.apache.org>
16. Rajan, A., Judy, M.V.: An Enhanced Map Reduce Framework for Improving the Performance of Massively Scalable Private Clouds. In: *International Journal of Computer Applications (IJCA)*, Proceedings on Amrita International Conference of Women in Computing, AICWIC 2013, January 24-26. Published by Foundation of Computer Science, New York (2013)
17. Running Hadoop on multi node environment, <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-multi-node-cluster>
18. Verma, A., Zea, N., Cho, B., Gupta, I., Campbell, R.H.: Breaking the Map Reduce stage Barriers. University of Illinois at Urbana-Champaign

# A Wavelet Transform Based Image Authentication Approach Using Genetic Algorithm (AWTIAGA)

Amrita Khamrui<sup>1</sup> and J.K. Mandal<sup>2</sup>

<sup>1</sup> Dept. of Computer Science and Engineering,  
Future Institute of Engineering & Management,  
Sonarpur Station Road, Sonarpur, Kolkata-150, West Bengal, India

<sup>2</sup> Dept. of Computer Science and Engineering,  
University of Kalyani,  
Kalyani, Nadia-741235, West Bengal, India  
{khamruiamrita, jkm.cse}@gmail.com

**Abstract.** Steganography is the technique of hiding confidential information within a media. In this paper a Genetic Algorithm based image authentication technique in transformed domain using haar wavelet transform is proposed. Wavelet transform is applied on sub mask of the source image to obtain transformed coefficients. Bits of the hidden image are embedded on each coefficients of sub mask with BPP of two. The hidden bit positions are chosen in such a way that there is no loss of precession during reverse wavelet transform. Resultant stego image is obtained through inverse transform which is input for the Genetic Algorithm. Genetic algorithm is applied on it to enhance a layer of security. New Generation followed by Crossover is applied on initial population. Reverse process is followed during decoding. The proposed algorithm obtains better image fidelity and high PSNR as compared to existing approach Huang et al [1].

**Keywords:** Steganography, Frequency Domain, Peak Signal to Noise Ratio (PSNR), Least Significant Bit (LSB), Image Fidelity (IF), Mean Square Error (MSE), BPP (Bit Per Pixel).

## 1 Introduction

Steganography offers an interesting alternative to image integrity and authenticity. Image Authentication provides a way of security/ protection for digital image data. The purpose of steganography is to hide secret message/ data within cover image without creating any visible changes to keep the intruder unaware of communication. Generally, a steganographic message may be picture, video, sound file and radio communication also [1], [7], [6]. A message is hidden in invisible way to ensure security. Information hiding [5] in the image is a nice approach for image authentication. Data security and image authentication is very important for digitized document especially legal document [2], [3]. Secret information is hidden by embedding of information within a host data set as message, image or video [4], [8].

A well known example of watermarking is that of a prisoner communicating with the outside world under the supervision of a warden. The data hiding is one alternative of the construction of a hypermedia document or image, which is very less convenient to manipulate. The hidden image is embedded by keeping the fidelity of the cover image intact. The most popular steganographic technique is least-significant bit (LSB) substitution method developed through [9] masking, filtering and transformations on the source image [6]. Present proposal is an algorithm which would facilitate secure message transmission through block based data hiding. Most of the works [10] used minimum bits of the hidden image for embedding in spatial domain, but the proposed algorithm embeds in transformed domain with a bare minimum distortion of visual property.

Rest of the paper is organized as follows. Section 2 describes the proposed technique. Results and comparisons are shown in section 3. For generating result and compare it with existing approach some benchmark image is taken from image database [11]. Concluding remarks are presented in section 4 and references are drawn at end.

## 2 The Technique

Embedding is initiated by taking a  $2 \times 2$  mask from the source gray scale image. Haar wavelet transformation is applied on sub mask to generate frequency coefficients. Two bits from hidden image are embedded on each coefficient of image mask with two bits per pixel. The dimension of the hidden image is embedded on first thirty two positions. Resultant stego image mask is transformed using inverse wavelet transform as post embedding operation. Stego image mask of size 32 bits are taken as initial population. New Generation followed by Crossover is applied on the initial population. New Generation is performed on the  $k$  bits by consecutive bitwise XOR operation on  $k$  steps, taking the MSB of the intermediate stream generated in each step. Crossover is applied on the consecutive two pixels. As a result rightmost two bits of two consecutive pixels are swapped. Genetic Algorithm is applied as post embedding operation to enhance a level of security.

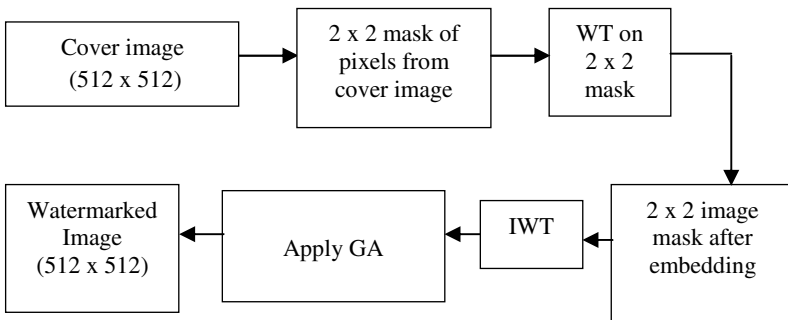


Fig. 1.1. The process to embed the Secret data into the source image

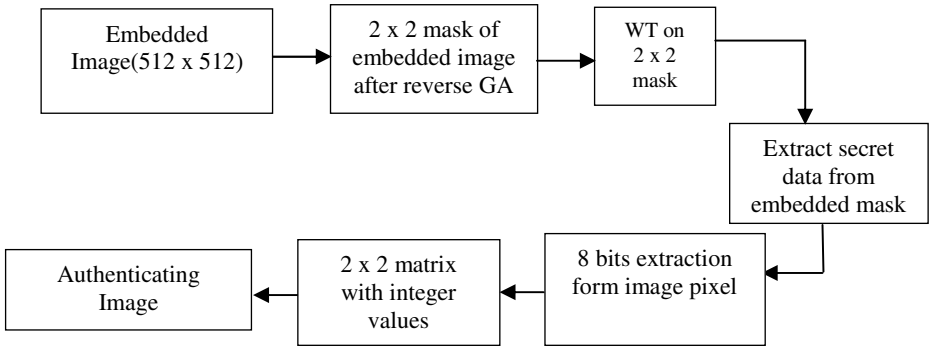


Fig. 1.2. The process to extract Secret data from the watermarked image

Fig. 1. Schematic diagram of AWTIAGA

Haar Wavelet is defined as follows

$$\text{har}(k, 0) = \lim_{\theta \rightarrow 0, \theta > 0} \text{har}(k, \theta),$$

$$\text{har}(k, 1) = \lim_{\theta \rightarrow 1, \theta < 0} \text{har}(k, \theta),$$

and at the points of discontinuity within the interior (0,1) of the interval [0,1].

$$\text{har}(k, \theta) = \frac{1}{2}(\text{har}(k, \theta - 0) + \text{har}(k, \theta + 0)).$$

A discrete Haar function is given in equation ...

$$r_n(t) = \sum_{k=0}^{2^n-1} \psi_{n,k}(t), \quad t \in [0, 1], \quad n \geq 0.$$

This is an orthogonal function. Haar Wavelet method is very high accuracy transformation and it has simplicity and small computation costs.

Figure1 show the diagram of the proposed technique out of which figure 1.1 shows encryption technique and that of figure 1.2 shows decryption technique. The process of embedding and extraction are given in section 2.1 and 2.2 respectively. A complete example has also been illustrated in section 2.3.

### 2.1 Algorithm for Insertion

The IAFDGA technique takes source gray scale image of size  $p \times q$ . Authenticating image size is  $m \times n$ . Haar wavelet transform is used for transformation and bit per pixel for embedding is two.

Input: Source image of size  $p \times q$ , hidden image of size  $m \times n$ .

Output: Embedded image of size  $p \times q$ .

Method: Insertion of hidden image bitwise into the source gray scale image.

- Step 1:* Extract dimension  $m \times n$  from the hidden image
- Step 2:* Wavelet transform is used to transform the image from spatial domain into spectral domain. Haar wavelet transform is taken to convert the sub image matrix of  $2 \times 2$  size from source image as preembedding operation
- Step 3:* Dimension is embedded onto the second and third position of each coefficients. First sixteen embedding positions are reserved for width and next sixteen are reserved for height
- Step 4:* Embed the hidden image bits onto the source image
- Step 5:* Apply inverse haar Transform as post embedding operation
- Step 6:*  $2 \times 2$  stego image mask of size 32 bits is taken as initial population. Perform New Generation operation on the initial population. For this operation rightmost 3 bits from each byte is taken as input. A consecutive bitwise XOR is performed on it for the 3 steps. It will form a triangular form and first bit from each step is taken
- Step 7:* For Crossover operation, rightmost two bits of two consecutive pixels are swapped
- Step 8:* Repeat step 2 to 7 for the whole cover image
- Step 9:* Stop.

## 2.2 Algorithm for Extraction

The embedded image is received in spatial domain is taken as the input and the hidden message/ image size, content are extracted from it.

Input: Embedded image of size  $p \times q$ .

Output: Host image of size  $p \times q$ , hidden image of size  $m \times n$ .

Method: Extract bits of hidden image from embedded image

- Step 1:* Reverse Crossover is performed on the rightmost 2 bits of two consecutive pixels of each  $2 \times 2$  mask. As a result rightmost two bits of the each pixel are swapped
- Step 2:* Reverse Generation is produced by consecutive bitwise XOR operation on the rightmost 3 bits of each byte in three steps. The first bit of each intermediate step is taken as the output
- Step 3:* Read reverse Generated image mask (of size  $2 \times 2$ ) in row major order. Apply Haar wavelet transform on the embedded image mask to transform the embedded sub image from spatial to frequency domain so that four frequency components are regenerated
- Step 4:* Extract the hidden bits from the second and third position of each coefficient. Replace hidden message/ image bit position in the block by '1'. For each eight extracted bits construct one image pixel of authenticating image.
- Step 5:* Repeat step 1 to 4 to regenerate hidden image as per size of the hidden image.
- Step 6:* Stop.

## 2.3 Example

Consider pixels of Jet image (figure 2a) to be inserted into each transformed mask of Lenna image (figure 2c). Figure 2b shows a sub mask of Lenna image in spatial

domain. Two bits of the Jet image are inserted on each Lenna image transformed coefficients. Insertion is done in the second and third position of each coefficient. Resultant mask after embedding is shown in Figure 2d in frequency domain and Figure 2e in spatial domain after inverse transform. Figure 2f shows New Generated Stego mask. Figure 2g shows Crossovered stego mask.

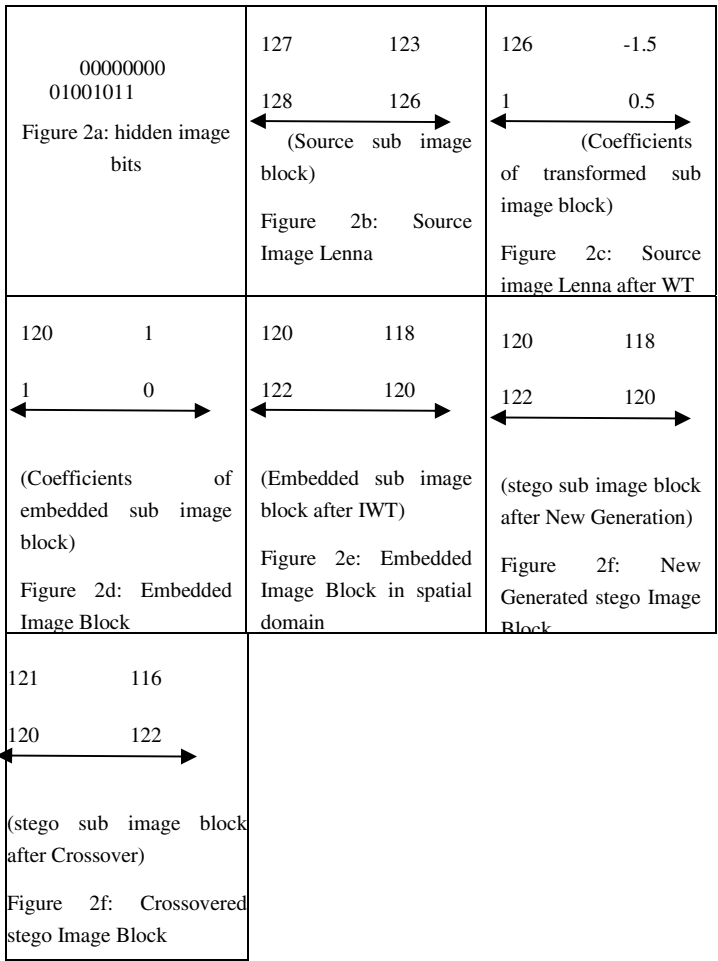


Fig. 2. Encoding process of AWTIAGA

### 3 Result Comparison and Analysis

Extensive analysis has been made on various images[11] using AWTIAGA technique. This section represents the results, discussions in terms of visual interpretation and peak signal to noise ratio. Figure 3a shows the host images Lenna, Tiffany,

Cameraman. Figure 3b shows embedded Lenna, Tiffany, Cameraman on embedding Jet image using AWTIAGA. Figure 3c is the authenticating image Jet. From figure 3b it is clear that there is no such visible change in the embedded image. Table 1 shows the PSNR value for each embedding against the source image. Results show high PSNR and image fidelity is almost closer to 1 which means there is not much deviation from the source image. From the table it is seen that the maximum value of the PSNR is 42.627369 and that of minimum value of the PSNR is 40.046883 which means PSNR is consistent for various images.

Table 2 shows the comparison with existing approach [1] for three images Jet, Lenna and Boat. From the table it is clear that the proposed technique obtain better PSNR as well as high embedding capacity compared to the existing [1]. The following formulas are used to calculate PSNR, MSE and IF (image fidelity).

$$PSNR = 10 \log(\max(I_{m,n}^2)/MSE)$$

$$MSE = \frac{1}{MN} * \sum_{m,n} (I_1(m,n) - I_2(m,n))^2$$

$$IF = 1 - \frac{\sum_{m,n} (I_{1,m,n} - I_{2,m,n})^2}{\sum_{m,n} I_{2,m,n}^2}$$

			
3.a.i. Host Lenna	3.a.ii.Host Tiffany	3.a.iii.Host Cameraman	3.b.i Embedded Lenna
			
3.b.ii Embedded Tiffany	3.b.iii Embedded Cameraman	3.c.i. Hidden Jet	

Fig. 3. Visual effect of embedding in AWTIAGA

**Table 1.** PSNR, MSE, IF values obtained for various images using AWTIAGA

Host Image	PSNR values	MSE Values	IF
Lenna	42.303341	3.826012	0.999762
Baboon	42.485359	3.668972	0.999803
Peppers	42.618370	3.558308	0.999800
Boat	42.627369	3.550941	0.999813
Cameraman	41.814621	4.281719	0.999763
Elaine	42.607716	3.567047	0.999828
Sailboat	42.605824	3.568600	0.999825
Jet	42.384212	3.755423	0.999890
Tiffany	40.046883	6.432682	0.999858

**Table 2.** Comparison of PSNR values obtained for various images using AWTIAGA and existing[1]

Host Image	PSNR values of AWTIAGA	Capacity of AWTIAGA	PSNR values of existing[1]
Jet	42.384212	45000	28.49
Lenna	42.303341	45000	29.68
Boat	42.627369	45000	28.74

## 4 Conclusion

The proposal is a novel embedding approach termed as, AWTIAGA based on Wavelet Transformation for gray scale images authentication. From experimental results it is clear that the proposed technique obtained high PSNR ratio along with good image fidelity for various images which conform that Wavelet-transformed based image steganography can obtain better visibility/quality. Greater payload works better compared to the existing approach.

**Acknowledgments.** The authors express deep sense of gratitude towards the department of CSE, University of Kalyani, India and PURSE scheme of DST, Govt. of India.

## References

1. Huang, H.-Y., et al.: A lossless data hiding based on discrete Haar wavelet transform. In: CIT 2010, 978-0-7695-4108-2/10 \$26.00 ©. IEEE (2010), doi:10.1109/CIT.2010.276
2. Bandyopadhyay, S.K., et al.: Genetic Algorithm based Substitution Technique of Image Steganography. Journal of Global Research in Computer Science 1(5) (December 2010) ISSN: 2229- 371X
3. Ghasemi, E., et al.: High Capacity Image Steganography using Wavelet Transform and Genetic Algorithm. In: Proc. IMECS 2011, Hong Kong, March 16-18, vol. 1 (1821) ISBN: 978-988-18210-3-4



4. Ghoshal, N., Mandal, J.K., et al.: Masking based Data Hiding and Image Authentication Technique (MDHIAT). In: Proceedings of 16th International Conference of IEEE on Advanced Computing and Communications ADCOM 2008, Anna University, December 14-17 (2008) ISBN: 978-1-4244-2962-2
5. Ghoshal, N., Mandal, J.K.: A Novel Technique for Image Authentication in Frequency Domain using Discrete Fourier Transformation Technique (IAFDZTT). *Malaysian Journal of Computer Science* 21(1), 24–32 (2008) ISSN 0127-9094
6. Ghoshal, N., Mandal, J.K.: A Bit Level Image Authentication / Secrete Message Transmission Technique (BLIA/SMTT). Association for the Advancement of Modelling and Simulation Technique in Enterprises (AMSE), *AMSE Journal of Signal Processing and Pattern Recognition* 51(4), 1–13 (2008)
7. Khamrui, A., et al.: An Authentication Technique for Image/ Legal Document. *J. Sign. Process. Syst.* 67, 187–199 (2012)
8. Khamrui, A., et al.: An Image Authentication Technique in Frequency Domain using Genetic Algorithm. *IJSEA* 3(5) (September 2012)
9. Bin, L., et al.: A Survey of Image Stegaography and Steganalysis. *Journal of Information Hiding and Multimedia Sigal Processing* 2(2) (April 2011) ISSN: 2073-4212
10. Wang, S., et al.: A Secure Steganography Method based on Genetic Algorithm. *Journal of Information Hiding and Multimedia Sigal Processing* 1(1) (January 2010) ISSN: 2073-4212
11. Allan, G.: Weber, The USC-SIPI Image Database: Version 5, Original release: Signal and Image Processing Institute, University of Southern California, Department of Electrical Engineering (October 1997), <http://sipi.usc.edu/database/> (last accessed on January 20, 2011)

# A New Approach to Monitor Network

Hemant Kumar Saini<sup>1</sup>, Anurag Jagetiya<sup>2</sup>, Kailash Kumar<sup>3</sup>,  
and Satpal Singh Kushwaha<sup>3</sup>

<sup>1</sup> Department of CSE, RTU, Kota (Rajasthan), India  
hemantrhce@rediffmail.com

<sup>2</sup> MLV Govt. Textile & Engineering College, Bhilwara(Rajasthan), India  
anurag.mlvttec@gmail.com

<sup>3</sup> Department of CSE, MITRC, Alwar(Rajasthan), India  
{maheshwari\_kailash2002, singh\_satpal25}@rediffmail.com

**Abstract.** Network-based attacks have become common and intervened. For this reason, detecting systems are now shifting their focus from host to network. Skimmers routinely perform “portscans” to search the vulnerable servers to intervene. Network Intrusion Detection Systems (NIDS) try to detect such behavior and flag them as malicious. An important requirement in such systems is instant response: the faster a NIDS detects malice, the lesser would be the resulting damage. At the same time, NIDS should not pseudo implicate the remote hosts as malicious. Balancing the promptness and accuracy in identifying malicious activity is a delicate and typical task. We develop detection system. TanceQi is a design which determines unfaithful processes and malicious services into our network in a fast and significant manner without having any tracking in route. This simplifies the work of admin and also performs monitoring faster and more accurately than the other than other current solutions.

**Keywords:** Security, Network, NetSTAT, LAN, NIDS, ethtool, iptraf, netperf, grep.

## 1 Introduction

The Internet today is a complex entity comprised of diverse networks, users, and resources. Most of the users are oblivious to the design of the Internet and its components and only use the services provided by their operating system or applications. However, there is a small minority of advanced users who use their knowledge to explore potential system vulnerabilities. Hackers can compromise the vulnerable hosts and can either take over their resources or use them as tools for future attacks. With so many different protocols and countless implementations of each for different platforms, the launch of an effective attack often begins with a separate process of identifying potential victims.

### 1.1 Popular Methods

One of the popular methods for finding susceptible hosts is port scanning. Port scanning can be defined as “hostile Internet searches for ‘open doors,’ or ports,

through which intruders gain access to computers.” [1] This technique emphasizes on posting a message to a port and wait for getting its response. The received message flags the port status and can be helpful in determining a hosts’ operating system and other information relevant to launching a future attack. In principle, this pattern of *port scanning* manifests quite differently from legitimate remote access, and thus holds promise for providing a means by which a network intrusion detection system (NIDS) can identify an attacker at a stage early enough to allow for some form of protective response to mitigate or fully prevent damage.

## 1.2 Problems

A number of difficulties arise, however, when we attempt to formulate an effective algorithm for detecting port scanning by defining a set of heuristics and applying them to the network for tracing the data, we would be able to separate suspicious packets and group them into scan trails. The first is that since networking is the hardest to monitor due to fact that the network is abstract so Ethernet settings has to be configured using “ethtool” to ensure that the specific system is synchronized at auto negotiation. The second is to test network throughput by sending traffic among systems and measure statistics like latency and speed using “iptraf” for local throughput, “netperf” for endpoint throughput and “iperf” to measure network efficiency.

## 1.3 Extension of Approach

Network-oriented intrusion detection systems can be roughly divided into distributed IDSs and network-based IDSs. A survey of network-oriented IDSs is given in [2] [15] Distributed IDSs are an extension of the original, single host intrusion detection approach to multiple hosts. So we give a smart idea for fast and accurate monitoring the network without any trail and application specific using techniques to isolate suspicious packets. TanceQi is aimed at real-time network-based intrusion detection. This approach extends the state transition analysis technique (STAT) [3] to network-based intrusion detection in order to represent attack scenarios in a networked environment.

With this approach we make the workings much easier, accurate and friendly for the user to control on remote system and also monitor its host as well as network.

## 2 Design

Tanceqi uses the NetSTAT[4] architecture and mapping (as discussed in above section) which displays various network related information such as network connections, routing tables, interface statistics, masquerade connections, multicast memberships etc as shown in Fig 1. Under Linux raw data can often be obtained from the `/proc/net/dev` and `/proc/net/tcp` to work around the printf output corruption arising in netstat's network interface statistics summary, netstat -i, until such time as the problem is corrected. The execution of model was based on the dialog software before running the desired model it will automatically detect the requirements and if

not then it will take automatically from its server and install its own. It has been designed in bash shell and has accommodate all the strategies to filter the results using different options like `-t`, `-nt` so it will be scanning port [5] and save the state in different file which we had taken as server name file. And for mapping with server ip we use the `nmap` [6] [7] to scan ports that has been developed to help system administrators (SAs) with the analysis of open ports in system(s). The aim is to determine unauthorized access, ill-use, and maltreat of computer systems by both system internal and external perforators. The intrusion detection problem is becoming a difficult task due to the rapid increase in computer networks due to increased connectivity of host's which gives greater access to outsiders and makes it easier for them to hide their identification. Intrusion detection systems (IDSs) [8] are designed with the view that an intruder's behavior is always different from the vindicated user so that many illegitimate actions are discoverable. Typically, IDSs employ statistical anomaly and rule based misuse models [16] in order to detect intrusions. Hence without any rules or any installation my proposed model will give the best results and therefore it will work independently. There is no need to make any database as done before in different techniques for network monitor.

In past time intruders use kernel level rootkits [9] [10] to hide their presence from the vindicate users and administrators of a compromised machine. Originally, rootkits mainly included modified versions of system auditing programs (e.g., `ps` or `netstat` on a UNIX system). So using this concept we trace kernel routing in my proposed model.

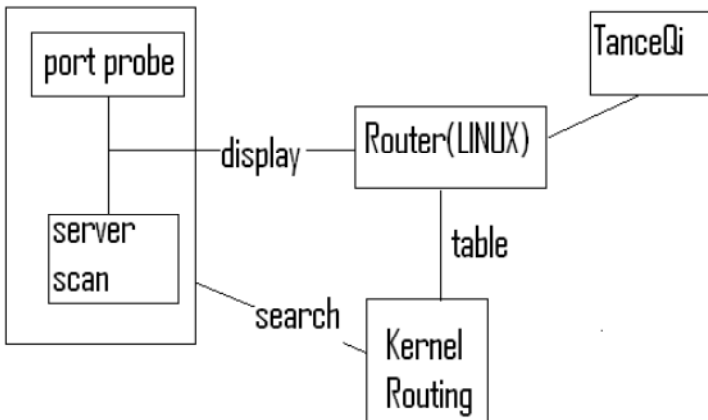


Fig. 1. TanceQi Design

### 3 Implementation

The main advantage of this design is it works on any Linux platform which is indulge in network server. It will be implemented with `dialog` [11] which is also detecting automatically. For its implementation the only pre-requisite is to execute it must have

a proper permission in binary otherwise it will be operated by shell. Here we are using only the binary execution for my experiment.

We have taken Red hat enterprise Linux 5.1 which had an already server configured. We just execute my idea as with root and it will ask for choice as we have mentioned in design.

For the choice and monitoring we use algorithm which is given in section 3.1 below.

```
Procedure monitor ($connected, $connected_from, $remotes,
$port, $srport, $ip, $state)
```

```
Input:
```

```
$connected ← array of choice for initialization
$connected_from ← selection mode for connection
$remotes ← variable for choices stored
$port, $srport ← temporary port number
$ip ← ip address
$state ← states of close establish and synchronize
Select choice for $connected to servers in $remotes
Return value of choice
```

```
If return value =0 then
```

```
Case choice selected $connected for connection to
```

```
Byname
```

```
Enter Name of Server in $name
```

```
Grep List of Connected $name
```

```
By Port
```

```
Enter IP in $port
```

```
Grep List of Connected $port
```

```
By State
```

```
Enter port of Server in $server_port Enter State in
```

```
EST or SYN or CLOSE in $state List of $server_Port
```

```
is connected by $State
```

```
esac
```

```
Else
```

```
Case choice selected for $connected_from
```

```
ByIP and Port
```

```
Enter $ip of Server
```

```
Enter $srport of Server
```

```
Grep list of $srport on $ip
```

```
ByPort
```

```
List (Connected Client) $srport and $state
```

```
Esac
```

```
fi
```

```
End Procedure
```

In above DIALOG [12] is a red hat package which is installed before. And the auditing trails will be removed after entire working by using trap.

```
trap "rm -f $connected $connected_from s $name $port
$server_port $state" 0 1 2 5 15
```

This all about monitoring and now the control is given on different server and trace them and check the loopholes and stop those. For that we use to restrict the port for that we use the algorithm as below.

**Procedure control** (stop\_services, sip, sport)

**Input :**

```
stop_services ← temporary data for stop services
sip ← temporary internet address
sport ←temporary ports
Select the service to stop in stop_services
Return value of choice
```

**If** Return value=0

then

**Case** stop\_services choice in Secure

```
Enter IP of Server into sip
Enter Port Of Server into sip Input sip and port
into iptables for firewall
Save firewall permanently
sip restricted for sport
Unsecure
iptables flush
save iptables
```

**esac**

**Else**

```
echo bye
```

**End Procedure**

This is how the restriction is allowed or denied and monitoring the entire LAN.

## 4 Results

The model gives the results according to choices of user respectively. As with the choice 1 for connected links to gives again three parameters for port, name and state (as given in Fig 2 and Fig 3) in which input the name of server like here we used ssh so it will give desired list of establishes tcp connections as given in Fig 4. The throughput of choice 1 (in fast scanning) with netstat responds in 15.6 seconds which is much better than the proposed models. Also with the choice of port it limits to hosts of 1024 with UDP or some TCP and scan finishes in 15 seconds. Its maximum scan sending rate is used 100 packets per second.

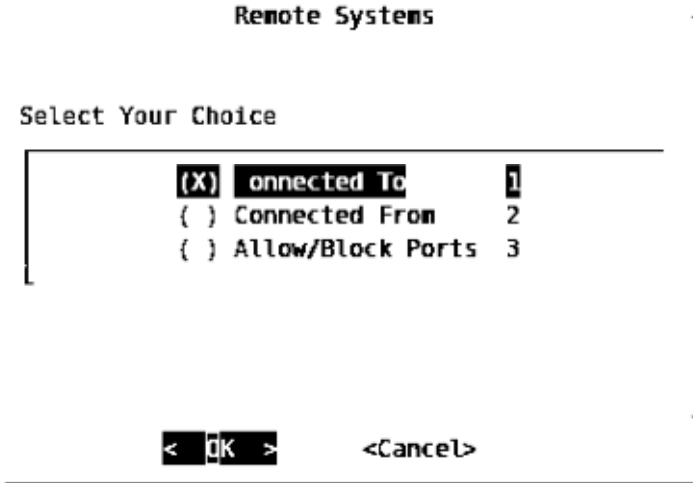


Fig. 2. Remote option

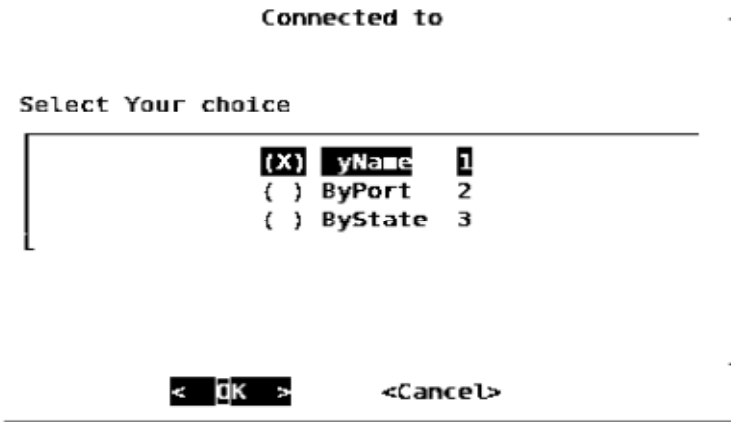


Fig. 3. Connections

```
      List of Connected ssh Server
tcp 0 0 localhost.localdomain:39946 localhost.localdomain:ssh ESTABLISHED
tcp 0 0 localhost.localdomain:39947 localhost.localdomain:ssh ESTABLISHED
tcp 0 0 localhost.localdomain:39945 localhost.localdomain:ssh ESTABLISHED
tcp 0 0 localhost.localdomain:ssh localhost.localdomain:39945 ESTABLISHED
tcp 0 0 localhost.localdomain:ssh localhost.localdomain:39946 ESTABLISHED
tcp 0 0 localhost.localdomain:ssh localhost.localdomain:39947 ESTABLISHED
```

Fig. 4. List of Servers Connected

After scanning different ports [13] we got results for fast scans by our approach which produce better throughput than before old scanners and facilitate with blocking port if find any suspect (as shown in Fig 5) until we improve the security so that our network would be secure. This is how fast we monitor and control the host with

network without any dependencies and accuracy within time measurements. Also it is user friendly just without having much commands grep all the information needed for our monitoring.

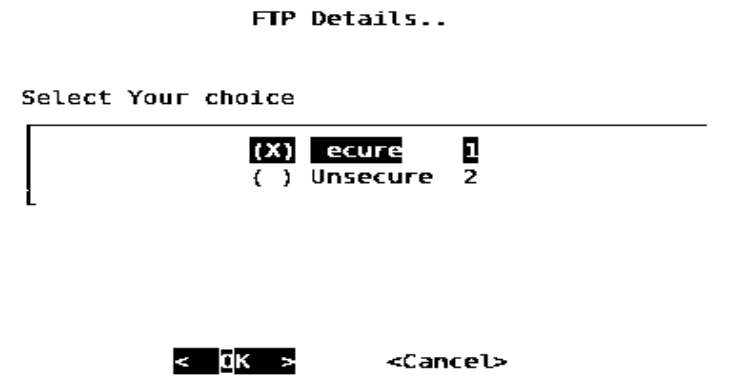


Fig. 5. Restriction

## 5 Conclusion

The word “TanceQi” means the detector. Before the IDS work was challenging one due to large auditing trails and user has to find in it which takes lot of time. So we have presented a model design to easy, interactive and independent monitoring which saves lot of time in scanning different ports with different ways in a well designed manner. This has an advantage of traps which does not leave any trail of its own that’s why no one else in network could find us.

## 6 Summary of Research Contributions

The extension of the STAT to network intrusion detection gives a better way for defending the network; it is possible to deploy a practical monitoring technique that improves on existing approaches. Network detection techniques benefit from the communications of hosts in the network. During evaluation our independent monitoring prototype was successful in scanning the ports in less time, easy and organized manner. We have provided specific recommendations for different settings especially for individual hosts and for specific local network. It is based on the observation that for auditing large log files our own work is also makes logs into system so it will also be taken which increases the latency so we make our design in such a way that it do not create any log which is helpful in r4educing the time and increases the performance of our approach.

Since the design has trapped all the states created by monitoring algorithm so the log created by our own design has been removed and the auditing trails has been halved which is an improvement over log based scanning techniques. Also it enables the user with the facility to interact in graphical and responsive way with user friendly design so that even the less trained person on Linux could easily be handle it.



## References

1. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K.: Surveying Port Scans and Their Detection Methodologies. *Comput. J.* 54(10), 1565–1581 (2011)
2. Mukherjee, B., Heberlein, L.T., Levitt, K.N.: Network Intrusion Detection. *IEEE Network*, 26–41 (May/June 1994)
3. Iglun, K., Kemmerer, R.A., Porras, P.A.: State Transition Analysis: A Rule-Based Intrusion detection System. *IEEE Transactions on software Engineering* 21(3) (March 1995)
4. Vigna, G., Kemmerer, R.A.: NetSTAT: A Network-Based Intrusion Detection Approach. In Proceedings. In: Proceedings of the 14th Annual Computer Security Applications Conference, ACSAC 1998, p. 25. IEEE Computer Society, Washington, DC (1998)
5. Gadge, J., Patil, A.A.: Port scan detection. In: 16th IEEE International Conference on Networks, pp. 1–6. ICON (2008)
6. Kocher, J.E., Gilliam, D.P.: Self port scanning tool: providing a more secure computing environment through the use of proactive port scanning. In: 14th IEEE International Workshops on Enabling Technologies Infrastructure for Collaborative Enterprise, June 13–15, pp. 139–143 (2005)
7. Nmap - Free Security Scanner For Network Exploration & Security Audits, <http://nmap.org/>
8. Mallisery, S., Prabhu, J., Ganiga, R.: Survey on Intrusion detection Methods. In: 3rd International Conference on Advances in Recent Technologies in Communication and Computing, November 14–15, pp. 224–228 (2011)
9. Gupta, S.: Logging and Monitoring to Detect Network Intrusions and Compliance Vs in the Environment. Secunia's Yearly Report 2010 (July 4, 2012)
10. Levine, J.G., Grizzard, J.B., Owen, H.L.: Detecting and Categorizing Kernel-Level Rootkits to Aid Future Detection. *IEEE Security and Privacy* 4(1), 24–32 (2006)
11. Aubert, S.: rkscan: Rootkit Scanner (2004), <http://www.hsc.fr/ressources/outils/rkscan/>
12. Dialog: An Introductory Tutorial, <http://www.linuxjournal.com/article/2807>
13. Dialog Tool box red hat packet manager for RHEL 5.1, [http://rpm.pbone.net/index.php3/stat/4/idpl/2392201/dir/redhat\\_5.x/com/dialog-0.6-12.i386.rpm.html](http://rpm.pbone.net/index.php3/stat/4/idpl/2392201/dir/redhat_5.x/com/dialog-0.6-12.i386.rpm.html)
14. Server security, [https://access.redhat.com/site/documentation/en-US/Red\\_Hat\\_Enterprise\\_Linux/3/html/Security\\_Guide/s1-server-ports.html](https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/3/html/Security_Guide/s1-server-ports.html)
15. Srivastava, R., Richhariya, V.: Survey of Current Network Intrusion Detection Techniques. *Journal of Information Engineering and Applications* 3(6) (2013) ISSN 2224-5782
16. Sonawane, S., Pardeshi, S., Prasad, G.: A survey on intrusion detection techniques. In: Proceedings of National Conference on Emerging Trends in Information Technology, vol. 2(3), pp. 127–133 (2012); *World Journal of Science and Technology*

# Detection and Mitigation of Misbehaving Vehicles from VANET

Megha Kadam<sup>1</sup> and Suresh Limkar<sup>2</sup>

<sup>1</sup> Department of Computer Engineering, GHRCEM, Pune, India

<sup>2</sup> Department of Computer Engineering, AISSMS IOIT, Pune, India  
{megha.desail,sureshlimkar}@gmail.com

**Abstract.** VANET means vehicular ad hoc network is nothing but the group of independent vehicles nodes which are moving throughout the wireless network freely. Such kind of networks are temporary as the vehicles and their positions are not fixed and hence the all the routing paths which are established in order to make the communication in between the source and destination are on demand and depends on the nodes movement into the network. The architecture is not at all needed for such kind of networks. Role of routing protocols is most important for the VANET which is used to route the data from source to destination, but they are also vulnerable to the many of the security attacks in the VANET. Due to the unprotected nature of the VANET networks routing protocols, such networks also unprotected from the malicious vehicles in the network itself. This paper presents new approach for not only the detection of malicious vehicles attack but also their prevention from the VANET. Proposed algorithm is referred as Detection and Prevention of Malicious Vehicles (D&PMV). The malicious vehicles detected using the monitoring process over the VANET, once they are detected, proposed algorithm is applied for the prevention of the same. The detection of malicious vehicles is based on DMV algorithm presented earlier.

**Keywords:** Abnormal behavior, Vehicular Ad Hoc Networks, Honest vehicle, Secure communication, Malicious vehicle, Detection, Prevention, MANET.

## 1 Introduction

Vehicular Ad Hoc Networks (VANETs) are appropriate networks that can be applied to intelligent transportation systems [3]. VANET is based on short-range wireless communication between vehicle- to- vehicle and some roadside infrastructure. Moreover, a large number of Certification Authorities (CAs) will exist, where each CA is responsible for the identity management of all vehicles registered in its region (e.g., national territory, district, country) [16].

Several types of messages are exchanged among vehicles such as traffic information, emergency incident notifications, and road conditions. It is important to forward correctly message in VANET; however, attacker nodes may damage the messages. Attackers or malicious vehicles perform in several ways and have different

objectives such as attackers eavesdrop the communication between vehicles, drop, and change or inject packets into the network.

To provide network services under the presence of misbehaving nodes, it is necessary to consider “fault tolerance” as a main objective at the design level of routing protocols. To address this concern, several secure routing protocols have been proposed recently. Some of these protocols handle attacks by malicious nodes but not the selfish nodes and some handle selfish nodes but not malicious nodes. At the best of our knowledge, there is no solution that handles all misbehaving nodes actions. So it’s necessary to provide a simulation study that measures the impact of misbehaving nodes in order to provide protocol designers with new guidelines that help in the design of fault / attack tolerant routing protocols for VANET.

### 1.1 VANET Communications

With an immense improvement in technological innovations, we find Vehicular Communication (VC) as a solution to many problems of our modern day communication system in roads. VC involves the use of short range radios in each vehicle, which would allow various vehicles to communicate with each other which is also known as (V-V) communication and with road side infrastructure(V-I) communication. These vehicles would then form an instantiation of ad hoc networks in vehicles, popularly known as Vehicular Ad Hoc Networks (VANET). It is a subset of Mobile Ad Hoc Networks (MANET). The similarity between these two networks is characterized by the movement and self organization of nodes. Also the difference between these ad hoc networks is that MANET nodes cannot recharge their battery power where as VANET nodes are able to recharge them frequently. We can understand VANETs as subset of MANET and best example of VANET is Bus System of any University which is connected. These buses are moving in different parts of city to pick or drop students if they are connected, make an Ad hoc Network.

VANET is mainly designed to provide safety related information, traffic management, and infotainment services. Safety and traffic management require real time information and this conveyed information can affect life or death decisions. Simple and effective security mechanism is the major problem of deploying VANET in public. Without security, a Vehicular Ad Hoc Network (VANET) system is wide open to a number of attacks such as propagation of false warning messages as well as suppression of actual warning messages, thereby causing accidents. This makes security a factor of major concern in building such networks. VANET are of prime importance, as they are likely to be amongst the first commercial application of ad hoc network technology. Vehicles are the majority of all the nodes, which are capable of forming self organizing networks with no prior knowledge of each other whose security level is very low and they are the most vulnerable part of the network which can be attacked easily. The capacity of VANET technology is high with a wide range of applications being deployed in aid of consumers, commercial establishments such as toll plazas, entertainment companies as well as law enforcement authorities. However, without securing these networks, damage to life and property can be done at a greater extent.

## 1.2 Technology

In VANET defines an intelligent way of using Vehicular Networking. InVANET integrates on multiple ad-hoc networking technologies[15] such as WiFi IEEE 802.11p, WAVE IEEE 1609, WiMAX IEEE 802.16, Bluetooth, IRA, ZigBee for easy, accurate, effective and simple communication between vehicles on dynamic mobility. Effective measures such as media communication between vehicles can be enabled as a well method to track the automotive vehicles is also preferred.

Vehicular Networks are envisioned of the Intelligent Transportation Systems (ITS). Vehicles communicate with each other via Inter-Vehicle Communication (IVC) as well as with roadside base stations via Roadside-to-Vehicle Communication (RVC). The optimal goal is that vehicular networks will contribute to safer and more efficient roads in the future by providing timely information to drivers and concerned authorities.

## 2 VANET Applications

Major applications of VANET include providing safety information, traffic management, toll services, location based services and infotainment. One of the major applications of VANET include providing safety related information to avoid collisions, reducing pile up of vehicles after an accident and offering warnings related to state of roads and intersections. Affixed with the safety related information are the liability related messages, which would determine which vehicles are present at the site of the accident and later help in fixing responsibility for the accident. There are several applications of VANET like Location-based services [14], Traffic optimization [17] and collision avoidance [7].

## 3 Security of VANET

To become a real technology that can guarantee public safety on the roads, vehicular networks need an appropriate security architecture that will protect them from different types of security attacks such as Denial of Service [19], Message Suppression Attack [19], Fabrication Attack [19], Alteration Attack [19] and Replay Attack [19]. VANET security should satisfy goals like Information authenticity, message integrity, source authentication, privacy and robustness of system. VANET should held Greedy drivers, Pranksters, industrial insiders and malicious attackers that affects on security of VANET.

## 4 Proposed Work

Our main aim with this is to simulate the misbehaving vehicles attack in the VANET by modifying existing routing protocol DSR (Dynamic routing protocol) in order to detect and prevent such attack in the network. Our research methods are implemented over the DSR for presenting the impact of presence of malicious vehicle in network.

Proposed DSR algorithm also addresses the all kinds of misbehaving nodes such as selfish as well as malicious nodes. Simulation study presents how proposed DSR detect the misbehaving nodes and how to prevent them from dropping the data. Following are the main objectives:

- To present the various significance of the VANET networks.
- To present the detailed study over VANET
- To analyze the malicious nodes attack in the VANET.
- To present approaches to provide security to the mobile ad hoc networks from malicious vehicle attacks.

In this proposed research work, we are presenting the design of a novel misbehavior detection scheme at the application layer, called DMV (Detection of Malicious Vehicles), and tag vehicles using their distrust values[6]. In DMV, verifiers operate independently from each other. In addition, DMV can improve the performance of verifier selection at high speeds. In DMV[6], a number of vehicles are located in a cluster. Each cluster has one main cluster-head and one spare cluster-head, where the spare cluster-head is the trustiest vehicle after the main cluster-head. Each vehicle is monitored by some of its trustier neighbors which are defined as verifier nodes. If verifier nodes observe an abnormal behavior from vehicle V (a node that drops or duplicates packets), they increase the distrust value of vehicle V. The identification code of vehicle V is reported to its Certificate Authority (CA) as a malicious node when its distrust value becomes higher than a threshold value.

### Algorithm Design

In this section, we introduce a monitoring algorithm to detect malicious vehicles and prevent them from honest ones. Following algorithm shows the monitoring process for vehicle V when it joins to a cluster.

#### *Algorithm 1: Malicious vehicle detection*

Step 1: Vehicle V joins the network

Step 2: Get the cluster keys

Step 3: Assigning the verifiers to newly joined vehicle V.

Step 4: Start monitoring behavior of vehicle V.

Step 5: If (verifier detecting the abnormal behavior of vehicle V)

Report to cluster\_head (CH)

goto step 6;

else

goto step 4;

Step 6: CH modifies the distrust value (d)

Step 7: if distrust value less than or equal to threshold (t) value (which is set for each new vehicle once that join the network) then update the white list of network else generation of warning message to cluster agents of CH.

if (d <= t)

```

goto step 4
else
goto step 8

```

Step 8: Alarm generation in order to provide the warning message among all the other vehicles under the same cluster head.

### ***Algorithm 2: Prevention of Malicious Vehicle***

Step 1: As once in algorithm 1 and step 8, alarm is generated for the detection of malicious vehicle in VANET; our second algorithm is executed for its prevention. Information retrieval of malicious vehicle is done at first.

Step 2: Parsing all communication paths in the network.

Step 3: Identify the presence of malicious nodes, if it's presented their in path, simply discard that path and use alternate communication path.

## **5 Research Methodology**

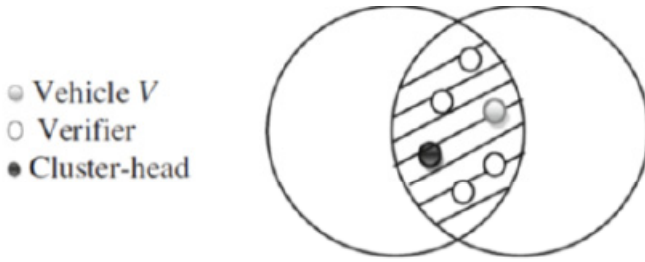
Research methodologies for the project work are related to the analysis of the misbehaving vehicles which are responsible for the black hole attacks in the VANET. Mainly in the Blackhole attack, traffics for the network are redirected to the mobile node in the network which not at all exists in the network. Thus in this case the network traffic is disappear into the one of the special mobile node such node is called as Blackhole node. The Blackhole attack has two characteristics: first one, the misbehave node advertise itself regarding to the information that it has shortest route to the destination with the intention of dropping the packets or intercepting packets in the routing protocols like DSR, AODV. Intercepted packets then consumed by the node: this is the second characteristics. For the simulation of the Blackhole attack, our research works on the misbehaving nodes simulation on the DSR protocol and prevention of it. We used Qualitative and Quantitative research methods.

## **6 Mathematical Model**

### **NP-Hard Problem**

As per from given algorithm, we are finding the malicious vehicle using the monitoring process. Here the process of monitoring is NP-Hard problem. Following is the mathematical model for this problem.

The purpose of monitoring is to collect information about the behavior of all vehicles in the network. Vehicles that perform the monitoring process are called "verifier" vehicles. A verifier is a vehicle which has a smaller or equal Td compared with the Td of vehicle V, and is located inside the region z (V, CH). This region shows the intersection area of both vehicle V and its CH (see Fig. 1).



**Fig. 1.** Illustration of region  $z(V,CH)$

Using this way, we are sure that all verifiers which monitor vehicle  $V$  are able to send their reports to CH [20]. Note that the area of CH is equal to its transmission range but the area of  $V$  is obtained from Eq. (1):

$$Area(V) = TR(V) - T_f(S_{max} - S_{min}) \tag{1}$$

Definition of general variables

- $\beta$  - Number of transmitted packets per second
- CA - Certificate Authority
- CH - Cluster head
- $I_v$  - Abnormal behavior rate
- $L$  - Number of verifiers
- $P_c$  - Probability of missing or duplicating a packet
- $P_{CA}$  - Radius of the CA
- $S_{max}$  - Legal maximum speed of the vehicle
- $S_{min}$  - Legal minimum speed of the vehicle
- $T_d$  - Distrust value
- $T_f$  - Packet Latency
- TR - Transmission range
- $t_{rem}$  - Maximum time that malicious vehicle requires to pass through the CA region
- U - Verifier
- V - Certain vehicle
- X - Number of missed or duplicated packets
- Y - Number of packets received to vehicle V
- $\gamma$  - Legal number of drops or duplicates packets
- $\lambda$  - Average number of dropped or duplicated packets by each vehicle before detection
- $\sigma$  - Threshold value on distrust

**NP-Complete Problem**

*Determining abnormal behavior of a vehicle by its neighbors:*

From a source vehicle point of view in VANET, another vehicle can play either destination role or relaying role. When vehicle  $V$  plays the relaying role, some of its verifier vehicles verify the behavior of  $V$  via monitoring. When  $u_i$  operates as verifier

of V, it counts the number of packets that V receives (denoted by parameter y) and those missed or duplicated by V that verifier  $u_i$  detects (denoted by parameter x). For this purpose verifier  $u_i$  sends its report to its CH. If after elapsing  $T_f$ , vehicle V does not forward a received packet or sends a packet twice,  $u_i$  considers it as an abnormal behavior and then increases the parameter x by one unit.

The parameter  $T_d$  is automatically mapped to each vehicle and can change when the vehicle performs either as a relaying vehicle or as a source vehicle. When  $T_d$  of vehicle V changes, the new  $T_d$  is broadcast to neighbors, and the neighbors update their white lists.

Vehicles cooperate with vehicle V if its  $T_d$  is lower than a threshold value  $\sigma$ . When  $T_d$  of vehicle V is larger than  $\sigma$ , its ID should be reported to the relevant CA as a malicious vehicle. Then, CA broadcasts the ID of malicious vehicle V to all vehicles and roadside infrastructure units. In order to update  $T_d$ , the CH calculates parameter  $I_v$  for vehicle V using Eq. (2):

$$I_v = \sum_{j=0}^L \frac{(1-P_c)^{y_j-x_j}(P_c)^{x_j}}{T_{dj}} \tag{2}$$

Where L is the number of verifier vehicles of malicious vehicle V. In Eq. (2), we assume that vehicle V forwards a received packet or misses or duplicates it. The parameter  $P_c$  is the probability of missing or duplicating a packet by a malicious vehicle. The parameters  $y_j$  and  $x_j$  show the number of packets received to vehicle V as well as those missed or duplicated by V, respectively, counted by the  $j^{th}$  verifier. According to Eq. (2), the CH computes a normalized  $I_v$  using coefficient in order to reduce the effect of verifiers with high distrust values. Using  $I_v$ , the CH computes new  $T_d$  of V using Eq. (2):

$$T_d(new) = T_d(old) + I_v \tag{3}$$

## 7 Performance Evaluation

### 1. Throughput

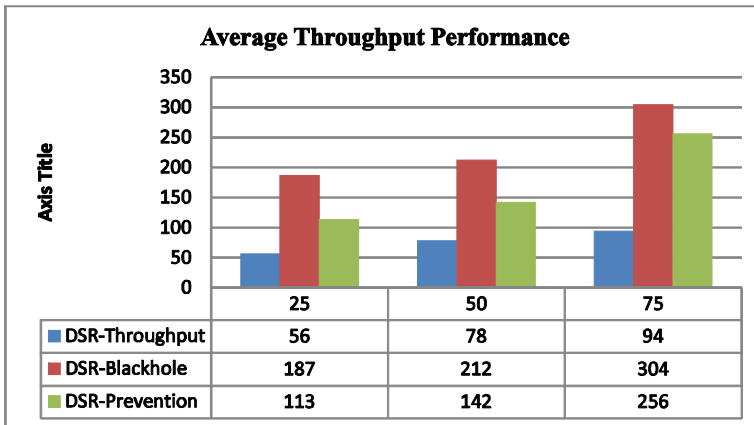


Fig. 2. Average Throughput



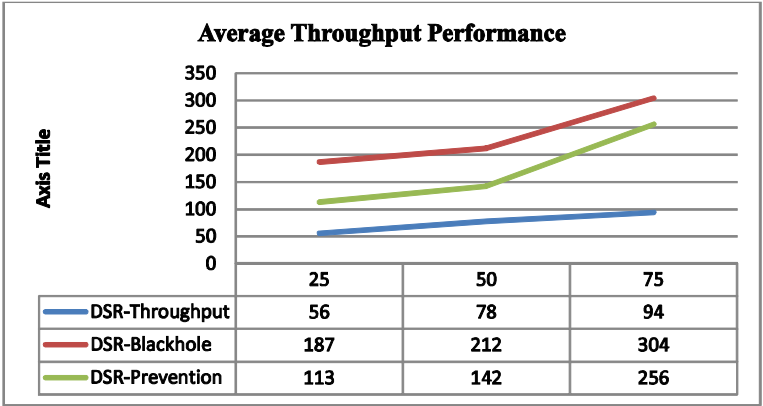


Fig. 3. Average Throughput (Packet/Sec)

2. Delay

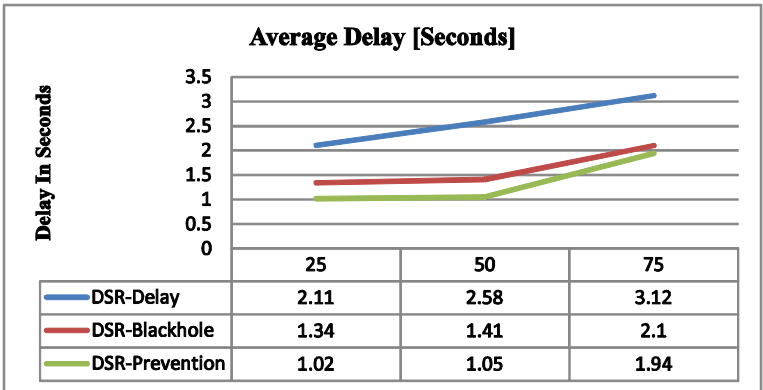


Fig. 4. Delay in Seconds

3. Jitter

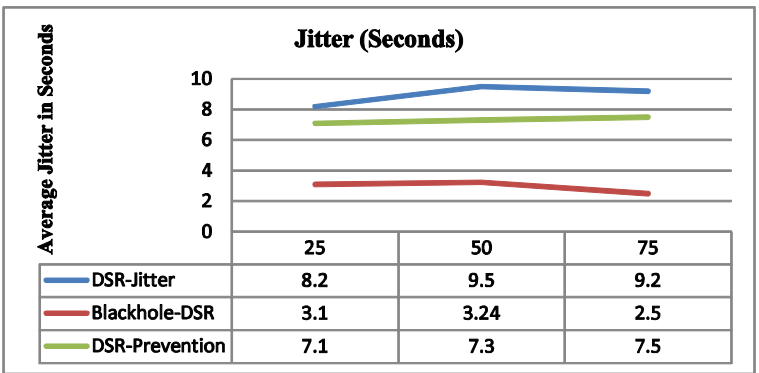


Fig. 5. Average Jitter in Seconds

### 4. Packet Dropped Ratio

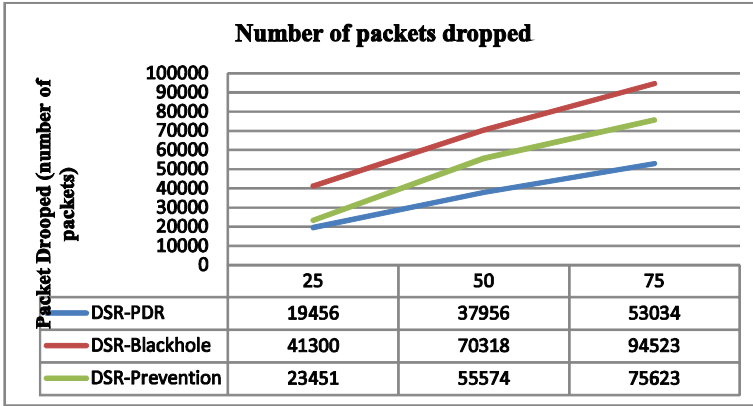


Fig. 6. Packet Drooped (Number of Packets)

From above graphs, it’s obvious that at the presence of blackhole attack in VANET network makes the extra packet drops as compared to normal working of DSR. Hence to reduce such packet drops in the network, we have introduced the new approach to mitigate the blackhole attack from the VANET. From above results, this approach reduced the effect of blackhole attack in the VANET.

## 8 Conclusion

In this research work, we have proposed and evaluated a misbehavior detection scheme for detecting vehicles that drop or duplicate received packets. The simulation results show that DMV can detect malicious vehicles before the number of dropped or duplicated packets becomes more than a given number of drops or duplicates ( $\gamma$ ). In addition, DMV can improve the performance of verifier selection at high speeds. In this algorithm, a distrust value has been introduced that can be used to determine trustiness value of each vehicle when it forwards messages. Therefore, vehicles can select trustier vehicles to forward messages based on the vehicles distrust value.

## References

1. Kadam, M., Limkar, S.: D &PMV: New Approach for Detection and Prevention of Misbehave/Malicious Vehicles from VANET. In: Satapathy, S.C., Udgata, S.K., Biswal, B.N. (eds.) FICTA 2013. AISC, vol. 247, pp. 293–303. Springer, Heidelberg (2014)
2. Kadam, M., Limkar, S.: Performance Investigation of DMV (Detecting Malicious Vehicle) and D&PMV (Detection and Prevention of Misbehave/Malicious Vehicles): Future Road Map. In: Satapathy, S.C., Udgata, S.K., Biswal, B.N. (eds.) FICTA 2013. AISC, vol. 247, pp. 379–387. Springer, Heidelberg (2014)

3. Abdulhamid, H., Tepe, K.E., Abdel-Raheem, E.: Performance of DSRC systems using conventional channel estimation at high velocities. *Int. J. Electron. Commun.*, 556–561 (2007)
4. Artimy, M.: Local density estimation and dynamic transmission-range assignment in Vehicular Ad Hoc Networks. *IEEE Trans. Intell. Transp. Syst.* 8(3), 400–412 (2007)
5. Bettstetter, C.: Smooth is better than sharp: a random mobility model for simulation of wireless networks. In: 4th ACM International Workshop on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM 2001), Rome, Italy (2001)
6. Detection of malicious vehicles (DMV) through monitoring in Vehicular Ad-Hoc Networks. Ameneh Daeinabi & Akbar Ghaffarpour Rahbar. Springer Science+Business Media, LLC (2011)
7. Raya, M., Hubaux, J.P.: The security of vehicular ad hoc networks. *J. Comput. Secur. Spec. Issue Secur Ad Hoc Sensor Netw.* 15(1), 39–68 (2007)
8. Wang, N.W., Huang, Y.M., Chen, W.M.: A novel secure communication scheme in vehicular ad hoc networks. *Comput. Commun.* 31(12), 2827–2837 (2008)
9. Fan, P., Haran, J.G., Dillenburg, J., Nelson, P.C.: Cluster-based framework in Vehicular Ad-Hoc Networks. In: Syrotiuk, V.R., Chávez, E. (eds.) ADHOC-NOW 2005. LNCS, vol. 3738, pp. 32–42. Springer, Heidelberg (2005)
10. Hu, Y.-C., Perrig, A.: Survey of Secure Wireless Ad Hoc Routing. *IEEE Security & Privacy* 2(3), 28–39 (2004), doi:10.1109/MSP.2004.1
11. Yan, G., Olariu, S., Weigle, M.C.: Providing VANET security through active position. *Comput. Commun.* 31(12), 2883–2897 (2008)
12. The Security of Vehicular Ad Hoc Networks Maxim Raya and JeanPierre Hubaux SASN 2005, November 7, Alexandria, Virginia, USA. Copyright, ACM 95932275/ 05/0011 (2005)
13. Raya, M., Papadimitratos, P., Hubaux, J.P.: Securing Vehicular Communications. *IEEE Wireless Communications* 13 (October 2006)
14. Golle, P., Greene, D., Staddon, J.: Detecting and Correcting Malicious Data in VANETs. In: VANET 2004, Copyright 2004, Philadelphia, Pennsylvania, USA, ACM 1581139225/04/0010 (October 1, 2004)
15. Abdalla, G.M.T.: SM Senouci Current Trends in Vehicular Ad Hoc Networks. In: Proceedings of UBIROADS Workshop (2007)
16. Raya, M., Jungels, D., Papadimitratos, P., Aad, I., Hubaux, J.P.: Certificate Revocation in Vehicular Networks, Laboratory for computer Communications and Applications (LCA) School of Computer and Communication Sciences, EPFL, Switzerland (2006)
17. Nadeem, T., Shankar, P.: A comparative study of data dissemination models for VANETs. *IEEE Mob. Ubiquitous Syst. Netw. Serv.*, 1–10 (2006)
18. Picconi, F., Ravi, N., Gruteser, M., Iftode, L.: Probabilistic validation of aggregated data in Vehicular Ad Hoc Networks. In: International Conference on Mobile Computing and Networking, Los Angeles, CA, USA, pp. 76–85 (2006)
19. Parno, B., Perrig, A.: Challenges in Securing Vehicular Networks
20. Raya, M., Papadimitratos, P., Aad, I., Jungels, D., Hubaux, J.P.: Eviction of Misbehaving and Faulty Nodes in Vehicular Networks. *IEEE J. Sel. Areas Commun.* 25(8), 1557–1568 (2007)

# Comparative Study of Hierarchical Based Routing Protocols: Wireless Sensor Networks

Pritee Parwekar

Department of Computer Science,  
Jaypee Institute of Information Technology,  
Noida, India  
pritee2000@gmail.com

**Abstract.** Wireless Sensor and Actor Networks (WSANs) comprise of sensors and actors to perform distributed sensing and acting tasks. Sensing of the environment using wireless sensor networks (WSN) has been one of the favorite areas of the research community. Conversion of this concept into real life implementable solutions would find interest only if a response mechanism is developed to complement the sensing. Introduction of the actor element in the WSNs makes it Wireless Sensor and Actor Networks (WSANs). The WSANs come with their set of issues having subtle differences from the WSN. The protocol design thus differs from the traditional WSNs. The issues like mobility of nodes, energy consumption, scalability, fault tolerance, response time have been highlighted in this paper and a survey and comparison are undertaken in respect to clustering based WSAN protocols.

**Keywords:** wireless sensor actor network, clustering protocol.

## 1 Introduction

Wireless Sensor Networks (WSNs) monitor the environment passively and collect the required information. The data is sent to a receiving station where the data is processed and stored. In many applications, however, only sensing the environment is not sufficient. It is also sometimes necessary to respond to the sensed events/data by performing corresponding actions in that environment.

The wireless sensor and actor networks (WSANs) enable the applications to sense, interact, and change the physical world. Every WSAN protocol aims at solving a problem statement which has one or more factors viz. energy consumption, mobility, scalability, reliability, fault tolerance, applicability, and response time. Finding best combination of these factors is the challenge being addressed by the various approaches in WSAN. Quest to design energy efficient sensor nodes and its energy efficient deploying methods in the network have given birth to an array of protocols which are principally categorized under the following heads[1]:-

- Flat-based routing - where all nodes are assigned equal roles or functionality.

- Cluster based (hierarchical-based) routing - nodes play different roles in the network.
- Location-based routing - sensor nodes' positions are exploited to route data in the network.
- QoS based – This set of protocols is based on quality of service requirement.
- Data-centric based – Desired information is sent based on query.
- Multipath Routing – nodes find alternative path to reach destination.

The former architecture of wireless sensor networks consisted of single central processing station that is connected to several sensor nodes [2]. With introduction of mobility, wireless network applications have redirected the focus towards distributed sensing nodes networks. The distributed nature of network has more energy consumption and is required to be placed closer since the exact sensor location is not known. Multiple sensing nodes further need to handle issues like information direction, information duplicity, network clogging, accuracy etc. [3]. Therefore among the existing protocols for data gathering, the cluster-based structure is considered by research community as an effective architecture for data-gathering in WSN. This paper aims at studying some representative cluster based routing protocols for Wireless Sensor Actor networks.

The remainder of this paper is organized as follows: In Section 2 some related work in the field of WSN is summarized. In Section 3, the architectural requirements of WSN versus the WSAN are shown. Further in section 4, some noteworthy clustering protocols in WSAN have been discussed. In Section 5, a comparison is made between the protocols, and this is the contribution work in the paper. Finally, in Section 6, analysis and performance is evaluated, in Section 7, a few inferences are made and the future work is discussed.

## 2 Related Work

Most of the research work has been done in WSN [4]. A approach proposed for data gathering in wireless sensor network works on energy efficiency of the sensor nodes. This has now become a principle design parameter in WSAN. A clustering algorithm known as LEACH (low energy adaptive cluster head algorithm) [5] is a localized algorithm in which probability of a node is decided on the basis of its energy. The node with a greater probability of action over its neighbors is elected as cluster head. LEACH is found to have a few errors in election of cluster heads. A new approach using fuzzy logic techniques [6] has been formulated to resolve the issues of CH election in LEACH. The fuzzy variables include energy of the nodes, concentration of sensor nodes at specific location and centrality positioning of a node with respect to the network. However this approach also has issues due to its centralized nature. An attempt to resolve these issues was undertaken in 2008[7] through a new, more efficient CH election mechanism by removing the involvement of base station. Another protocol, fully based on LEACH algorithm for clustering in WSN was proposed [8], where the CH burden is reduced by dividing its burden over few neighbor nodes that are called their associates. Because of these

associates, the energy consumption of the CH reduced and they stay for longer time in the network. However, the main drawback is that CH election is fully dependent on base station. The self selecting reliable path routing protocol (SRP) [9] for wireless sensor networks is based on SSR (self selecting routing) protocol which is a memory-less protocol. The major advantage of SRP protocol is that it converges traffic to a reliable and shortest route in the terms of number of hops in the route as well as link failure probability. This protocol also memorizes the traversed routes. The event to sink directing clustering (ESDC) protocol [10] for wireless sensor networks is a very efficient method of cluster formation. The clustering involves route discovery towards the sink which occurs every time an event is triggered. The ESDC protocol uses this technique to save energy of the nodes. An energy-efficient secure routing protocol for WSNs uses the location information of a stationary sink and builds sink-oriented grids to ensure the path availability from the source to the sink. Pre-shared secret keys, are used for authentication of every new joining node in the network thereby making an energy efficient, robust and a secure protocol. INSENS [11] is an Intrusion-tolerant routing protocol for WSNs uses redundant multipath routing for intrusion tolerance by bypassing malicious nodes. Two Tier Secure Routing Protocol for Heterogeneous Sensor Networks [12] uses a secure intra-cluster routing scheme using shortest path tree and a secure inter-cluster routing scheme in relay cells method. This two tier routing scheme makes it resistant to spoofed routing information, selective forwarding, sink-hole and wormhole attacks.

### 3 Clustering Protocols in WSN

Clustering is a formation of small group of sensor and actor nodes within the WSN, to collaborate and to respond to a sensed stimulus. Clustering protocols permit breaking down a larger network into smaller clusters which can be independently organized. This micro level organization capability within the network allows tweaking the various parameters gainfully towards a more effective network [13]. Further, these clusters can be scaled to a required geographical area and thus introduce a very desirable quality of scalability. Study of clustering WSN protocols therefore become a popular choice. The five protocols have been chosen as representatives of the sub-technology within the clustering protocols for discussions in the succeeding paragraphs. These are further evaluated on a qualitative matrix.

#### 3.1 Event-Driven Clustering Model [14]

The event-driven clustering model uses an event in the environment to be monitored as a starting point of its sensing, communication and response activity. Once the event is triggered, the event data is broken down into parts which are processed and transmitted to the actors. The actors, in turn gather, re-process, and eventually reconstruct the event data for performing actions. Thus the cluster formations have active participation of the sensors and the actors for handling an event. This protocol has comprehensively dealt with integrated networks of sensors and actors, and also

provided a unified framework for communication and coordination problems in WSANs. However, the protocol restricts itself to immobile actors that can act on a limited area defined by their action range. This issue has been handled in a revolutionary way by the Ripples protocol.

### **3.2 Ripples Protocol [15]**

Ripples protocol can best be explained using an analogy of throwing pebbles in a lake. Due to the pebble, ripples are formed on the surface of the lake. These ripples move as concentric circles away from the spot where the pebble is thrown. Similarly, in this algorithm, clusters are made in concentric circles which grow outwards. The actor in the network sends an 'Actor Association Announcement (AAA). The sensor nodes, which are within the range, hear this announcement and elect a leader amongst them. The leader then forwards the announcement to the next circle. Sensor nodes in that layer form suitable clusters based on their locations and then forward the announcement. This process continues until there can be no more clusters formed. In case of multiple actor nodes, they all throw their own pebbles. The individual ripples proceed until they meet each other. The Ripples protocol permits mobile actors and is optimally designed for a scalable, but homogenous, area of operation. To bring in heterogeneity, the RAT protocol has assigned self-sufficient clusters with a fixed area of responsibility. This also allows tasking different areas with different functionalities, thereby permitting heterogeneity.

### **3.3 Routing by Adaptive Targeting in Wireless Sensor/Actor Networks: RAT [16]**

RAT consists of two methods as Delay-constrained geographical-based routing (DC-GEO) and Integrated Pull/Push (IPP) coordination. DC-GEO sends the packets in greedy mode in such a way that delay constraint are met and energy of forwarding nodes is balanced. In IPP, actor nodes subscribe to specific events of their interest in the field and sensor nodes disseminate the event readings to subscribed actor for a time period of subscription life. So the actor nodes do not require sending a query every time they need event readings. The sensor nodes push the data as long as there is a subscribed actor interested for the event. RAT uses the mobility of actor nodes to form dynamic liability clusters that ensure a specific response time to emergencies.

In an attempt to alleviate data loss and real time reconfiguration of the network in an event of a network breach / error, the VDSPT protocol has come up with a novel method of reconfiguring clusters during the network operation.

### **3.4 Voronoi Diagram and Shortest Path Tree VDSPT[17]**

Similar to the LEACH protocol the VDSPT protocol configures clusters in every iteration. The protocol considers that the sensors and actors are already deployed. The process of network operation is divided into rounds, and each round consists of the two phases.

- **Cluster Set-up Phase:** Every actor is the cluster head of the clusters and passes the information of their residual energy to the sink. The sink in turn constructs the Actor weighted Voronoi diagram. The sink then sends the actor weighted Voronoi diagram to every actor. Every actor sends Voronoi diagram to the sensors in its cluster. Finally, each sensor sends acknowledgement information to the actor in its cluster with single-hop transmission and thus the entire network has the Actor weighted Voronoi diagram.
- **Data Transmission Phase:** After the cluster set-up phase, network nodes are divided into different clusters having one actor and a few sensors. The shortest path tree from sensors to actor in a cluster is calculated using the Dijkstra algorithm with the product of the maximum consumption energy of the two nodes in link and the consumption required to send data package as the weight. Then actor passes the SPT structure to all sensors in the cluster. Every sensor can transmit data package along the path of the shortest path tree in its cluster. The operation of VDSPT protocol is divided into rounds. In each round, the shortest path trees are configured and data package is transmitted from sensors to the cluster-head. The process is repeated until the expiry of the cluster lifetime that is when the residual energy of a sensor in the cluster is exhausted.

To concentrate the energy of the network to where it matters, the ADCP protocol has evolved a methodology to form the clusters based on the qualities of the actor.

### **3.5 HERO: A Hierarchical, Efficient and Reliable Routing Protocol for Wireless Sensor and Actor Networks[18]**

Most hierarchical routing and clustering protocols are designed to be efficient when the data is sent from the sensor nodes to their cluster-head but not when it is sent the opposite way. The HERO protocol uses a simple mechanism to form clusters in an efficient way by using metadata (stored in the nodes) and thus refrains from the energy hungry flooding or multicast methods. The metadata information is used to make the routing tables. A subtle difference from the Fuzzy logic based cluster election protocol [6] is that the protocol is not centralized in the cluster-head, and the nodes are in charge of joining in the clusters through their neighbors. This provides flexibility to any node in the cluster region to participate in cluster formation and do not have to wait for the cluster head to execute a discovery trigger. This protocol introduces concepts of “memory path” and “clue node”. The memory path helps the cluster head to answer an event received from a sensor node. The clue node is used by the cluster head to identify its area of interest. This results in multi-hop and fault-tolerant routing protocols that transport the data from sensor nodes to their cluster-head and vice-versa consuming minimum energy. This approach has allowed developers to establish the desired reliability level in a quantitative way between two nodes.

## **4 Comparison of the Discussed Protocols**

The comparison of the protocols (A) Hierarchical, efficient and reliable routing protocol for WSN (HERO) (B) Voronoi Diagram and Shortest Path Tree (VDSPT),



has been carried in form of a qualitative matrix. (C) Routing by Adaptive Targeting in Wireless Sensor/Actor Networks(RAT), (D) Ripples, (E) Event-driven clustering model, The qualities considered are the principle design characteristics which every protocol tries to optimize [18][17][14]. It may be noted that these characteristics are application specific and an optimum tradeoff between them would be based on the application in which they are used. These characteristics are listed below:-

- **Fault tolerance:** Ability of the algorithms to detect fault, repair and re-establish the network.
- **Nodes coordination:** A framework that provide efficient sensor-actor and actor-actor communication.
- **Energy Consumption:** Minimal use of the most premium resource of energy.
- **Mobility of Nodes:** Whether the sensors or the actors are mobile.
- **Response time:** Time lapse between the sensed stimulus and the response.
- **Scalability:** Ability to deploy network based on any geographical area.
- **Heterogeneity:** Ability to perform in an environment having different physical characteristics.

**Table 1.** Comparison of Protocols

Protocols	Fault tolerance	Nodes co-ordination	Energy consumption	Mobility of nodes	Response Time	Scalability	Heterogeneity
A	High	Med	Low	Med	high	high	Med
B	high	Med	low	Low	high	high	high
C	Med	Med	low	high	high	low	high
D	Low	Med	Med	Med	Low	Med	Low
E	Low	Med	low	Low	Low	Low	Low

## 5 Analysis and Performance Evaluation

### 5.1 Assumptions

For the simulation we have considered fixed number of sensor nodes for all the protocols. The protocols are evaluated with respect to simulation time.

### 5.2 Simulator

We have evaluated all the protocols in contending flow, data gathering scenario. We primarily focused on total throughput in bytes and mean packet delay in seconds. TOSSIM (TinyOS-2.1.1 Simulator) is used to evaluate the protocols. The plots of total throughput and mean packet delay are shown in Fig-1 and Fig-2 respectively. The simulation for total throughput has been carried for 20 nodes for a range of

simulation time. The time ranges between 50 to 400ms. It may be seen from the graph (Fig-1) that the plots for protocol A and C are clearly distinguishable. The other three protocols have similar total throughput and hence the plots are seen as overlapping. The Fig-2 which is a plot of mean packet delay over time, the response of all the five protocols are clearly distinguishable indicating a wide range of mean packet delay over time.

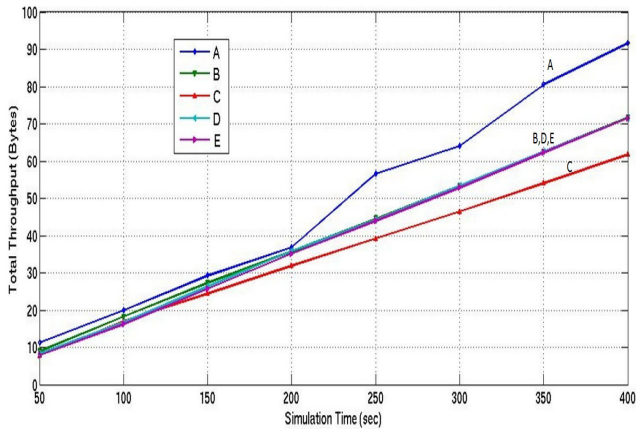


Fig. 1. Comparison of protocols - Total throughput vs simulation time

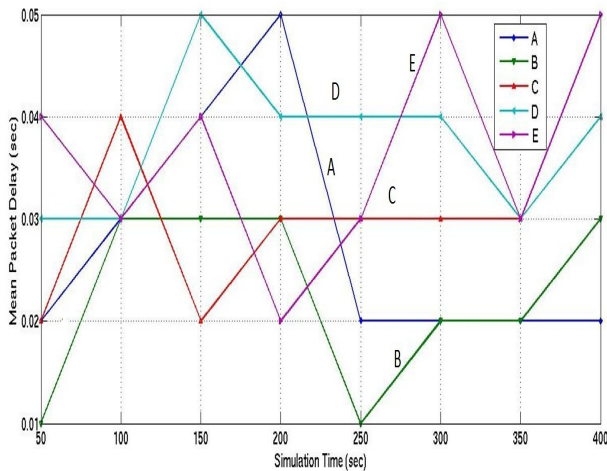


Fig. 2. Comparison of protocols – Mean packet delay vs simulation time

The main factor to judge a network is to have higher throughput with lower delay. From the simulation results it is seen that the protocol (A) Hierarchical, efficient and reliable routing protocol for WSN (HERO) is having better throughput with respect to time as compared to other protocols. Even though protocol (A) shows more delay is

the initial stage, but over the time the delay is less as compared to the other protocols, and is almost stable over the simulation time. The protocol B shows a lower delay compared to remaining three protocols with reasonable throughput.

## 6 Conclusion and Way Ahead

Cluster based (hierarchical-based) routing has emerged as the most optimal method for wireless sensor networks. This paper has surveyed some of the noteworthy cluster based protocols for WSN. A comparison study has revealed that no single protocol is self sufficient in achieving all the desirable characteristics of the network. Further, on evaluating the protocols for performance in terms of total throughput and mean packet delay vs simulation time, on a standard simulator TOSSIM, the Hierarchical, efficient and reliable routing protocol for WSN (HERO) has emerged as the most optimized amongst all. Further study would pave a path towards designing and proposing a cluster protocol based on the strengths gathered from the above study.

## References

1. Al-Karaki, J.N., Kamal, A.E.: Routing techniques in wireless sensor networks: a survey. *Wireless Communications*. IEEE 11(6) (2004)
2. Bonivento, A., Fischione, C., Sangiovanni-Vincentelli, A., Graziosi, F., Santucci, F.: Seran: A semi random protocol solution for cluster wireless sensor networks. In: *Proceedings of MASS, Washington D.C.* (November 2005)
3. Dasgupta, K., Kalpakis, K., Namjoshi, P.: An efficient clustering-based heuristic for data gathering and aggregation in sensor networks. In: *Proceedings of the IEEE Wireless Communications and Networking Conference, WCNC 2003* (2003)
4. Akyildiz, I.F., Su, W., Sankarasubramanian, Y., Cayirci, E.: *Wireless Sensor networks: A survey*. *Computer Networks* 38 (2002)
5. Heintzelman, W.R., Chandarkasan, A., Balakrishnan, H.: Energy Efficient Communication Protocol for Wireless Microsensor Networks. In: *Proceedings of the 33rd Hawaii International Conference on System Sciences* (2000)
6. Gupta, D.R., Sampalli, S.: Cluster-head Election using Fuzzy Logic for Wireless Sensor Networks. In: *Communication Networks and Services Research Conference* (May 2005)
7. Kim, J.-M., Park, S.-H., Han, Y.-J., Chung, T.-M.: CHEF: Cluster Head Election mechanism using Fuzzy logic in Wireless Sensor Networks, February 17-20 (2008)
8. Hussain, S., Matin, A.W.: Base Station Assisted Hierarchical Cluster-based Routing. In: *IEEE International Conference on WSN 0-7695-26 29-31* (July 2006)
9. Bhibbitt, T.A., Morrel, C., Zymanski, B.K., Branch, J.W.: Self-selecting reliable paths for Wireless Sensor Network routing. Elsevier (May 9, 2008)
10. Bereketli, A., Shah, G.A., Akan, O.B.: Event -to-Sink Directed Clustering in wireless Sensor Networks. In: *Proc. IEEE WCNC 2009, Budapest, Hungary* April (2009)
11. Deng, J., Han, R., Mishra, S.: INSENS: Intrusion-Tolerant Routing in Wireless Sensor Networks, University of Colorado, Department of Computer Science Technical Report CU-CS-939-02
12. Protocol for Heterogeneous Sensor Networks. *IEEE Transactions on Wireless Communications* 6(9) (September 2007)

13. Khan+, M.A., Shah, G.A., Ahsan, M., Sher, M.: An Efficient and Reliable Clustering Algorithm for Wireless Sensor Actor Networks (WSANs). IEEE (2010)
14. Melodia, T., Pompili, D., Gungor, V.C., Akyildiz, I.F.: A Distributed Coordination Framework for Wireless Sensor and Actor Networks. In: *MobiHoc 2005*, May 25-27. ACM, Urbana-Champaign (2005)
15. Trivedi, N.: A Message-Efficient, Distributed Clustering Algorithm for Wireless Sensor and Actor Networks. In: *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems* (2006)
16. Bozyigit, M., Aksoy, D.: RAT: Routing by Adaptive Targeting in Wireless Sensor/Actor Networks. In: *2nd International Conference on Communication Systems Software and Middleware, COMSWARE 2007*, January 7-12 (2007)
17. Dai, Z., Hubei, W., Wang, B., Li, Z., Yin, A.: VDSPT: A Sensor-Actor Coordination Protocol for Wireless Sensor and Actor Network Based on Voronoi Diagram and Shortest Path Tree. IEEE (2009)
18. Cañete, E., Manuel, D., Llopis, L., Rubio, B.: HERO:A hierarchical,efficient and reliable routing protocol for wireless sensor and actor networks. *Science Direct* (April 2012)

# Privacy Based Optimal Routing in Wireless Mesh Networks

T.M. Navamani<sup>1</sup> and P. Yogesh<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering,  
Easwari Engineering College, Chennai-89

<sup>2</sup> Department of Information Science and Technology,  
Anna University, Chennai-25

navsmi2001@yahoo.com, yogesh@annauniv.edu

**Abstract.** User privacy and security have been primary concerns in Wireless Mesh Networks (WMNs) due to their peer-to-peer network topology, shared wireless medium and highly dynamic environment. Nowadays, anonymity has received increasing attention due to the users' awareness of their privacy. Anonymity provides protection for users to enjoy network services without being traced. The multi-hop architecture of WMN lays a foundation for most of the routing attacks like packet dropping and packet misdirecting attacks. In this paper, we propose Privacy Based Optimal Routing (PBOR) which provides security and user anonymity while maintaining communication efficiency in WMN. To address privacy and anonymity, security card based routing scheme is implemented. In order to provide security against packet dropping and misdirecting attacks, strong protection is needed and it can be achieved by enhancing the routing metrics during route discovery. Although there are several schemes are proposed to defend against these attacks, there occurs more routing overhead and delay. To overcome these issues, we propose privacy based optimal routing protocol to provide an optimal path for data transmission. Our implementation results show that the proposed scheme increases the overall performance of the network substantially.

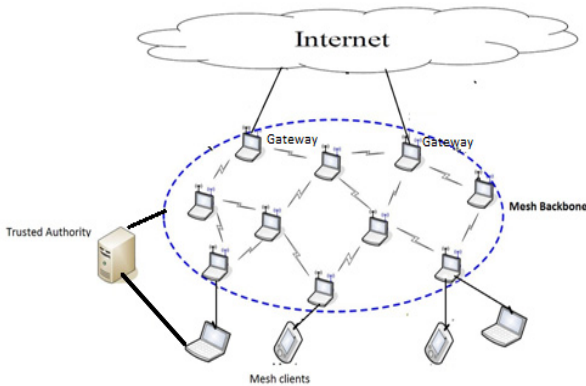
**Keywords:** Wireless Mesh Networks (WMN), Security, Privacy, Anonymity, Routing.

## 1 Introduction

Wireless mesh networking has emerged as a promising concept to meet the challenges in next-generation wireless networks such as providing flexible, adaptive, and reconfigurable architecture while offering cost-effective solutions to service providers. WMNs are multi-hop wireless networks formed by mesh routers (which form wireless mesh backbone) and mesh clients. Mesh routers are typically stationary and do not have power constraints. However, the clients are mobile and energy-constrained. Some mesh routers are designated as gate-way routers which are connected to the internet through a wired backbone. A gateway router provides access to conventional clients and interconnects ad hoc, sensor, cellular, and other networks

to the Internet. Wireless Mesh Networks are of three types. They are, i) Infrastructure based WMN ii) Hybrid WMN iii) Client WMN. Fig. 1 shows an architecture of infrastructure based wireless mesh network.

The broadcast nature of transmission and the dependency on the intermediate nodes for multi-hop communications lead to several security vulnerabilities in WMNs. The attacks can be external as well as internal in nature. External attacks are launched by intruders who are not authorized users of the network. For example, an intruding node may eavesdrop on the packets and replay those packets at a later point of time to gain access to the network resources. On the other hand, the internal attacks are launched by the nodes that are part of the WMN. One example of such attack is an intermediate node dropping packets which it was supposed to forward. To prevent external attacks in vulnerable networks such as WMNs, strong authentication and access control mechanisms should be in place for practical deployment and use of WMNs. As in other wireless networks, a weak authentication scheme can easily be compromised due to several reasons such as distributed network architecture, the broadcast nature of the wireless medium, and dynamic network topology.



**Fig. 1.** Infrastructure based Wireless Mesh Network

Many communications in WMNs contain various kinds of sensitive user information like personal identities, activities, location information, movement patterns, financial information, transaction profiles, and social/business connections. Hence, securing these communications is of paramount importance in WMNs.

Keeping these problems in mind, this paper describes several schemes for WMNs that addresses privacy and security. Here we are extending our previous work, Secured Identity Based Routing (SIBR) which was discussed in [1] to enhance the routing path with additional routing metrics. To address privacy and anonymity in WMNs, SIBR scheme is implemented in our proposed protocol. Routing between each node is constructed based on anonymous security card approach. Each mesh client has to register with nearest mesh router in order to get access to the network. Mesh client then registers it with the nearest mesh router. The registered identities are

protected from adversaries and then used for route discovery within the mesh backbone. Since the routing in mesh backbone is known to mesh routers, we make use of pair wise secret keys along with security cards to keep mesh clients anonymous from mesh routers. To protect the network against packet dropping and misdirecting attacks such as wormhole, black hole, gray hole and jellyfish attacks, we introduce Neighbor Acknowledgement Mechanism (NAM). Our proposed protocol also provides an optimal path for data transmission by computing bandwidth of the path using cross layer information exchange and minimizes the routing overhead by reducing the broadcasting of control packets. Specifically, our major contribution in this paper includes 1) Extending our previous work Secured Identity Based Routing (SIBR) scheme [1] to provide anonymity 2) Neighbor Acknowledgement Mechanism (NAM) is introduced to prevent against packet dropping and misdirecting attacks, which also helps to reduce routing overhead by reducing broadcasting of control packets 3) Finding optimal path for efficient data transmission by providing better bandwidth path using cross layer information exchange. By implementing these schemes we can achieve privacy and security in wireless mesh networks.

The rest of this paper is structured as follows. In Section 2, we discuss about related works on privacy and security for wireless mesh networks. In section 3, we describe about our proposed protocol. Security and privacy analysis are presented in section 4. In Section 5, Implementation and performance analysis are discussed. Finally in section 6, we conclude the paper with future enhancements.

## 2 Related Works

Since security and privacy are two extremely important issues in any communication network, researchers have worked on these two areas extensively. However, as compared to MANETs and Wireless Sensor Networks (WSNs), WMNs have received very little attention in this regard. This section briefly discusses some of the existing mechanisms for ensuring security and privacy in WMNs.

Zinguo Wan et al. [2] proposed a scheme called Anonymous user communication for privacy protection in wireless metropolitan mesh networks. This paper proposed two solutions for providing privacy and security in wireless mesh networks. One is based on group signature scheme, and the other is based on pair wise secret keys to provide strong anonymity and unlinkability in mesh networks. Hui Lin et al. [3] designed a privacy aware secure hybrid wireless mesh protocol for IEEE 802.11s wireless mesh networks. Combination of a new dynamic reputation mechanism based on subject logic, uncertainty scheme and multi-level security technology are implemented to provide privacy and security for WMNs. Zhiguo Wan et al. [4] proposed a new protocol for providing anonymity, unlinkability and unobservability to the nodes and packets in the mobile ad hoc networks. Xiangfang Li et al. [5] proposed distributed security architecture to manage the mesh clients and a secure routing scheme that preserves the privacy of the end-users. Ben Salem and Hubaux [6] discussed specifics of WMNs and identified fundamental network operations that need to be secured. Siddiqui and Hong [7] surveyed the threats and vulnerabilities

faced by WMNs and also identified a number of security goals. Capkun *et al.* [8] have given a privacy-preserving scheme for the so-called hybrid ad hoc networks, which are actually WMNs. This scheme provides anonymity and location privacy for mobile nodes.

Shafiullah Khan *et al.* [9] designed a Secure Routing Protocol for IEEE 802.11 Infrastructure Based Wireless Mesh Networks. In this scheme, a new routing metric called Unreliable Value (UV) is proposed, which is capable of searching the shortest secure path by computing UV of the neighbors by implementing two-hop passive acknowledgment scheme. Shafiullah Khan *et al.* [10] introduced a secure route selection scheme in wireless mesh networks. This scheme is based on two hop passive acknowledgement mechanism to prevent the network from packet dropping and packet misdirecting attacks. In [11], Divya Bansal *et al.* introduced a Secure Routing Protocol for Hybrid Wireless Mesh Network which uses cryptographic extensions to provide authenticity and integrity of HWMP routing messages and prevents unauthorized manipulation of mutable fields in the routing information elements.

Zhang *et al.* in [12] have come up with attack resilient security architecture for multi hop wireless mesh networks. They have modeled WMN architecture as credit card based e-commerce system and showed that a mesh client need not to be bound to a specific WMN operator, can get ubiquitous network access by a universal pass issued by a third-party broker. Liu *et al.* [13], a neighborhood authentication protocol has been proposed that allows neighboring nodes to authenticate each other without revealing their identities. As destination ID needs to reveal for route discovery, only conditional anonymity is achieved for destination. Other general privacy-aware authentication schemes are described in [14], [15], [16].

After reviewing the previous work, we propose a new scheme for optimal routing with better privacy and security in wireless mesh networks. This can be achieved by implementing Secure Identity Based Routing, Neighbor Acknowledgement Mechanism (NAM) and Cross layer information exchange mechanisms.

### 3 Privacy Based Optimal Routing (PBOR)

In this section, we propose Privacy Based Optimal Routing (PBOR) for wireless mesh networks. The following subsections describe functions of the proposed scheme and specifications taken in detail.

#### 3.1 Network Model

We consider an infrastructure based wireless mesh network which consists of mesh routers and mesh clients as shown in Fig. 1. An Infrastructure WMN comprises of a set of Mesh Routers (MRs) and Mesh Clients (MCs), where MRs are connected to the Internet backbone through the Internet gateways (IGWs). Mesh clients can directly connect to the nearest mesh router. We further assume that all traffic from source client node to destination client node passes through the routers present in the mesh backbone. To enjoy WMN access, each network user/client has to first register with the Trusted Authority (TA) which in turn issues a security card to the users. In



addition, TA also provides necessary revocation capabilities and cryptographic means to network users to protect their communication against eavesdropping, altering and also other sophisticated attacks aimed to compromise privacy. We also consider a scenario that a global passive attacker or an active attacker which can able to eavesdrop all network communications, which can compromise and control a small number of users and mesh routers subject to his choices.

### 3.2 Notations

We denote Trusted authority and WMN Operator as  $T_i$  and  $O_i$  respectively. We use  $MC_{i,j}$  to indicate the unique identifier of client  $j$  registered in  $T_i$ .  $MR_{i,j}$  refers to the unique identifier of mesh router  $j$  of  $O_i$ . We indicate the security card for  $MC_{i,j}$  as  $MC_{i,j}$ -card and  $K_{MC_{i,j}}$  is a card-based key both are issued by  $T_i$  to  $MC_{i,j}$ . Likewise,  $MR_{i,j}$ -card and  $K_{MR_{i,j}}$  are used to denote the router card and the card-key respectively.

### 3.3 Neighbor Acknowledgement Mechanism (NAM)

Each node in the network maintains a temporary routing table which consists of node's one hop neighbors and two hop neighbors. Each node calculates Distrust Value (DV) for its one hop neighbors by forwarding probe packets to its one hop neighbors and getting acknowledgement from its two hop neighbors. DV is calculated by using the following equation.

$$DV = \text{Packets forwarded to one hop neighbors} - \text{Passive acknowledgement from two hop neighbors.}$$

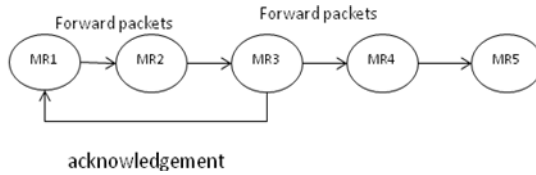
### 3.4 Cross Layer Information Exchange

The exchange of cross layer information is to use many parameters from different layers for overall optimization of protocols across the communication stack and can be used to tune various aspects of the wireless communication. Here it is used to get the value of link bandwidth from MAC layer.

### 3.5 Routing Metrics

Our proposed scheme intends to enhance the routing metrics to defend against packet dropping and misdirecting attacks and also to select an optimal path in WMNs. The routing metrics such as Distrust Value (DV), Bandwidth metric (BM) in addition to hop count are implemented during Route discovery. Route information in the routing table varies depending on the routing algorithm and the metric used. There may be multiple paths for a single destination; the selection of the best path is done on the basis of the routing metric. Each node in the network calculates DV of its one hop neighbors using Neighbor Acknowledgement Mechanism (NAM) and updates it in its temporary routing table. Higher values of DV mean that the node is more unreliable. At the same time, each node computes bandwidth of link (BM) to neighbor node by getting information from MAC layer using cross layer information exchange. The

above metrics are computed and updated at each node present in the network. The Route discovery process uses these metrics for selecting an optimal path from source to destination. The DV is computed on the basis of Neighbor Acknowledgement Mechanism as shown in Fig. 2. Source MR1 sends packets to MR2, but receives passive acknowledgement from MR3 to validate the reliability of MR2.



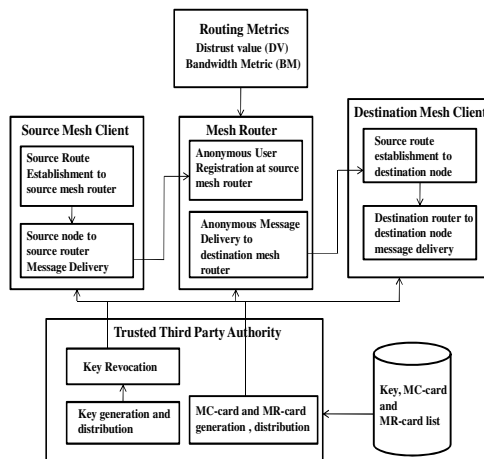
**Fig. 2.** Neighbor Acknowledgement Mechanism

$$DV_{MR2} = \text{No. of packets forward to MR2 from MR1} - \text{No. of Passive Acknowledgements from MR3 to MR1}$$

The more DV, the less secure is the node. The Neighbor Acknowledgement Mechanism is resilient to packet dropping and misdirecting attacks like black hole, gray hole and wormhole attacks.

### 3.6 Route Discovery

The route discovery process of the proposed protocol is discussed in this section. In wireless mesh networks, communication occurs mainly through mesh clients, mesh routers and gateways. Fig. 3 shows the functional architecture of our proposed scheme.



**Fig. 3.** Functional Architecture

An anonymous route between a mesh client and its nearest mesh router is established by registering the client to the mesh router. Initially each mesh router that needs to access the network sends request to the nearest mesh router that reaches under the protection of key pairs. Then the mesh router registers the node and puts it into its user list in the second step. This information is exchanged among mesh routers so that every mesh router knows how to reach a specific node. Next, the mesh router sends a reply to the source node, and the route is constructed when the reply successfully reaches the source node.

Now each node presents in the network including all mesh routers and mesh clients calculates Distrust Value (DV) for its one hop neighbor nodes and store it in its temporary routing table. Higher the distrust value, then the node is more unreliable and this node can be a malicious node. At the same time, each node presents in the network computes bandwidth of the link (BM) between its neighbor nodes using cross layer information exchange. Each node stores these two values in its temporary routing table. Route from source mesh client to destination mesh client is established by Route Request (RREQ) and Route Reply (RREP) process. The Route Request and Route Reply process of the proposed protocol are discussed in detail.

**Route Request (RREQ):** Mesh Router MR periodically broadcasts a beacon as shown in (1) via the single-hop downlink to announce its presence. The beacon should include MR-Card<sub>i,j</sub>, Cert<sub>O<sub>i</sub></sub>, and a fresh timestamp  $t_1$  signed with its MR-Cardkey  $K_{MR_{i,j}}$ . The beacon can be received by all mesh clients in router R's coverage. Cert<sub>O<sub>i</sub></sub> specifies the trusted WMN Operator's identity. MR sends the following beacon message to its neighbors:

$$MR_{i,j} \rightarrow: MR\text{-Card}_{i,j}, Cert_{O_i}, K_{MR_{i,j}}(t_1, \text{Other Info}) \quad (1)$$

Where  $MR\text{-Card}_{i,j} := (MR_{i,j}, \text{expiration-time})$  and  $K_{MR_{i,j}} = \alpha^{O_i} H^{O_i}(MR\text{-Card}_{i,j})$ . Here,  $\alpha^{O_i}$  is operator  $O_i$ 's domain secret and  $H^{O_i}$  is the hash function specified. The mesh client upon receipt of the beacon message, finds route to that router through the intermediate nodes. MC will broadcast the following undistinguishable RREQ packet as shown in (2), to MR by encrypting the MC-Card<sub>i,j</sub>, Seqno, SIG and Bandwidth Metric (BM) which is initially set with value zero, within its neighborhood:

$$E_{pk}(RREQ, MC\text{-Card}_{i,j}, MR, \text{seqno}, SIG, \text{pad}, BM) \quad (2)$$

Where  $MC\text{-Card}_{i,j} := (MC_{i,j}, \text{expiry-time})$ . Here, expiry-time specifies the expiry time of MC-Card<sub>i,j</sub> before which MC<sub>i,j</sub> has to renew itself. Thus Route Request from Mesh Client (MC) reaches Mesh Router (MR) finally. Now the mesh router has to forward it to other nodes. Before broadcasting the RREQ, it checks its temporary routing table for the Distrust Value (DV) of its one hop neighbors. If any of its one hop neighbors has distrust value greater than threshold value, then it simply drops RREQ. It will not broadcast it to its neighbors. This reduces routing overhead at the time of route discovery process which also considerably reduces route discovery time. On receiving RREQ and after successful verification, each intermediate mesh router updates BM

with the value already stored in its routing table. It may be noted that MR may receive more than one route request that originates from the same source having same sequence number, but MR replies to the first arriving message and drop others.

**Node Registration:** Upon successful receipt and verification of a routing request message, the mesh router proceeds to register the requesting node. Let S be the source node requesting registration, then MR knows that S is currently within its service coverage and puts S into its current user list but only for a predetermined period of time. S has to periodically reregister itself to MR to maintain the active status otherwise S's registration automatically expires in MR's user list. The real identity of a network user is not disclosed to the current service mesh router because only the user's card id is used for registering anonymously at mesh router.

**Route Reply (RREP):** Route Reply messages (RREPs) are unicast and it is sent from destination mesh client to the source mesh client through the path selected as given in (3). Destination mesh client receives several RREQs and it selects the best path according to the value of BM present in the RREQ message. Destination node selects lowest available bandwidth of the route from received RREQs and sends RREP through this path. Due to this optimal path selection, throughput has been increased and delay is reduced. Once the reply message reached the source node, message is transmitted through the established route. MR performs the padding operation as appropriate and replies the following undistinguishable Route Reply packet to the mesh client MC:

$$MR, E_{pk}(RREP_R, MR, MC-Card_{i,j}, seqno, pad) \quad (3)$$

Now the optimal route is established between source nodes to destination node by implementing cross layer based routing metrics. The data that is transferred from source to destination in this route is secure from packet dropping and misdirecting attacks and also privacy of the nodes is preserved to other nodes present in the network.

## 4 Security and Privacy Analysis

### A. Security Analysis

The security of the proposed protocol is based on keys constructed during card generation and key distribution phase. The key exchange scheme makes use of card based keys to provide message authentication while keeping the users' identities anonymous. In the following, we will discuss the security features of our proposed protocol.

- **Packet dropping and misdirecting attacks:** PBOR is resistant against packet dropping and misdirecting attacks like worm hole, black hole, gray hole and jellyfish attacks. All these attacks can be mitigated using PBOR protocol by

keeping in view of two hop neighbor's information and next hop passive acknowledgement. Two hop neighbor information kept in temporary routing table of each node helps to avoid wormhole and node isolation attacks and next hop passive acknowledgement verifies intermediate nodes and also helps to mitigate black hole, gray hole and jellyfish attacks.

- **Access Control:** Our protocol ensures that only legitimate users can gain access to mesh networks. To be able to access the mesh network, a mobile user has to obtain a security card along with the card key and successfully register with the mesh router. An adversary cannot easily forge the valid signature of TA of his choice which would not be authenticated by the mesh router.
- **Preventing Route Disruption:** Route disruption attack is caused by the malicious behavior of a node through modification of a mutable field and dropping routing information elements. It may be noted that, in our scheme, only authenticated nodes can participate in the route discovery phase. Moreover, routing information elements are authenticated and verified per hop. So, it is not possible to launch a route disruption attack.

## B. Privacy Analysis

Here we will analyze how the identity information of the communicating participants are protected by our proposed protocol.

- **Identity preservation:** Most people would like to remain anonymous while roaming in WMNs for privacy reasons. In this proposed scheme, session key is generated between each node with its neighbors. By using this secret session key, route discovery process is carried out between source and destination. Since pseudonyms and nonces are used in this route discovery process, the nodes present in the network cannot able to reveal sender's identity.
- **Against the Global Eavesdropper:** In the proposed protocol, all types of packets such as route request, route reply and normal message packets are indistinguishable from each other to the outsider nodes. This is because these packets are not carrying the real identity information of the source or destination nodes in plain text and they are of more or less same format. Also the security card based approach reveals no information about the network nodes. Hence it is impossible for an eavesdropper to obtain the source or destination node identity information of any communication session.
- **Against Mesh Routers and the WMN Operator:** In the proposed protocol, mesh routers cannot have access to the real identity information of the network users due to anonymous user registration. Each mesh router on the routing path also can not able to obtain the identity information of the two communicating nodes. Each node can get access to the network by receiving card based key as explained in the third section. The WMN operator also knows that the session is going on between two anonymous communicating entities and no other identity information is known to him. In summary, the proposed protocol satisfies the privacy requirements.

## 5 Implementation and Analysis

The proposed privacy based optimal routing (PBOR) has been implemented and analyzed using the network simulator NS2. The proposed scheme utilizes a network topology comprising of 25 wireless nodes that provides communication to other networks. We evaluate the performance of PBOR in terms of packet delivery ratio, throughput and average end to end delay and comparing it with the existing privacy preserved secure routing protocol SIBR which was proposed by us in [1].

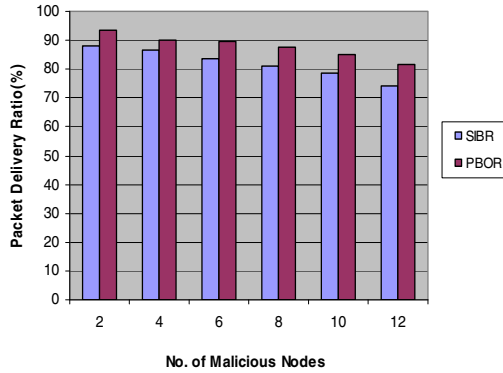


Fig. 4. Packet Delivery Ratio

Fig. 4 shows the comparison of packet delivery ratio at different number of malicious nodes with SIBR and PBOR protocol. Although both the protocols implement the same privacy preserving mechanisms, PBOR shows better performance than SIBR due to routing metrics enhancement and the selected routing path is more optimal. Since the trusted nodes are forwarding the packets, there will be less chance for dropping or loss of packets. PBOR is secure privacy based optimal routing protocol and every node follows the alternate path in case of the presence of intruders. Hence most of the packets are not dropped and the throughput is increased. When the number of malicious nodes is less, both the protocols show more or less the same performance with little difference. When the number of malicious nodes is more, PBOR gives better performance compared to SIBR as shown in the Fig. 4.

Fig. 5 shows the comparison of throughput at different simulation times. In addition to privacy preserving mechanisms, PBOR implements cross layer information exchange to get the bandwidth metric and neighbour hop acknowledgement mechanism to defend against packet dropping attacks. Hence the proposed protocol shows better performance compared to SIBR when the simulation time is increased. Fig. 6 shows the average end to end delay ratio of PBOR and comparison with SIBR. In PBOR, end to end delay is reduced considerably since broadcasting of control packets are minimized during route discovery process. This minimizes the overhead of control packets during route discovery. Hence end to end delay is reduced when compared to SIBR.

Thus the overall performance evaluation of the proposed protocol shows that it provides better packet deliver ratio and throughput and also lowering the end to end delay than existing SIBR protocol. Hence the performance of PBOR is more effective than the existing SIBR protocol.

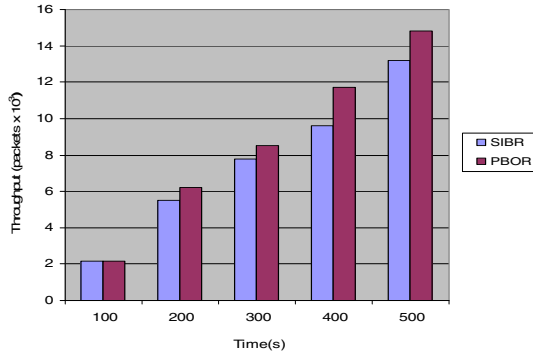


Fig. 5. Throughput

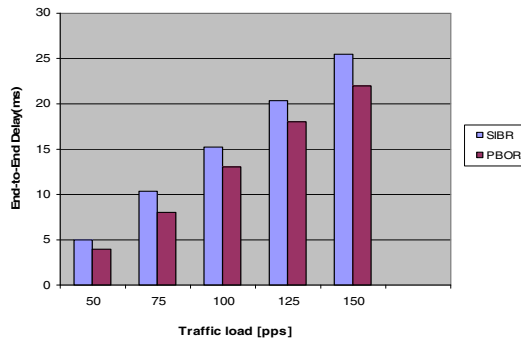


Fig. 6. End to End delay

## 6 Conclusion and Future Work

In this paper, we have presented the design and evaluation of PBOR, a new privacy based optimal routing protocol to provide security, anonymity, and authentication. The privacy and security analysis demonstrate that proposed PBOR protocol provides strong privacy and security to the network users by providing secured card based routing and routing metric enhancements. The proposed protocol provides optimal performance by addressing security, privacy issues and also the network performance. By implementing neighbor acknowledgement mechanism during route discovery, packet dropping and misdirecting attacks are prevented. The improvement in network performance is achieved by implementing cross layer information exchange such as

bandwidth and minimizing the broadcasting of control packets. The simulation results show that PBOR has better performance in terms of packet delivery ratio, throughput and end to end delay than existing SIBR protocol. In future, it is planned to enhance the proposed protocol such that it should provide strong protection against Denial of Service attacks and also to provide optimal performance.

## References

1. Ramya, R., Navamani, T.M., Yogesh, P.: Secured Identity Based Routing and privacy Preservation in Wireless Mesh Networks. In: ICRTIT, vol. 26, pp. 521–526 (2011)
2. Wan, Z., Zhu, K.R.B., Preneel, B., Gu, M.: Anonymous User Communication for Privacy Protection in Wireless Metropolitan Mesh Networks. *IEEE Transactions on Vehicular Technology* 59, 519–532 (2010)
3. Lin, H., Ma, J., Hu, J., Yang, K.: PA-SHWMP: a privacy-aware secure hybrid wireless mesh protocol for IEEE 802.11s wireless mesh networks. *EURASIP Journal on Wireless Communication and Networking* 6 (2012)
4. Wan, Z., Ren, K., Gu, M.: USOR: An Unobservable Secure On-Demand Routing Protocol for Mobile Ad Hoc Networks. *IEEE Transactions on Wireless Communications* 11, 1922–1932 (2012)
5. Li, X., Qian, L., Kamto, J.: Secure Anonymous Routing in Wireless Mesh Networks. In: *IEEE International Conference on E-Business and Information System Security*, vol. 64, pp. 1–5 (2009)
6. Ben Salem, N., Hubaux, J.P.: Securing Wireless Mesh Networks. *IEEE Wireless Communications* 13(2), 50–55 (2006)
7. Siddiqui, M., Hong, C.: Security Issues in Wireless Mesh Networks. In: *IEEE International Conference on Multimedia and Ubiquitous Engineering*, pp. 717–722 (April 2007)
8. Capkun, S., Hubaux, J., Jakobsson, M.: Secure and Privacy preserving communication in hybrid ad hoc networks. In: *Swiss Fed. Inst. Technol.,-DI-ICA, Lausanne, Switzerland* (2004)
9. Khan, S., Loo, K.-K., Mast, N., Naeem, T.: SRPM: Secure Routing Protocol for IEEE 802.11 Infrastructure Based Wireless Mesh Networks. *Springer Journal* 18, 190–209 (2010)
10. Khan, S., Alrajeh, N.A., Loo, K.-K.: Secure route selection in wireless mesh networks. *Elsevier* 56(2), 491–503 (2011)
11. Bansal, D.: Sofat: Secure Routing Protocol for Hybrid Wireless Mesh Network (HWMN). In: *International Conference on Computer and Communication Technology (ICCCCT)*, vol. 25, pp. 837–843 (2010)
12. Zhang, Y., Fang, Y.: ARSA: An Attack Resilient Security Architecture for Multihop Wireless Mesh Networks. *IEEE Journal on Selected Areas in Communications* 24(10) (October 2006)
13. Zhang, Y., Liu, W., Luo, W.: Anonymous Communication in Mobile Ad Hoc Networks. In: *Proceedings of INFOCOM*, vol. 3, pp. 1940–1951 (2005)
14. Ren, K., Lou, W., Kim, K., Deng, R.: A Novel Privacy Preserving Authentication and Access Control Scheme for Pervasive Computing Environment. *IEEE Transactions on Vehicular Technology* 55(4), 1373–1384 (2006)



15. Ren, K., Lou, W.: Privacy-Enhanced, Attack-Resilient Access Control in Pervasive Computing Environments with Optional Context Authentication Capability. *ACM Mobile Networks and Applications (MONET) (Special Issue on Wireless Broadband Access)* 12, 79–92 (2007)
16. Zhang, Y., Ren, K.: On Address Privacy in Mobile Ad Hoc Networks. *ACM/Springer Mobile Networks and Applications (MONET)* 14(2), 188–197 (2009)

# Security Issues and Its Counter Measures in Mobile Ad Hoc Networks

Pritee Parwekar and Sparsh Arora

Department of Computer Science,  
Jaypee Institute of Information Technology,  
Noida, India

pritee.parwekar@jiit.ac.in, sparsh7054@gmail.com

**Abstract.** The dynamic topology of MANETS has made them vulnerable to security threats. This research work focuses on the Security Issues in Mobile Ad-Hoc Network, the different types of attacks in MANETS, the existing security mechanisms and the solutions approaches in the security of different applications of MANETS. The prime objective of the work is to provide a generic solution approach which can be implemented before initiating the working on any MANET security algorithm, so as to provide better results and security. NS-2 simulator is used to make the network topology on MANETS and analyse them using the trace files obtained when using different routing protocols like AODV, DSDV. An algorithm is proposed to detect the popular nodes, the critical nodes, the potentially malicious nodes in a network, for every rounds of time. The cluster based approach is used by which all the details of parameters evaluated in terms of critical nodes, potentially malicious nodes and popular and non popular nodes can be send to the cluster heads for performing further action using the other existing security algorithms.

**Keywords:** MANETS, AODV, DSDV, Network Security.

## 1 Introduction

Ad hoc is a Latin word meaning “for the purpose”. Ad – hoc networks is a field of multiple networks which interact and is of interest to most software developers and programmers [1]. Ad hoc networks are used when there is lack of infrastructure to develop a proper network or wired link. Mobile ad hoc networks are used on the platforms when the ad hoc nodes are mobile devices. These have a strong potential for applications in the areas where there are no base station to communicate among mobile nodes like earthquake prone areas or flood affected areas, military warships. Mobile Ad-hoc Network (MANETS) is an appealing technology that has attracted lots of research efforts over past years. The concept of wireless, without the boundaries of structure, dynamic networks is very attractive and yet there are some major flaws that have been preventing commercial viability [2]. Security is one of these main barriers MANETS are known to be particularly vulnerable to security attack. . The security

challenges in the MANETS arise due to its dynamic topology, vulnerable wireless link and nomadic environment.

## **2 Background Study**

There are some of the existing fields on which currently there are developers working and international communities are doing on-going research. Existing solution approaches are more application specific and are addressing a particular security issue of MANETS. The existing approaches use concepts of Secure Multiparty communication, EODMRP[10], Pro-active and reactive approaches and Ariadne algorithm[9], symmetric encryption techniques[11]. These solutions are mostly application and specific security issue dependent. The NIST information technology laboratory under its department of Computer Security Division is currently performing a number of studies in security issues in pervasive networking they have done a number of research projects and also provides software MLAB for developers to perform security implementation test without having a need of actual mobile node and platforms.

## **3 Security in MANETS**

### **3.1 MANETS**

MANETS are vulnerable to attacks both internal and external and they require enhancement in the protocols on several security requirements. MANETS involve dynamic infrastructure less environment and no centralized system of control hence these challenges along with scalability add on to the likeliness of the network attacks. Attacks and security requirements of MANETS are somehow little different as attacks on normal ad hoc network or any other network. Every security algorithm can be selected on considering the balance of trade off between security level, implementation complexity and its energy efficiency. Some of the constraints with respect to MANETS which are to be considered while designing a security algorithm for MANETS are slow computation powers due to a peanut CPU, Limited energy resource (battery life), limited memory, relatively high latency especially when performing periodic turn on-turn off for devices, power conservation, dynamic topology which has large number of weak links with no central node to monitor traffic, link breakages, no predefined Boundary. Further the MANETS are scalable and therefore it is difficult to predict the number of nodes in future and there is no predefined security Architecture.

### **3.2 Mobile Ad Hoc Networks: Line of Attacks**

Attacks on MANETS can be broadly classified into External and Internal. The External Attacks are attacks on the communication network. This is done by attacking the weak links or attacks on the routing information. These attacks essentially limit

the availability, confidentiality, anonymity, authorization and authentication of user for a network. The Internal Attacks concentrate on the security algorithm or the functioning of internal nodes through impersonation or making them selfish nodes. There may also be certain device oriented attacks.

### 3.3 Some Common Attacks on MANETS

#### 3.3.1 Denial Of Service (DOS) [3]

This type of attack may occur due to all the network attacks that take place on topology of the network, in which the attacker floods the network which results in unavailability of the response from the destination node. Any intermediate node in the routing can drop packets to cause denial of service at the base station or server. In the lack of getting back the response the source node keeps on sending the requests hence the network keeps on flooding further.

#### 3.3.2 Routing Table Overflow Attack [4]

Routing Table Overflow attack involves employing a non-existent node data over the network, thereby corrupting and degrading the rate at the time of updation of routing tables. The attacker sends RREQ messages in the network to non-existent nodes. The nodes under attack are overwhelmed with data and its routing table is full and cannot accept any further entries.

#### 3.3.3 Wormhole Attacks [4]

Wormhole attack is a more severe attack in which two attackers manage to place themselves at a vantage point in the network to misguide the network by advertising their as the shortest path for the transmitting their data. Thus, the wormhole attacker is able to create a tunnel in order to records the ongoing communication and traffic at one network position and channel them to another position in the network.

#### 3.3.4 Black Hole Attack

In black hole attack [4], a malicious node uses its routing protocol in order to advertise itself for having the shortest path to the destination node or to the packet it wants to intercept. As a consequence malicious nodes interrupt the route discovery process, resulting in traffic and multiple route requests in network.

**Table 1.** Classification of MANET attacks based on network layer

<u>S.No</u>	<u>Network Layer</u>	<u>Type of vulnerability, attack</u>
	Application layer	Malicious code, Repudiation, Impersonation
	Transport layer	Session Hijacking, Flooding.
	Network layer	Sybil, Flooding, Black hole, Grey Hole, Worm Hole, Link spoofing, Link withholding, Location disclosure.
	Data link layer	Malicious Behavior, Selfish Behavior
	Physical layer	Interference, Traffic Jamming, Eavesdropping

## 4 Problem Statement

Security algorithms for MANETS should consider the constraints mentioned in section 3 above. The solution approach should use initial phase of a security algorithm and by reducing the complexity of algorithm a security solution on selected number of critical nodes should be implemented.

## 5 Solution Approach

There is a good scope of improvement in the existing solution approaches in terms of efficiency, and lightness of protocol. One existing such piece of improvement is by introducing a generic security solution, with some algorithmic steps to classify the network nodes into critical nodes, non-critical nodes, popular nodes based upon some parameters like redundancy ratio and mathematical concepts and the use of rounds of time based clustering which helps in intrusion detection by the use of cluster heads.

The solution approach is to implement a basic ad hoc network on NS-2 using AODV, DSR, DSDV protocol using O-TCL as the language, the analysis and solution approach proposed is done in C/C++. Libraries available in NS-2 are used to add on features and parameters to judge the energy decay and to insert the malicious behaviour of the node AODV.CC file is changed and modified as per the malicious node behaviour. To analyse the .Tr file of the output of topology, solution in the form of algorithmic steps are used which are implemented using C as programming language.

Cluster based intrusion using TDMA is used to get the parameters for critical node behaviors and to set the number of rounds for which the parameters will be evaluated. Critical nodes, popular nodes and non-popular nodes are selected by implementing the underlined proposed solution algorithm which performs analysis of network parameters for every round as entered by the user and displays results for deviation, redundancy ratio and flags the nodes with exceptionally high sequence numbers.

### 5.1 Implementing Proposed Solution Approach

This work has implemented AODV, DSR and DSDV algorithms in NS-2 as a wireless ad hoc network and simulation results on NAM are recorded. This is done because for addressing the security issues on network layer which can be done either on routing algorithm or on packet forwarding intrusion of a malicious node is needed, which will drop the incoming packets without sending it forward. The malicious node behaviour is added by changing the header files and procedure file Aodv.h and Aodv.cc.

Hence the prime objective for implementing these routing protocols was to study the computations inside the procedures declared in MANETS for these routing mechanisms.

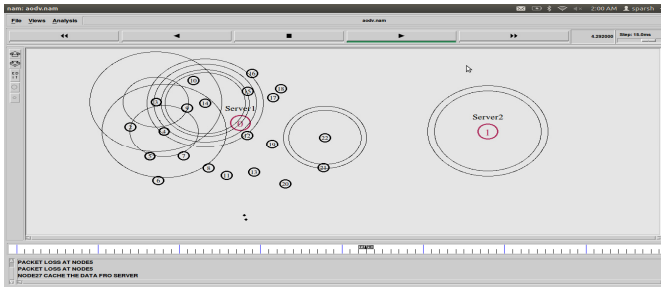


Fig. 1. Malicious node behaviour: packet loss at nodes

### 5.2 Implementation of AODV, DSDV Routing in NS-2

Ns-2 simulator shows the simulation results in trace files and NAM file as an output for visualization. The nodes are connected using a connection oriented network like TCP or a connection less network like UDP. Flush trace is used to dump these traces to the respective files. Droptail is declared to determine the buffer capacity of the queue whose default value is 50. All the objects are instantiated using set command and the objects computes time to live for each received packet. A different network protocol has different flavours in ns-2 as TCP has tahoe, reno, and vegas. The network agent is attached to the source node and to the sink node. The source node is the producer node as it produces the packets and the sink node is the consumer node, as it consumes the packets.

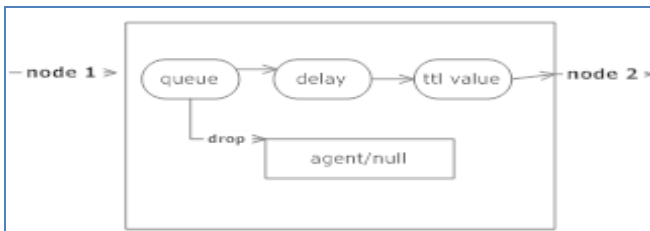


Fig. 2. The general node to node communication procedure in NS-2

The traffic flow patterns can also be designed using FTP and CBR (constant bit rate). Ns-2 is an event based simulator and we have to define in the script that when the event has to occur. For implementing aodv ns-2 has some classes and pre defined procedures in its ns2.lib which are found in the base directory of ns-2 as aodv.cc, aodv.h, aodv.packet.h, aodv.rqueue.cc, aodv.rtable.h.

### 5.3 Secure Architecture for MANETS: Implementation Approach

#### 5.3.1 Intrusion Detection

A cluster based approach in which every node participates in detecting the intrusion of malicious nodes. Formation of cluster heads which will act as leaders for intrusion detection based upon the parameters of classification of topology.

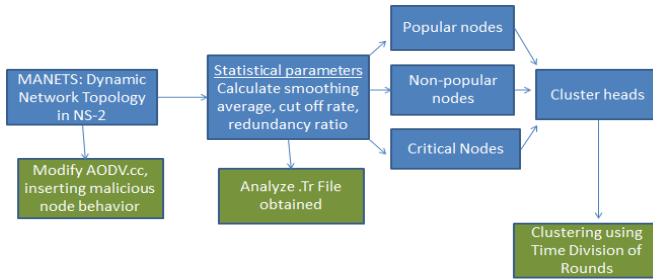


Fig. 3. Secure Architecture for MANETS

### 5.3.2 Clustering

Network topology is divided into clusters, each cluster will have a cluster head based upon Time division, and the cluster head will initiate the existing solution approach of security depending upon the application on MANET (based upon information of critical, popular and non-popular nodes received from the solution approach proposed in this research).

### 5.3.3 Classification of Nodes into Critical Nodes, Popular Nodes, Non Popular Nodes

Each node maintains a counter for every RREQ, for each round. At the end of each round, the node records the rate at which it has been receiving the requests. Smoothing average is calculated and RREQ deviations is calculated from the smoothing average, if the RREQ deviation is much greater than the smoothing average, that round or a particular node is considered as invalid.

### 5.3.4 Popular Nodes and Non-popular Nodes

For every round, consider the REQ's having exceptionally high sequence numbers, mark an identity to a node which sends a packet with exceptionally high sequence number as a potential malicious node for that round (i). In case same node is detected as potential malicious for round (i+1), that node is considered to be a malicious node.

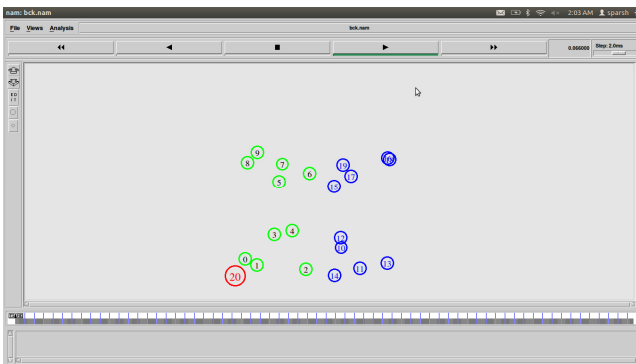


Fig. 4. Clustering (cluster based approach for intrusion detection)

### 5.3.5 Redundancy Ratio Mechanism

Calculating redundancy ratio for every round for each node:

**Redundancy ratio = N (number of transmitted packets by node) / M (Number of received requests by nodes).**

A node with exceptionally smaller value of Redundancy ratio is considered as malicious. A node with malicious intent will not reply for requests within the specified TTL. The TDMA based rounds is applied for cluster formation. TDMA reduces single point of failures, establishes power management in an efficient manner.

For every round in TDMA, the number of requests and replies on each node are calculated, applying following operations:

Calculating the smoothing average of rate of requests:

- $\text{rate}(i) = \text{RREQ}/dt$
- $\text{delta}(i) = \text{rate}(i) - s.\text{avg}$
- $\text{Node rate}(i) = [\text{Total RREQ}(i)/ dt]$
- $\text{Node avg} = \text{node rate}(i) - \text{delta}(i)$
- $\text{Node avg}(i) = \text{node avg}(i-1) + g * \text{delta}(i)$
- $\text{Node dev}(i) = \text{node dev} + h * (\text{delta}(i) - \text{node dev})$
- $\text{Cut off rate}(i) = \text{node avg} + 2 * \text{node dev}$

**A node is critical if cut off rate < s.avg**

All critical nodes, popular nodes, non-popular nodes applying TDMA for specified number of rounds is listed and analyzed for a particular time period as mentioned by the user.

## 6 Conclusion and Way Ahead

In pervasive computing environment where there is limitations for resources it is very much required to consider the security aspects of MANETS, to preserve the usage, and to get optimum results. Security of MANETS and the security related issues are slightly different to those as of normal Wireless Network, as the limitations and prerequisites of MANETS are changed, the security concerns and the possible lines of attacks also changes. Existing security algorithms are very specific and provides solution for one or two aspects of security, algorithms provides countermeasures by using the techniques of randomization, cryptographic and network based techniques and it is clear that the MANETS are vulnerable to internal attacks and external attacks as classified. There is a good scope of improvement in the existing solution approaches in terms of efficiency, and lightness of protocol. One existing such piece of improvement is by introducing a generic security solution, with some algorithmic steps to classify the network nodes into critical nodes, non-critical nodes, popular nodes based upon some parameters like redundancy ratio and mathematical concepts and the use of rounds of time based clustering which helps in intrusion detection by



the use of cluster heads. The solution proposed provides cluster based formation of topology, with a feature of selecting a cluster head in time-round based approach, and classifying the network nodes into popular nodes, non-popular nodes, critical nodes (which are potentially malicious nodes) based upon mathematical parameters like redundancy ratio and smoothing average and addresses some issues of security in MANETS. Any existing solution approach in the security of MANETS can adopt this process before initiating the security protocol which provides optimum results and energy efficiency.

There is always a scope of improvement in terms of efficiency, the solution approach can be made more efficient in terms of robustness and checking the simulation for hundreds of mobile moving nodes, the solution can be tested on some real time mobile ad-hoc application for its efficiency and deviation with the simulator system based results.

## References

1. Hu1, X., Grechniko, E.: On the Connectivity in One-Dimensional Ad Hoc Wireless Networks with a Forbidden Zone, University of Moscow
2. Berton, S., Yin, H., Lin, C.: Secure, Disjoint, Multipath Source Routing Protocol (SDMSR) for Mobile Ad-Hoc Networks. In: Proceedings of the Fifth International Conference on Grid and Cooperative Computing. IEEE (2004)
3. Andhare, A., Patil, A.B.: Denial-of-Service Attack Detection Using Genetic-Based Algorithm. International Journal of Engineering Research and Applications (IJERA) 2(2), 094–098 (2012), <http://www.ijera.com> ISSN: 2248-9622
4. Ullah, I., Rehman, S.U.: Analysis of Black Hole attack on MANETS Using different MANETS Routing Protocols. Master Thesis Blekinge Institute of Technology, Sweden, Thesis no: MEE-2010-2698 (June 2010)
5. Sreenath, N., Amuthan, A., Selvigirija, P.: Countermeasures against Multicast Attacks on Enhanced-On Demand Multicast Routing Protocol in MANETS. In: IEEE International Conference on Computer Communication and Informatics (ICCCI 2012), Coimbatore, India, January 10-12 (2012)
6. Kim, Y., Sankhla, V., Helmy, A.: Efficient Traceback of DoS Attacks using Small Worlds in MANETS Vehicular Technology Conference. In: IEEE VTC 2004-Fall (2004)
7. Sheikh1, R., Chandee, M.S., Mishra, D.K.: Security issues in MANETS: a review. In: 2010 Seventh International Conference Wireless and Optical Communications Networks, WOCN (2010)
8. Sankareswary, P., Suganthi, R., Sumathi, G.: (Under the able guidance of Mr. Julian Benedict M.E.) Impact of Selfish Nodes in Multicast Ad Hoc on demand Distance Vector Protocol. In: International Conference Wireless Communication and Sensor Computing, ICWCSC (2010)
9. Burmester, M., de Medeiros, B.: On the Security of Route Discovery in MANETS. IEEE Transactions on Mobile Computing, 1180–1188 (September 2009)

10. Sreenath, N., Amuthan, A., Selvigirija, P.: Countermeasures against Multicast Attacks on Enhanced-On Demand Multicast Routing Protocol in MANETS. In: IEEE International Conference on Computer Communication and Informatics (ICCCI 2012), Coimbatore, India, January 10-12 (2012)
11. Umaparvathi, M., Dharmishtan, D.: Evaluation of Symmetric Encryption Algorithms for MANETS. In: IEEE International Conference Computational Intelligence and Computing Research, ICCIC (2010)

# Finding a Trusted and Shortest Path Mechanism of Routing Protocol for Mobile Ad Hoc Network

Rabindra Kumar Shial, K. Hemant Kumar Reddy, and Bhabani Sankar Gouda

Department of CSE,  
National Institute of Science and Technology  
rkshial@yahoo.com, khemant.reddy@gmail.com,  
bhabani012@rediffmail.com

**Abstract.** A mobile ad-hoc network (MANET) is a contemporarily positioned in wireless network where A mobile ad-hoc network (MANET) is a contemporarily positioned in wireless network where nodes able to interact with each other without forming any frame such as access points or base stations. Nodes can unite and split the network at their convenience and they are free to move randomly. Nodes can even organize themselves randomly. This is a resistant of the changeable behavior of MANET, it is possible that there could be little malicious and selfish nodes that try conciliating the efficiency of the routing protocol and construct MANET defenseless to security assail. Therefore an any armed security-motivated AODV (Ad hoc On-demand Distance Vector Routing) routing protocol devised as R-AODV (Reliant Ad hoc On-demand Distance Vector Routing). The accomplishment of this work is tailored and relied on trusted mechanism known as direct and recommendations trust model which is integrated into AODV. This consents the AODV for discovering a shortest path along with its highest degree of reliability. The diversely designed modus operandi restricts the data to go through malicious nodes that intend to avert. The R-AODV protocol has been implemented and simulated on NS-2. Based on the simulation consequences, it can be shown that R-AODV does offer a steadfast data hopping in context to the customary AODV if there are malicious nodes in the MANET.

**Keywords:** Trust, mobile ad-hoc network, routing protocol. Security.

## 1 Introduction

Mobile ad hoc network (MANET) is a network compiled by only nodes, these nodes do not have fixed casing or any converge controller such as access point or server that can ascertain the route of paths. Thus every single node in an ad hoc network has to wait for each other in order to advance the packets and there is a requisite to master a specific mutual aid mechanism to advance packet from hop to hop before it attains compulsory destination by means of routing protocol. Instances of available routing protocol for ad hoc network are ad hoc on demand distance vector (Aodv) destination sequenced distance vector (DSDV) and dynamic source routing(DSR).The major hypothesis of these routing protocols is to locate the minimized path in the source

destination routes assortment. These routing protocols can face confrontation by numerous spices of attacks such as black hole, denial of service DoS and worm hole. These attacks tend to amend or fabricate routing of packets.

In this paper we proposed a trust mechanism in the Ad-hoc On-demand Distance Vector routing protocol AODV calls R-AODV that include both trust and shortest path. R-AODV can reduce the risk on MANET that caused by some of attacks such as DoS and black hole attacks. It also increases the packet delivery fraction by selection the best path. AODV routing protocol is described in section 2. R-AODV is explained in details in section 3. In section 4, we describe the simulation environment to evaluate the performance of R-AODV with three performance metrics (Packet Delivery Fraction, Average End-to-End delay and Normalized Routing Load).

## 2 Related Work

Trusted AODV (TAODV) [16] is a good mechanism for AODV security. It uses subjective logic proposed by Pirzada [9], which is based on subjective believes about the world. In this protocol the trust about a specific node is calculated by using opinion of other nodes. The proposed framework has four basic modules; basic routing protocol, trust model, trusted routing protocol and self-organized key management mechanism. For trust calculation it uses trust combination. Trust is not transitive in nature and this scheme does not take care of this fact while calculating trust based on the opinion of other nodes [11].

In [16], Trust embedded AODV (TAODV) protocol, it calculates end-to-end secure route free of malicious nodes through collaboration of nodes in the path. The header of RREQ contains an extra trust-level field. When an intermediate node receives the RREQ it modifies the trustlevel field to include the trust level of the node which generates RREQ. Every node in the path checks back broadcasted RREQ from its next node to see whether it has provided the proper information [15]. In T-AODV an intermediate node cannot send the route reply and route selection is performed by utilizing the trust level through a third party involvement. This violates the concept of flexible configuration in ad hoc networks. T-AODV protocol can be attacked by colluding malicious nodes. SAODV [14], uses cryptographic methods to secure the routing information in the AODV protocol. SAODV uses digital signatures to authenticate non-mutable fields and hash chains to authenticate the hop-count field in both RREQ and RREP messages.

## 3 Ad-Hoc On-Demand Distance Vector Routing Protocol (AODV)

In AODV, the network is silent until a connection is needed. At that point the network node that needs a connection broadcasts a request for connection. Other AODV nodes forward this message, and record the node that they heard it from, creating an explosion of temporary routes back to the needy node. When a node receives such a

message and already has a route to the desired node, it sends a message backwards through a temporary route to the requesting node. The needy node then begins using the route that has the least number of hops through other nodes. Unused entries in the routing tables are recycled after a time [10]. When a link fails, a routing error is passed back to a transmitting node, and the process repeats. Much of the complexity of the protocol is to lower the number of messages to conserve the capacity of the network. Another such feature is that if a route request fails, another route request may not be sent until twice as much time has passed as the timeout of the previous route request [7]. The advantage of AODV is that it creates no extra traffic for communication along existing links. Also, distance vector routing is simple, and doesn't require much memory or calculation. However AODV requires more time to establish a connection, and the initial communication to establish a route is heavier than some other approaches.

## 4 Trust Models

### 4.1 Direct and Recommendation Trust Model

Direct trust as the trustworthiness of a node depending on the evidence collected from the one-to- one interactions with the node. The evidence is collected by forwarding a packet to a next-hop and then monitoring the next-hop's successive forwarding of the same packet. The evidence collected for the next-hop is evaluated to a positive value, *pos (packet)*, only if the packet has been forwarded without any modification.

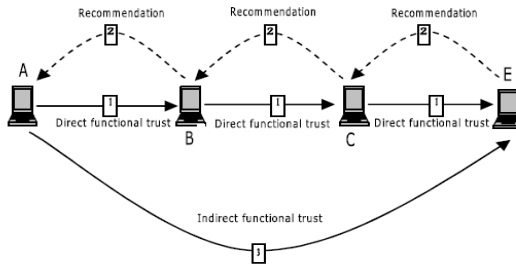


Fig. 1. Direct and Recommendation trust model

When node A wants to send information to node E, node A needs to authenticate node E since node A does not have direct trust with node E. In this case, node A will ask other nodes' recommendations. Node C who has direct trust for node E, can recommend node E to node B which has direct trust for node C. As a result, Node A can trust node E based on the recommendation reply from node B.

#### A. Proposed Approach for Recommendation

It is feasible to *deduce* from an explicit recommendation, whether the recommender will forward packets on behalf of the recommended node or not. In our approach,

nodes communicate recommendations by following the inverse of abovementioned deduction process such that the recommendations are free from the associated issues. Consider the scenario from Figure to understand our approach. In the scenario, node D *derives* an implicit recommendation from its neighbour C (recommender) for node N (recommended node) depending on whether C has forwarded packets on behalf of its previous-hop N. Node D captures the evidence for the recommendation from the route contained in a received packet. Finally, node D computes its opinion for the derived recommendation depending on its trust for the recommender, i.e., node C.[11]

## B. Fuzzy Logic Trust Model

**Fuzzy Logic:** In a mobile ad-hoc network (MANET) environment, intermediate nodes on a communication path are expected to forward packets of other nodes so that the mobile nodes can communicate beyond their wireless transmission range.

**Fuzzy Trust Model:** The proposed scheme could be employed in any MANETs routing protocols to enforce cooperation among nodes and counter with non-cooperative nodes in a MANET environment.

**Basic Assumptions:** This scheme requires the following features to accomplish its functions properly:

- (i) All nodes could operate in promiscuous mode for neighbor monitoring.
- (ii) All links are bidirectional and transmission ranges of nodes are rather equal.
- (iii) All nodes use omni-directional transceivers.
- (iv) Misbehaving nodes are selfish and are not malicious.
- (v) The network is a multi-hop network which means that packets exchanged between any two nodes (that are not within each other's communication range) are forwarded by other nodes.

## Fuzzy Direct Trust Model

To manage a collection of trust related activities across domains, we need to understand trust itself. From different points of views, trust can be categorized into different classes: direct trust and indirect trust. In Fig.2 we say node  $j$  is trustworthy or untrustworthy for node  $i$ , there is a trust model between node  $i$  and node  $j$ . If this statement is based on node  $i$ 's direct experiences with node  $j$  completely, this model is called the direct trust model. The direct trust relation just has fuzzy properties. We can use the fuzzy theory to describe the direct trust model.

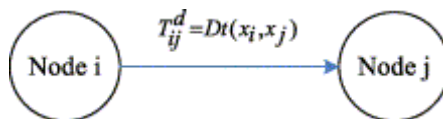


Fig. 2. Fuzzy Direct Trust Model

According to the data forwarding table, node  $i$  can make the trust evaluation to node  $j$  denoted as source, destination,(i,j),where *source* is the trust evaluation nodes, *destination* is the evaluated nodes, *index* is the trust value to destination at  $t$ , which stands for time, and  $T$  is the Time To Life (TTL) of trust. The fuzzy membership function of the direct trust model is defined as:

$$T_{ij}^d = D t (x_i, x_j) = \frac{H F_{ij}(t)}{H F_{ij}(t) + \alpha [R F_{ij}(t) - H F_{ij}(t)] + \lambda}$$

Represents the weights of the past negative behavior which can be regulated to punish the selfish node action, that is, the bigger the value of  $\alpha$ ,the greater the degree of punishment. Constant  $\lambda$  is an uncertainty trust for the weight value, adjusted for the decline in the failure rate. That is, the greater of the value  $\lambda$ , the slower the decline pace of the failure. As a node behavior is not always constant but often changes with time and volatility, therefore, the recent experience is more credible than the general historical experience. That is to say the interactive information takes on the time-forgotten effect. This paper combines the historical events with the recent events to update  $HF_n(t)$ :  $HF_{ij}(t)=rHF_{ij}(t-1)+HF_{ij}(\Delta t)$ .  $r$  presents the weight of past behavior for direct rating. Our computation uses the entire history, but as time progresses the impact of old history is diminished. So we have an interaction threshold value of interaction times  $Hd$ . Then the fuzzy direct trust model membership function can be revised as:

$$T_{ij}^d = D t (x_i, x_j) = \begin{cases} 0.5 + \frac{[2 H F_{ij}(t) - R F_{ij}(t)]}{2 H_d}, & R F_{ij}(t) < H_d, \\ \frac{H F_{ij}(t)}{H F_{ij}(t) + \alpha [R F_{ij}(t) - H F_{ij}(t)] + \lambda}, & R F_{ij}(t) > H_d. \end{cases}$$

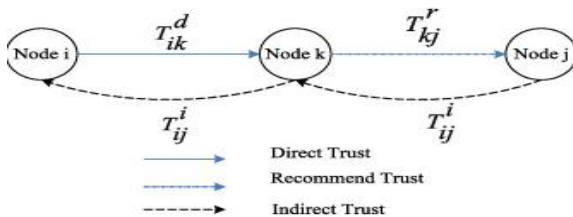


Fig. 3. Fuzzy Indirect and R Recommended Trust Model

## 5 Reliant On-Demand Distance Vector Routing Protocol (RAODV)

AODV can be modified to select better path (best path ( $Bp$ )) during the route discovery cycle based on the trust and number of hops (trusted and shortest). When the route request and route reply (R-RREQ and R-RREP) messages in Reliant R-AODV are generated or forwarded by the nodes in the network, each node appends its own trust to the trust accumulator (trust summation accumulator  $S[t]$ ) on these route discovery messages. Each node also updates its routing table with all the information contained in the control messages. As the R-RREQ messages are broadcast, each intermediate node that does not have a route to the destination forwards the R-RREQ packet after appending its trust to the trust accumulator in the packet which is computed by

$$S[t] = \sum_{i=1}^n trust_{value}(i)$$

Where:

$n$  : number of hop counts received in one path.

$S[t]$ : the trust summation accumulator.

$trust_{value}(i)$ : trust value of neighbors node in routing table

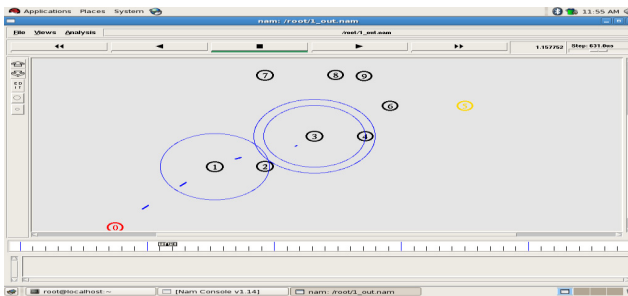


Fig. 4. RAODV Route Forward Request

## 6 Simulation Model

A detailed simulation model based on Network Simulation version 2.33 NS2 [14] is used in the evaluation. The Linux-Red hat Operating System was used because it is a user friendly platform and easy to manage and to setup a simulator. The Distributed Coordination Function (DCF) of IEEE 802.11[15] for wireless LANs is used as the MAC layer protocol. An unspotted carrier sense multiple access (CSMA) technique with collision avoidance (CSMA/CA) is used to transmit the data packets. Fig.4 shows the procedure chart to execute simulation on NS2.



**Table 1.** Simulation Parameters

Map Size 500 m x 500 m	Max Speed 25 m/s
Mobility Model Random way point	Traffic Type Constant bit rate (CBR)
Packet Size 512 bytes	Connection Rate (Nominal
Radio Range)	4pkts/sec
Pause Time 20,40,60,80,100 second	Number of Connection 5
bandwidth of links 2Mbit	MAC layer type IEEE 802.11
Number of Nodes 50 nodes	Simulation Time 900 seconds
Map Size 500 m x 500 m	Max Speed 25 m/s
Mobility Model Random way point	Traffic Type Constant bit rate (CBR)
Packet Size 512 bytes	Connection Rate (Nominal
Radio Range)	4pkts/sec
Pause Time 20,40,60,80,100 second	Number of Connection 5
bandwidth of links 2Mbit	MAC layer type IEEE 802.11
Number of Nodes 50 nodes	Simulation Time 900 seconds

## 6.1 Performance Metrics

Three performance metrics are evaluated in our simulation:

1) Packet delivery Fraction The packet delivery fraction is the ratio of the total number of data packets received by destinations over the total number of data packets transmitted by sources.

2) Average end-to-end delay – The average end-to-end delay is the average of delays for all received data packet along the path (from the sources to destinations).

3) Normalized routing load– The normalized routing load is the total number of routing control packets (RREQ, RREP, and RERR) over the total number of received data packets

## 7 Performance Analysis and Results

### A. Packet Delivery Fraction

The packet delivery fraction of AODV without drop packet (AODV-wo-drop), AODV with drop packet AODV-w-drop) and Reliant AODV with drop packet (R-AODV-w-drop) shown in Fig 5. In this simulation, when pause time set to 20

second which is close to 0 (continuous motion), each of them obtained lower packet delivery fraction. The reason is that changing of the topology of network caused by high motion of nodes due to less pause time. As the pause time reaches 100 second (no motion), packet delivery frication for R-AODV-w-drop will be increased to 70% due to the stable network. In summary, for node equal to 50, high values of pause time and simulation area 500m x 500m, R-AODV shows much better performance in the team of packet delivery fraction compare with the one with drop (AODV-w-drop).

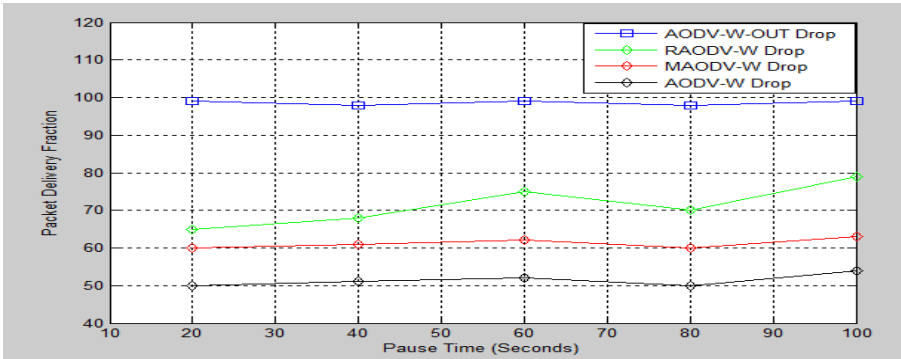


Fig. 5. Packet Delivery Fraction

**B. Average End-to-End Delay**

The average end-to-end delays of original AODV without drop (AODV-wo-drop), AODV with drop (AODV-wo-drop) and R-AODV with drop (R-AODV-w-drop) are shown in Fig.6. Average end-to-end delay (seconds) is the average time it takes a data packet to reach the destination including the queue in send buffer. As routes break, nodes have to discover new routes which lead to longer end-to-end delays and source packets are buffered at the source during route discovery. As a result, the Average end-to-end delay in R-AODV is decreasing with the increasing of the pause time. For fix number of nodes equal to 50 and small spaces, for example 500m x 500m, R-AODV perform well but with a slightly delay.

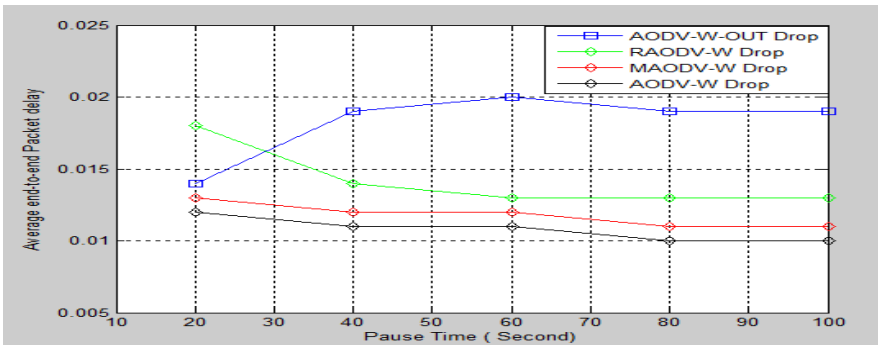


Fig. 6. Average End-to-End Delay

### C. Normalized Routing Load

The comparison among R-AODV with drop with AODV without drop and AODV with drop shows different performance regardless on the degree of mobility Fig.7. When pause time set to 20 second each of AODV-wo-drop, AODV-w-drop and R-AODV-w-drop shows high routing load due to routes break or change in route directory due to the high motion of nodes. In this case, nodes need to send RRER message to notify other nodes about that routes break. And then nodes have to discover new routes by sending other RREQs messages. Once these RREQs received, destinations will generate RREPs. However, The experimental results for fix number of nodes equal to 50 and small spaces, for example 500m x 500m, R-AODV perform well and show better result in terms of normalized routing load compare with AODV-w-drop.

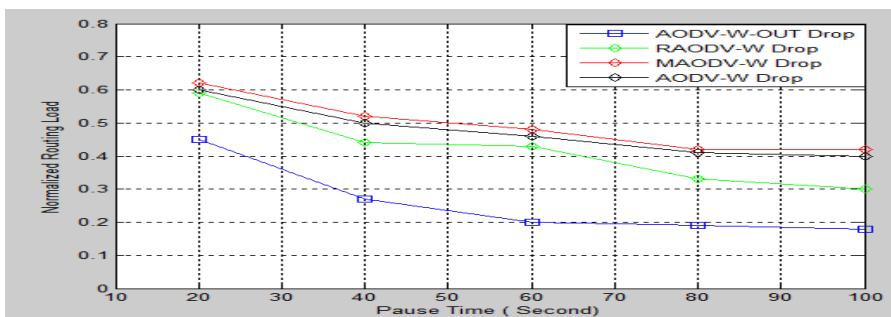


Fig. 7. Normalized Routing Load

## 8 Conclusion and Future Works

In this paper, we proposed a new MANET routing algorithm called Reliant R-AODV which is basically an extension to the AODV routing protocol that incorporates a trust mechanism to enhance its security. The proposed algorithm was implemented and simulated using the NS-2 network simulator. In the simulation used, each node is given a trust value and this value is associated with the possibility of the node to perform a packet drop. With the inclusion of trust mechanism, it is expected that using R-AODV would result in a higher percentage of successful data delivery as compared to AODV.

## References

- [1] Asokan, R., Natarajan, A.M., Venkatesh, C.: Ant based dynamic source routing protocol to support multiple quality of service (QoS) metrics in mobile ad hoc networks. *International Journal of Computer Science and Security* 2(3), 48–56 (2008)
- [2] Perkins, C.E., Bhagwat, P.: Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. *ACM SIGCOMM Computer Communication Review* 24(4) (1994)

- [3] Jacquet, P., et al.: Optimized link state routing protocol for ad hoc networks. In: Proceedings. IEEE International Technology for the 21st Century Multi Topic Conference, IEEE INMIC 2001. IEEE (2001)
- [4] Gouda, B.S., Shial, R.K., Reddy, K.H.K.: An optimal path finding routing protocol for mobile Ad-Hoc network using random mobility model. In: Emerging Trends in Science, Engineering and Technology, INCOSSET (2012)
- [5] Johnson, D.B., Maltz, D.A.: Dynamic source routing in ad hoc wireless networks. International Series in Engineering and Computer Science, pp. 153–179 (1996)
- [6] Perkins, C.E., Royer, E.M.: Ad-hoc on-demand distance vector routing. In: Proceedings of the Second IEEE Workshop on Mobile Computing Systems and Applications, WMCSA 1999. IEEE (1999)
- [7] Park, V.D., Corson, M.S.: A highly adaptive distributed routing algorithm for mobile wireless networks. In: Proceedings of the Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 1997, vol. 3. IEEE (1997)
- [8] Ko, Y.-B., Vaidya, N.H.: Location-Aided Routing (LAR) in mobile ad hoc networks. Wireless Networks 6(4), 307–321 (2000)
- [9] Pirzada, A., McDonald, C.: Reliable routing in ad hoc networks using direct trust mechanisms. In: Cheng, M., Li, D. (eds.) Advances in Wireless Ad Hoc and Sensor Networks, pp. 133–159 (2008)
- [10] Guang, L., Assi, C.: Vulnerabilities of ad hoc network routing protocols to MAC misbehavior. In: IEEE International Conference on Wireless And Mobile Computing, Networking And Communications (WiMob 2005), vol. 3. IEEE (2005)
- [11] Tamilselvan, L., Sankaranarayanan, V.: Prevention of impersonation attack in wireless mobile ad hoc networks. International Journal of Computer Science and Network Security (IJCSNS) 7(3), 118–123 (2007)
- [12] Fall, K., Vardhan, K. (eds.): Ns notes and documentation (1999), <http://wwwmash.cd.berkeley.edu/ns/>
- [13] Borgia, E., Anastasi, G., Conti, M., Gregori, E.: IEEE 802.11 ad hoc networks: protocols, performance and open issues. In: Basagni, S., Conti, M., Giordano, S., Stojmenovic, I. (eds.) Ad hoc Networking. IEEE Press, Wiley, New York (2003)
- [14] Anastasi, G., Borgia, E., Conti, M., Gregori, E.: IEEE 802.11 ad hoc networks: performance measurements, icdcs. In: 23rd International Conference on Distributed Computing Systems Workshops (ICDCSW 2003), p. 758 (2003)
- [15] IEEE, Wireless LAN Medium Access Control (MAC) and Physical layer (PHY) Specifications. IEEE Std. 802.11 (1997)
- [16] A Trusted AODV Routing Protocol for Mobile Ad Hoc Networks. PhD thesis, Department of Computer Science and Engineering, The Chinese University of Hong Kong, <http://www.cse.cuhk.edu.hk/lyu/student/phd/gigi/term4/>

# Fusion Centric Decision Making for Node Level Congestion in Wireless Sensor Networks

N. Prabakaran<sup>1,\*</sup>, K. Naresh<sup>1</sup>, and R. Jagadeesh Kannan<sup>2</sup>

<sup>1</sup> SCSE, VIT University, Vellore

<sup>2</sup> CSE, RMKEC, Chennai, India

dhoni.praba@gmail.com, naresh.k@vit.ac.in, dr\_rjk@hotmail.com

**Abstract.** The data-centric wireless sensor networks comprise numerous autonomous tiny nodes forms random topology in nature. Applications oriented WSNs immensely used for monitoring harsh environment. Its unique constrains are distinguishing it from traditional networks by energy, lifetime, fault tolerance, scalability and computational power. When they are deployed randomly sensing & generating vast amount of data, for which they are being used. Network meets Congestion, if huge volume of data passed. To eradicate it, misbehaving nodes are identified and skilled for self-healing without human intervention. Existing approaches focus on controlling link level congestions not node level. Our proposed fusion-centric scheme controls node congestion. To achieve this, selectively concentrate on intermediate level. Misbehaving nodes are identified by using their historic data based on fault level occurred. Lifetime is determined by allocation-rate and more radio signal usage. Our scheme keeps the network without consuming much resource. Node level control is needed for self-configuring WSNs.

**Keywords:** WSNs, Congestion Control, Node deployment, self-healing.

## 1 Introduction

WSNs have been used greatly in Area monitoring, environmental and health monitoring, location tracking and agriculture etc it might not be possible in other networks and configuration is done without human interruption [1][2]. Generally they are self battery powered with target node being most powerful and used as access point to any other network for data processing. Before their energy exhausted, it should be used effectively. In many of the applications, less cost sensors are used and placed in the unattended region [1] [3]. The ad-hoc network routing protocols are not applicable [4], because of scalability. The network structure also unpredictable since it is changing dynamically. During the quick communication flow towards the gateway, the traffic is introduced between the nodes and causes the congestion. Congestion in WSNs has depressing impacts on network performance and application

---

\* Corresponding author.

purpose, i.e., packet loss, increased packet delay, node energy wastage and severe reliability degradation.

Once the nodes started to misbehave or deviate from communication flow, then it leads to traffic. However, the source nodes will sense large volume of data and intermediate nodes will propagate them by forwarding it. Congestion is possible, only if the node is trying to forward more than its capacity [3]. Similarly malfunction or misbehave also will lead to congestion. If the node becomes faulty, it affects the entire cluster even network also [4].

Existing fault management system is approached with different architecture, protocols, detection algorithms and decision algorithms too [5]. Congestion notification bit is set based on source level data generation and fairness among the source nodes [6]. The main factors deciding congestion in WSNs can be Channel Contention and link failure and packet Collision. It can be detected by two main factors such as queue overflow and link collisions. Similarly the fault system can be detected by explicit and implicit detection which can be managed by centralized and distributed approaches. The remaining part of this paper is continued with related works, models and problems, proposed scheme, evaluation and finally the conclusion.

## **2 Related Works**

Literature survey has been conducted on congestion detection, congestion control and communication models. EDCAM in WSNs allocates the multiple queue, they are used for storing the sequential information [2]. But they are not leveling or controlling any type of congestion in the node. Recently many congestion controlling techniques have been introduced. The congestion can be detected easily by checking queue occupancy and channel occupancy. Mitigating Congestion in Wireless Sensor Networks (Bret Hull in 2004) uses hop by hop and node level congestion control. But the fairness of the network is under progress and performance comparison is left out for future work. On The Interdependence of Congestion and Contention In Wireless Sensor Networks (Mehmet C. Vuran Vehbi C in 2004) investigated widely about congestion in wireless sensor networks and their experiment result says interaction between the nodes directly affects the performance. The interaction among the nodes only determines the well established communication.

## **3 Models and Problems in Communication**

### **3.1 Node Locality**

Fault management can be obtained easily from the node based on its location. Geographically or logically centralized node is chosen and this central node normally has resource to execute wide range of signals. WSNs have large scalability and the central node become the single point data traffic for fault detection and creating funneling effect. This subsequently causes huge volume of radio message traffic and rapid energy depletion in certain region of the network especially the node close to

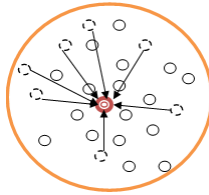
the central node. The solution could be configuring the distributed fault management. SPIN protocol uses routing discovery and updates the stage with neighbor nodes so they need additional radio communication and using energy more.

### 3.2 Node Selection

The fault management is (i.e. the sink, central coordinator, manager, base station, gateway, etc.) acting crucial role in the wireless sensor network. It primarily aggregates all the data from various intermediate nodes and energy level is more important than any other nodes. Always it is maintaining more messages through radio signals with others to identify faulty. Forwarding all the packets through itself and fault node's location is easily traced by simple divide and conquer method in centralized approach. On the other hand, distributed approach handles with  $n$  number of cluster heads and selective transmission of signal with each other. It identifies the suspicious node from the group, in which a node has large variation against with neighborhoods [6][7].

### 3.3 Congestion Detection

The following schema represents the way congestion happens.



**Fig. 1.** Centralized approach in WSNs

Normally funneling effect leads to congestion in centralized approach, in which all the nodes could not seize the channel simultaneously to update their information. In distributed approach many cluster heads are seeking to send their updates to the sink. If a node from the cluster is not properly working then it fails to forward the message. Now the cluster head communicates with all members and sending message, checking their status.

### 3.4 Model

In networking concept, network congestion is possible in centralized as well as in distributed approach also. In order to identify the silent node (failed node) compare the readings that have huge difference against neighbors. The following diagram depicts the ways of communication in the distributed approach and centralized approach. The practical problem is, if half or more of the sensor nodes are faulty then it is difficult to determine the faulty nodes. This approach requires the network to be

pre-configured consequently and may end-up routing into a new neighbor that has also failed. Resource management and traffic control are the techniques handled to control the congestion. The resource management tells to increase the resources which are used in the communication and it is a big challenge in WSNs. The figure 2 denotes the way of possible congestion points.

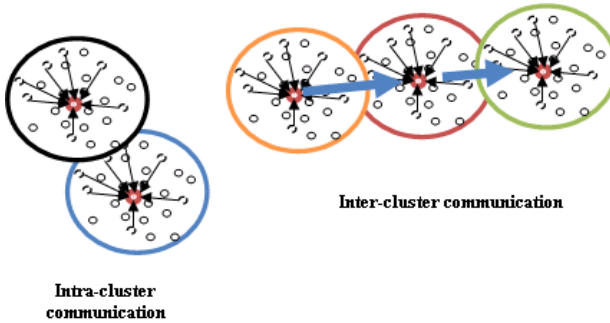


Fig. 2. Communication model

## 4 Fusion Centric Decision- Making Scheme Description

A node which is located commonly between different clusters can be termed as common node, which can be the neighbor of cluster head of two different clusters. General procedure to find the fault in distributed environment is,

- Collect the updated messages from the cluster members.
- Identify the status (successful/varying)
- Identify the whole cluster status with neighbor's coordination.
- Indicate the fault node with an event to all the members for avoiding useless radio message transmission.

WSN is implemented using distributed for finding fault detection easily and making a node to take decisions on faults such as physical malfunction of sensing devices. A sensor node in distributed environment expected to self monitor not frequently consulting with central node.

### 4.1 Fusion Sensors Coordination

A node is skilled to take decision after aggregating decisions from the nodes in the network. A fusion node (normally cluster head) detects abnormal readings of a node and takes decision, after comparing data from a set of neighbor nodes. Construct the fusion center as single node and skilling it to aggregate the previous decisions.

Arrived historic decisions are injected into the fusion node and introduce a technique called fusion sensors coordination and composing self decision from the historic decisions by single server (cluster head/ fusion center).



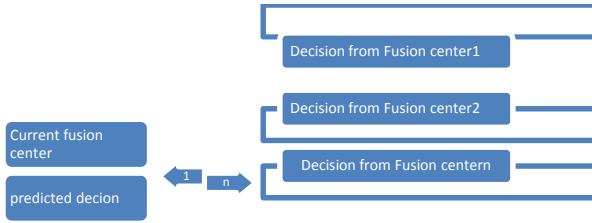


Fig. 3. Decision composition

### 4.2 Fault Diagnosis

To control the congestion, wireless sensor network needs strict fault management. Software portion is not creating any fault in the sensor node since it occupies less participation.

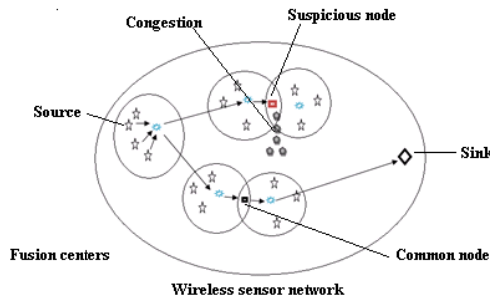


Fig. 4. Logic Scheme for node level congestion control

Sensor and actuator are the hardware are prone to failure or malfunction mostly. We classify them according to possible kind of faults with the following scenario.

**Sharing the gateway with other gateways:** - if none of the gateways have received status updates from certain gateway, it clearly indicates that the node is not able to transmit any data results transceiver failure. But existing proposals indicates them as misbehaving node and eliminates them from communication.

**Historic data:** - using ‘a-priori’ knowledge discover model, a node can predict the future possible data. This model can predict and identify errors in the recovered data packet. But existing proposal is not using the transceiver is failed node’s historic data.

**Faultless path:** - Using more historic data, the sink node can choose faultless routing path to transmit the data towards destination. By using silent node’s history we can make them to anticipate in successful transmission.

**Redundancy:** - Multiple copies of sensed data are transmitted to the sink via separate paths based on decision made. If received 2 copies are not identical, then the sink node decides inconsistency may have occurred. Now the same copy of data is transmitted via third path to identify fault nodes. Now sink sends packets back to

source through correct and faulty paths and counter is embedded with packet format for CNB. For the suspicious node it is incremented for every error message. If counter exceeds maximum times count, then its faulty node.

The following flow diagram represents the fusion center along with cluster members and states such as fault or successful routing path. The nodes are not choosing the route path via fault node and this scheme offers the ability to accept aggregated decisions from the neighbor fusion centers. The flow diagram explains the flow about how the decision is taken since the hardware failure node is still with energy.

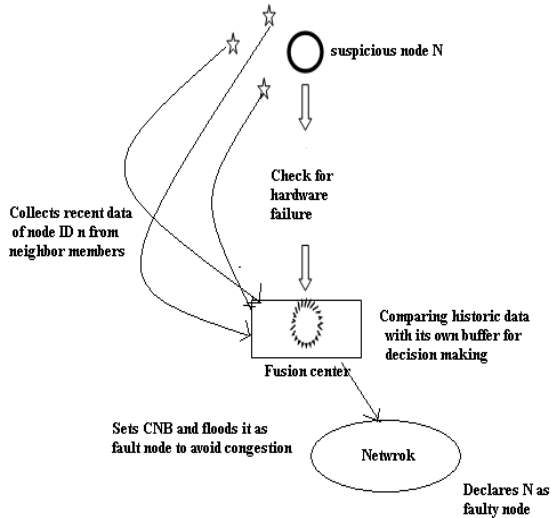


Fig. 5. Flow of the fusion center decision making

The following algorithm illustrates that, how this scheme works

**Algorithm:**

1. Check the kind of fault diagnosed from the node 'n'.
2. If it is hardware failure, then determine the present buffer occupancy level ( $\alpha$ ), since it is not energy drained.
3. Fusion center (corresponding Cluster head) can access neighborhood link nodes of node 'n' and collect ' $\alpha$ ' capacity of recently used packets for node ID 'n' from neighbors.
4. Fusion center refers its buffer and compares ' $\alpha$ ' capacity of aggregated recent packets of node ID 'n' from neighborhood link nodes.
5. Use this historic data for decision and set congestion notification bit on the node 'n' and update the status as failed node.
6. Every node in the network avoids the transmission path using node 'n' and prevents the node level congestion.

## 5 Performance Evaluation

Combining both centralized and distributed approach into single and the centralized manager concentrates on performance, whereas cluster head or fusion centre looks nodes with its sub region. The proposed scheme draws the benefits of both technologies so the accuracy and correctness of the fault is identified easily. Cluster is constructed with a fusion center which has unlimited capacity and multiple cluster heads or fusion centers are coordinated with each other. Every fusion node is aggregating the previous decisions to identify the faultless routing path without consuming much resource. The arrival rate can be represented as  $\lambda_0$  and service rate as  $\mu_0$ . For a fusion center the average service rate should be higher than the average arrival. The traffic identity or channel utilization ( $\rho$ ) can be represented as the rate between the service and arrival rates. For a single cluster an M/M/1 model is followed [8][9],

$$\rho = \lambda_0 / \mu_0$$

For M/M/S model, in entire network there is n number of clusters. The fusion centers are indicated by the term ‘s’ and the number of incoming packets are indicated by ‘n’. The arrival rate is  $\lambda_n$  and the service rate is  $\mu_n$ . If the incoming packets are less than available cluster head’s capacity then there will be no congestion or queue overflow and vice versa can be indicated as,

$$\mu_n = \begin{cases} n\mu, & n < s \\ s\mu, & n \geq s \end{cases}$$

In order to avoid the queue congestion, the expected number of packets to be in the individual cluster (M/M/1) is  $P_n = (\lambda_0 / \mu_0)^n \cdot P_0$  where  $P_0 = (1 - \rho)$   $P_0$  indicates that packets are no need to wait in the queue. Similarly for the model with several cluster heads or fusion centers (M/M/S),

$$P_n = \begin{cases} 1/n! \left( (\lambda/\mu)^n \right) P_0, & n < s \\ (1/s! (s)^{n-s}) (\lambda/\mu)^n P_0, & n \geq s \end{cases}$$

Where,

$$P_0 = \left[ \sum_{n=0}^{s-1} 1/n! \left( (\lambda/\mu)^n \right) + (1/s!) (\lambda/\mu)^s s\mu/s\mu - \lambda \right]^{-1}$$

With the given queue model, the network performance can be determined by the factor called waiting time i.e. how long the packet or data will be waiting in the queue to get the service. For M/M/1 model the expected waiting time ( $w_q$ ) can be determined by [9],  $w_q = \lambda/\mu + L_q$  Where,  $L_q = \lambda^2/\mu(\mu - \lambda)$

The length of the overflow queue is denoted by  $L_q$ , it says that expected number of packets in the queue. Nodes are taken into an account to test their efficient by NS2.29 simulator along with its extended package ‘mannasim’. The result tells that the queue length is increased while increasing the network load. If Nodes detect any unusual

events, during the event detection, the intermediate nodes are struggling to forward the data continuously and queue size is rapidly increased. Packet drop is possible if queue length is limited. The efficiency is computed from traffic load because it is inverse proportional. Arrived packets waiting time to get service is determining the node level congestion.

The congestion can be measured by  $L_q$ . Similarly for M/M/S, the average or expected waiting time is determined by,

$$wq = Lq/\mu$$

Where  $L_q$  is,

$$Lq = \left( \frac{(\lambda/\mu)^{s+1}}{s \cdot s! (1 - (\lambda/\mu s)^2)} \right) P_0.$$

Once a queue becomes idle, the arriving packets are no need to wait. In M/M/S model, the probability of arrival has to wait is given by [9],

$$P(n \geq s) = \left( \frac{(\lambda/\mu)^s \cdot P_0}{s! (1 - \lambda/s\mu)} \right)$$

## 6 Conclusion

We have proposed this scheme to control the node level congestion in WSNs and to achieve reasonable balance between fault detection, accuracy and efficient energy usage of network resources. The efficiency is computed in terms of node communication cost, precision, detection accuracy and number of faulty sensor nodes tolerated in the network. It adopts cluster technology for data aggregation and avoids redundant data. To have traffic free environment, this scheme is opted. It alleviates to control congestion at node level particularly at intermediate nodes.

## References

1. Wang, Q.: Traffic Analysis, Modeling and Their Applications in Energy-Constrained Wireless Sensor Networks-On Network Optimization and Anomaly Detection. Mid Sweden University, Sweden (2010) ISBN 978-91-86073-64-0, ISSN 1652-893x
2. Deshpande, V., Sarode, P., Sarode, S.: EDCAM-Early Detection Congestion Avoidance Mechanism. Inter. Journal of Comp. Application 7(2), 11–14 No 18, Article 6
3. Sankarsubramaniam, Y., Bakan, O., Akyildiz, I.F.: ESRT: Event to sink reliable transport in Wireless Sensor Network. In: Proc. of ACM MobiHoc 2003 (2003)
4. Kulik, J., Rabiner, W., Balakrishnan, H.: Adaptive protocols for information dissemination in wsn. In: Proc. ACM MobiCom, Seattle, WA, pp. 174–185 (August 1999)

5. Hull, B., Jamieson, K., Balakrishnan, H.: Mitigating congestion in wireless sensor networks. In: Proc. 2nd ACM Conf. Embedded Netw. Sensys, Baltimore, MD (November 2004)
6. Chen, S., Fang, Y., Xia, Y.: Lexicographic maxmin fairness for data collection in wireless sensor networks. *IEEE Trans. Mobile Comput.* 6(7) (July 2007)
7. Shih, E., Cho, S.: Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. In: Proc. ACM MobiCom, Rome, Italy, pp. 272–287 (July 2001)
8. Wan, C.-Y., Eisenman, S.B., Campbell, A.T.: CODA: Congestiondetection and avoidance in sensor networks. In: Proc. ACM SenSys (November 2003)
9. Chen, H., Yao, D.D.: *Fundamentals of Queuing Networks: Performance, Asymptotics, and Optimization. Applications of Mathematics* 46 (2001)

# An Intelligent Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks

K.S.S. Anupama<sup>1</sup>, S. Sri Gowri<sup>2</sup>, B. Prabakara Rao<sup>3</sup>, and T. Satya Murali<sup>1</sup>

<sup>1</sup> Department of EIE, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh

<sup>2</sup> Department of ECE, SRK Institute of Technology, Vijayawada, Andhra Pradesh

<sup>3</sup> Department of ECE, JNTU, Kakinada, Andhra Pradesh

**Abstract.** The next generation wireless systems will consist of heterogeneous networks from cellular, Wi-Fi, WiMAX to other emerging access technologies. Vertical handoff occurs when a mobile terminal decides to switch between networks. One of the challenging problems during vertical handoff is the selection of an optimal network that maximizes end users satisfaction. This paper presents an intelligent vertical handoff decision algorithm that selects the target network based on the traffic class of the mobile user. The algorithm uses two modules, to estimate the handoff requirement and to select the optimal network. These modules utilize Fuzzy Logic and Genetic algorithm to make an intelligent vertical handoff decision.

## 1 Introduction

In the next generation networks also called as heterogeneous networks, various wireless and cellular technologies are seamlessly integrated to provide ubiquitous coverage and high quality of service anywhere and anytime. The integration and interoperation of heterogeneous wireless networks poses several challenges. One of the major challenges is Vertical Handoff. Handoff is a process of transferring an ongoing call or data session from one access point to another in a homogeneous or heterogeneous network. Vertical handoff takes place between wireless networks of diverse technologies [1] [2]. Since the diverse networks overlap each other many metrics such as network conditions, system performance, application requirements cost and user preferences should be considered in vertical handoff decision. When the aforementioned requirements are regarded, one can easily deduce that design of vertical handoff algorithm is complex and challenging.

In this paper an Intelligent Vertical Handoff Decision (IVHD) algorithm is proposed to perform handoff to the most suitable network depending on the traffic class of the Mobile Subscriber (MS). The IVHD algorithm has two modules: Handoff initiation and Handoff decision. Handoff initiation module uses fuzzy inference system to decide whether handoff is required or not. The handoff decision module uses Genetic algorithm for assigning weights to various input parameters and computes the Optimal Network Selection Factor (ONSF). Finally the network with highest ONSF is selected as the target network for vertical handoff. The

remaining part of the paper is organized as follows: Section 2 reviews related work. In section 3 the proposed scheme is explained. Section 4 discusses the simulation results. Finally, concluding remarks are drawn in Section 5.

## 2 Related Work

A number of proposals have been found for vertical handoff decision algorithms in the literature. In [3], a cost function based vertical handoff decision algorithm for multi services was proposed. The network that has the lowest cost function value is chosen as the target network. But the quality of service (QoS) requirements for each traffic class is not considered.

In [4], a preference value-based cell selection scheme is proposed to maintain QoS requirements of a call request maximize accommodated number of calls and minimize handoff occurrence frequency. In [5], different fuzzy MADM algorithms are compared but the selected weights and some of the results are disputed. In [6], a vertical handoff decision algorithm that enables the access network to balance load among all attachment points was proposed.

In [7], the authors proposed a vertical handoff decision algorithm between WLAN and CDMA networks. In handoff initiation RSS is used for the avoidance of unnecessary handoffs. In [8], a multi criteria decision making algorithm based on fuzzy theory for access network selection is presented. But it only gave theoretic description for mobility management in heterogeneous networks.

In [9], a fuzzy logic based algorithm that adopts the RSS threshold, bandwidth and cost values as input parameters were proposed. The weight of each QoS metric is adjusted to trace the network condition.

The challenges of Vertical handoff decision in heterogeneous networks considering multiple criteria is quite complex. Sometimes, the trade-off of some criteria should be considered. Therefore, in this paper heuristic approach based on Genetic Algorithms (GA) and Fuzzy Logic (FL) is proposed to deal with imprecise and contradictory attributes of candidate networks and improve quality of service to the end users.

## 3 Intelligent Vertical Handoff Decision Algorithm (IVHD)

The application scenario for the proposed IVHO scheme is a heterogeneous network consisting of Wi-Fi, WiMAX and CDMA networks. WiMAX network supports several applications in one platform with large coverage areas and provides good bandwidth. Wi-Fi offers relatively high data rates in smaller areas at low cost and CDMA network supports low data rates in a much wider area of coverage. The complimentary characteristics of Wi-Fi, WiMAX and CDMA networks make it attractive to integrate these wireless technologies as heterogeneous wireless networks. Vertical handoff decision is crucial to the integration of heterogeneous wireless networks.

The proposed IVHD algorithm is shown in Fig.1. The IVHD algorithm has two modules: Handoff initiation and Handoff decision. In this paper, Received Signal

Strength (RSS), Bandwidth (BW), Cost of service (C), User Preference (UP) and Velocity (V) of the mobile are considered as the vertical handoff decision criteria. The criteria are normalized to a value between 0 and 1 to achieve fair and effective comparison between received parameters of different networks. Normalization is required to ensure that the values in different units are meaningful.

RSS is the strength of the signal received from the candidate network. The normalization function for RSS is given by equation (1):

$$\begin{aligned}
 N_{RSS} &= 0, \text{ for } 0 \leq RSS \leq RSS_{th} \dots \dots \dots (1) \\
 \text{For } RSS > RSS_{th} & \quad 0 \leq RSS \leq RSS_{th} \\
 N_{RSS} &= (RSS_x - RSS_{th}) / (RSS_{max} - RSS_{th})
 \end{aligned}$$

Where  $RSS_x$  is the actual strength of the signal received from the candidate base station.  $RSS_{th}$  is defined as the threshold signal strength.  $RSS_{max}$  is the maximum RSS that can be received from a candidate base station.

BW is the available bandwidth of the candidate network. The normalization function for BW [10] is given by equation (2):

$$\begin{aligned}
 N_{BW} &= 0, \text{ for } B_x > B_{max} \dots \dots \dots (2) \\
 &= B_x / B_{max}, \text{ for } 0 \leq B_x \leq B_{max}
 \end{aligned}$$

Where  $B_x$  is the available bandwidth of the base station and  $B_{max}$  is the maximum bandwidth that can be provided by the base station.

The normalization function for cost [10] is given by equation (3):

$$\begin{aligned}
 N_C &= 0, \text{ for } C_x > C_{th} \dots \dots \dots (3) \\
 &= 1 - (C_x / C_{th}), \text{ for } 0 \leq C_x < C_{th}
 \end{aligned}$$

Where  $C_x$  is the operating cost of the network to which the candidate base station belongs to.  $C_{th}$  is the threshold cost, above which it is considered that the network is expensive.

The normalization function for velocity is given by equation (4).

$$\begin{aligned}
 N_V &= 1, \text{ for } V_x > V_{max} \dots \dots \dots (4) \\
 &= (V_x / V_{max}), \text{ for } 0 \leq V_x \leq V_{max}
 \end{aligned}$$

$V_x$  is the velocity with which the mobile is moving and  $V_{max}$  is the maximum velocity supported by the candidate base station.

The normalization function for user preference is given by equation (5).

$$\begin{aligned}
 N_{UP} &= 0, \text{ for } P_x > P_{max} \dots \dots \dots (5) \\
 &= P_x / P_{max}, \text{ for } 0 \leq P_x \leq P_{max}
 \end{aligned}$$

From user point of view the normalized value of RSS, BW, UP and V should be as high as possible and that of C should be as low as possible.

### 3.1 Vertical Handoff Initiation Module

A key issue in vertical handoff is the ability to correctly decide at any given time whether or not to carry out vertical handoff. The handoff necessity estimation is



important in order to keep the unnecessary handoffs and their failures at a low level. In this module, RSS from the source network and velocity of the mobile are considered as handoff metrics. RSS is the most widely used criterion because it is easy to measure and is directly related to the service quality. Velocity of the mobile should also be considered. Because in heterogeneous networks, handing off to small cells while travelling at high speeds is discouraged, since a handoff back to the original network would occur very shortly afterwards.

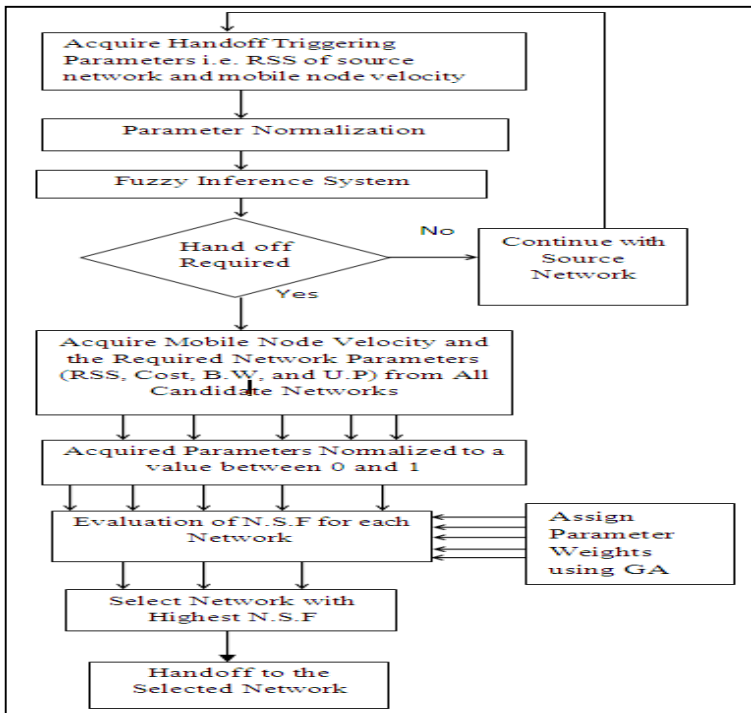


Fig. 1. IVHD algorithm

The Handoff Initiation module uses a Mamdani fuzzy logic inference system (FIS) to process the vertical handoff initiation parameters. The velocity of the mobile and RSS from the current and available networks are mapped into normalized values by using equations (4) and (1). The normalized values are then fed to the fuzzifier. The role of the fuzzifier is to transform real time measurements into fuzzy sets using membership functions. Trapezoidal membership functions are chosen due to their capability of achieving better performance especially for real time applications [11].

In this module the fuzzifier transforms RSS and Velocity into one of the following fuzzy sets: Very Low, Low, Medium and High. The fuzzy set very low is defined from 0 to 0.0625 with maximum membership function from 0 to 0.0375. The fuzzy set low is defined from 0.0375 to 0.4375 with maximum membership function from 0.0625 to 0.3125. The fuzzy set medium is defined from 0.3125 to 0.75 with

maximum membership function from 0.4375 to 0.625. The fuzzy set high is defined from 0.625 to 1 with maximum membership function from 0.6875 to 0.9375. The fuzzy sets are then fed to a fuzzy inference engine which outputs a linguistic variable using a set of IF-THEN rules. Some of the IF-THEN rules are defined as follows:

- IF RSS from current network is Very Low, and RSS from available network is Low, and velocity is Low THEN handoff not required.
- IF RSS from current network is Low, and RSS from available network is High and velocity is Medium THEN handoff partially required.
- IF RSS from current network is Very Low, and RSS from available network is High, and velocity is High THEN handoff highly required.

The output of the inference engine is a single fuzzy variable which is passed on to the defuzzifier, to be converted into a precise quantity called Handoff Requirement Factor (HRF). The range of HRF is from 0-100.

The output of the fuzzy system, HRF is then compared to a certain threshold constant of 35 to make handoff decision. If HRF value is greater than the threshold constant, handoff decision process is triggered. The algorithm then enters the VHO handoff decision module, where the target network for the future connection is determined. Changing the threshold value is a tradeoff factor which helps to balance the system.

### 3.2 Vertical Handoff Decision Module

In this module an optimal network selection factor (ONSF) is defined for all alternative target networks. The ONSF is a measure of the improvement in quality of service gained by mobile users, by handing off to a particular network. The network that provides the highest ONSF value is selected as the best network for handoff.

In the first stage of vertical handoff decision module, the input parameters RSS, Bandwidth, Velocity, User preference and Cost of service from all the surrounding networks are normalized by the fuzzy inference system. The Genetic algorithm (GA) then assigns weights to the normalized inputs based on the specifications of the traffic class.

The Genetic algorithm is a search method for solving both constrained and unconstrained optimization problems that is based on the principles of natural selection and genetics [12]. Genetic algorithms operate on encoded representations of the solutions also called as chromosomes. Each solution is associated with an objective function that reflects how good it is compared with other solutions in the population. In this algorithm two objective functions are defined, one for each traffic class: streaming and voice. Here a function minimization problem is considered. Hence, a good solution is one that has low relative fitness. Once the fitness functions are defined, the GA proceeds to initialize a population of solutions and then improve it through the following steps:

**Selection:** The primary objective of the selection process is to emphasize the good solutions and eliminate the bad solutions in a population while keeping the population

size constant. For selecting potentially useful solutions recombination roulette wheel selection technique is used.

Crossover: is a process of taking more than one parent solutions and producing a child solution from them. Two point crossovers are used in this algorithm.

Mutation: For maintaining diversity between the populations of chromosomes, uniform mutation function is used.

The current population is replaced with the children created by selection, crossover, and mutation to form the next generation. For obtaining an effective output, the optimization of any objective function is carried out for 300 generations. In this module the algorithm is programmed to stop if there is no improvement in fitness value for 300seconds or for 100generations continuously. All other GA options like population size, crossover probability and mutation probability are set to default values.

The objective functions are optimized with Genetic algorithm to assign weights to normalized inputs. The optimization of streaming objective function is carried out in such a way that more weight is assigned to the bandwidth parameter, when compared to other input parameters. Similarly the optimization of voice objective function assigns relatively more weight to RSS parameter. The weights generated by the Genetic algorithm are used for computing ONSF value of Wi-Fi, WiMAX and CDMA networks.

$$\text{ONSF} = \sum_{n=1}^5 N_n * W_n$$

Where  $N_n$ =normalized value of the nth parameter

$W_n$ = weight vector of the nth parameter

$n$ = RSS, BW, V, UP, C

The network that provides the highest ONSF value is selected as the target network for vertical handoff.

## 4 Experimental Results and Observations

The performance of the intelligent vertical handoff decision algorithm is tested within the framework of three diverse networks Wi-Fi, WiMAX and CDMA that cover the whole simulation area.

In this work, considered a situation where MS is moving through the heterogeneous network. When the MS enters into the overlapping area, a new link will be detected and the MS will start the vertical handoff process. An optimal network should be selected depending on the ongoing application of the MS.

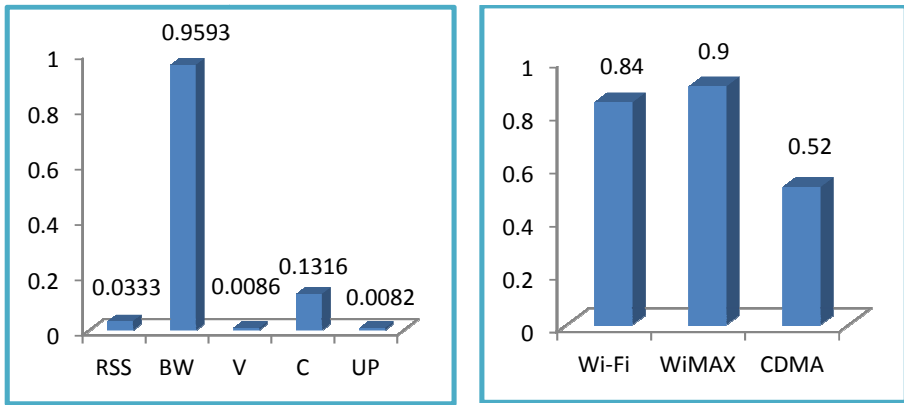
Initially the MS is connected to WiMAX network and is moving with a velocity of 35kmph. The inputs acquired by MS are normalized and are listed in Table 1.

In the case of streaming traffic the handoff initiation module generated a HRF of 50. As this value is greater than the threshold, handoff decision process is triggered. For streaming traffic, band width is the key parameter. Accordingly the Genetic algorithm assigns more weight to the bandwidth parameter, when compared to other normalized input parameters. Then ONSF is computed for various networks. For

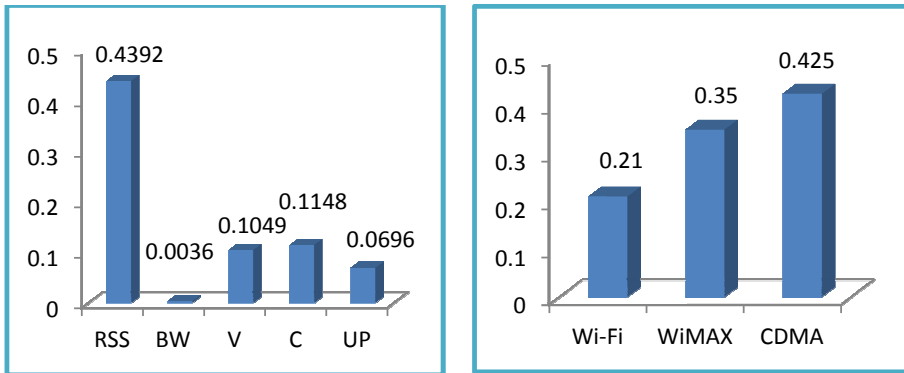
**Table 1.** Normalized input parameters

Network	RSS	BW	Cost	UP	Velocity (actual)
Wi-Fi	0.3	0.74	0.2	0.2	35kmph
WiMAX	0.5	0.75	0.4	0.6	35kmph
CDMA	0.5	0.4	0.7	0.3	35kmph

streaming application the weights assigned by Genetic algorithm and ONSF values of various networks are shown in Fig.2. From Fig.2, it is concluded that WiMAX network with maximum ONSF value is selected as the target network.



**Fig. 2.** Weights and ONSF values for streaming traffic



**Fig. 3.** Weights and ONSF values for voice traffic

For voice application, RSS is the key parameter. Accordingly the Genetic algorithm assigns more weight to the RSS parameter, when compared to other normalized input parameters. Then ONSF is computed for various networks. For voice application the weights assigned by Genetic algorithm and ONSF values of various networks are shown in Fig.3. From Fig.3, it is concluded that CDMA network with maximum ONSF value is selected as the target network.

## 5 Conclusions

Vertical handoff decision making is one of the most challenging issues in heterogeneous wireless networks. This paper has proposed an Intelligent Vertical Handoff Decision algorithm for optimal network selection. Parameters like RSS, Bandwidth, and Velocity, Cost, and User preference are chosen as the decision criteria. The fuzzy inference system evaluates RSS of the source and target network and velocity of the mobile to initiate handoff. Genetic algorithm assigns weights to the parameters depending on the traffic class of an ongoing session. Finally ONSF is computed. The network with highest ONSF is selected for handoff. The experimental results indicate that the proposed IVHD algorithm can make accurate handoff decisions thereby maximizing end users satisfaction. Future work includes the use of multi objective genetic algorithm to optimize the weights of input parameters for different traffic classes.

## References

1. Makaya, C., Pierre, S.: An Architecture for Seamless Mobility Support on IP-Based Next Generation Wireless Networks. *IEEE Transactions on Vehicular Technology* 57(2), 1209–1225 (2008)
2. Ferrus, R., Sallent, O., Agush, R.: Interworking in Heterogenous Wireless Networks: Comprehensive Framework and Future Trends. *IEEE Wireless Communications*, 22–31 (April 2010)
3. Zhu, F., McNair, J.: Optimizations for Vertical Handoff Decision Algorithm. In: *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 2 (March 2004)
4. Chang, C.J., Hsieh, C.Y., Chen, Y.H.: A Preference Value-Based Cell Selection Scheme in Heterogeneous Wireless Networks. In: *IEEE Wireless Communications and Networking Conference (WCNC)*, Sydney, Australia, pp. 1–6 (April 2010)
5. Zhang, W.: Handover Decision Using Fuzzy MADM in Heterogeneous Networks. In: *WCNC IEEE Communications Society*, pp. 653–658 (March 2004)
6. Lee, S., Sriram, K., Kim, K., Golmie, N.: Vertical Handoff Decision Algorithms for Providing Optimized Performance in Heterogeneous Wireless Networks. *IEEE Transactions on Vehicular Technology* 58(2) (February 2009)
7. Jang, K.S., Kim, K.S., Shin, H.J., Shin, D.R.: A Novel Vertical Handoff Strategy for Integrated IEEE 802.11 WLAN/CDMA Networks. In: *Fifth IEEE International Symposium on Network Computing and Applications (NCA 2006)*, pp. 221–230 (July 2006)

8. Hongyan, B., Chon, H., Lingee, J.: Intelligent Signal Processing of Mobility Management for Heterogenous Networks. In: IEEE International Conference Neural Networks and Signal Processing, China, pp. 1578–1581
9. Xia, L., Lingase, J., Chen, H., Wei, L.H.: An intelligent Vertical handoff Algorithm in Heterogenous Wireless Networks. In: Wireless Communications, Networking and Mobile Computing, WICOM 2008, pp. 1–3 (2008)
10. Liao, H., Tie, L., Du, Z.: A Vertical Handover Decision Algorithm Based on Fuzzy Control Theory. In: Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences, IMSCCS 2006 (2006)
11. Ling, Y., Li, B., Ziu, Q.: Vertical Handoff Decision Strategy in Wireless Overlay Networks. In: Proc. WICOM 2009, pp. 1–3. IEEE Press (September 2009)
12. Goldberg, D.E.: Genetic Algorithm in Search Optimization and Machine learning. Addison-Wesley, Reading (1989)

# Energy Efficient and Reliable Transmission of Data in Wireless Sensor Networks

Chilukuri Shanti and Anirudha Sahoo

Department of Computer Science and Engineering, Indian Institute of Technology, Bombay,  
Powai, Mumbai, India  
{shanti06, saho0}@cse.iitb.ac.in

**Abstract.** Reliable transmission of data over the lossy wireless medium can be achieved by using an acknowledgement (ack) scheme. However, acks can be quite energy-consuming and should be used sparingly in energy-constrained networks. The simplest of acknowledgement schemes is the stop-and-wait ARQ (swack). Recent work proposes an implicit acknowledgement scheme (iack) which seems to be energy-saving, as there are no separate acknowledgements sent. IEEE 802.11e proposes a hop-by-hop block ack (eack). We modify this scheme to suit WSNs and propose that the choice of the ack scheme that would keep the network intact for the longest time depends on the tree overlay and the data and ack packet lengths. We theoretically compare the maximum energy spent for sending and receiving acknowledgements for these three types of ack schemes. We then give a guideline to choose the most energy-saving of these three types of acknowledgements for convergecast in a wireless sensor network (WSN) applications with a routing tree overlay and continuous event generation rate.

## 1 Introduction

Recently, there has been a lot of interest in designing energy-efficient protocols for Wireless Sensor Networks (WSNs). Due to the inherent lossy nature of the wireless medium and the harsh environment in which WSNs are often deployed, such networks also have high packet error rates ([4], [6]). Most applications of WSNs like [13] and [15] require reliable data delivery. Several schemes exist for reliability of data delivery in lossy scenarios. Reliability is often achieved by acknowledgements at the transport or MAC layer and hence, comes at the cost of energy. Hop-by-hop stop-and-wait (SW) automatic repeat request (ARQ) protocols at the MAC level are simple and have low buffer requirements, but consume a lot of energy and increase the latency of data delivery. A better scheme is the iack (implicit acknowledgement) scheme, where the broadcast nature of the wireless medium is exploited to save energy. In this scheme, a node overhears when its data packets are relayed by the next hop node in a multihop scenario and treats this as an acknowledgement. Hence there are no separate acknowledgements sent explicitly, resulting in saving a lot of power. The iack scheme may seem to be very energy efficient as no acknowledgements need to be sent. However, considerable energy is spent in overhearing the data that is forwarded (*iacks*). Power used for reception of data is in fact higher than the power spent for transmitting data for some nodes ([1]). In this paper, we consider a hop-by-hop explicit block ack scheme (eack) using a bitmap at the MAC level. In doing so, we exploit the fact that WSNs usually have overlay of

a routing tree in its topology, where the sink is the root and all other nodes transmit data to the sink in one or more hops, along the edges of the tree. The eack scheme is somewhat similar to the IEEE 802.11e block ack scheme, with the difference that we consider a tree overlay and a single eack is sent to all the children of a node. The format of the eack is different from that of IEEE 802.11e. Further, in IEEE 802.11e, block acks are sent only after a block ack request (BAR) message is received from the transmitter of data, as the default ack mechanism of IEEE 802.11e is not the block ack scheme. BAR messages consume some energy. Hence, we consider a hypothetical system where the block ack is the default ack scheme for a network and there is no need for BAR messages. This is similar to 802.11n, which can use block ack as the default. However, as [9] points out, the size of a block has a significant impact on the throughput. We first propose a mechanism to determine the block size in a WSN with a tree overlay. Then, we derive expressions to calculate the maximum power spent in transmitting and receiving acknowledgements when the stop- and-wait, implicit ack or eack schemes are used. We note that the choice of the scheme that is most energy-efficient is dependent on the network topology and the data and ack packet lengths. We substantiate our theory with simulation results run for different topologies. To the best of our knowledge, these ack schemes had not been compared till now, especially based on the power spent. The application considered by us is a constant rate data streaming application like that in [10].

## 2 Related Work

Acknowledgements have been considered to provide reliable data transfer by many authors in the past. [14] gives a comprehensive comparison of various acknowledgement schemes. The authors conclude that for energy efficiency, hop-by-hop acknowledgements are better than end-to-end acknowledgements. Hence, we consider hop-by-hop acknowledgements in our scheme. [12] proposes a selective NACK-based transport protocol which is tightly bound to the directed diffusion routing protocol. In this paper, Heidemann et al. propose that for a WSN with high error rates, a NACK-based transport protocol running over a selective ARQ based MAC layer is an appropriate solution. The eack scheme proposed by us is essentially a selective ARQ-based MAC layer. In [8], the authors propose an energy efficient, reliable transport protocol (ERTP) that uses hop-by-hop iack approach and duplicate detection. For this, they consider a low rate data streaming application. In [15], the authors propose a windowless block ack scheme for bursty convergecast. The goal of this paper is to improve the channel utilisation and packet delivery delay for realtime bursty traffic, but energy efficiency is not considered by the authors. In [4] and [9], the authors analyse the performance of the 802.11e block ack mechanism in a noisy channel. These papers conclude that the the block ack scheme achieves higher throughput than the legacy DCF of 802.11e. [10] compares the energy efficiency of the stop-and-wait ARQ and the iack schemes at the MAC level on a hop-by-hop basis. The main contribution of this paper is to propose ways to reduce the avalanche of iacks that may ensue as a result of a lost iack, with the aim of reducing the energy spent. We compare the eack scheme with one of the solutions (oriented iack) proposed in this paper.



### 3 Topology

We model the WSN as a connectivity graph with vertices representing nodes and edges representing the wireless links between the nodes. We assume an overlay of routing tree on this network, with the base station (or sink) at the root of the tree. Nodes within reception radius of the root are the direct children of the root and are said to be at level one. A node is said to belong to the  $i^{th}$  level if the node is at a distance of  $i$  hops from the root in the routing tree. This tree overlay needs to be formed only once, during the initial configuration phase by using appropriate algorithm, e.g., algorithms presented in [11] and [7]. Leaf nodes transmit one packet of sensed data at a time. A non-leaf node receives these packets of data from its children. It then transmits all the packets received from the children and the packet of data sensed by it as a single burst of packets to its parent. Hence, the maximum burst of a node at  $level_i$  essentially consists of one packet sent by each descendant node and one packet of its own. Descendant nodes of a given node are all the nodes in the subtree rooted at that node. Sending a burst of packets like this reduces the energy spent in channel contention if a CSMA MAC is used [2]. The  $i^{th}$  level has  $M^i$  nodes. Each node at  $level_i$  is denoted by  $node_{i,j}$ ,  $j > 0$ . We assume that they are indexed from left to right in the tree overlay. We denote the number of children of the  $j^{th}$  node of  $level_i$  as  $n_{i,j}$  and the number of packets transmitted by this node as  $a_{i,j}$ . Note that  $a_{i,j}$  is also the number of nodes in the subtree of  $node_{i,j}$ . Let the maximum number of levels in the tree excluding the root be  $H$ . For leaf nodes which are at  $level_H$ ,  $a_{H,j}$  is 1, since each node transmits only the data sensed by it. The maximum number of children of any node at  $level_i$  is denoted by  $N_i$ .

Source addr. (16 bits)	Dest. addr. (16 bits)	Seq. No. (16 bits)	ack bit (1 bit)	Payload
---------------------------	--------------------------	-----------------------	--------------------	---------

Fig. 1. Format of the Data Packet

addr <sub>1</sub>	len <sub>1</sub>	seq <sub>1</sub>	bitmap <sub>1</sub>	addr <sub>2</sub>	...	addr <sub>nl</sub>	len <sub>nl</sub>	seq <sub>nl</sub>	bitmap <sub>nl</sub>
-------------------	------------------	------------------	---------------------	-------------------	-----	--------------------	-------------------	-------------------	----------------------

Fig. 2. Format of the Bitmap Block Ack

### 4 Acknowledgement Schemes

Acknowledgement schemes provide reliability and can be of various types. In this paper, we propose a selective repeat block ack (eack) scheme and compare it with a simple stop-and-wait ack (swack) scheme and an implicit acknowledgement (iack) scheme. [14] gives a comprehensive comparison of various acknowledgement schemes and concludes that for energy efficiency, hop-by-hop acknowledgements are better than end-to-end acknowledgements. Hence, we consider hop-by-hop acknowledgements in all the three schemes. The notation used is given in Table 1.

#### 4.1 The Implicit Acknowledgement Scheme

In this scheme, the nodes take advantage of the wireless medium to do away with explicit acknowledgements. A node at  $level_i$  transmits a packet to its parent, sets the retransmission timer and stays awake to overhear this packet being relayed by the par-

**Table 1.** Notation used

Notation	Definition	Notation	Definition
$a_{i,j}$	max. number of packets in a burst sent by $node_{i,j}$	$M_i$	number of nodes of level $i$
$A$	max. number of packets in a burst by any node	$n_{i,j}$	number of children of $node_{i,j}$
$N_i$	max. number of children of any node at the $i^{th}$ level	$\alpha_{i,j}$	max. length of block ack sent by $node_{i,j}$
$N_{max}$	max. number of children of any node	$\beta$	length of the block ack field for a given network
$data$	length of the data packet in bits	$b$	length of the node address field
$swack$	length of the ack packet in the stop-and-wait scheme	$c$	length of the $len$ number field in bits
$d$	length of the sequence number field in bits	$\rho$	number of nodes in the network

ent. The overheard data packet serves as the implicit acknowledgement (iack). If the node does not hear the packet being relayed by the parent node before the retransmission timer goes off, the packet is retransmitted. The base station is the only node that sends an explicit acknowledgement to its children, as it does not relay the data further. Nodes save energy spent for ack transmission, but have to stay awake to listen for iacks, spending considerable energy.

## 4.2 The Explicit Acknowledgement Scheme

In this section, we propose a block acknowledgement scheme for WSNs. As per this scheme, a node transmits an explicit block acknowledgement (eack) using a bitmap just before transmitting its burst of data packets. The burst of data packets consists of one sensed data packet of the node itself and data packets received by the node from its children to be relayed to the next hop. The eack is received by all the children of this node and can serve as an acknowledgement for all the packets transmitted by each child in its burst of data. The data packet format considered by us is given in Figure 1. The format of the eack is as shown in Figure 2. The  $addr$  field is the address of the child whose packets are acknowledged. We denote the number of children of  $node_{i,j}$  to be  $n_{i,j}$  and  $addr_k$  is the address of the  $k^{th}$  child of  $node_{i,j}$ . The length of the bitmap ack (in bits) to this child is denoted as  $len_k$ . Let the first child of  $node_{i,j}$  be the  $m^{th}$  node of  $level_{i+1}$ . Hence, the  $n_{i,j}$  number of children of this node are nodes  $m$  to  $m + n_{i,j} - 1$  of the  $(i + 1)^{th}$  level and the  $k^{th}$  child of this node is  $node_{i+1,m+k-1}$ . Since the  $k^{th}$  child of a node in  $level_i$  belongs to  $level_{i+1}$ , it transmits a maximum of  $a_{i+1,m+k-1}$  packets in each burst. Hence, the maximum value of  $len_k$  is  $a_{i+1,m+k-1}$ . The sequence number of the first packet being acknowledged is  $seq_k$  and  $bitmap_k$  is the block ack to the  $k^{th}$  child in the form of a bitmap. The bitmap is formed with a 1 in the  $l^{th}$  ( $l > 0$ ) bit of the bitmap if the packet with sequence number  $seq_k + l - 1$  has been received successfully; otherwise the bit is set to 0. Hence, an eack with  $seq_k$  and  $len_k$  serves as an acknowledgement for  $len_k$  packets with consecutive sequence numbers starting from  $seq_k$ . An eack has such bitmap acks for  $n_{i,j}$  children each of which sends a burst of  $a_{i+1,m+k-1}$  packets. Each eack thus serves as an acknowledgement for a burst of data packets. This greatly reduces the energy consumption compared to the simple stop-and-wait scheme where each data packet requires a separate acknowledgement. Since a node of  $level_i$  transmits maximum of  $a_{i+1,m+k-1}$  packets of each of its  $n_{i,j}$  children and one of its own data packet in a burst, we have

$$a_{i,j} = 1 + \sum_{k=1}^{n_{i,j}} a_{i+1,m+k-1}. \quad (1)$$

Note that  $a_{i,j}$  is also the number of nodes in the subtree of  $node_{i,j}$ . The  $addr$ ,  $len$  and  $seq$  fields are of fixed length for a given routing tree overlay. The length of address field depends on the number of nodes in the network,  $\rho$ .  $seq$  depends on the sequence number space. The lengths of  $len$  and  $seq$  fields are design parameters of a network. The number of children of each node is not the same across the entire network. Also, the length of the bitmap for each child varies depending on the number of packets successfully received. Hence, the number of bits in  $len$  also varies. In order to have the same ack size across the entire network, we consider the node with the maximum number (say  $N$ ) of children across the entire network. To calculate the size of the bitmap field, we consider the maximum number of packets in a burst (say  $A$ ). Hence,

$$N = \max_{1 \leq i < H} \{N_i\} \quad \text{and} \quad A = \max_{1 \leq i < H, 1 \leq j \leq M_i} \{a_{i,j}\}.$$

When a node transmits its burst of packets, it sets the retransmission timer and waits for the eack. The format of the eack is such that minimum energy is spent in sending and receiving the eack. When a node receives an eack, it checks if its address is in  $addr_1$ . If so, it reads the  $seq_1$  and  $len_1$  fields and then scans the next  $len_1$  bits for the bitmap ack. Packets with sequence numbers  $seq_1 + l$  are not acknowledged if the  $l^{th}$  bit of the bitmap is 0, and have to be retransmitted. If  $addr_1$  is not the address of the node, it has to skip  $x_1$  number of bits, where

$$x_1 = (b + c + d + len_1) \quad (2)$$

In this equation,  $b$ ,  $c$  and  $d$  are the lengths of the  $addr$ ,  $len$  and  $seq$  fields respectively. After skipping  $x_1$  bits, the node checks for its address in the next  $addr$  field and repeats the process till it finds its address in the eack or it has finished parsing the eack. If it receives the eack and notes that a few of the packets have been lost, it can retransmit only the lost packets. In case the eack is not received before the timer goes off or the node does not find its address in the eack, it retransmits all the  $len_j$  packets. Though the maximum size of the eack is based on the node that has the maximum number of children, the actual length of the eack sent depends on the number of children from which a node has received packets in a burst. Let  $node_{i,j}$  have  $n_{i,j}$  children and the first child starts from index  $m$  (in  $level_{i+1}$ ). Hence, the maximum size of the eack sent by the  $j^{th}$  node at the  $i^{th}$  level is given by

$$\alpha_{i,j} = \sum_{k=1}^{n_{i,j}} (b + c + d + a_{i+1,m+k-1}), \quad 2^{(b-1)} < \rho \leq 2^b, \quad 2^{(c-1)} < A \leq 2^c \quad (3)$$

The length of the eack for the network is determined by the level that needs the longest ack. If  $M_i$  is the number of nodes at  $level_i$ ,  $\beta = \max \{\alpha_{i,j}\}, 1 \leq i < H, 1 \leq j \leq M_i$ .

## 5 Energy Consumption

In this section, we calculate the maximum energy spent by any node in the network when different ack schemes are deployed. Since the energy spent by a node for transmission and reception of data is the same in all the three schemes, we consider only the energy spent for transmitting and receiving acknowledgements to compare the three schemes. We derive expressions for the power consumed to transmit/receive a packet. The energy consumed is directly proportional to the power, as it can be obtained by multiplying the power with the bit duration (which is fixed for a given data transmission rate). In the event of loss of data packets, all three schemes *selectively repeat* only the lost packet, after timeout (in the stop-and-wait or iack schemes) or upon parsing the bitmap ack (in the eack scheme). In the event of loss of an acknowledgement, the stop-and-wait and iack schemes cause only the packet whose ack (explicit or implicit) has been lost to be retransmitted. However, loss of a block ack in eack scheme causes the nodes child to retransmit the entire burst of packets sent by it.

Let *swack* be the number of bits in the ack when the stop and wait scheme is used.  $node_{i,j}$  transmits a maximum of  $a_{i,j}$  data packets and receives a maximum of  $(a_{i,j} - 1)$  data packets in a burst. The maximum power spent by  $node_{i,j}$  for transmitting and receiving acks when the stop-and-wait scheme is used is

$$P_{sw}^{i,j} = swack * (P_{tx} * (a_{i,j} - 1) + P_{rx} * a_{i,j}) \quad (4)$$

In the iack scheme, since there are no separate acks to be transmitted, nodes spend power only for receiving the iacks. The minimum number of transmissions that a node of  $level_i$  has to overhear to know if the packets have been received is zero (when all packets are lost). However, when all the packets are transmitted successfully, the node has to overhear the  $a_{i,j}$  data packets which were transmitted by it, when they are being forwarded by its parent. Hence, the power spent by it for receiving iacks is as given below:

$$P_{iack}^{i,j} = P_{rx} * data * a_{i,j} \quad 2 \leq i \leq H \quad (5)$$

Equation 5 cannot be applied to the nodes in  $level_1$  since those nodes (children of the sink) receive one explicit acknowledgement (from the sink) for each data packet sent by them. Let the length of this explicit ack be the same as the ack in the stop-and-wait scheme (*swack* bits). Hence, the maximum power spent by  $node_{1,j}$  of  $level_1$  for receiving acks is

$$P_{iack}^{1,j} = P_{rx} * swack * a_{1,j} \quad (6)$$

When the eack scheme is used, each node spends energy to transmit acks and also to receive acks for the packets that are transmitted by it. In the worst case of energy expenditure,  $node_{i,j}$  receives  $a_{i+1,k}$  packets from each of its  $n_{i,j}$  children and transmits  $a_{i,j}$  packets. The maximum power spent by  $node_{i,j}$  to transmit the eack is

$$P_{eack\_tx}^{i,j} = P_{tx} * \alpha_{i,j} \quad (7)$$

The node spends power to listen to the acks for the data transmitted by it. The parent of a node of  $level_i$  belongs to  $level_{i-1}$ . There are a maximum of  $a_{i-1,p}$  nodes in this

parent's subtree and hence, it sends an eack of maximum length  $\alpha_{i-1,p}$ . Hence, the maximum power spent to receive the ack is

$$P_{eack\_rx}^{i,j} = P_{rx} * \alpha_{i-1,p} \quad (8)$$

Hence, the maximum power spent by  $node_{i,j}$  in sending and receiving eacks is given by

$$P_{eack}^{i,j} = P_{eack\_rx}^{i,j} + P_{eack\_tx}^{i,j} \quad (9)$$

From Equations 3, 7 and 8, it can be seen that the energy spent by a node to transmit and receive an eack depends on the number of children ( $n_{i,j}$ ), the number of nodes in the subtree rooted at the node ( $a_{i,j}$ ), the number of children of the parent node ( $n_{i-1,p}$ ) and the number of nodes in the subtree of the parent of the node ( $a_{i-1,p}$ ).

The energy spent by  $node_{i,j}$  for transmitting and receiving acks can be computed using Equations 4, 5, 6 or 9 (based on the ack scheme used). Let this be denoted as  $P_{ack}^{i,j}$ . The maximum power spent by any node in the network is given by

$$P_{max} = \max_{1 \leq i \leq H, 1 \leq j \leq N_i} (P_{data}^{i,j} + P_{ack}^{i,j}) \quad (10)$$

For each ack scheme maximum power spent by a node is computed using Equation 10. Then the ack scheme that consumes least power should be chosen, since this would result in the longest time till a node failure.

## 6 Simulation Results

We set up simulation to study the energy spent by the three acknowledgement schemes discussed in this paper. Various system parameters used in the simulation are given in Table 2. The MAC used is the TDMA MAC presented in [11]. When this MAC is used, each child knows the exact slots in which its data is relayed by the parent and can choose to stay awake only in these slots to overhear iack. In this case,  $node_{i,j}$  overhears only the  $a_{i,j}$  data packets which were transmitted by it, when they are being forwarded by its parent. Hence, the power spent by it for receiving iacks is as given below, instead of Equation 5.

$$P_{iack}^{i,j} = P_{rx} * data * a_{i,j}, \quad 2 \leq i < H \quad (11)$$

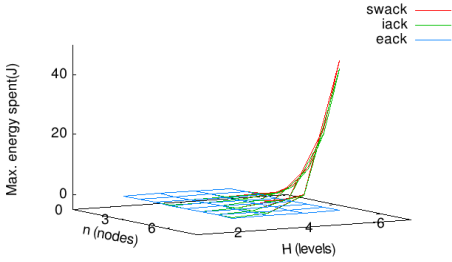
The values of the physical parameters of the nodes are taken from [5], which contains representative values for amps sensor nodes. For easy comparison of the ack

**Table 2.** Physical parameters used in simulation

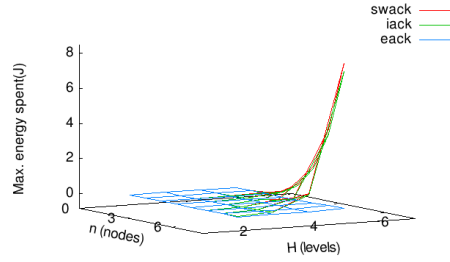
Parameter	Value	Parameter	Value	Parameter	Value
$P_{rx}$	63mW	$P_{tx}$	30mW	<i>TransitionPower</i>	30mW
$P_{idle}$	30mW	<i>TransitionTime</i>	2.45ms	<i>InitialNodeEnergy</i>	54,000J
$P_{sleep}$	0.003mW	Data transmission rate	19.2Kbps		

**Table 3.** Number of nodes below which eack is better for various cases

$p$	0%	10%	0%	50%	0%	50%	10%	50%
$H$	2	4	3	2	4	3	2	4
$n$	6	129	132	4	113	118	5	121



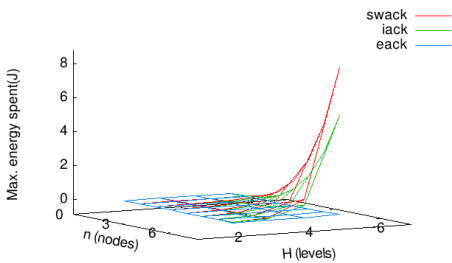
**Fig. 3.** Energy Consumption versus Number of Children and Subtree Size (data packet=64 bytes)



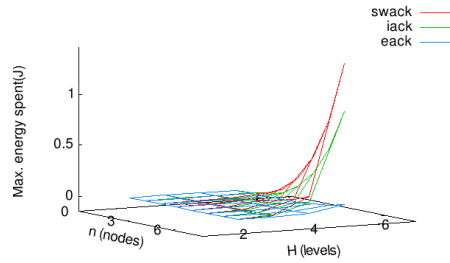
**Fig. 4.** Energy Consumption versus Number of Children and Subtree Size (data packet=8 bytes)

schemes, we considered a tree overlay where all non-leaf nodes have the same number of children, i.e.,  $n_{i,j}$  is the same across the network (let this be  $n$ ). Since the total number of levels in the tree is  $H$ , the number of levels below  $level_i$  is  $(Hi)$ . The size of  $(Hi)$  the subtree (including  $node_{i,j}$ ) can be given by recursion as  $a_{i,j} = \sum_{l=0}^{H-1} (n^l)$ .

We take  $H$  and  $n$  as two of our simulation parameters, as the subtree size depends on these. For each value of  $H$  and  $n$ , we find the maximum amount of energy spent by any node in the network using Equation 10. We first consider an ideal channel condition with zero packet loss probability. This means that there will be no retransmission of packets in the network. Figure 3 depicts the energy spent when different ack schemes are used, for number of levels in the tree varying from 2 to 5 and number of children varying from 1 to 8, when the data packet size is 64 bytes. This packet size is typical of TinyOS [3]. We use data packet size of 64 bytes for simulation. The acknowledgement packet size is taken to be 48 bytes. For these data and ack packet lengths, it can be seen



**Fig. 5.** Energy Consumption versus Number of Children and Subtree Size, when listening time is known (data packet=64 bytes)



**Fig. 6.** Energy Consumption versus Number of Children and Subtree Size, when listening time is known (data packet=8 bytes)

that the maximum energy consumed by a node in the network with the eack scheme is always lesser than that for the iack and the swack scheme.

Figure 4 plots the same performance metric, but for data and ack packet lengths of 8 bytes. Figures 5 and 6 depict the energy spent when the node knows when exactly its data is relayed by the parent. For these graphs, the energy spent by the iack scheme is given by Equations 6 and 11. In this case also, it can be seen that the eack scheme always consumes lesser energy than the swack or iack scheme when the data packet size is 64 bytes. However, for data packet size of 8 bytes, as the number of children increases, the iack scheme comes closer to the eack scheme and is less energy-consuming than the eack scheme beyond certain number of children per node. This is because the length of the block acknowledgement (relative to the packet length) increases with the number of children, resulting in the eack scheme consuming more energy.

As described in Section 5, the three schemes handle loss of ack or data packets differently. To demonstrate the effect of retransmissions as a result of loss of data or ack packets on the energy consumption, we simulated the three schemes with a packet loss probability (chosen randomly anywhere in the network) of 10% and 50%. We considered a low event generation rate such that nodes do not have fresher packets before the retransmission timer goes off, allowing packets to be retransmitted. It has been seen that for both data packet sizes considered (64 and 8 bytes), the eack scheme consumes lesser energy than the iack or swack schemes. When the nodes know when to listen, for data packet length of 64 bytes, eack is always less energy-consuming. For data packet length of 8 bytes, for higher number of children per node, the eack is relatively more energy-consuming, due to the increased size of the block ack. When the event generation rate is high such that all nodes have fresher packets before the retransmission timer goes off, the maximum burst sent by  $node_{i,j}$  is always of  $a_{i,j}$  packets. The energy consumption for any packet loss probability in this case would be same as the case of zero packet loss probability and is independent of packet loss probability. The energy consumption of nodes for such cases is identical to that shown in Figures 3, 4, 5 and 6. Table 3 shows the cross-over point (number of children per node,  $n$ ) below which the eack scheme is more energy-saving, for various loss probabilities ( $p$ ) and levels of the tree ( $H$ ). Note that all these results are for data packet length of 8 bytes and for the case where the child knows exactly when to listen to the iack. In the absence of this knowledge and for higher data packet lengths, the eack scheme is always less energy-consuming.

## 7 Conclusion

We proposed an explicit block ack scheme for WSNs with tree overlays and gave expressions for the energy spent when the traditional stop-and-wait, the implicit ack and our proposed eack schemes are used. The energy spent by each of the schemes depends on the topology of the tree overlay, the data and ack packet sizes. We presented simulation results that support our theory, when a TDMA MAC like that discussed in [11] is used. For larger packet sizes the eack scheme consumes lesser energy than the other two schemes for any network topology and packet loss probability considered by us (Figure 3). For very small data and ack packet sizes, we observe that the eack scheme is again better than the swack and iack schemes (Figure 4), when the child does not

know when to listen. Only in the case of iack when the child knows when to overhear, iack is better when the number of children is larger (Figure 6). Based on the number of children per node and packet loss probability, the methodology presented in this paper can be used to choose between the eack or iack scheme. In this study we found that the swack is a clear loser in terms of energy-consumption for all data and ack packet lengths, loss probabilities and topologies considered by us.

## References

1. Chipcon. Chipcon AS SmartRF(R) CC2420 Preliminary Datasheet (rev. 1.2) (June 2004)
2. IEEE 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications
3. TinyOS Community Forum — An open-source OS for the networked sensor regime, <http://www.tinyos.net/>
4. Dinh, T.L., Hu, W., Sikka, P., Corke, P., Overs, L., Brosnan, S.: Design and deployment of a remote robust sensor network: Experiences from an outdoor water quality monitoring network. In: LCN 2007: Proceedings of the 32nd IEEE Conference on Local Computer Networks, pp. 799–806 (2007)
5. Shih, E., Cho, S.-H., Ickes, N., Min, R., Sinha, A., Wang, A., Chandrasekharan, A.: Physical layer driven algorithm and protocol design for energy-efficient wireless sensor networks. In: IEEE International Conference on Mobile Computing and Networking (Mobicom), pp. 272–287 (2001)
6. Fitz, S., Gonzalez-Velazquez, A., Henning, I., Khan, T.: Experimental investigation of wireless link layer for multi-hop oceanographic-sensor networks. *Electronic Letters* 41, 1310–1311 (2005)
7. Kulkarni, S., Iyer, A., Rosenberg, C.: An address-light, integrated MAC and routing protocol for wireless sensor networks. *IEEE/ACM Transactions on Networking* 14(4), 793–806 (2006)
8. Le, T., Hu, W., Corke, P., Jha, S.: Ertp: Energy-efficient and reliable transport protocol for data streaming in wireless sensor networks. *Comput. Commun.* 32(7-10) (2009)
9. Li, T., Ni, Q., Xiao, Y.: Investigation of the block ack scheme in wireless ad hoc networks. *Wireless Communications and Mobile Computing* 6, 877–888 (2006)
10. Rosberg, Z., Liu, R.P., Dong, A.Y., Le, T.D., Jha, S.: Arq with implicit and explicit acks in wireless sensor networks. In: GLOBECOM, pp. 50–55 (2008)
11. Shanti, C., Sahoo, A.: Dgram: A delay guaranteed routing and mac protocol for wireless sensor networks. *IEEE Transactions on Mobile Computing* 9(10), 1407–1423 (2010)
12. Stann, F., Heidemann, J.: Rmst: Reliable data transport in sensor networks. In: Proceedings of the First International Workshop on Sensor Net Protocols and Applications, pp. 102–112 (2003)
13. Xu, H., Huang, L., Wu, J., Wang, Y., Xu, B., Wang, J., Wang, D.: Wireless fire monitoring system for ancient buildings. In: Proceedings of the 2nd International Conference on Scalable Information Systems (InfoScale 2007), pp. 1–4 (2007)
14. Rosberg, Z., Liu, R., Tuan, L.D., Jha, S., Dong, A.Y., Zic, J.: Energy Efficient Statistically Reliable Hybrid Transport Protocol for Sensed Data Streaming. CSIRO ICT Centre Pub. no. 07/213 (June 2007), <http://fairflows.com/rosberg/papers/eRDC.pdf>
15. Zhang, H., Arora, A., Choi, Y.R., Gouda, M.G.: Reliable bursty convergecast in wireless sensor networks. *Comput. Commun.* 30(13) (2007)



# Inter-actor Connectivity Restoration in Wireless Sensor Actor Networks: An Overview

Sasmita Acharya<sup>1</sup> and C.R. Tripathy<sup>2</sup>

<sup>1</sup> Department of Computer Applications, VSS University of Technology, Burla  
talktosas@gmail.com

<sup>2</sup> Department of Computer Science and Engineering, VSS University of Technology, Burla  
crt.vssut@yahoo.com

**Abstract.** Wireless Sensor and Actor Networks (WSANs) consist of powerful actors and resource constraint sensors linked by wireless medium. In WSANs, the sensors gather information about the physical environment while the actors take decisions and perform appropriate actions depending on the sensed data. In some applications, actors must communicate with each other to make appropriate decisions and perform the coordinated actions. Maintaining inter-actor connectivity is extremely important in critical WSAN applications, where the actors need to quickly plan for optimal coordinated response to events detected by the sensors. The failure of a critical actor partitions the inter-actor network into disjoint segments and hinders the network operation. Under such circumstances, the network no longer becomes capable of giving a timely response to a serious event. So, recovery from an inter-actor connectivity failure is of utmost importance. This paper reviews different approaches for restoring the inter-actor connectivity in WSANs.

**Keywords:** Wireless Sensor and Actor Networks, inter-actor connectivity restoration, disjoint segments, recovery.

## 1 Introduction

Wireless Sensor and Actor Networks (WSANs) are gaining growing interest because of their suitability for mission critical applications that require autonomous and intelligent interaction with the environment. Some of the important application areas for WSANs include forest fire monitoring, disaster management, battlefield surveillance, factory automation, and oil and gas pipeline monitoring [1]. The WSANs employ a number of sensor nodes that report an event of interest to one or multiple actors. These actors respond to various events like fire, earthquake, disasters, etc. In most application setups, the actors need to co-ordinate and collaborate with each other, plan an optimal response and select the most appropriate subset of actors to execute the plan. For example, in forest monitoring applications, actors like fire trucks and flying robots need to collaborate with each other to control the fire effectively. For this, they should be reachable to each other. So, a connected inter-actor network needs to be maintained at all time.

However, the harsh application environments of WSAWs make actors prone to physical damage and component malfunction. The failure of an actor may partition the inter-actor network into disjoint segments and make the network incapable of delivering a timely response to a serious event. So, the recovery from an actor failure is of utmost importance for ensuring proper functioning of a WSAW. This recovery should be a self-healing process for the network and should be performed in a distributed manner with minimal overhead [8]. This paper gives an overview of different inter-actor connectivity restoration approaches for WSAWs.

The Section 2 provides the necessary background. Three distributed self-healing approaches to restore inter-actor connectivity along with illustrations are discussed in Section 3. The Section 4 presents a comparative study of the three approaches. The Section 5 concludes the paper.

## 2 Background

There are two types of nodes in WSAWs – sensors and actors. The sensors are inexpensive, highly energy- constrained and are having limited data processing capabilities, whereas, the actors are more powerful nodes than the sensors in terms of energy, communication and computation. Both the sensors and the actors are deployed randomly in an area of interest [2].

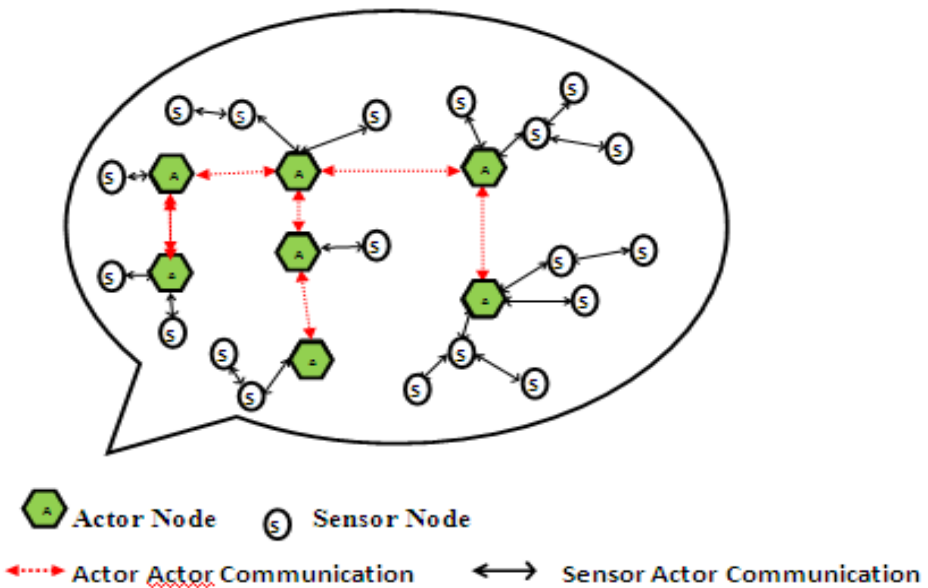


Fig. 1 A Wireless Sensor Actor Network setup

After deployment, the actors are assumed to discover each other and form a connected inter-actor network [3]. An actor is assumed to be able to move on demand and is aware of the positions of its one-hop and two-hop neighbors. The Fig. 1 illustrates a WSAW setup. The oval-shaped nodes represent the sensors and the

hexagonal nodes represent the actors. The job of sensors is to probe the environment and send the sensed data to nearby actors. The figure also shows communication between sensor and actor nodes (sensor-actor communication) and between two actor nodes (actor-actor communication).

The impact of an actor’s failure in a connected inter-actor network depends on the position of that actor in the network topology. For example, the loss of a leaf node or non-critical node does not affect the inter-actor connectivity. But the failure of a critical or cut-vertex node partitions the network into disjoint segments. In order to tolerate critical node failure, three approaches are identified. They are pro-active, reactive and hybrid. The proactive approaches [4] establish and maintain a bi-connected topology in order to provide fault tolerance. A bi-connected topology refers to a network where each sensor sends data to multiple actors and each actor receives sensed information from multiple sensors in an event area. This guarantees event notification but maintaining such a level of connectivity requires a large actor count that leads to higher cost and becomes impractical. In reactive approaches [3],[8-9] and [11-12], the network responds only when a failure occurs. In hybrid approaches [5] and [10], each critical actor proactively designates another appropriate actor as its backup to handle its failure when such a contingency arises.

### 3 Connectivity Restoration through Actor Movement

This section discusses different approaches to restore inter-actor connectivity in WSAN in the event of failure of an actor node.

#### 3.1 The Connectivity Restoration Problem

It deals with different approaches to repair the network topology to its pre-failure state and restore back the connectivity between actor nodes[6-7]. An example of a connected inter-actor network topology is shown in Fig. 2. The node ‘F’ is a cut-vertex node with a node degree of 4 as it has four independent links to actors B, D, J and K.

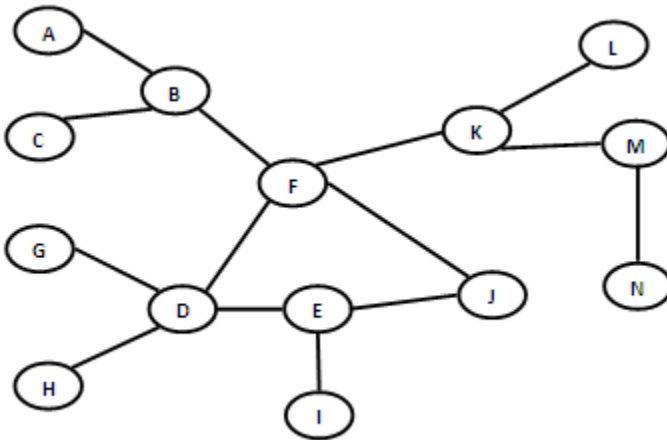


Fig. 2. A connected inter-actor network topology.

The failure of actor node 'F' causes the partitioning of the network into three disjoint sub-networks, namely, {A, B, C}, {D, E, G, H, I, J} and {K, L, M, N} as shown in Fig. 3. As a result, the network loses its connectivity. The pre-failure connectivity needs to be restored for the proper functioning of the network by reconnecting the disjoint partitions.

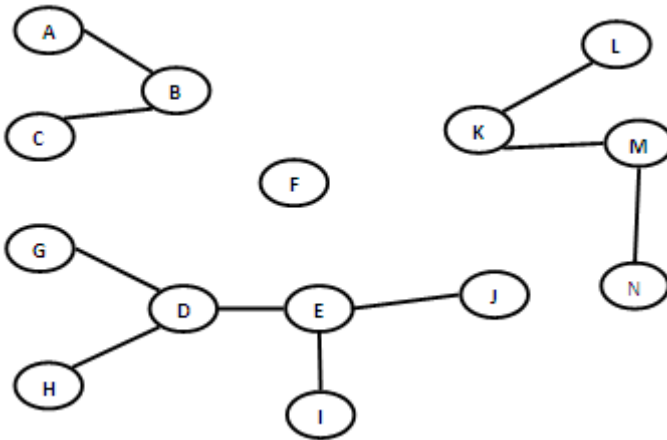


Fig. 3. Partitioning of network of Fig. 2 due to failure of actor 'F'.

### 3.2 Inter-actor Connectivity Restoration

This paper discusses three different distributed self-healing approaches to restore the inter-actor connectivity in case of failure of an actor node. They are Distributed Actor Recovery Algorithm ( DARA ) [3], Recovery through Inward Motion ( RIM ) [8] and Partitioning Detection and Connectivity Restoration ( PCR ) [10]. DARA and RIM are reactive approaches whereas PCR is a hybrid approach.

#### 3.2.1 Distributed Actor Recovery Algorithm

The approach in [3] is called a Distributed Actor Recovery Algorithm. It requires that each actor in the network maintains a list of its one-hop and two-hop neighbors.

##### 3.2.1.1 Algorithm

1. A two-hop neighbor table is created for each actor node.
2. The *heartbeat messages* exchanged between an actor and its one-hop neighbors are constantly monitored.
3. The failure of an actor node is indicated by consecutive missing heartbeat messages.
4. A best candidate is selected from among the one-hop neighbors of the failed actor based on least node degree and least node distance.
5. The selected candidate replaces the failed actor and re-establishes the connectivity.
6. The algorithm is applied recursively for the child nodes of the selected candidate actor.

3.2.1.2 Description

The one-hop neighbors of failed actor ‘F’ are nodes B, K, D and J. The node degrees of neighbor nodes B, K, D and J are respectively 3, 3, 4 and 2. So, based on least node degree, the neighbor node ‘J’ is selected as the best candidate. It then simply replaces the failed node ‘F’ and restores back the lost links as shown in Fig. 4.

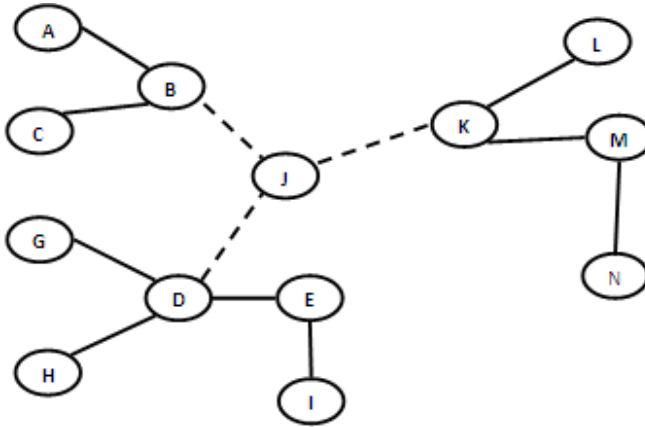


Fig. 4. Node J takes the place of failed actor ‘F’ and restores connectivity.

3.2.2 Recovery through Inward Motion Algorithm

The approach in [8] is called the Recovery through Inward Motion algorithm. It requires only one-hop neighbor information in order to recover from a node failure. The restoration process is done locally and execution is done in a distributed fashion requiring no co-ordination among the neighbor nodes of a failed actor.

3.2.2.1 Algorithm

1. A one-hop neighbor table is created for each actor node.
2. The heartbeat messages exchanged between an actor and its one-hop neighbors are constantly monitored.
3. The failure of an actor node is indicated by consecutive missing heartbeat messages.
4. All the one-hop neighbors of the failed actor move a distance of ‘ $r / 2$ ’ units from their original position to form a connected sub-network where ‘ $r$ ’ represents the communication range of an actor node.
5. If the children of the one-hop neighbors of the failed actor are disconnected from their parent nodes due to the movement described in Step 4, they move a distance of ‘ $r$ ’ units to get connected to their parent nodes.

3.2.2.2 Description

The Fig. 5(a) through Fig. 5(d) illustrate the RIM restoration process using the example WSAN in Fig. 2. The partitioning of the inter-actor network of Fig. 2 due to the failure of critical actor ‘F’ is shown in Fig. 5(a). Then , the one-hop neighbors of the failed actor ‘F’, that is, actors B, K, D and J move a distance of ‘ $r/2$ ’ units towards ‘F’ to form a connected sub-network as shown in Fig. 5(b).

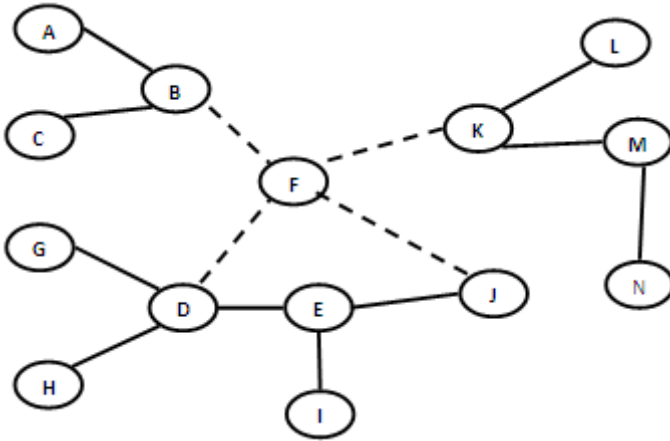


Fig. 5. (a) Partitioning of inter-actor network of Fig. 2.

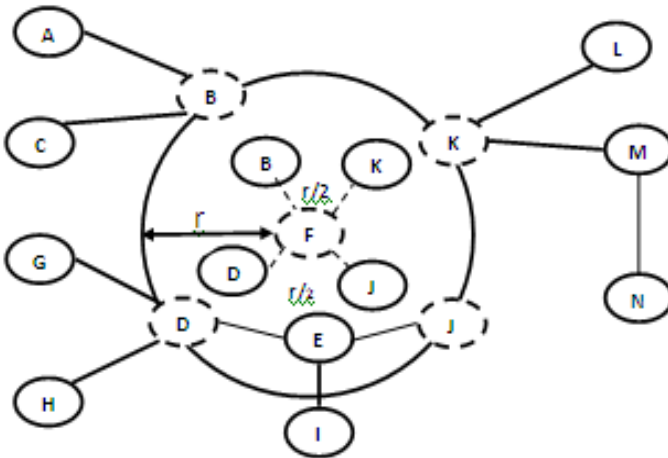
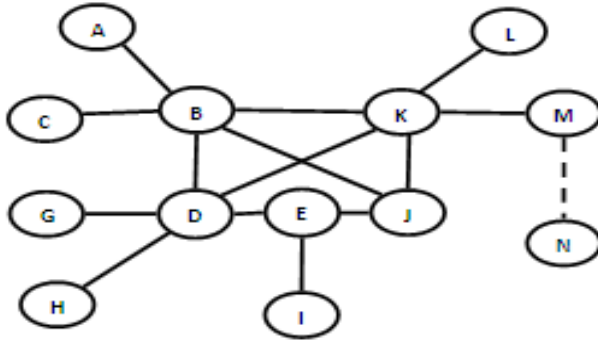


Fig. 5. (b) Nodes B, D, K and J move ' $r/2$ ' units towards the failed actor 'F'.

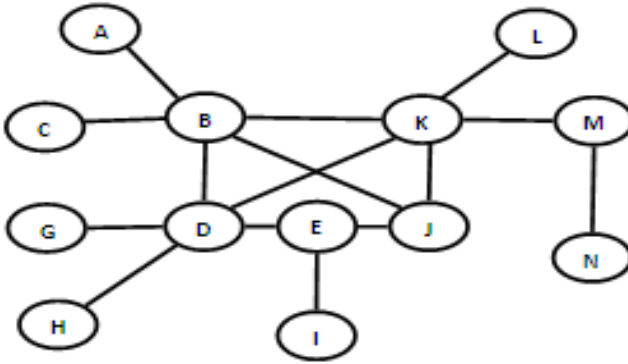
Fig. 5(c) illustrates the movement of children nodes 'L' and 'M' by ' $r$ ' units towards their parent node 'K' to get connected due to which the child node of 'M', that is, node 'N' gets disconnected from its parent. Fig. 5(d) shows the movement of child node 'N' by ' $r$ ' units towards its parent node 'M' to get connected. This completes the RIM restoration process.

### 3.2.3 Partitioning Detection and Connectivity Restoration Algorithm

The approach in [10] is called the Partitioning Detection and Connectivity Restoration Algorithm. It consists of two parts – proactive and reactive. In the proactive part, the critical actors are determined using a localized cut-vertex detection algorithm [13] that only requires one-hop positional information. In the reactive part, a backup node



**Fig. 5.** (c) Children nodes ‘L’ and ‘M’ move ‘r’ units to get connected to their parent ‘K’ whereas ‘N’ gets disconnected.



**Fig. 5.** (d). Child node N moves ‘r’ units to get connected to its parent ‘M’.

initiates the recovery process when the primary fails. The backup replaces the primary and cascaded relocations are performed until complete recovery.

### 3.2.3.1 Algorithm

1. The critical nodes in the network are first identified using a localized *cut-vertex detection algorithm* [13].
2. A *backup actor* is designated for each of the critical nodes identified in step 1 from among their one-hop neighbors.
3. neighbors.
4. The *heartbeat messages* exchanged between an actor and its designated backup are constantly monitored.
5. The failure of an actor node is indicated by consecutive missing heartbeat messages.
6. In case of failure of a non-critical actor (leaf node), the failed actor node is simply removed from the topology and there are no cascaded relocations (movement of children nodes to restore connectivity).

- In case of failure of a critical actor (cut-vertex node identified in step 1), the backup actor replaces the failed actor, designates a backup for itself (using step 2) and re-establishes the connectivity. In this case, cascaded relocation is needed for the children of the backup actor.

3.2.3.2 Description

Fig. 6(a) shows the backup actors for critical actors – B, D, F, K and M identified by procedure [13]. The 1<sup>st</sup> prefix denotes that it is a backup actor and the second prefix stands for the actor ID. For example, ‘J<sub>BF</sub>’ is the backup actor for critical actor ‘F’.

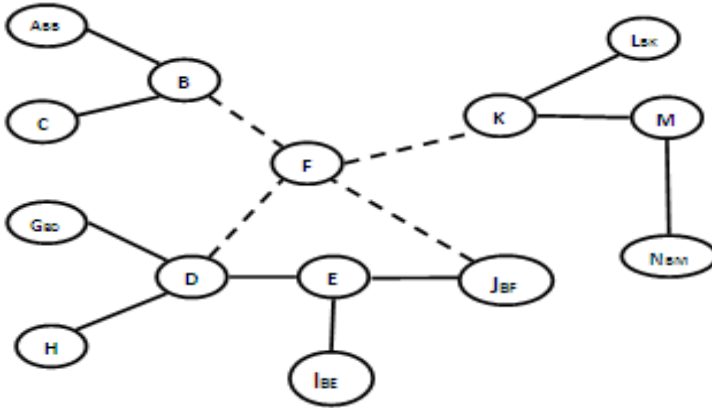


Fig. 6. (a) Backup actor selection for critical actors of Fig. 2 when node ‘F’ fails.

Fig. 6(b) shows the recovery process where the backup actor ‘J<sub>BF</sub>’ replaces the failed actor ‘F’, designates one-hop neighbor actor ‘B’ as its backup (which now becomes B<sub>Bj</sub>) and re-establishes links with actors B, D and K.

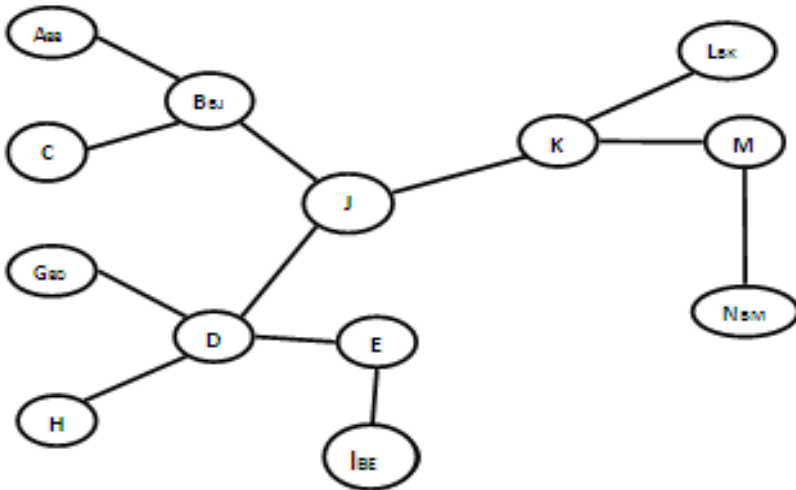


Fig. 6. (b) Recovery process initiated by backup actor J.



## 4 Comparative Analysis of Inter-actor Connectivity Restoration Approaches

This section does a comparative study of the three approaches for inter-actor connectivity restoration [3], [8] and [10]. The demerit of the DARA approach [3] was that it required the knowledge of the two-hop neighbors for each actor that needed to be updated every time with each actor movement. This in turn increased the messaging overhead. The RIM approach [8] overcame this problem by maintaining only one-hop neighbor information. However, the complexity of the RIM approach increased significantly with an increase in the communication range as more and more number of neighbor actors participated in the recovery process. The PCR approach [10] overcame this problem by designating a backup actor for each critical actor in the network. This backup actor took the overall responsibility of failure detection and recovery.

## 5 Conclusion

This paper presented an overview of three different distributed, self-healing inter-actor connectivity restoration approaches – DARA, RIM and PCR. It also made a critical comparative analysis of the three approaches. The PCR inter-actor connectivity approach outperforms DARA and RIM approaches with respect to scalability, scope of recovery, movement and messaging overhead though it also involves an additional overhead in terms of identification of critical actor nodes in a network and designation of appropriate backups.

## References

1. Akyildiz, F., Kasimoglu, I.H.: Wireless Sensor and actor networks: Research Challenges. *Ad Hoc Networks* 2, 351–367 (2004)
2. Younis, M., Akkaya, K.: Strategies and Techniques for Node Placement in Wireless Sensor Networks: A Survey. *Ad-Hoc Networks* 6(4), 621–655 (2008)
3. Abbasi, A., Akkaya, K., Younis, M.: A Distributed Connectivity Restoration Algorithm in Wireless Sensor and Actor Networks. In: 32nd Conference on Local Computer Networks, Dublin, Ireland, pp. 496–503 (2007)
4. Ozaki, K., et al.: A Fault-Tolerant Model for Wireless Sensor-Actor System. In: IEEE HWISE 2006, Vienna, Austria (2006)
5. Zamanifar, A., Kashefi, O., Sharifi, M.: A Hybrid Approach to Actor-Actor Connectivity Restoration in Wireless Sensor and Actor Networks. In: 8th International Conference on Networks, pp. 76–81 (2009)
6. Basu, P., Redi, J.: Movement Control Algorithms for Realization of Fault-Tolerant Ad Hoc Robot Networks. *IEEE Networks* 18(4), 36–44 (2004)
7. Das, S., Liu, H., Kamath, A., Nayak, A., Stojmenovic, I.: Localized Movement Control for Fault Tolerance of Mobile Robot Networks. In: First IFIP International Conference on WSANs, Albacete, Spain (2007)

8. Younis, M., LKee, S., Gupta, S., Fisher, K.: A Localized Self-healing Algorithm for Networks of Moveable Sensor Nodes. In: IEEE Global Telecommunications Conference (Globecom 2008), New Orleans, LA (2008)
9. Akkaya, K., Thimmapuram, A., Senel, F., Uludag, S.: Distributed Recovery of Actor Failures in Wireless Sensor and Actor Networks. In: Wireless Communications and Networking Conference, Los Vegas, USA, pp. 2480–2485 (2008)
10. Imran, M., Younis, M., Said, A.M., Hasbullah, H.: Partitioning Detection and Connectivity Restoration Algorithm for Wireless Sensor Actor Networks. In: IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, pp. 200–207 (2010)
11. Tamboli, N., Younis, M.: Coverage-aware Connectivity Restoration in Mobile Sensor Networks. In: IEEE International Conference on Communications (ICC 2009), Dresden, Germany (2009)
12. Imran, M., Younis, M., Said, A.M., Hasbullah, H.: Volunteer-instigated Connectivity Restoration Algorithm for Wireless Sensor and Actor Networks. In: IEEE International Conference on Wireless Communication, Networking and Information Security (WCNIS 2010), Beijing, China, pp. 679–683 (2010)
13. Jorgic, M., Stojmenovic, I., Hauspie, M., Simplot-Ryl, D.: Localized Algorithms for Detection of Critical Nodes and Links for Connectivity in Ad Hoc Networks. In: 3rd IFIP MedHoc, Bodrum, Turkey, pp. 360–371 (2004)

# MANFIS Approach for Path Planning and Obstacle Avoidance for Mobile Robot Navigation

Prases Kumar Mohanty, Krishna K. Pandey, and Dayal R. Parhi

Robotics Laboratory, Department of Mechanical Engineering,  
National Institute of Technology, Rourkela, Odisha, India  
pkmohanty30@gmail.com, kknitrkl@yahoo.in, dayalparhi@yahoo.com

**Abstract.** Path planning and obstacle avoidance are very crucial issues for an Autonomous mobile robot. In this research paper an intelligent hybrid approach MANFIS (Multiple Adaptive Neuro-Fuzzy Inference system) has been implemented for mobile robot navigation. The adaptive neuro-fuzzy inference system (ANFIS) has taken the advantages of expert knowledge of fuzzy inference system and learning capability of artificial neural network. The inputs to the MANFIS controller include the front obstacle distance, the left obstacle distance, the right obstacle distance and the target angle and outputs from the controller are left wheel velocity and right wheel velocity of the mobile robot. In order to validate the proposed hybrid technique a series of simulation experiments using MATLAB were performed and it was found that the proposed navigational controller is capable to avoid obstacle and reach the destination successfully. The experimental results also have been compared with simulation results to prove the authenticity of the developed navigational controller MANFIS.

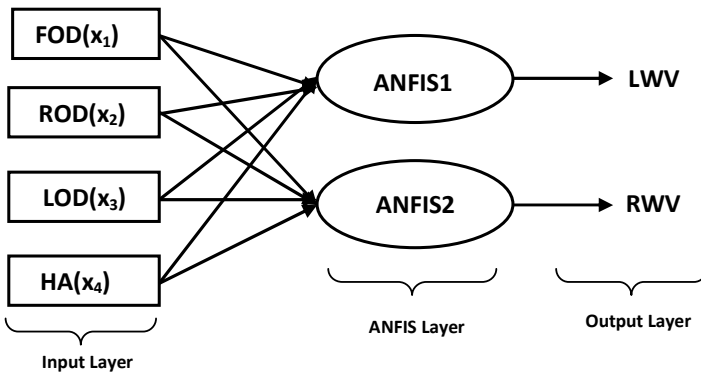
**Keywords:** Neuro-Fuzzy, Obstacle avoidance, Mobile robot, Navigation.

## 1 Introduction

Nowadays autonomous mobile robots are being deployed in various field of engineering such as aerospace research, production industry, military operation, nuclear research etc. The primary strategy of a mobile robot is to navigate among the obstacles without hitting them and reach the specified target point. The sensor based path planning techniques can be classified into two main categories, the global and local path planning depending upon the surrounding environment. Global path planning needs the environment to be completely known and the terrain should be static; on other side local path planning means the environment is completely or partially unknown for the mobile robot. So various sensor devices are used to locate the shape and position of the obstacles present in the environment and navigate the robot safely to reach the target.

In literature review, many researchers have been developed various interesting research work for navigation of autonomous mobile robots. Many researchers have considered a motion planner with complete knowledge of the environment [1-2].Due

to the intricacy and uncertainty of the navigation problem, classical path planning methods, such as Visibility Graph [3], Voronoi diagrams [4], Grids [5], Cell decomposition [6], artificial potential field [7], are not meet for path planning in dynamic environments. The use of the above methods for path finding for mobile robot requires more time and the finding of this path will not completely feasible for real-time movement. In recent times intelligent soft computing techniques such as fuzzy inference system (FIS) [8-11], artificial neural network (ANN) [12-14] and adaptive Neuro-fuzzy inference system (ANFIS) [15] have been implemented for path planning of mobile robot. Navigation of mobile robots using adaptive neural-fuzzy system discussed by Nefti et al. [16]. In this model different sensor based information they have given to the Sugeno-Takagi type fuzzy controller and output from the controller are the robot orientation. Experimental results settle the importance of the methodology when dealing with navigation of a mobile robot in unknown or partially unknown environment. Path planning of mobile robot using ANFIS presented by Sudhagar et al. [17]. In this paper ANFIS is used for obtain the control architecture to navigate the backward motion of a mobile robot with back propagation algorithm. Autonomous parallel parking of a car-like mobile robot using Neuro-fuzzy technique proposed by Demirli et al. [18]. In this proposed model uses the data from three sonar sensors mounted in the front left corner of the car to decide on the steering angle. A Neuro-Fuzzy Controller for mobile robot navigation addressed by Kim and Trivedi [19]. In this study they implemented neural integrated fuzzy controller to control the mobile robot motion in terms of steering angle, heading direction, and speed. Navigation of multiple mobile robots using Neuro-fuzzy technique addressed by Pradhan et al. [20]. In this design, output parameters from the neural controller are given as input to fuzzy controller to control the mobile robot successfully in the clutter environment. Experimental verifications also have been done with the simulation results to prove the validity of the developed technique. A Neuro-fuzzy approach for obstacle avoidance for a mobile robot addressed by Marichal et al. [21]. This technique is able to extract automatically the fuzzy rules and membership functions in order to control a mobile robot in a maze environment.



**Fig. 1.** Proposed MАНFIS (Multiple ANFIS) controller for Mobile Robot Navigation

In this article, we introduce a new intelligent hybrid motion planner MANFIS to drive a mobile robot in an unknown or partially known environment containing static obstacles. Finally, simulation results are performed to test the efficiency of the proposed navigational controller in various environments. Experimental verifications also have been done with simulation results to prove the validity of the developed scheme.

## 2 Design of Adaptive Neuro-Fuzzy Inference System for Mobile Robot Path Planning

Adaptive network-based fuzzy inference system (ANFIS) is one of hybrid intelligent neuro-fuzzy system and it functioning under Takagi-Sugeno-type FIS, which was developed by Jang [15] in 1993. ANFIS has a similar configuration to a multilayer feed forward neural network but links in this hybrid structure only specify the flow direction of signals between nodes and no weights are connected with the links. There are two learning techniques are used in ANFIS to show the mapping between input and output data and to compute optimized of fuzzy membership functions. These learning methods are back propagation and hybrid. Parameters associated with fuzzy membership functions will modify through the learning process.

As for the prediction of left wheel velocity (LWV) and right wheel velocity (RWV) for mobile robot we assume that each adaptive Neuro-Fuzzy controller under consideration of four inputs i.e. Front obstacle distance (FOD) ( $x_1$ ), Right obstacle distance (ROD) ( $x_2$ ), Left obstacle distance (LOD) ( $x_3$ ), Target angle (HA) ( $x_4$ ), and each input variable has five bell membership functions (MF) such as  $A_1$ (Very Near),  $A_2$ (Near),  $A_3$ (Medium),  $A_4$ (Far) and  $A_5$ (Very Far),  $B_1$ (Very Near),  $B_2$ (Near),  $B_3$ (Medium),  $B_4$ (Far) and  $B_5$ (Very Far),  $C_1$ (Very Near),  $C_2$ (Near),  $C_3$ (Medium),  $C_4$ (Far), and  $C_5$ (Very Far),  $D_1$ (Very Negative),  $D_2$ (Negative),  $D_3$ (Zero),  $D_4$ (Positive) and  $D_5$ (Very positive) respectively, then a Takagi-Sugeno-type fuzzy inference system if-then rules are set up as follows;

*Rule: if  $x_1$  is  $A_i$  and  $x_2$  is  $B_i$  and  $x_3$  is  $C_i$  and  $x_4$  is  $D_i$ , then*

$$f_n(\text{wheel velocity}) = p_n x_1 + q_n x_2 + r_n x_3 + s_n x_4 + u_n$$

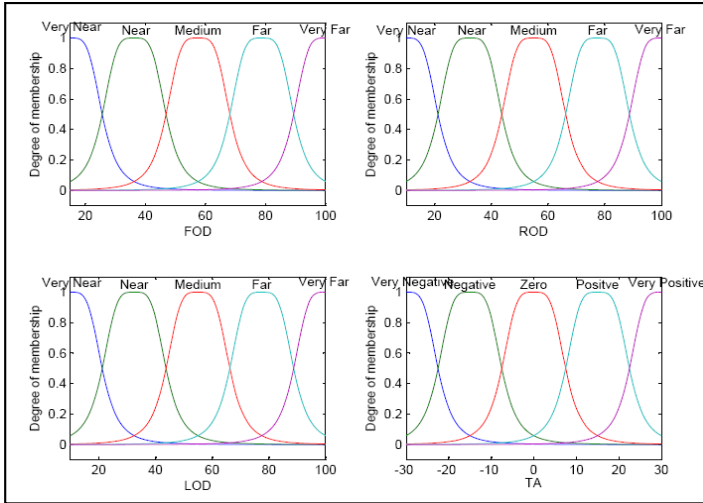
A, B, C, and D are the fuzzy membership sets for the input variables  $x_1, x_2, x_3$  and  $x_4$  respectively.

where,  $i=1-5$  and  $p_n, q_n, r_n, s_n$  and  $u_n$  are the linear parameters of function  $f_n$  and changing these parameters we can modify the output of ANFIS controller.

The function of each layer in ANFIS structure is discussed as follows:

**Input Layer:** In this layer nodes simply pass the incoming signal to layer-1. That is

$$\left. \begin{aligned} O_{0,FOD} &= X_1 \\ O_{0,ROD} &= X_2 \\ O_{0,LOD} &= X_3 \\ O_{0,TA} &= X_4 \end{aligned} \right\} \quad (2.1)$$



**Fig. 2.** Membership functions for MANFIS controller

**First Layer:** This layer is the fuzzification layer. Neurons in this layer complete fuzzification process. Every node in this stage is an adaptive node and calculating the membership function value in fuzzy set. The output of nodes in this layer are presented as

$$\left. \begin{aligned}
 O_{1,i} &= \mu_{A_i}(X_1) \\
 O_{1,i} &= \mu_{B_i}(X_2) \\
 O_{1,i} &= \mu_{C_i}(X_3) \\
 O_{1,i} &= \mu_{D_i}(X_4)
 \end{aligned} \right\} \quad (2.2)$$

Here  $O_{1,i}$  is the bell shape membership grade of a fuzzy set  $S( A_i , B_i ,C_i \text{ and } D_i )$  and it computing the degree to which the given inputs (  $X_1,X_2,X_3$ and  $X_4$ ) satisfies the quantifier  $S$ . Membership functions defined as follows;

$$\mu_{A_i}(x) = \frac{1}{1 + \left[ \left( \frac{x_1 - c_i}{a_i} \right)^2 \right]^{b_i}} \quad (2.2(i))$$

$$\mu_{B_i}(x) = \frac{1}{1 + \left[ \left( \frac{x_2 - c_i}{a_i} \right)^2 \right]^{b_i}} \quad (2.2(ii))$$

$$\mu_{C_i}(x) = \frac{1}{1 + \left[ \left( \frac{x_3 - c_i}{a_i} \right)^2 \right]^{b_i}} \tag{2.2(iii)}$$

$$\mu_{D_i}(x) = \frac{1}{1 + \left[ \left( \frac{x_4 - c_i}{a_i} \right)^2 \right]^{b_i}} \tag{2.2(iv)}$$

$a_i, b_i$  and  $c_i$  are parameters that control the Centre, width and slope of the Bell-shaped function of node ‘i’ respectively. These are also known as premise parameters.

**Second Layer:** It is also known as rule layer. Every node in this layer is a fixed node and labeled as  $\pi_n$ . Every node in this stage corresponds to a single Sugeno-Takagi fuzzy rule. A rule node receives inputs from the respective nodes of layer-1 and determines the firing strength of the each rule. Output from each node is the product of all incoming signals.

$$O_{2,n} = W_n = \mu_{A_i}(X_1) \cdot \mu_{B_i}(X_2) \cdot \mu_{C_i}(X_3) \cdot \mu_{D_i}(X_4) \tag{2.3}$$

where  $W_n$  represents the firing strength or the truth value, of nth rule and  $n=1, 2, 3 \dots 636$  is the number of Sugeno-Takagi fuzzy rules.

**Third Layer:** It is the normalization layer. Every node in this layer is a fixed node and labeled as  $N_n$ . Each node in this layer receives inputs from all nodes in the fuzzy rule layer and determines the normalized firing strength of a given rule. The normalized firing strength of the nth node of the nth rule’s firing strength to sum of all rules’s firing strength.

$$O_{3,n} = \bar{W}_n = \frac{W_n}{\sum_{n=1}^{625} W_n} \tag{2.4}$$

The number of nodes in this layer is the same the number of nodes in the previous layer that is 625 nodes. The output of this layer is called normalized firing strength.

**Fourth layer:** Every node in this layer is an adaptive node. Each node in this layer is connected to the corresponding normalization node, and also receives initial inputs  $X_1, X_2, X_3$  and  $X_4$ . A defuzzification node determines the weighted consequent value of a given rule define as,

$$O_{4,n} = \bar{W}_n f_n = \bar{W}_n [p_n(X_1) + q_n(X_2) + r_n(X_3) + s_n(X_4) + u_n] \tag{2.5}$$

Where  $\bar{W}_n$  is a normalized firing strength from layer-3 and  $p_n, q_n, r_n, s_n, u_n$  are the parameters set of this node. These parameters are also called consequent parameters.

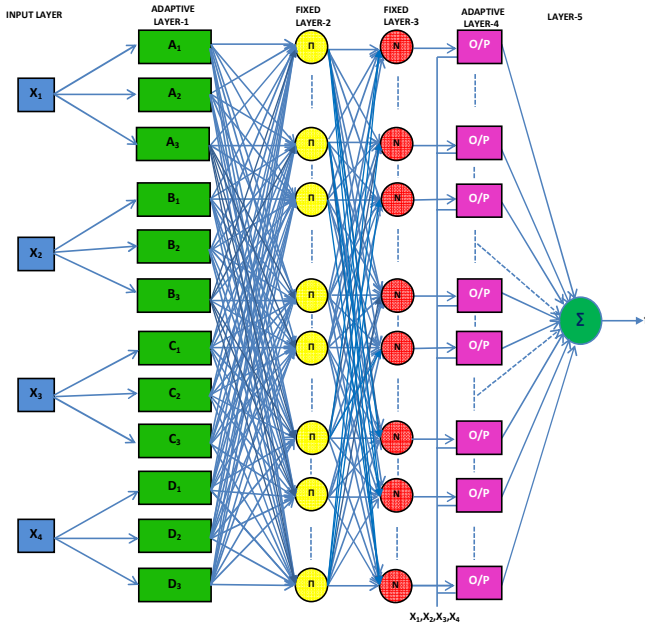


Fig. 3. The structure of ANFIS 1 network

**Fifth layer:** It is represented by a single summation node. This single node is a fixed node and labeled as  $\Sigma$ . This node determines the sum of outputs of all defuzzification nodes and gives the overall system output that is wheel velocity.

$$O_{5,1} = \sum_{n=1}^{625} \bar{W}_n f_n = \frac{\sum_{n=1}^{625} W_n f_n}{\sum_{n=1}^{625} W_n} \tag{2.6}$$

### 3 Simulation Results and Discussion

We have tested our proposed algorithm in two dimensional path planning through series of simulation experiments under unknown or partially known environment. Our implementation was compiled using MATLAB R2010a processing under Windows XP. All the simulation results were applied on PC with Intel core2 processor running at 3.0GHz, 4 GB of RAM and a hard drive of 320 GB.

In current path planning model, we have been designed two main reactive behaviors: one to reach the target and the other avoiding obstacles. The simulated robot path planning algorithm has been created and set obstacles at different position



of the environment. When a robot is close to an obstacle, it must change its speed to avoid the obstacle. If a target is sensed by a mobile robot, it will decide whether it can reach that target, i.e. it will judge whether there are obstacles that will obstruct its path. If the path leading to the source is clear, the robot will turn and proceed towards the source. In Fig. 4 shows the path created for mobile robot motion in various environments with considering different start and goal positions. It can be found that, using sensory information, the mobile robot can reach successfully at the goal by efficiently using multiple types of reactive behaviors with proposed navigational controller. Experimental verifications also have been done with simulation results to prove the validity of the developed algorithm.

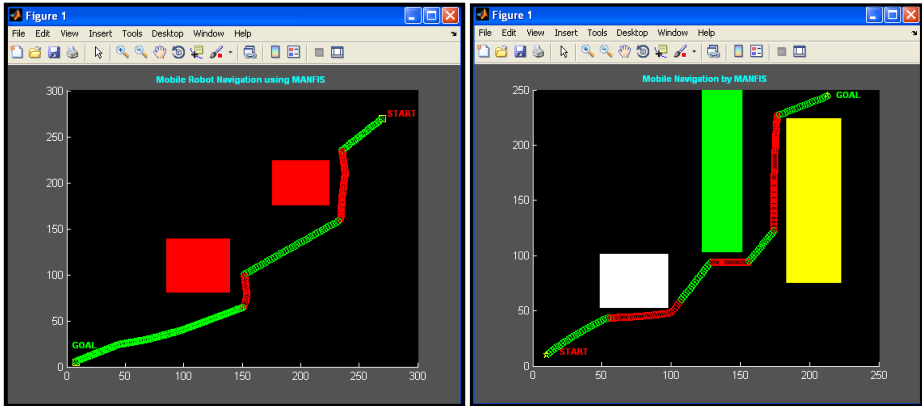


Fig. 4. (a-b) Navigation by single mobile robot using current analysis

### 4 Experimental Results

To show the effectiveness of the proposed control system and authenticity of the navigation technique, a variety of real time experiments were conducted using co-robot(CL-4)) mobile robot.

Two different cases in Fig.4 which are already verify in the simulation link, have verified experimentally in (Fig.5) to show the robustness of the designed MANFIS controller. Table-1 shows the path length cover by the robot in simulations and in the experimental tests scenario during target searching. The path length traced by the experimentally nearly same as the path traced by the mobile robot during simulation link.

**Table-1**

No.	Path length in simulation	Path length in real time experiment
Scenario-1	1.74m (Fig. 4 a)	1.88m (Fig. 5 d)
Scenario-2	2.77m (Fig. 4 b)	2.98m (Fig. 5 h)

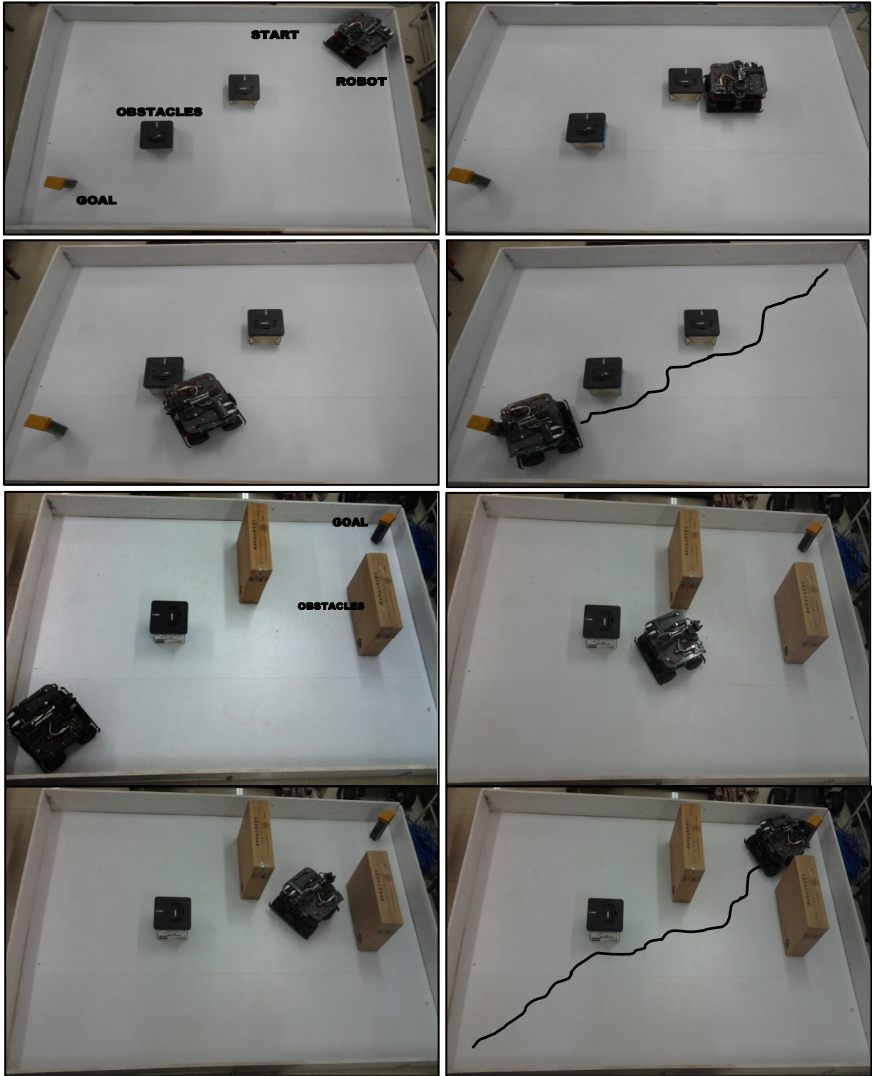


Fig. 5. (a-h) Real time experiment conducted by Co-Robot using MANFIS

## 5 Conclusion and Future Work

In this article, path planning and obstacle avoidance problem for single mobile robot in an unknown or partially unknown environment populated by variety of static obstacles was addressed. It has been found that the multiple adaptive neuro-fuzzy (MANFIS) motion planner is capable of avoiding obstacles and effectively guiding the mobile robot moving from the start point to the desired destination point with shortest path length. The authenticity of the proposed navigation technique has been

verified and proven by simulation experiments using MATLAB. The proposed hybrid technique has been also compared with real time experimental results and the settlement in results show the efficiency of the navigational controller. In future multiple robots are to be considered instead of a single mobile robot for solving the path planning problem in cluttered environment.

## References

1. Latombe, J.C.: Robot Motion Planning. Kluwer Academic Publishers, New York (1990)
2. Canny, J.E.: The Complexity of Robot Motion Planning. MIT Press, Cambridge (1988)
3. Lozano-Perez, T.: A simple motion planning algorithm for general robot manipulators. *IEEE Journal of Robotics and Automation* 3, 224–238 (1987)
4. Leven, D., Sharir, M.: Planning a purely translational motion for a convex object in two dimensional space using generalized voronoi diagrams. *Discrete & Computational Geometry* 2, 9–31 (1987)
5. Payton, D., Rosenblatt, J., Keirse, D.: Grid-based mapping for autonomous mobile robot. *Robotics and Autonomous Systems* 11, 13–21 (1993)
6. Regli, L.: Robot Lab: Robot Path Planning. Lectures Notes of Department of computer Science. Drexel University (2007)
7. Khatib, O.: Real time Obstacle Avoidance for manipulators and Mobile Robots. *IEEE Conference on Robotics and Automation* 2, 505 (1985)
8. Huq, R., Mann, G.K.I., Gosine, R.G.: Mobile robot navigation using motor schema and fuzzy content behavior modulation. *Application of Soft Computing* 8, 422–436 (2008)
9. Selekwa, M.F., Dunlap, D.D., Shi, D., Collins Jr., E.G.: Robot navigation in very cluttered environment by preference based fuzzy behaviors. *Autonomous System* 56, 231–246 (2007)
10. Abdessemed, F., Benmahammed, K., Monacelli, E.: A fuzzy based reactive controller for a non-holonomic mobile robot. *Robotics Autonomous System* 47, 31–46 (2004)
11. Pradhan, S.K., Parhi, D.R., Panda, A.K.: Fuzzy logic techniques for navigation of several mobile robots. *Application of Soft Computing* 9, 290–304 (2009)
12. Velagic, J., Osmic, N., Lacevic, B.: Neural Network Controller for Mobile Robot Motion Control. *World Academy of Science, Engineering and Technology* 47, 193–198 (2008)
13. Singh, M.K., Parhi, D.R.: Intelligent Neuro-Controller for Navigation of Mobile Robot. In: *Proceedings of the International Conference on Advances in Computing, Communication and Control*, Mumbai, Maharashtra, India, pp. 123–128 (2009)
14. Castro, V., Neira, J.P., Rueda, C.L., Villamizar, J.C., Angel, L.: Autonomous Navigation Strategies for Mobile Robots using a Probabilistic Neural Network (PNN). In: *33rd Annual Conference of the IEEE Industrial Electronics Society*, Taipei, Taiwan, pp. 2795–2800 (2007)
15. Jang, J.S.R.: ANFIS: Adaptive network-based fuzzy inference system. *IEEE Transaction on System, Man and Cybernetics –Part B* 23, 665–685 (1993)
16. Nefti, S., Oussalah, M., Djouani, K., Pontnau, J.: Intelligent Adaptive Mobile Robot Navigation. *Journal of Intelligent and Robotic Systems* 30, 311–329 (2001)
17. Sudhakar, K., Noorul, A., Selvaraj, T.: Neuro-Fuzzy Based navigation for Truck like Mobile Robot. *International Journal of Soft Computing* 5, 633–637 (2007)
18. Demirli, K., Khoshnejad, M.: Autonomous parallel parking of a car-like mobile robot by a neuro-fuzzy sensor-based controller. *Fuzzy Sets and Systems* 160, 2876–2891 (2009)

19. Kim, C.N., Trivedi, M.M.: A Neuro-Fuzzy Controller for Mobile Robot Navigation and Multirobot Convoying. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics* 28, 829–840 (1998)
20. Pradhan, S.K., Parhi, D.R., Panda, A.K.: Neuro-fuzzy technique for navigation of multiple mobile robots. *Fuzzy Optimum Decision Making* 5, 255–288 (2006)
21. Marichal, G.N., Acosta, L., Moreno, L., Mendez, J.A., Rodrigo, J.J., Sigut, M.: Obstacle avoidance for a mobile robot: A neuro-fuzzy approach. *Fuzzy Sets and Systems* 124, 171–179 (2001)
22. Mohanty, P.K., Parhi, D.R.: Path Planning Strategy for Mobile Robot Navigation using MANFIS Controller. In: 2013 International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), Bhubaneswar (accepted 2013)
23. The Math Works Company, Natick, MA, ANFIS Toolbox User's Guide of MATLAB
24. Parhi, D.R.: Navigation of multiple mobile robots in an unknown environment. Doctoral Thesis, Cardiff School of Engineering. University of Wales, UK (2000)

# The Effect of Velocity and Unresponsive Traffic Volume on Performance of Routing Protocols in MANET

Sukant Kishoro Bisoy<sup>1</sup>, Prasant Kumar Patnaik<sup>2</sup>, and Tanmaya Kumar Swain<sup>2</sup>

<sup>1</sup> SOA University, Bhubaneswar, India  
sukantabisoyi@yahoo.com

<sup>2</sup> School of Computer Engineering, KIIT University, Bhubaneswar, India  
{patnaikprasantfcs,tanmayafcs}@kiit.ac.in

**Abstract.** Mobile ad hoc networks (MANETs) form a random network by consists of mobile nodes where node share information with each other while moving. Due to presence of mobility in the MANET, the interconnections between nodes are likely to change, resulting in frequent changes of network topology. Therefore there is a need of identifying efficient dynamic routing protocol to provide call services in such network. In this paper, the effect of node velocity and unresponsive traffic volume is explored on performance of three routing protocols i.e. DSR, AODV and DYMO via NS2 simulator of ad hoc network of 100 mobile nodes. The performance is measured based on traffic admission ratio, packet delivery ratio (PDR), routing overhead, and average end-to-end delay. We observe that DYMO performs better in terms of PDR and traffic admission ratio and DSR has least overhead than others irrespective of node velocity, traffic volume and number of connection.

**Keywords:** MANET, AODV, DSR, DYMO, Mobility, Traffic admission ratio and NS2.

## 1 Introduction

A Mobile Ad hoc Network is an autonomous collection of mobile nodes forming a random network and communicating over wireless links. This kind of network is more suitable where networking infrastructure is not available and set up time is very less and temporary network connectivity is required. Such limitation make routing is more challenging where each node performs the functionality of router and host. There are many routing protocols available in MANET. Among them ad hoc on demand distance vector (AODV) [1] and dynamic source routing (DSR) [2] and dynamic manet on-demand (DYMO)[3] are reactive routing protocols. Due to dynamic topology change wireless link suffers packet loss problem which is more complicated for MANET. The reason for packet loss may due to due to transmission errors; route does not exist, link breakage or may be congestions, etc.

The rest of the paper is organized as follows. In section 2 we present the related work. We explain the overview of routing protocol in MANET in section 3. Section 4 present the simulation model and section 5 explains the result and analysis. Finally, we conclude our work in section 6.

## 2 Related Work

Author in [4] provides a comparison study of proactive and reactive protocol by varying the node mobility in MANET. The outcome of the study is the performance of reactive protocol is better than proactive protocol. DSR show better PDR while AODV shows lower delay.

In [5] the effect of packet size on AODV protocol is studied on homogeneous and heterogeneous MANET. The effect of packet size in homogeneous network is better than heterogeneous network.

In [6] author studied the behavior of routing protocol over different traffic source by varying different node speed and the outcome is CBR traffic performs better than pareto and exponential traffic.

## 3 Routing Protocol in MANET

Routing is one of the key issues in MANET [7] due to their highly dynamic and distributed nature. Almost all proposed routing protocols are based on minimum hops in mobile ad hoc network. Author in [8] studied the inter layer interaction between MAC and physical layer and demonstrated that even though DSR and AODV share a similar behavior, the differences in the protocol mechanics can lead to significant performance differentials. Author in [9], evaluated the performance of AODV, DSR and OLSR routing protocols in MANETs under CBR traffic with different network conditions.

Generally routing protocols for ad-hoc networks can be classified in two different classes: pro-active and re-active protocols based on how they discover the route. Many reactive routing protocols are available for ad hoc network, including AODV, DYMO and DSR.

### 3.1 Ad Hoc On- Demand Distance Vector (AODV)

In AODV [1] source node initiate the route discovery process if it desired to send a message to some destination. It discovers route on demand when a packet needs to be send by a source. Route discovery process starts by sending route request (RREQ) packet to their neighbors. Then neighbor forward the RREQ to their neighbor and so on. This sending process is continued by every neighbor node until the destination gets the message or they have a route to destination. On either case nodes reply back with a route reply (RREP) message. In case of route breakage the intermediate node discover another new route or send a route error (RERR) message to the source. Upon receiving RERR the source node tries to get new route by invoking again route discovery process.

### 3.2 Dynamic Source Routing (DSR)

The key concept of DSR [2] protocol is the use of source initiated routing. Through the route cache method the sender knows the hop-by-hop information to the every destination. In fact the packet header carries the source route for a destination. It starts

discovering the route by flooding the RREQ packets. Upon receiving the RREQ every node rebroadcasts it, until it has a route to the destination in its cache or it is the destination. Then it replies with a RREP packet that is routed back to the original source. The RREQ packet accumulates a path traversed so far to the destination. The RREP packet uses the accumulated path of the RREQ in the backward path to reach the destination. Then the source node caches the route for future use. In case of route breakage the intermediate node discovers another new route or sends a route error (RERR) message to the source. Upon receiving RERR the source node tries to get a new route by invoking the route discovery process again.

### 3.3 Dynamic Manet On-Demand (DYMO)

Route discovery and route management are the basic operations of the DYMO protocol. Like AODV, the source initiates the route discovery process by disseminating the RREQ for an entire network to get the route to the destination. Each intermediate node records the route during the dissemination process. The destination node responds with a RREP upon receiving the RREQ packet. Each intermediate node that receives the RREP creates a route to the target, and then the RREP is unicast hop-by-hop toward the source. The route between the source and destination can be established in both directions after getting the RREP from the receiver. During route management, the node monitors the link over which the traffic is flowing to cope with changes in network topology. Upon route breakage the source node is notified with a RERR message. A RERR is sent toward the source to indicate that the current route to a particular destination is invalid or missing. The source node deletes the route and performs route discovery if it still has packets to deliver to that destination.

## 4 Simulation Model

We use the discrete-event simulator NS-2.34 [10] with a LINUX platform to investigate the effect of node velocity and unresponsive traffic (CBR) volume on the routing protocol (AODV, DSR and DYMO) of MANET. We measure the performance with various metrics like PDR, routing overhead, end-to-end delay and traffic admission ratio (TAR) using AWK scripts [11].

### Environment

Normally, the performance of routing protocols is studied by three major parameters: node velocity, traffic volume, and node density. We create a random network of size 100 with a specified simulation area of 1000m x 1000m. Each connection stays for 300 seconds long. We use the random way point (RWP) mobility pattern [12] to define the movement of mobile nodes. The pause time is set to zero to indicate continuous movement of nodes, and node velocity is shown in table 1. We use unresponsive traffic (UDP) between the source and destination pair. Every source is associated with a CBR traffic generator. Each source sends packets of 512 bytes at a different rate of 2 (low), 4, 6, 8, 10 and 12 (high) packets per second. Three major parameters are discussed in detail in the following.

### Node Velocity

The node velocity determines the frequency of link breakage and corresponding routing overhead for a network. The routing protocol behaves differently with different mobility model used [13]. The performance of a network depends on behavior of routing protocols and the movement pattern of particular mobility model [14]. The traffic pattern of network also severely impact on the performance of routing protocol in MANET [15].

### Traffic Volume

The communication model of a network describes the number of source and traffic volume and other parameters. We used the communication model which is included with RWP mobility model in the ns2 simulator of version-2.34. Our simulation model comprised of the parameters node velocity (S) and Traffic volume (V). The traffic volume describes the aggregate packet rate from all CBR sources in the network. The packet rate per source (K) is calculated as  $K = N / V$  packets/sec, where V is the traffic volume and N is the number of source. The traffic volume measurement depends on number of connection (C) per source, N and K. Previously, the communication model is studies in [16][17] to compare the performance of routing protocol by varying the parameter N. In order to change the value of V, the packet rate for our simulation is set as 2, 4, 6, 8, 10 and 12 packets/sec. The value for the node velocity and number of connection are shown in table 1.

### Node Density

Node density represent to the total number of nodes placed in the network. Average hop length of route is increases with increase of node density. For network of  $n$  number of nodes the average hop length of route is  $\Theta(\sqrt{n})$ [18]. This impact may increase with increases of node density. We keep the node density to constant value of 100 nodes.

**Table 1.** Simulation Parameter

<i>PARAMETER</i>	<i>VALUES</i>	<i>PARAMETER</i>	<i>VALUES</i>
Mobility Model	RWP	Packet size	512 bytes
Channel type	Wireless channel	Packet type	CBR
Antenna Model	Omni-directional	Channel Bandwidth	2 Mbps
MAC	802.11	Packet Rate	2, 4, 6, 8, 10 and 12 Packets/sec
Routing Protocol	AODV, DYMO and DSR	Number of Connections	10 and 20
Number of nodes	100	Transmission Range	250 m
Pause Time	0 Sec	Simulation Terrain	1000 m X 1000 m
Node Speed	5,10,15,20 and 25 m/s	Time of simulation	300 Sec.



## 5 Result and Analysis

In this section, we provide the results obtained from the number of experiments to estimate the desired true characteristic in ad hoc network. The scenarios are varying by node velocity and number of connections and unresponsive traffic (CBR) volume. We analyzed the experimental results contained in generated output trace files by using the AWK command. We carry out simulation to evaluate the how the routing protocol behaves according to node velocity, different traffic volume and number of connections. As [19] mentioned that the performance of routing protocol may vary dramatically according to the mobility model and performance ranking. Similar work also has been done in [20], but our mobility model, routing protocol and environment are different from them.

### 5.1 The Impact of Node Velocity

Velocity is an important parameter that can influence the proactive or reactive protocol performance like DYMO, DSR, and AODV. The node velocity determines the rate at which link fails and routing overhead required for route maintenance in reactive protocol. First, we study the effect of node velocity (mobility) on the ad hoc routing protocols. We found DYMO shows better performance in terms of PDR and traffic admission ratio so as to get more throughput than AODV and DSR (figure 1). Traffic admission ratio(TAR) is the ratio between packet send by the source to packet generated by source.

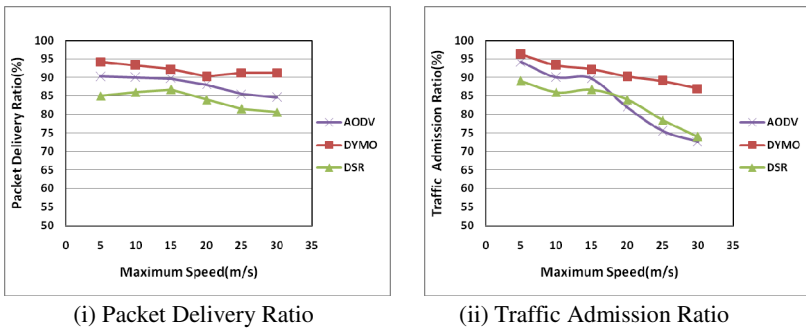
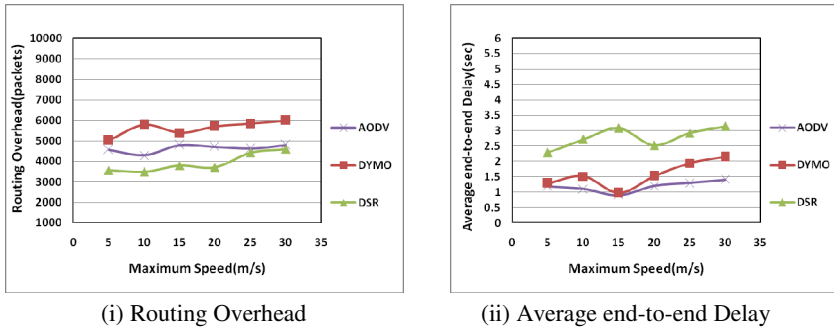


Fig. 1. AODV, DYMO and DSR with connection = 10 (i) PDR (ii) TAR

Next we measure routing overhead and average end to end delay with different node mobility. As figure 2 suggests DSR has lower routing overhead and AODV achieves lower delay as compared to others. DYMO protocol has worst routing overhead as compared to other with the increased node velocity. When node velocity of a network increases, the occurrence of link failure becomes frequent this causes more packet loss and decreases PDR. Whenever a link failure occurs, it initiate route discovery in DYMO and AODV and frequency of route discovery increases.

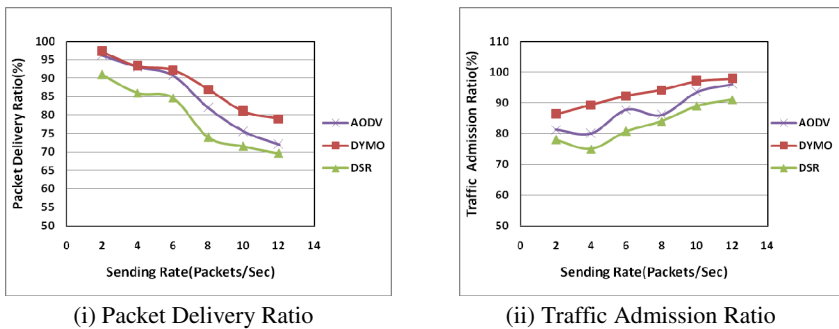


**Fig. 2.** AODV, DYMO and DSR with connection = 10 (i) Routing Overhead (ii) Average end-to-end Delay

So RREQ brings more routing overhead in AODV and DYMO. On the other hand DSR has less overhead due to route discovery. Due to inefficient route maintenance, average end to end delay is the largest for DYMO (see figure 2(ii)).

### 5.2 The Impact of Traffic Volume

In this section we examine the impact of increasing traffic volume on these routing protocols. As shown in figure 3(i) the PDR of DYMO is better than AODV and DSR. The traffic admission ratio becomes very high (see figure 3(ii)) for network with heavy traffic load. In such environment the packet collision ratio increase and as a result degrades the performance. We examine the system performance with 10 and 20 number of connection with transmission rate at 2, 4, 6, 8, 10 and 12 packets per second. The performance of DSR protocol is affected by increased traffic volume as compared to AODV and DYMO. DSR protocol has huge MAC load [17] than others. However, regardless of node velocity or traffic volume, DSR protocol produce less routing overhead than others as shown in figure 2(i) and 4(i). On the other hand DSR has higher delay as compared to others as shown in figure 4(ii).



**Fig. 3.** AODV, DYMO and DSR with connection = 10 (i) PDR (ii) TAR

Then we doubled the number of connection from 10 to 20. Comparing figure 1 and figure 5, we conclude that the effectiveness of all schemes decreased with the number of connection increase. The number connection creates more data session (i.e. more TAR) and causes more packet collisions. DSR performs worse than AODV when we double the number of data sessions from 10 to 20. It is due to data forwarded through alternate paths may collide with data transmitted via the main route.

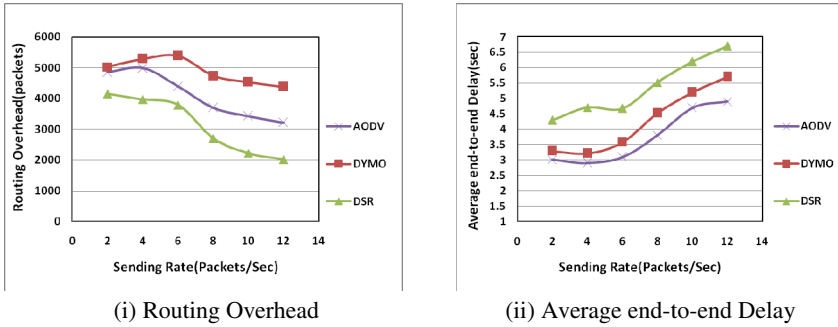


Fig. 4. AODV, DYMO and DSR with connection = 10 (i) Routing Overhead (ii) Average end-to-end Delay

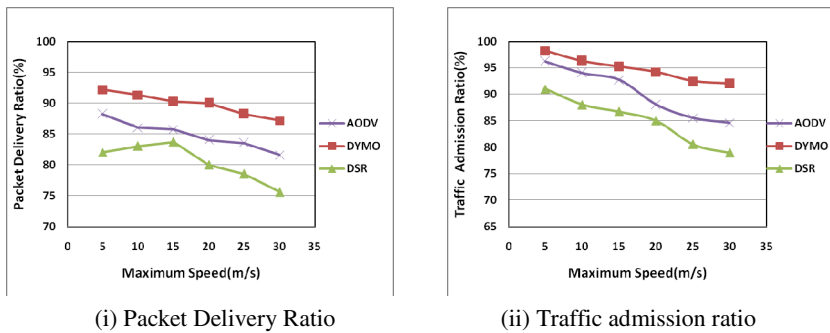


Fig. 5. AODV, DYMO and DSR with connection = 20 (i) PDR (ii) TAR

## 6 Conclusions

In this, we analyzed the impact of node velocity, connection and traffic volume on performance of MANET routing protocol with a thorough analysis. It is observed that the node velocity affect the link condition and topology on the performance of MANET routing protocols. This shows interaction among parameters like node velocity, connectivity, and performance. The simulation results shows that DYMO has better PDR and higher traffic admission ratio than DSR and AODV. On the other hand regardless of mobility, number of connection or traffic volume, DSR protocol has less routing overhead than AODV and DYMO. The PDR of these protocols

decreases for increasing the node velocity in MANET. Along with the behavior of routing protocols node velocity, traffic volume and number of connection per source determines the overall performance.

## References

- [1] Perkins, C.E., Belding-Royer, E., Das, S.R.: Ad-hoc on-demand distance vector (AODV) routing. IETF RFC 3561 (2003)
- [2] Johnson, D.B., Maltz, D.: Dynamic source routing in ad hoc wireless networks. In: Imielinski, T., Korth, H. (eds.) *Mobile Computing*, pp. 153–181. Kluwer Academic Publishers, Dordrecht (1996)
- [3] Chakeres, I., Perkins, C.: Dynamic MANET On-demand (DYMO) Routing. IETF Internet Draft 19 (2010)
- [4] Kushwah, S.S., Tomar, G.S.: Investigation of Effects of Mobility on Routing Protocols in MANET. In: *International Conferences on Ubiquitous Computing and Multimedia Applications (UCMA)*, Daejeon, pp. 82–84 (2011)
- [5] Ismail, Z., Hassan, R.: Effects of Packet Size on AODV Routing Protocol Implementation in Homogeneous and Heterogeneous MANET. In: *Third International Conference on Computational Intelligence, Modeling (CIMSIM)*, Langkawi, pp. 351–366 (2011)
- [6] Pala, A., Singh, J.P., Dutta, P.: The Effect of speed variation on different Traffic Patterns in Mobile Ad Hoc Network. In: *2nd International Conference on Computer, Communication, Control and Information Technology (C3IT)*, pp. 743–748 (2012)
- [7] Imrich, C., Marco, C., Jennifer, J.N.L.: Mobile ad hoc networking: imperatives and challenges. *Ad Hoc Networks* 1, 13–64 (2003)
- [8] Lin-zhu, W., Ya-qin, G., Min, S.: Performance comparison of Two Routing Protocols for Ad Hoc Networks. In: *WASE International Conference on Information Engineering*, pp. 260–262 (2009)
- [9] Mbarushimana, C., Shahrabi, A.: Comparative study of reactive and proactive routing protocols performance in mobile ad hoc networks. In: *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW 2007)*, pp. 679–684 (2007)
- [10] The Network Simulator, Ns2, <http://www.isi.edu/nsnam/ns>
- [11] Robins, A.D.: *GAWK:an effective AWK programming*, 3rd edn. (2010)
- [12] Camp, T., Boleng, J., Davies, V.: A survey of mobility models for ad hoc network research. *Wireless Communication and Mobile Computing Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications* 2, 483–502 (2002)
- [13] Camp, T., Boleng, J., Davies, V.: A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing* 2(5), 483–502 (2002)
- [14] Bisoy, S.K., Panda, M.R., Pallai, G.K., Panda, D.: Analysing the Interaction between Mobility Model and Unipath Routing Protocols in Mobile Ad Hoc Networks. *International Journal of Application or Innovation in Engineering & Management* 2(6), 449–456 (2013)
- [15] Pucha, H., Das, S.M., Hu, Y.C.: The performance impact of traffic patterns on routing protocols in mobile ad hoc networks. In: *Proc. of ACM MSWiM* (2004)
- [16] Broch, J., Maltz, D.A., Johnson, D.B., Hu, Y.C., Jetcheva, J.: A performance comparison of multi-hop wireless ad hoc network routing protocols. In: *Proc. of ACM MobiCom* (1998)

- [17] Das, S.R., Perkins, C.E., Royer, E.M.: Performance comparison of two on-demand routing protocols for ad hoc networks. In: Proc. of IEEE INFOCOM (2000)
- [18] Gupta, P., Kumar, P.R.: The capacity of wireless networks. *IEEE Transactions on Information Theory* 46(2), 388–404 (2000)
- [19] Bai, F., Sadagopan, N., Helmy, A.: IMPORTANT: A framework to systematically analyze the Impact of Mobility on Performance of Routing protocols for Adhoc Networks. In: *IEEE International Conference on Computer and Communications*, pp. 825–835 (2003)
- [20] Divecha, B., Abraham, A., Grosan, C., Sanyal, S.: Impact of node mobility on MANET routing protocols models. *Journal of Digital Information Management* 5(1), 19–24 (2007)

# Improving the QOS in MANET by Enhancing the Routing Technique of AOMDV Protocol

Ankita Sharma and Sumit Vashistha

Dept. of CSE, Sagar Institute of Research and Technology, Bhopal, India  
{as20121988, SUMITVBPL}@gmail.com

**Abstract.** A MANET is an interconnection of mobile devices by wireless links forming a dynamic topology without much physical network infrastructure such as routers, servers, access points or centralized administration. The multipath routing protocols establish efficient communication within the network by discovering multiple routes between a pair of source and destination in order to have load balancing to satisfy Quality of Service (QoS) requirements. The aim of this work is to modify the existing MANET reactive Multipath routing protocol Ad hoc On-demand Distance Vector (AOMDV) using drop minimization under MAC error control technique like collision minimization, dynamic queue scheme etc. In AOMDV routing we apply multipath base data sending scheme but that protocol not fulfill QoS so here we apply drop analysis base and find out pitfall of the AOMDV and after that we apply various drop minimization technique like rate base data sending scheme, queue base congestion control etc. and improve the performance of the network. A narrative Multipath QoS Aware Routing Protocol based on AOMDV is proposed to support packet delivery ratio, routing overhead minimization and UDP analysis constraints.

**Keywords:** MANET, QOS, AOMDV, QOS- AOMDV, Dynamic Queue length.

## 1 Introduction

All a mobile Ad-hoc network is a self configuring network of mobile devices. The provision of Quality of Services (QoS) guarantees is much more challenging in mobile Ad-hoc Networks (MANETs) [1] than that of wire-line networks. This is mainly due to the mobile nature of MANET nodes, and other characteristics of MANETs such as Multi hop communication and lack of central coordination. Therefore, designing a MANET routing protocol that guarantees the desired QoS is challenging. Many routing protocols for MANETs have been proposed.

Depending on the time of route discovery MANET routing protocols are divided into two categories; table-driven (proactive) routing protocol and on-demand (reactive) routing protocol. In table-driven routing protocols the routes are discovered

and refreshed periodically. All routing related information is stored in routing tables at each node. Whenever a traffic source needs a route, it uses the route available in the routing table. In contrast to this, the traffic source initiates route discovery process when it need route in case of on demand routing protocols. MANET nodes are mobile and, hence route failure probability is greater. The route discovery flood is associated with significant latency and overhead.

Among the on-demand routing protocols, MANET multipath routing protocols have relatively greater ability to reduce the route discovery frequency than MANET single path routing protocols. On-demand multi-path routing protocols discover multiple paths between a source-destination pair, in a single route discovery[2]. So a new discovery is needed only when all these paths fail. In contrast, a single path routing protocol has to invoke a new route discovery whenever the only path from source to destination fails.

QoS based routing becomes challenging in MANETs, as nodes should keep an up-to-date information about link status. Also, due to the dynamic nature of MANETs, maintaining the precise link state information is very difficult. Finally, the reserved resource may not be guaranteed because of the mobility caused path breakage or power depletion of the mobile hosts. QoS routing should rapidly find a feasible new route to recover the service.

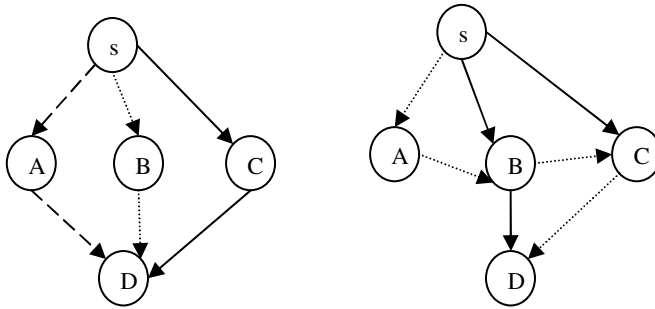
In this paper we improves the performance of the AOMDV routing protocol by applying random queue length based scheme to improves QoS of MANET.

The rest of this paper is organized as follows. In section 2 we discuss AOMDV routing protocol in section 3 we discuss different QoS issues for MANET. Section 4 has presents the related work and Section 5 has problem statement and in section 6 we discuss proposed scheme after that in section 7 we analyze the network behavior under network simulator-2. Finally in section 8 present the conclusion with future work.

## 2 AOMDV Routing

Ad Hoc On-demand Multipath Distance Vector (AOMDV) is a multi-path routing protocol [2, 3]. It is an extension to AODV [4] and provides two main services i.e. route discovery and maintenance. Unlike AODV, every RREP is being considered by the source node and thus multiple paths can be discovered in one route discovery. Being the hop-by-hop routing protocol, the intermediate node can maintain multiple path entries in their respective routing table.

AOMDV is considered more efficient in terms of creating less overhead Number of paths in any given source and destination is directly proportional to the number of nodes in entire network. AOMDV works more efficiently in dense and heavy networks. Fig. 1 is the Route discovery in Ad hoc On-demand Multipath Routing:



**Fig. 1.** 3 Node disjoint paths SAD, (b) 2 Link disjoint SBD and SCD SBD and SABCD

AOMDV is an on-demand routing protocol that builds multiple routes using request/reply cycles. When the source needs a route to the destination but no route information is known, it floods the ROUTE REQUEST (RREQ) message to the entire network. Because this packet is flooded, several duplicates that traversed through different routes reach the destination. The destination node selects multiple disjoint routes and sends ROUTE REPLY (RREP) packets back to the source via the chosen routes. The purpose of computing alternate paths for a source node is that when the primary path breaks due to node movement, one of the alternate paths can then be chosen as the next primary path and data transmission can continue without initiating another route discovery.

To discover distinct paths, AOMDV suppresses duplicate route requests (RREQs) at intermediate nodes. Such suppression comes in two different variations, resulting in either node (shown in Fig. 1 (a)) or link (shown in Fig. 1(b)) disjoint. AOMDV can be configured to either discover the link (no common link between any given pair of nodes) or node (in addition to link disjoint, common intermediate nodes are also excluded between any given pair of nodes) disjoint paths.

### 3 QoS Issue in MANET

Providing better QoS in MANET is challenging due to the following issues:

#### 3.1 Node Mobility

MANET nodes are move freely in network. This makes the topology dynamic. This means that topology information has a changing frequently and must be updated frequently and balance load in network allow data packets to be routed to their destinations. This updating means more routing overheads. Also, due to the node mobility packet losses increase, the end to end delay gets also affected.

#### 3.2 Lack of Central Control

The principal advantage of MANET is that it is deployed without planning in unknown terrains, hazardous conditions and its members can change dynamically. This



makes it difficult to have any centralized control. Hence the controlling activities will be distributed among the nodes, which require lot of information exchange. This also adds up to the routing overheads. [2, 3].

## 4 Related Work

M. Stewart B. G, Shahrabi A and Vallavaraj A in [5] have presented an load balancing approach. They find a Node disjoint and Link disjoint path. They select a path which has no loop. Source node initiates route request procedure. They broadcast RREQ messages to its adjacent nodes. Adjacent node each time save the request number and compare the number to previous RREQ message if it finds the number of request is small to previous one they accepted that request otherwise discards that RREQ message. They basically compares the advertise hop count number and selects lower advertise hop count number for path establishment.

Soundararajan, S. and Bhuvaneshwaran R.S. in [6] have presented load balancing mechanism. They used a concept of battery power of each node in their approach. They used threshold concept to calculate battery power of each node. They defined some threshold value for power if power is above defined threshold value then that node take part in communication and if the power is less than the defined threshold value simply that node rejected. In the end source to destination multiple paths are build. The energy level of those paths is above threshold value. Source select best path for data sending and set as a primary path and rest of that paths are secondary paths. If primary path is congested or break during communication then load distribute to the alternate paths.

Tekaya Mohamed, Tabbane Nabil and Tabbane Sami in [7] have proposed mechanism for load balancing. They apply changes in route discovery phase. In starting when source initiates route request. They chooses route which have less hop count, means traversing of nodes is less and congestion on that nodes is also less and high throughput performance. They chooses that path as primary path and rest of paths are secondary paths used for backups. They used same route maintenance procedure as in AOMDV protocol.

C.Wu et al .[8] presents an ad hoc on-demand multipath routing protocol which provides quality of service (QoS) support(Q-AOMDV) in terms of bandwidth, hop count and end-to-end delay in mobile ad hoc networks(MANET). The protocol uses path preference probability that is calculated by delays, bandwidth, and hop-count to select the path to transmit the packet. To discover the route to the destination, the source node floods the ROUTE REQUEST (RREQ) message to the entire network. Because this packet is flooded, several duplicates that traversed through different routes reach the destination. The destination node selects multiple disjoint routes and sends ROUTE REPLY (RREP) packets back to the source via the chosen routes.

This paper [9] discusses an extension of the on-demand DSR protocol. It consists of a scheme to distribute traffic among multiple routes in a network. Its performance in terms of delay degrades (reaches to 2.2 Seconds) as the traffic increases i.e. 40 and above.

This paper [10] states that for a QoS AODV routing protocol, problems would rise when the node density of the network is high. The reason is that the QoS AODV routing protocol uses the control message to exchange information between neighbors. When the node density is too high, the sending of control will cost much available data rate. As a result, the network will be ruined and traffic will be delayed more since control messages have higher priority than data packets.

In MANETs, node mobility often results in frequent topology changes, which presents a significant challenge when designing QoS routing protocols. High node mobility can make satisfying QoS requirements unreachable. Consequently, it is required that the network be combinatorically stable in order to achieve QoS support [11].

QoS is an agreement to provide guaranteed services, such as bandwidth, delay, delay jitter and packet delivery rate, to users. Supporting more than one QoS constraint makes the QoS routing problem NP-complete [12]. Therefore, we only consider the delay constraint when studying QoS-aware routing for supporting real-time video or audio transmission.

In this paper, we employ the facility to determine multiple routes to a host and switch between them to expand the definition of AOMDV [13]. Enabling a QoS constrained route from source to destination is one of the objectives of the routing protocol. The route chosen by the protocol must send packets with minimum bandwidth and end-to-end latency, without facing congestion. The protocol should satisfy the above constraints and also select the most robust among all possible candidate routes.

## 5 Problem Statement

Mobile Ad-hoc network is a form of dynamic with un-control way movement of node motion and decentralize control system that is bigger challenging issue for better performance, so that point to encourage to work in the field of quality control under multipath routing using AOMDV protocol, in this paper we design a system for minimization congestion and increasing quality of service of the network.

## 6 Proposed Methodology

Our aim to improve the Network Performance and decreases the drop packet in the MANET network through the QoS (quality of service) parameter base, we find out various drop reason in MANET like Collision, Media Packet Error , MAC Busy , Route not existed (update routing table) , Time to live Zero, Drop if Queue full, Drop if MAC Invalid Destination etc. according to that define Drop reason we minimize the drop via Quality of service base and increase Throughput, packet delivery ratio, TCP performance and UDP performance. All the work done through the network simulator -2 and we take routing protocol as AODV.

## 6.1 Proposed Algorithm

```

1) Generate test traffic through TCL script
2) Analysis trace file for finding drop reason and
   improving network performance
3) If (drop_reason==collision)
   {set MAC = CSMA/CA //carrier sense multiple access
with collision avoidance
   Send RTS and CTS message and avoid collision;
   }
   Else if (drop_reason ==MAC Busy)
   { set routing = AOMDV; //for multiple path
       find alternative path where channel ideal ;
   Or sender waits for ideal time ;
   }
   Else if (route ==Null)
   { wait for next RTT //round trip time;
   }
   Else if (queue_limit ==max)
   { sender use altranative path and send data to
destination;
       Or dynamic change queue limit
       Update queue length ;
   }
4) Create updated TCL script
5) Set mobile node = M; // M number of mobile node
6) Set MAC = CSMA/CA // carrier sense multiple access
with collision avoidance
7) Set RP = AOMDV // for removing MAC busy
8) Compute_Route (S,R, rr) //sender, receiver and radio
range
   a. If (next hop != null && radio range<=250)
{ create rtable;
Forward (route packet)
If (receiver ==true)
   {Receives route packet;
   Send's acknowledgment through each existing path;
   }
Else {route not exist;
   Wait next RTT;
   }
   }
   Else {node out of range}
9) Analyze updated trace file
10) Stop analysis

```

MANET form dynamic topology with each node contains routing functionality but MANET survives through various crises from physical layer to application layer because that work under the wireless communication with dynamic nature and no any centralize controller.

For Physical and MAC data dropping resolve through RTS (request to send) and CTS (Clear to send method) both resolve collision problem and MAC Busy, next we minimize queue full case drop through the alternative path mechanism, alternative path provides the communication between source to destination through more than one route if one route heavy loaded so we use alternative route and transmit data to destination that case we modified AODV routing protocol after this approach we improve the performance through transport layer, in communication transport layer work as gent between actual data and routing layer but also that layer provide guarantee of data delivery if we apply TCP protocol.

In TCP also survive through data drop, congestion, and retransmission time out problem, all those problem solve through bandwidth estimation technique, in that case we apply the mechanism for acknowledgment based bandwidth information in particular time interval so that sender send data according to bandwidth and if available bandwidth is less than the required bandwidth than sender minimize the data rate or increases the delay between data or alternative path use. All the above approach gathers into single module and improves the network quality of service privation.

## 7 Simulation Environment

The NS-2 [14] is a discrete event driven simulator developed at UC Berkeley. NS-2 is suitable for designing new protocols, comparing different protocols and traffic evaluations. It is an object oriented simulation written in C++, with an OTcl interpreter as a frontend. Simulation of protocols is performed on windows operating system using ns-2.31 version and also install "cygwin" to provide Linux environment in windows. The NS instructions can be used to define the topology structure of the network and the motion mode of the nodes, to configure the service source and the receiver etc. is mentioned in table 1.

### 7.1 Performance Metrics

We evaluate the following key performance metrics:

**Packet Delivery Fraction (PDF):** The ratio of the data packets delivered to the destinations to those generated by the CBR sources, i.e., Packet delivery fraction (pdf%) = (Received Packets / Sent Packets) \*100.

**Average end-to-end delay of data packets:** This includes all possible delays caused by buffering during route discovery latency, queuing at the interface queue, retransmission delays at the MAC, and propagation and transfer time.

**Normalized Routing load (NRL):** The number of routing packets transmitted per data packet delivered at the destination. Each hop-wise transmission of a routing packet is counted as one transmission, i.e. Normalized routing load = (routing packets sent) / receives.

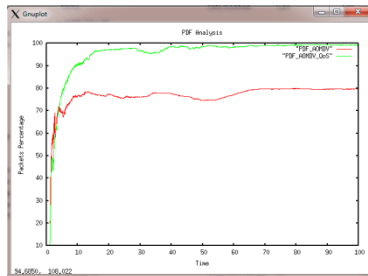
**Table 1.** Simulation parameters will uses for simulation

Simulator Used	NS-2.31
Number of nodes	50
Dimension of simulated area	800m×800m
Routing Protocol	AOMDV
Simulation time	100 sec.
Traffic type (TCP & UDP)	FTP & CBR
Packet size	512 bytes
Number of traffic connections	15
Node movement at maximum Speed	random & 20 m/s
Transmission range	250m

**7.2 Simulation Results**

**PDF Analysis in AOMDV and AOMDV-QoS**

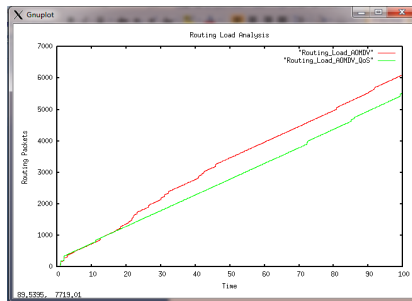
This graph represents the analysis of PDF (Packet Delivery Fraction) to measure parentage of packets are successfully deliver in network. The PDF in case of normal AOMDV routing are about 80% but in case of improving performance of AOMDV (AOMDV-QoS) are about 99% this is much better than normal AOMDV. The improvement in routing is enhancing the routing capability of protocol by that the receiving quantities of packets are increases in network in between sender and receiver. Here the dynamic queue length mechanism is improving the performance of AOMDV by that the buffering capacity of nodes is increase according to the number of packets.



**Fig. 2.** PDF Analysis

**Routing Load Analysis in AOMDV and AOMDV-QoS**

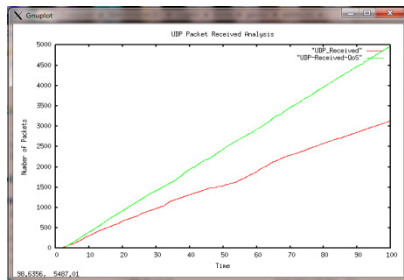
Routing load analysis is based on the number of control packets or routing packets the are complete the connection establishment procedure for data delivery in network. If the receiver is found by sender then in that case it also reply to sender for sending data. This graph represents the routing load analysis in case of AOMDV and improved AOMDV-QoS. The greater quantities of routing load are degrades the performance of network by that the data packets delivery is affected. This graph represents the routing packets analysis in case of AOMDV and AOMDV-QoS. The performance of normal AOMDV are not satisfactory here means above 6000 routing packets are deliver in network and in case of improved Quality of Service based AOMDV-QoS only about 5500 packets are deliver in network. It means the overhead in case of AOMDV are more and proposed dynamic queue based scheme are minimized it.



**Fig. 3.** Routing Load analysis

**UDP Packet Received Analysis in AOMDV and AOMDV-QoS**

UDP (User Datagram Protocol) packet analysis is very important to measure the performance of network. This graph represents the UDP packets analysis in case of AOMDV and AOMDV-QoS. The main problem in UDP packet is not acknowledged to sender about packet receiving in network. The numbers of packets are received in case of AOMDV is about 5000 but in case of proposed AOMDV-QoS routing about 3100 packets are received in network it means the dynamic queue management scheme are improved the packet receiving and enhance the routing capability.



**Fig. 4.** UDP Packet Received Analysis

**Table 2.** Overall Analysis

Parameters	AOMDV	AOMDV-QoS
SEND	9561.00	8221.00
RECV	7620.00	8150.00
ROUTINGPKTS	6115.00	5530.00
PDF	79.70	99.14
NRL	0.80	0.68
DROPPTS	44.00	40.00
No. of dropped data	1941	71
Actual Performance	(23296)85.21%	(21901)98.73%

**Table 3.** Different Drop Reasons

Drop Reasons	AOMDV		AOMDV-QoS	
Drop from COL	0	0.00%	0	0.00%
Drop from ARP	23	0.08%	20	0.09%
Drop from IFQ	1555	5.69%	49	0.22%
Drop from CBK	403	1.47%	73	0.33%
Drop from TOT	49	0.00%	0	0.00%
Drop from NRT	28	0.18%	0	0.00%
Drop from END	0	0.10%	29	0.13%
Drop from DUP	0	0.00%	0	0.00%
Drop from RET	0	0.00%	0	0.00%
Drop from BSY	0	0.00%	0	0.00%
Drop from SAL	0	0.00%	0	0.00%
Drop from ERR	0	0.00%	0	0.00%
Total Drop Via Congestion	1985	7.26%	111	0.50%
Total Drop	4043	14.79%	182	1.27%

### Overall Analysis in Case of AOMDV and AOMDV-QoS

This overall analysis represents the performance of both the protocols in table 1. Here we clearly notice that the performance of improved AOMDV-QoS is better. And the actual performance of network in proposed scheme is also improved about 13% more than AOMDV.

### Analysis of Different Packet Drop Reasons

This table 2 represents the analysis of different drop reasons in case of AOMDV and AOMDV-QoS. Here we notice that in case of proposed scheme the performance of proposed protocol are much better than normal AOMDV and also reduces the loss by different reasons.

## 8 Conclusion and Future Work

In this work, we present a new multipath QoS routing protocol for MANET with congestion control as well as quality control mechanism. In this paper we analyze packet delivery ratio, routing overhead and drop analysis and compare all the result through existing AOMDV routing protocol and finally we get our proposed QoS base AOMDV technique gives better performance and fulfill network QoS requirement.

The QoS metrics packet delivery ratio, routing load and drop minimization are better as we required in our proposal.

Here we simulate our result under one scenario in future we elaborate our network through number of different scenario and get result, we also enhance our work using ant like optimization and swarm intelligence technique and compare it.

## References

- [1] Kute, V.B., Kharat, M.U.: Survey on QoS for multi-path routing protocols in mobile ad-hoc networks. In: 3rd International Conference on Machine Learning and Computing (ICMLC 2011), vol. 4, pp. 524–528 (2011)
- [2] Patil, V.C., Biradar, R.V., Mudholkar, R.R., Sawant, S.R.: On-demand multipath routing protocols for mobile ad hoc networks issues and comparison. *International Journal of Wireless Communication and Simulation* 2(1), 21–38 (2010)
- [3] Marina, M.K., Das, S.R.: Ad hoc on-demand multipath distance vector routing. *Wireless Communication Mobile Computing* 6, 969–988 (2006)
- [4] Perkins, C.E., Royer, E.M., Das, S.: Ad hoc On-Demand Distance Vector (AODV) Routing, RFC 3561. IETF (2003)
- [5] Ali, M., Stewart, B.G., Shahrabi, A., Vallavaraj, A.: Multipath routing backbones for load balancing in Mobile Ad hoc Networks. In: 16th IEEE Mediterranean on Electrotechnical Conference (MELECON 2012), pp. 749–752 (2012)
- [6] Soundararajan, S., Bhuvaneshwaran, R.S.: Adaptive Multi-Path Routing for Load Balancing in Mobile Ad Hoc Networks. *Journal of Computer Science* (2012)
- [7] Mohamed, T., Nabil, T., Sami, T.: Multipath routing mechanism with load balancing in ad hoc network. In: IEEE International Conference on Computer Engineering and Systems (ICCES 2010) pp. 67–72 (2010)



- [8] Wu, C., Zhang, F., Yang, H.: A Novel QoS Multipath Path Routing in MANET. *JDCTA: International Journal of Digital Content Technology and its Applications* 4(3), 132–136 (2010)
- [9] Beaubrun, R., Molo, B.: Using DSR for Routing multimedia traffic in MANETs (January 2010)
- [10] Akana, C.M.V.S., Kumar, S., Divakar, C.: QoS for Real time transmission on MANETs. *International Journal of Advanced Networking and Applications* 02(03), 679–685 (2010)
- [11] Chen, K., Shah, S.H., Nahrstedt, K.: Cross layer design for data accessibility in mobile ad hoc networks. *J. Wireless Commun.* 21, 49–75 (2002)
- [12] Chen, S.: Routing support for providing guaranteed end-to-end quality-of-service. Ph.D. dissertation, Univ. of IL at Urbana-Champaign (1999)
- [13] Marina, M.K., Das, S.R.: Ad hoc On-demand Multipath Distance Vector Routing. Computer Science Department, Stony Brook University (2003)
- [14] <http://www.isi.edu/nsnam/ns/>

# PRMACA: A Promoter Region Identification Using Multiple Attractor Cellular Automata (MACA)

Pokkuluri Kiran Sree<sup>1</sup>, Inampudi Ramesh Babu<sup>2</sup>, and S.S.S.N. Usha Devi Nedunuri<sup>3</sup>

<sup>1</sup> Department of Computer Science & Engineering,  
Jawaharlal Nehru Technological Universtiy Hyderabad, Kukatpally  
profkiransree@gmail.com

<sup>2</sup> Department of Computer Science & Engineering, Acharya Nagarjuna University

<sup>3</sup> Dept of CSE, Jawaharlal Nehru Technological Universtiy Kakinada

**Abstract.** Promoter region identification from sequences of DNA has gained a remarkable attention in recent years. Even though there are some identification techniques addressing this problem, the approximate accuracy in identifying the promoter region is closely 70% to 72%. An automated procedure was evolved with MACA (Multiple Attractor Cellular Automata) for identifying promoter regions. We have tested the proposed classifier ENCODE benchmark datasets with over three dozens of modern competing predictors shows that proposed algorithm (PRMACA) provides the best overall accuracy that ranges between 77% and 88.7%. PRMACA can identify promoter region with DNA or Amino acid sequences as inputs.

**Keywords:** Promoter region, Cellular Automata, MACA.

## 1 Introduction

Promoter provides control for the gene transcription with some specific regulation, which is located in gene upstream. Promoters consist of sequences of DNA [1] factors which are predictable by proteins which are known as transcription factors. Promoter of type prokaryotes consists of two sequences which are short at -35 and -10 positions from the transcription start site at upstream. Eukaryotic promoters [2] are very difficult to characterize and exceptionally diverse. They typically lie upstream of the gene and can have regulatory elements several kilo bases away from the transcriptional start site.

Promoters are molecules with macro region that are responsible for a wide range of vital biochemical functions, which includes acting as oxygen, nutrient transport and building up muscle fibers. Specifically, the Promoters are chains of amino acids and DNA sequences, of which there are 20 different types, coupled by peptide bonds [2]. The structural hierarchy possessed by Promoters is typically referred to as primary and tertiary region. Promoter Region Predication from sequences of amino acid gives tremendous value to biological community. This is because the higher-level and secondary level [1], [2] regions determine the function of the Promoters and consequently, the insight into its function can be inferred from that.

As genome sequencing projects are increasing tremendously. The ENCODE databases [3],[4] of primary Promoter regions are expanding tremendously. Promoter Data Banks are not growing at a faster rate due to innate difficulties in finding the levels of the regions. Region determination[5], [6] procedure experimental setups will be very expensive, time consuming, require more labor and may not applicable to all the Promoters. Keeping in view of shortcomings of laboratory procedures in predicting the region of Promoter major research have been dedicated to Promoter identification of high level regions using computational techniques. This is usually called as Promoter folding problem which is the greatest challenge in bioinformatics. This is the ability to predict the higher level regions from the amino acid sequence.

## 2 Related Works in PROMOTER Region Identification

Reese MG al [2] has proposed a Neural Network Model for predicting the promoter region. Steen Knudsen al [3] has used statistical classifiers to identify promoter regions. Techniques for region identification include, but are not limited to, constraint programming methods, statistical approaches to predict the probability of an amino acid being in one of the structural elements, and Bayesian network models. Nearest neighbor techniques attempt to predict the region of a central residue, within a segment of amino acids[21] [24], based on the known regions of homologous segments. In, a technique based on multiple linear regressions was presented to predict region. We also proposed an algorithm [16] to identify the promoter regions using CA with text clustering with a lesser accuracy.

## 3 Cellular Automata

Cellular Automata (CA) is a simple model of a spatially extended decentralized system, made up of a number of individual components (cells). The communication among constituent cells is limited to local interaction. Each individual cell is in a specific state that changes over time depending on the states of its neighbors. From the days of Von Neumann who first proposed the model of Cellular Automata (CA)[4],[5], to Wolfram's recent book 'A New Kind of Science', the simple and local neighborhood region of CA has attracted researchers from diverse disciplines. It has been subjected to rigorous mathematical and physical analysis [18] [19] for past fifty years and its application has been proposed in different branches of science - both social and physical.

Definition: CA is defined a four tiple  $\langle G, Z, N, F \rangle$

Where  $G \rightarrow$  Grid (Set of cells)

$Z \rightarrow$  Set of possible cell states

$N \rightarrow$  Set which describe cells neighborhoods

$F \rightarrow$  Transition Function (Rules of automata)

The evolution process is directed by the popular Genetic Algorithm (GA) with the underlying philosophy of survival of the fittest gene. This GA framework can be adopted to arrive at the desired CA rule region appropriate to model a physical system. The goals of GA formulation are to enhance the understanding of the ways CA performs computations and to learn how CA may be evolved to perform a specific computational task and to understand how evolution creates complex global behavior in a locally interconnected system of simple cells.

## 4 Design of MACA Based Pattern Classifier

An  $m$ -bit MACA with  $k$ -attractor basins can be viewed as a classifier which was evolved naturally. It classifies a given set of patterns into  $q$  number of distinct MACA classes, each class containing the set of states in the attractor basin. To enhance the classification accuracy of the machine, most of the works have employed MACA to classify patterns into two classes (say I and II). The following example in Fig 1 illustrates an MACA [5], [6],[14] based two class pattern classifier.

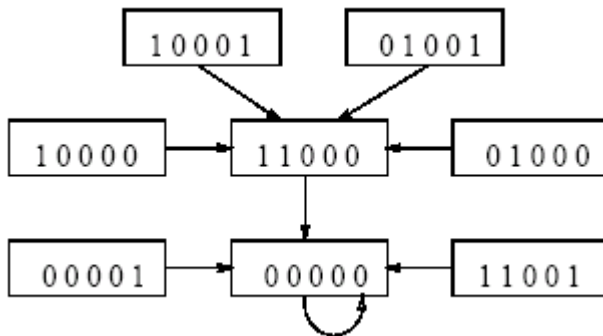


Fig. 1. Example of MACA with basin 0000

### 4.1 Random Generation of Initial Population

#### Algorithm

Input: Pattern set  $P$  to be memorized, Maximum Generation ( $G_{\max}$ ).

Output: Dependency String (DES) and associated information.

begin

Step 1: Generate 50 new chromosomes for initial population (IP1).

Step 2: Initialize counter for generation  $CG = \text{zero}$ ;  $PP1 \leftarrow IP1$ .

Step 3: Compute fitness value  $F$  for each chromosome of  $PP1$ .

Step 4: Store DES, and corresponding information for which the fitness value  $F = 100\%$ .

Step 5: If  $F = 100\%$  for at least one chromosome of  $PP1$ , then go to Step 12.

- Step 6: Rank the chromosomes in order of fitness.
- Step 7: Increment counter for generation (CG)
- Step 8: If  $CG > G_{max}$  then go to Step 11.
- Step 9: Form NP by selection, crossover and mutation.
- Step 10:  $PP1 \leftarrow NP$ ; Go to Step 3.
- Step 11: Store DS, and corresponding information for which fitness value is maximum.
- Step 12: Stop.

### 4.2 PRMACA Tree Building

Input : Training set  $S = \{S1, S2, \dots, SK\}$

Output : PRMACA Tree.

Partition(S, K)

- Step 1 : Generate a PRMACA with k number of attractor basins.
- Step 2 : Distribute training set S into k attractor basins (nodes).
- Step 3 : Evaluate the distribution in each attractor basin
- Step 4 : If all the examples (S') of an attractor basin (node) belong to a single class, then label the attractor basin.
- Step 5 : If examples (S') of an attractor basin belong to K' number of MACA classes, then, Partition (S', K').
- Step 6 : Stop.

## 5 Experimental Results

In this section we present the results on using PRMACA for ENCODE dataset. Values are given for the percentage accuracy on test set promoter sequences and the percentage accuracy on test set non promoter sequences. The prediction accuracies are reported in table 1. The interface is provided in figure 2.

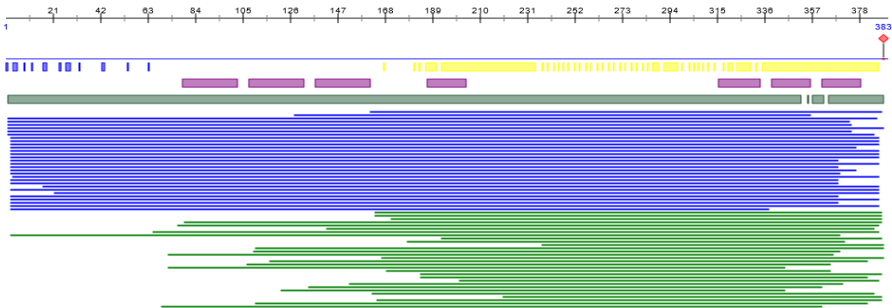


Fig. 2. Promoter Region Identification Interface (Green as Promoter)



**Table 1.** Predictive Accuracy

Algorithm	DNA Sequence	Amino Acid Sequenc
Dicodon Usage	61%	57%
Bayesian	51%	46%
Normal ID-CA	68%	72%
Neural Networks	74%	68%
PRMACA	79%	72.5%

## 6 Conclusion

PRMACA predicts the promoter regions from DNA sequence as well as Amino Acid sequences and provides the best overall accuracy that ranges between 77% and 88.7%. To provide a more thorough analysis of the viability of our proposed technique many experiments were conducted. Existing region-identification methods can predict the region with 70% to 72% accuracy. Our extensive results indicate that such a level of accuracy is attainable, and can be potentially surpassed with our method.

## References

1. Sandvej, K., Gratama, J.W., Munch, M., Zhou, X.-G., Bolhuis, R.L., Andresen, B.S., Gregersen, N., Hamilton-Dutoit, S.: Sequence analysis of the Epstein-Barr virus (EBV) latent membrane protein-1 gene and promoter region: identification of four variants among wild-type EBV isolates. *Blood* 90(1), 323–330 (1997)
2. Lavrovsky, Y., Schwartzman, M.L., Levere, R.D., Kappas, A., Abraham, N.G.: Identification of binding sites for transcription factors NF-kappa B and AP-2 in the promoter region of the human heme oxygenase 1 gene. In: *Proceedings of the National Academy of Sciences*, vol. 91(13), pp. 5987–5991 (1994)
3. Horikawa, I., LouAnn Cable, P., Afshari, C., Carl Barrett, J.: Cloning and characterization of the promoter region of human telomerase reverse transcriptase gene. *Cancer Research* 59(4), 826–830 (1999)
4. Miskimins, W.K., Roberts, M.P., McClelland, A., Ruddle, F.H.: Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. In: *Proceedings of the National Academy of Sciences*, vol. 82(20), pp. 6741–6744 (1985)
5. Huang, Q.R., Morris, D., Manolios, N.: Identification and characterisation of polymorphisms in the promoter region of the human Apo-1/Fas (CD95) gene. *Molecular Immunology* 34(8), 577–582 (1997)
6. Okuyama, Y., Ishiguro, H., Nankai, M., Shibuya, H., Watanabe, A., Arinami, T.: Identification of a polymorphism in the promoter region of DRD4 associated with the human novelty seeking personality trait. *Molecular Psychiatry* 5(1), 64–69 (2004)
7. Bauer, C.E., Young, D.A., Marrs, B.L.: Analysis of the *Rhodobacter capsulatus* puf operon. Location of the oxygen-regulated promoter region and the identification of an additional puf-encoded gene. *Journal of Biological Chemistry* 263(10), 4820–4827 (1988)

8. Mitra, D., Smith, M.: Digital Sequences Processing in Promoter Region Identification. *Innovations in Applied Artificial Intelligence Lecture Notes in Computer Science* vol. 3029, pp. 40–49 (2004)
9. Reese, M.G.: Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.* 26(1), 51–56 (2001)
10. Abagyan, R., Batalov, S., Cardozo, T., Totrov, M., Webber, J., Zhou, Y.: Homology Modeling With Internal Coordinate Mechanics: Deformation Zone Mapping and Improvements of Models via Conformational Search. *Promoters: Region, Function and Genetics* 1, 29–37 (1997)
11. Maji, P., Pal Chaudhuri, P.: FMACA: A Fuzzy Cellular Automata Based Pattern Classifier. In: Lee, Y., Li, J., Whang, K.-Y., Lee, D. (eds.) *DASFAA 2004. LNCS*, vol. 2973, pp. 494–505. Springer, Heidelberg (2004)
12. Kiran Sree, P., Babu, I.R.: Investigating an Artificial Immune System to Strengthen the Promoter Region Identification and Promoter Coding Region Identification using Cellular Automata Classifier. *International Journal of Bioinformatics Research and Applications* 5(6), 647–662 (2009)
13. Kiran Sree, P., Babu, I.R.: Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering. *The International Arab Journal of Information Technology (IAJIT)* 7(1), 75–78 (2010)
14. Kiran Sree, P., Babu, I.R.: A Novel Promoter Coding Region Identifying Tool using Cellular Automata Classifier with Trust-Region Method and Parallel Scan Algorithm (NPCRITCACA). *International Journal of Biotechnology & Biochemistry (IJBB)* 4(2), 177–189 (2008)



# Analyzing Statistical Effect of Sampling on Network Traffic Dataset

Raman Singh<sup>1</sup>, Harish Kumar<sup>1</sup>, and R.K. Singla<sup>2</sup>

<sup>1</sup> University Institute of Engineering and Technology, Panjab University, Chandigarh, India

<sup>2</sup> DCSA, Panjab University, Chandigarh, India

raman.singh@ieee.org, {harishk,rksingla}@pu.ac.in

**Abstract.** In sampling of huge network traffic dataset, some packets are chosen out of total packets. Leftover packets may have effect on statistical characteristics of the data. In this paper effect of sampling on statistical characteristics is discussed. A well-known benchmarked NSL KDD network traffic dataset is used. Three sampling techniques namely - random, systematic and under-over sampling are used. Various attributes of dataset considered are duration, src\_bytes, dst\_bytes, wrong\_fragment, num\_compromised, num\_file\_creations and srv\_count. Parameter of statistical characteristics like range, mean and standard deviation is used for analysis purpose. Result shows that sampling has considerable statistical effect on network traffic dataset.

**Keywords:** Sampling, Network traffic dataset, Intrusion detection system.

## 1 Introduction

Intrusion detection system (IDS) based on anomaly detection techniques process network traffic to find out abnormal traffic patterns. This requires huge computational power to process continuous streams of traffic. This computational requirement can be reduced by considering samples out of this traffic data. Sampling is the processes to pick some instances (says samples) out of total population. Method of picking of samples differs with various types of sampling techniques. Choosing only some packets of dataset may have effect on statistical characteristics. In this paper statistical effects of sampling on network traffic datasets have been analyzed. This paper has five sections. Section -1 introduces the topic, Section 2 discusses about sampling, types and various issues, and Section 3 describes research model and dataset used for this study. In section 4 results obtained from this study are discussed. Section 5 concludes the study, followed by references.

## 2 Sampling and Its Issues

There are many types of sampling which can be used for sub-set selection out of traffic data. This may introduce some issues in the dataset for example there may be

possibility of rejecting those samples which have influences on statistical characteristics. Widely used sampling methods and their issues are discussed as below:

## **2.1 Random Sampling**

Some samples are picked randomly from the total population [1]. This sampling assumes that every member is homogeneous and does not differentiate between varying statistical nature of network traffic instances. Due to which there may be variation of statistical nature between sampled dataset and full dataset. In network traffic dataset, packets are not homogeneous.

## **2.2 Systematic Sampling**

Samples are picked systematically. For examples samples are picket at a regular interval (say every 10<sup>th</sup>) from some starting point. This technique like Random sampling also assumes homogeneous samples. There may be variation of statistical properties between sampled and un-sampled dataset in this case too [2].

## **2.3 Under and Over-sampling**

In Under-sampling, samples are picked with fewer rates, if the population is large. In Over-sampling, samples are picked with higher rates for small population [3]. This technique is used when dataset is imbalanced in nature like Network traffic. For majority classes randomly or systematically, less samples are selected while for minority classes samples may be synthetically generated to balance the contribution of each class. Major issue with this sampling is loss of useful information. Synthetically generation of samples from minor classes may change statistical properties [4][5].

# **3 Material and Methods**

## **3.1 Network Traffic Dataset**

Intrusion detection dataset is prepared by MIT Lincoln Laboratory with DARPA sponsorship. Captured network traffic dataset can be used for performance analysis of IDS. Various network traffic services like e-mail, telnet, web was used to generate large amount of traffic and benchmarked dataset is prepared to develop new techniques against security threats. Different types of scenarios like users, managers, programmers were used [6]. There are other network datasets available like PU-CAN [7], but NSL KDD [8] dataset is used as a benchmark to evaluate performance. Total number of packets of un-sampled dataset used is 125973. To consider various sample sizes, different percentages of total dataset are taken into account.

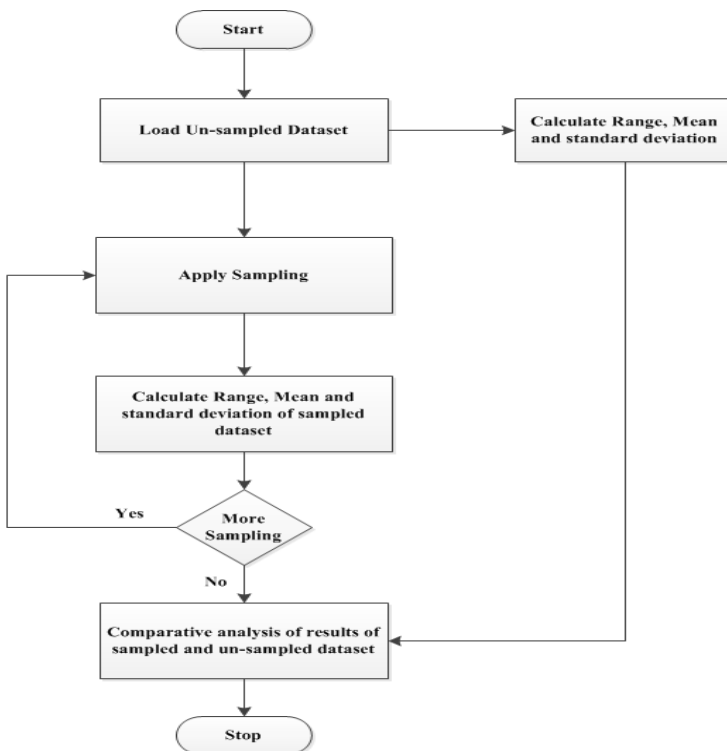
**Table 1.** List of dataset attributes used in the study

Attribute No.	Attribute	Attribute No.	Attribute
1	duration	5	num_compromised
2	src_bytes	6	num_file_creations
3	dst_bytes	7	srv_count
4	wrong_fragment		

Total seven attributes of dataset are used in this study as listed in table 1. Various performance evaluation measures used in this study are – range, mean and standard deviation.

### 3.2 Research Methodology

Purpose of this study is to analyze statistical effect of sampling on network traffic dataset. Figure 1 shows flow chart of experiments carried out.

**Fig. 1.** Research methodology used for the study

In the first phase, range, mean and standard deviation of un-sampled dataset are calculated. In the second phase three sampling methods are applied and three sampled dataset is prepared. Then Range, Mean and standard deviation for these sampled dataset is calculated. In the next phase, comparative analysis of un-sampled and sampled dataset is carried out.

## 4 Results and Discussion

Comparative study of un-sampled and various sampled dataset are discussed in sub-sections below.

### 4.1 Random Sampling

Four experiments are carried out by selecting 10%, 25%, 50% and 75% of full dataset using random sampling. Comparative values for various attributes and different parameters using randomly sampled and un-sampled dataset are shown in table 2.

**Table 2.** Values of un-sampled and randomly sampled network traffic dataset

Sub-Set Att. No.		Full Set (Un-sampled)	10% of full set	25% of full set	50% of full set	75% of full set
1	R <sup>+</sup>	42908	42804	42837	42908	42888
	M <sup>*</sup>	287.14	269.24	312.08	286.42	267.27
	SD <sup>-</sup>	2604.52	2461.10	2747.14	2623.39	2457.41
2	R	1379963888	21945520	693375640	621568663	693375640
	M	45566.74	14983.23	58594.14	26370.47	28152.07
	SD	5870331.18	382514.84	4981342.01	2936011.18	3440518.70
3	R	1309937401	5153460	1309937401	1309937401	1309937401
	M	19779.11	3517.13	44429.81	23840.29	25549.49
	SD	4021269.15	93328.90	7381795.94	5220048.91	4643220.84
4	R	3	3	3	3	3
	M	0.023	0.025	0.024	0.023	0.022
	SD	0.254	0.268	0.264	0.254	0.250
5	R	7479	691	1739	7479	7479
	M	0.279	0.243	0.351	0.260	0.325
	SD	23.94	10.65	15.97	30.81	26.73
6	R	43	40	40	40	43
	M	0.013	0.017	0.014	0.011	0.015
	SD	0.484	0.605	0.520	0.447	0.538
7	R	510	510	510	510	510
	M	27.74	27.97	27.80	27.58	26.98
	SD	72.64	73.51	72.72	71.93	70.54

Att. No. -> Attribute Number, + Range, \* Mean, - Standard Deviation

From table 2, it can be concluded that that there is difference between full set (Un-sampled) and various random sampled sub-sets. It is found that there is considerable variation between range and mean values of un-sampled and sampled dataset. It implies that range and mean values of various attribute changes while random sampling is used to reduce computations. Further, it is also observed that there are considerable differences in standard deviation of un-sampled and sampled dataset. Hence, random sampling may change the distribution of network traffic dataset impacting the wrong detection of traffic patterns by IDS. So it is not advisable for network traffic dataset.

### 4.2 Systematic Sampling

Four experiments are carried out by selecting 1<sup>th</sup> (2<sup>nd</sup>, 3<sup>rd</sup>, 5<sup>th</sup> and 10<sup>th</sup>) packet from un-sampled dataset. Comparative values for various attributes and different parameters using systematic sampled and un-sampled dataset are shown in table 3.

**Table 3.** Values of un-sampled and systematic sampled network traffic dataset

I Att. No.		Full Set ( Un-sampled)	l = 2	l = 3	l = 5	l = 10
1	R <sup>+</sup>	42908	42837	42804	42888	42616
	M <sup>*</sup>	287.14	275.61	282.20	283.36	287.86
	SD <sup>-</sup>	2604.52	2520.52	2588.37	2573.40	2622.73
2	R	1379963888	693375640	693375640	693375640	693375640
	M	45566.74	39966.17	34827.92	38397.50	66839.14
	SD	5870331.18	4108269.95	3868858.6	4376257.51	6183262.67
3	R	1310000000	5155468	5151385	5150938	5150938
	M	19779.11	2768.73	3387.33	3295.68	3105.02
	SD	4021269	60738.8	83218.63	79855.15	71573.47
4	R	3	3	3	3	3
	M	0.023	0.023	0.024	0.024	0.026
	SD	0.254	0.256	0.258	0.260	0.273
5	R	7479	7479	1739	1043	1043
	M	0.279	0.369	0.206	0.297	0.336
	SD	23.94	32.46	12.35	13.35	15.15
6	R	43	43	43	28	26
	M	0.013	0.015	0.014	0.012	0.010
	SD	0.484	0.557	0.560	0.416	0.385
7	R	510	510	510	510	510
	M	27.74	27.53	27.94	27.66	27.68
	SD	72.64	72.05	73.20	72.86	72.10

Att. No. -> Attribute Number, + Range, \* Mean, - Standard Deviation

From table 3 it is found that there are variations of ranges while using systematic sampling. So, minimum and largest values for each attribute may change during systematic sampling impacting range. Means of un-sampled and systematic sampled

dataset also have large gap. Variations of standard deviation show that systematic sampling may change the distribution of network traffic dataset.

### 4.3 Under and Over Sampling

Based on network protocol and its services, dataset is divided into different 72 stratum before conducting four experiments by selecting 'n' packets (175, 437, 700 and 874) from each stratum. This creates sub dataset of size 10%, 25%, 40% and 50% of full dataset. For example, 874 samples are picked from each stratum. If the number of packets are greater than 874 for some stratum, under sampling is applied and randomly any 874 packets are selected. If the number of packets are smaller than 874 for any stratum, again randomly packets are selected up to 874 packets. As 72 stratum are created, total 62986 packets are selected. (Around 50% of total 125973 packets of full dataset). Comparative values for various attributes and different parameters using under-over sampled and un-sampled dataset are shown in table 4.

**Table 4.** Values of un-sampled and under-over sampled network traffic dataset

n		Full set (Un-sampled)	n = 175	n = 437	n = 700	n = 874
Att. No.						
1	R <sup>+</sup>	42908	42636	41802	41802	42658
	M <sup>*</sup>	287.14	233.19	300.62	284.47	308.47
	SD <sup>-</sup>	2604.52	2231.68	2592.99	2515.30	2663.23
2	R	1379963888	5131424	381709090	5133876	381709090
	M	45566.74	6981.10	54495.82	6953.25	12869.30
	SD	5870331.18	172737.87	4306504.33	174031.55	1530300.65
3	R	1310000000	276683	5150180	273545	5150772
	M	19779.11	1363.85	1664.03	1316.92	1401.93
	SD	4021269	6541.317	41630	5992.449	22338.39
4	R	3	3	3	3	3
	M	0.023	0.042	0.039	0.037	0.039
	SD	0.253	0.344	0.338	0.328	0.335
5	R	7479	462	462	83	462
	M	0.279	0.078	0.056	0.032	0.055
	SD	23.94	4.19	2.76	0.65	2.90
6	R	43	8	10	13	15
	M	0.013	0.012	0.009	0.008	0.009
	SD	0.484	0.281	0.250	0.223	0.253
7	R	510	510	510	510	510
	M	27.74	20.07	19.95	20.50	19.99
	SD	72.64	53.47	52.94	55.08	54.05

Att. No. -> Attribute Number, + Range, \* Mean, - Standard Deviation

Due to random or systematic sampling within each stratum largest and smallest values of various attributes changes effecting range. It is found that there are considerable variation of range and mean values of un-sampled and various under-over sampled dataset. Table 4 also shows large variation of standard deviation which means distribution of sampled dataset considerably changes. As stratum of different protocol-services are created loss of information in sample is minimized but under-over sampling fails to protect the statistical characteristics of dataset.

## 5 Conclusion

Network traffic dataset is huge and IDS require large computational power to process it. Sampling may be used to decrease the size of dataset but this may introduce some loss of information. In this paper statistical effect of random, systematic and under-over sampling is discussed. Range, mean and standard deviation parameters are used for performance evaluation. Results shows that random and systematic should not be used for network traffic dataset as there are considerable variations of range, mean and standard deviation. Result also shows that due to sampling, distribution of dataset also changes. In under and over sampling for each stratum statistical characteristics changes and hence it is not advisable for statistical analysis of dataset. Results clearly show that in case of under and over sampling too, distribution of dataset changes. Some statistical characteristics balancing method should be developed for network traffic dataset which can handle statistical variations of sampled dataset.

## References

1. He, G., Hou, J.C.: On sampling self-similar Internet traffic. *Computer Networks* 50(16), 2919–2936 (2006)
2. Mahmood, A.N., Hu, J., Tari, Z., Leckie, C.: Critical infrastructure protection: Resource efficient sampling to improve detection of less frequent patterns in network traffic. *Journal of Network and Computer Applications* 33(4), 491–502 (2010)
3. Liu, J.G., Martin, C.: Generative oversampling for imbalanced datasets. In: *International Conference on Data Mining (DMIN)*, Las Vegas, Nevada, USA, June 25-28, pp. 66–72 (2007)
4. Kotsiantis, S., Kanellopoulos, D., Pintelas, P.: Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering* 30(1), 25–36 (2006)
5. Liu, Y., Yu, X., Huang, J.X., An, A.: Combining integrated sampling with SVM ensembles for learning from imbalanced datasets. *Information Processing & Management* 47(4), 617–631 (2011)

6. Lippmann Richard, P., Fried David, J., Isaac, G., Haines Joshua, W., Kendall Kristopher, R., David, M., Dan, W., Webster Seth, E., Dan, W., Cunningham Robert, K., Zissman Marc, A.: Evaluating Intrusion Detection Systems: The 1998 DARPA Off-line Intrusion Detection Evaluation. In: DARPA Information Survivability Conference and Exposition, Hilton Head, South Carolina, January 25-27, pp. 12–26 (2000)
7. Singh, R., Kumar, H., Singla, R.K.: Traffic Analysis of Campus Network for Classification of Broadcast Data. In: 47th Annual National Convention of Computer Society of India, International Conference on Intelligent Infrastructure, Science City, Kolkata, December 1-2, pp. 163–166 (2012)
8. KDD dataset, <http://nsl.cs.unb.ca/NSL-KDD>



# Localization of Information Dissemination in Agriculture Using Mobile Networks

Lokesh Jain<sup>1</sup>, Harish Kumar<sup>2</sup>, and R.K. Singla<sup>3</sup>

<sup>1</sup> School of Elect. Engg. & Inform. Tech., Punjab Agricultural University, Ludhiana

<sup>2</sup> UIET, Panjab University, Chandigarh, India

<sup>3</sup> DCSA, Panjab University, Chandigarh, India

lokjain@pau.edu, {harishk,rksingla}@pu.ac.in

**Abstract.** Usage of Information and Communication Technology (ICT) in agriculture focuses on improvement in productivity by timely and precise dissemination of appropriate information to all stakeholders. Web and mobile communication technology has quickly become the world's most common way of transmitting data and services in the developing countries. As most of the applications developed are either web based and having static information over the time, the need is to have a dynamic model for the information dissemination, which is continuously changing with respect to time and location. Localization of the information is important, as the generic information provided may not be relevant to the farmer's location. In this work, an effort has been made to conceptualize a system based on local parameters like location, language etc.

**Keywords:** ICT, Web application, Mobile Application, Localization, Dynamic Model.

## 1 Introduction

India is an agriculture-based economy with 69% of its population living in rural areas. Most of them are involved in the agricultural and allied activities. A large amount of research and extension work is being done to increase the quality and quantity of agricultural produce by the Indian Council of Agricultural Research (ICAR), State Agricultural Universities (SAUs) and various other organizations in public and private sector. It is also very important that the technology thrust should lay greater emphasis on the dissemination of scientific and technological information from the research laboratory and test fields to its actual stakeholders like farmers. Print and electronic media such as libraries and information centers (like Kissan Call Centers) are playing an important role in providing information to the farmers. These centers are using information systems and/or extension workers etc., to reach the farmers. India's mobile phone network is one of the largest networks in world with high penetration in the rural areas. Mobile phone tariff is also one of the lowest in the India as compared to other countries in the world. These networks have large potential to be used as media to disseminate information to the farmers.

In this paper, it has been tried to figure out the method to disseminate the agriinformation to a farmer on the basis of his/her geographical location. Various ways are explored to provide information the farmers using multimedia, web and mobile applications etc. Use of mobile-based communication to localization the information has also been explored in the sub-subsequent sections.

## **2 Changing Agricultural Scenario and Information Needs**

In past, pamphlets, posters, radio, television etc. are used for dissemination of information as per needs of farmers. There is large time gap between information transmitted and received by appropriate quarters. Need for providing accurate agriculture related information to farmers at appropriate time leads to application of Information and Communication Technology (ICT) in agriculture. Concepts of virtual community, E-publications of agricultural literature, various information networks, institutional repositories, social networks sites and information kiosks etc. is being employed for dissemination of agricultural information. Various ICT initiatives/tools like multimedia, web and mobile applications are being deployed for dissemination of information. Following section discusses some of the ICT tools, which can be used to disseminate the information to farming community.

### **2.1 Use Multimedia in Agriculture**

Technologies developed within laboratories and research field, must reach farmers at earliest [1]. Multimedia is one of the effective media for dissemination of the information to farmers among all other ways of communication [2]. For example photographs of symptoms of diseases, insect pest damage and of insects can be made available using multimedia for their effective monitoring and management. Quality of extension material and process can be supplemented time to time by multimedia and virtual reality. For information dissemination, interactive multimedia CDs, DVDs are useful tools [3] to transmit information to farming community. The Digital Green system [4] disseminates targeted agricultural information to small and marginal farmers in India through digital video. It works on the concept of participatory learning which includes a digital video database produced by farmers and experts. These videos are distributed to the farmers through a repository. Farmers are motivated and trained by these recorded experience. Indian society of agribusiness professional [5] has taken up an initiative to create videos and animation clips for good agriculture practices and adoption of new technology to be used during farmer extension and awareness program. Similarly other organizations, NGOs are disseminating the knowledge to farmers using multimedia.

### **2.2 Web and Mobile Applications**

To disseminate agriculture related information to farmers, mobile and Internet based communication are precise, fast and cheaper mode to transfer the information from

research farms and laboratories. It has reached almost all the districts and administrative blocks of India [6]. Farmers are eager and become concerned to get rapid, true and reliable information in globally changing agricultural scenario. Transfer of requisite and latest agricultural information to the farmers in scattered locations at diverse geographical situation in India is a very tough job. Still it can be said that development of technology and its dissemination to the framers is a continuous process [7]. This has been tried to be accomplished by the various web portals [8] as mentioned in Table 1:

**Table 1.** Some Agricultural Information systems and their hyperlinks

<b>Agricultural Information System</b>	<b>Web Address</b>
aAQUA (almost All QUestion Answered)	<a href="http://www.aaqua.org">http://www.aaqua.org</a>
KISSANKERLA	<a href="http://www.kissankerla.net">http://www.kissankerla.net</a>
Ministry of Agriculture : Farmer portal	<a href="http://farmer.gov.in">http://farmer.gov.in</a>
TNAU Agritech Portal	<a href="http://www.agritech.tnau.ac.in">http://www.agritech.tnau.ac.in</a>
DACNET	<a href="http://dacnet.nic.in">http://dacnet.nic.in</a>
e-Krishi	<a href="http://www.ekrishi.org">http://www.ekrishi.org</a>
InDG (India Development Gateway portal)	<a href="http://www.indg.in">http://www.indg.in</a>
RKMP (Rice Knowledge Management Portal)	<a href="http://www.rkmp.co.in">http://www.rkmp.co.in</a>
Agropedia	<a href="http://agopedia.iitk.ac.in">http://agopedia.iitk.ac.in</a>
Mahindra’s Kisan Mitra	<a href="http://www.mahindrakisanmitra.com">http://www.mahindrakisanmitra.com</a>
IFFCO Agri-Portal	<a href="http://www.iffco.nic.in">http://www.iffco.nic.in</a>
Agriwatch Portal	<a href="http://www.agriwatch.com">http://www.agriwatch.com</a>
mKrishi, Tata Consultancy Services Limited	<a href="http://www.tcs.com/offerings/technology-products/mKRISHI/Pages/default.aspx">http://www.tcs.com/offerings/technology-products/mKRISHI/Pages/default.aspx</a>

### 3 Need for Mobile Phone Based Communication in Agriculture

Mobile phones have such a huge impact on society that Jeffrey Sachs, Director of the Earth Institute at Columbia University, has described the mobile phone as the ‘single most transformative tool for development’ [9]. Mobile phone technology is simple, inexpensive and convenient to use. Mobile networks access is now widely available at district and block level as well as in remote areas. This has opened up wide opportunities for agricultural scientists and extension workers to work with the agricultural community in more focused and personalized manner. Some of the gaps for dissemination of information in agricultural based activities are:

### **3.1 Lack of Mobile Network Usage for Agricultural Information Dissemination**

Various expert, decision support and fuzzy logic based systems for agricultural activities and processes have been developed. But most of these are standalone or web based system or related to agricultural marketing activities like commodity exchanges, AGMARKNET ([www.agmarknet.nic.in](http://www.agmarknet.nic.in)), e-Choupal etc. IT knowledge is the basic requirement for a farmer to use these kinds of systems. It is felt that the mass understandable and economical technology needs to be exploited for dissemination of agricultural related information. The mobile phone network is one such popular technology that can be exploited easily for such information dissemination. Mobile phone networks are available in almost all the geographical areas of India and used by majority of farmers [10].

### **3.2 Localization of the Desired Agricultural Activity**

In some cases we are not able to guess the exact location of the agricultural activity under review. Therefore it is required to use the telecommunication infrastructure to accurately define the location of agricultural field under consideration. By getting this accurate information, the soft computing based information systems can give localized information for better decision making. It is possible to exploit the power of mobile technology to transfer this knowledge to the concerned at any time and at any location 24x7x365. The knowledge should also be made available in the native language of the farmers.

## **4 Mobile Based Systems for Localization**

Technological advances are opening up huge opportunities. With new technologies, unmatched developments, and booming economies, the face of the world of telecommunications & IT is changing rapidly. Integrating the villages into this change and adopting the technologies for maximizing the impact in rural areas is a challenging task [11]. Mainly mobile technologies have created new channels to communicate with others in a well-located way. The penetration of mobile applications in the rural market is increasing day by day due to reducing cost of smart phones, Personnel Digital Assistant (PDA), tablets etc.

The mobile application developer can use the Application Programme Interface (API) of these phones/technologies to develop various applications for providing localized services to the farmers. For getting the location information at the micro-macro level that is administrative block level, two techniques can be employed,

- Selection based techniques
- Automatic location information.

In the selection based techniques, a database is populated with the information of the state, district and blocks of the country, for example In MNREGA database, desired

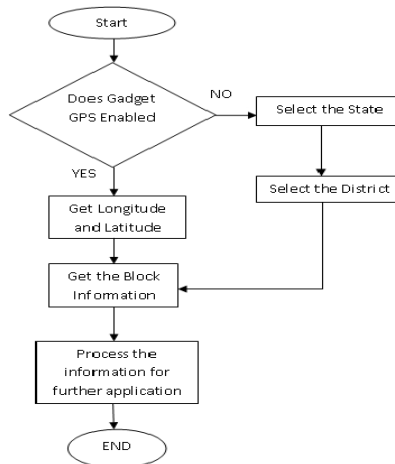
information can be obtained by selecting the state, district and administrative blocks under observation. There are 28 states and 7 union territories in India and having overall 640 districts and 6451 administrative blocks (Census of India, 2011). Thus, the whole nation is divided into smaller units called administrative blocks. These blocks having the diverse information related to the geography, geology, administration, ecology etc. Thus localized information can be obtained if stored in the digital format.

In automatic location information technique, Global Positioning System (GPS) of gadget can pass information related to longitude and latitude to the server. Location API on server can process this information to get the geographical data of administrative blocks in which gadget is located. Then based on this geographic location agricultural data can be extracted and disseminated to this gadget. The server Location API can access the GIS maps, which is imported in the database supporting spatial information e.g. MySQL Database.

Let us assume that in MySQL database, there is 'shptbl' named table which contains the column named 'shape' which contains blocks geometry along with other columns like 'area', 'perimeter', 'block-name' and 'block-id' etc. Let the automatic location technique passes two parameters named 'Longi' for longitude and 'Latti' for Latitude to the server, then the following syntax will help to find the location of the administrative block from the 'shptbl' table where the spatial information is stored using 'MBRContains' function of MySQL.

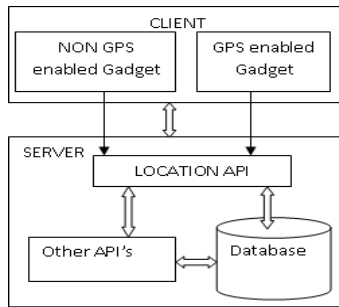
```
“Select * from shptbl where MBRContains(shptbl.shape, GeomfromText(Longi,Latti))”
```

This syntax can provide the whole information that is contained in the selected row of the table. This information can be used further for processing of the other parameters. Flow of both techniques has been described in fig. 1.



**Fig. 1.** Flow Chart for Implementation of Location based search

The Geo-location API features of the HTML5 can help to find the latitude and longitude of the user location. This location specific information can be used for activity under consideration to mine the information at local level (a location unit) for example Block level or District level. This can help the experts and extension personnel at the Agriculture Development office, Block level development offices, SAU's, Krishi Vigyan Kendras etc. to find out where the activity under interest is prevalent and the action plan can be prepared to cop up the situation at the location unit. Fig. 2 describes the way an electronic gadget whether a GPS enabled or not to access the Location API and the uses the other API's to process the location information. This information is used in the IS's for getting the localized knowledge. Storing the information about the local spoken language can further localize it. Based on this input, data can be disseminated in the localized format.



**Fig. 2.** Client-server architecture using the Location API

These kind of applications in the form of IS can be used anywhere by accessing the server through Internet. These applications can collect the information in the form of some questionnaire, images etc. and pass this collected pieces of information to the server for processing. The processing at the server level can be done using various techniques like fuzzy logic, artificial neural networks, genetic algorithms etc. and the result thus obtained can be passed to the farmer's mobile phone requesting the desired data.

## 5 Conclusion

Generic Information and knowledge furnished to farmers may not be relevant to them at their respective geographical locations. Dissemination of local level information through mobile networks can help them to improve their productivity and hence socioeconomic aspects. The experts can get the information about the location unit where a particular agricultural activity is more prevalent and it can help them to research that location unit. The information being requested can be processed for use in further sessions to get improvement in decision level and this can make the decisions dynamic based on previous experiences. Information dissemination in local language may also be explored.

## References

1. Philip, H., Vennila, M.A., Gomathy, M.: Multimedia: An Effective Communication Technology for Extension. In: International Conference on Communication for Development in the information Age: Extending the Benefits of Technology for All, Varanasi, India (September 07, 2003)
2. Brun, C., Mangstl, A.: Worldwide access to ICT: the Digital Divide. In: Virchow, D., von Braun, J. (eds.) Villages in the future: Crops, jobs and livelihood, pp. 259–262. Springer, Berlin (2001)
3. Senthil, K.M., Chandrakandan, K., Padma, C., Padma, S.R.: Modern communication technologies for sustainable farming in globalize era. In: National Seminar on ‘Responding to Changes and Challenges: New Roles of Agricultural Extension’, Nagpur (2003)
4. Gandhi, R., Veeraraghavan, R., Toyama, K., Ramprasad, V.: Digital Green: Participatory video for agricultural extension. In: International Conference on Information and Communication Technologies and Development, ICTD 2007, Bangalore, December 15–16, pp. 1–10 (2007)
5. Indian Society of Agribusiness Professionals (ISAP), <http://www.isapindia.org/isap/agripractices.php> (last seen August 03, 2013)
6. Chauhan, N.M.: Expectations of the Farmers from ICT in Agriculture. *Indian Res. J. Ext. Edu.* 10(1), 42–45 (2010)
7. Mehta, P.: Information Technology in Agriculture: Reaching the Unreached. In: National Workshop on ICT for Agriculture and Rural Development, Ahmadabad (2003)
8. Saravanan, R.: e-Agriculture Prototype for Knowledge Facilitation among Tribal Farmers of North-East India: Innovations, Impact and Lessons. *The Journal of Agricultural Education and Extension* 19(2), 113–131 (2013)
9. Vodafone Group Plc: Connected Agriculture: The role of mobile in driving efficiency and sustainability in the food and agriculture value chain, [http://www.vodafone.com/content/dam/vodaphone/about/sustainability/2011/pdf/connected\\_agriculture.pdf](http://www.vodafone.com/content/dam/vodaphone/about/sustainability/2011/pdf/connected_agriculture.pdf) (last seen August 03, 2013)
10. Jain, L., Kumar, H., Singla, R.K.: A Review of Fuzzy Rule Promotion Techniques in Agriculture Information System. *Journal of Computer Applications. IJCA, Special Issues on IP Multimedia Communications*, 55–60 (2011)
11. Nanda, S., Arunachalam, S.: Reaching the Unreached - Community based Village Knowledge Centres & Village Resource Centres. Jamsetji Tata National Virtual Academy, M S Swaminathan Research Foundation, Chennai (2010) ISBN : 978-81-88355-15-0

# Video Traffic in Wireless Sensor Networks

Shilpa Pandey, Dharm Singh, Naveen Choudhary, and Neha Mehta

Department of Computer Science and Engineering, College of Technology and Engineering,  
Maharana Pratap University of Agriculture and Technology, Udaipur, Rajasthan, India  
{ershilpa8,nehamehta.291187}@gmail.com,  
dharm@mpuat.ac.in, naveenc\_121@yahoo.com

**Abstract.** The primary goal of clustering in wireless sensor networks is to efficiently manage the energy consumption of among the sensory nodes using multi-hop communication with particular cluster. This proposed approach will reconsider the concept of leveling, sectoring and adaptive clustering for video traffic. Partitioning the entire network into clusters would manage the video traffic required for the data dissemination to the base station using multi-hop communication. Making both the cluster head selection and cluster size dynamic has been the central idea of this routing protocol. In this paper the simulation for transmission of videos over wireless networks is performed in MATLAB and the results are analyzed on the basis evaluation of configuring parameters such as PSNR, Network Life time, and Average Power Consumption. The Proposed mechanism improved the PSNR, Network Life time, and Average Power Consumption.

**Keywords:** WSNs, Video over Wireless Network, Two part sectoring, Leveling.

## 1 Introduction

Wireless sensor networks (WSNs) consists of thousands or millions of wireless sensor nodes deployed to gather the information from their surroundings and drive the sensed data to the Base Station (BS) or sink node through wireless links. A sensor node due to small size leads to limited battery power and lower hardware performance. A base station aggregates and summarizes the collected data and sends these to a user or to the remote host. The goal of Wireless sensor networks is how to collect data in energy efficient way since the energy is limited. The most commonly used technique for maximizing network lifetime is clustering [2]. Network is partitioned into small region called cluster. Clustering will facilitate scalability and each cluster has a cluster head (CH). The members of the cluster send the sensed data to the cluster head (CH) which then aggregate the sensed data for transmitting it to the base station. The cluster head (CH) can directly transmit the data to the base station if it near to the base station otherwise can send to the intermediate cluster heads. The election of cluster head may depends on various methods like random selection of cluster head, inter or intra cluster head selection, based on threshold value [8]. Cluster



head used for data aggregation and transmission may drain its energy earlier than the other members of cluster so the cluster head is reelected after each round. Today most of the existing routing algorithms in clustering are constraint to leveling which are deliberated to be energy efficient. But it may not be reliable as all sensors, which tend to receive data from the upper level are capable of forwarding the packets to next lower level and are not considered as energy proficient, when directed to the base station. Here the selection of cluster head is done by the member nodes. At higher level of clustering, the cluster head close to base station will act as cluster head for majority of lower level cluster heads which results in early dying of nodes close to base station.

## 2 Literature Review

Power Aware Sectoring Based Clustering Algorithm for Wireless Sensor Networks (PASCAL) -This protocol addresses some of the issues such as: Reducing the number of nodes that are used for transfer of data through leveling and sectoring, avoids flooding in the network, energy efficient path used to transfer data to base station and hence power aware., reliability of gateways increased and complexity of network is decreased [4]. LI-QING, GUO *et al.* proposed two hop LEACH protocol improves network lifetime, an adaptive algorithm is used for multi hop transmission and cluster head election. According to distance from BS nodes are tagged as near or far, all the near nodes belong to one cluster while the far nodes use the Greedy K means algorithm for division into different clusters. To balance the energy load the node having maximum residual energy is selected as the CH. To communicate with BS far CH uses near CH, if it is alive else direct communication is used [6]. Haosong Gou *et al.* Proposed partition based LEACH algorithm (p-LEACH), In this paper a partition based LEACH algorithm is proposed (p-LEACH) in which firstly the network is partitioned into sectors then a CH is elected in each sector using centralized calculations, the node having highest energy is elected as the head node [3].

## 3 Proposed Efficient Routing Mechanism

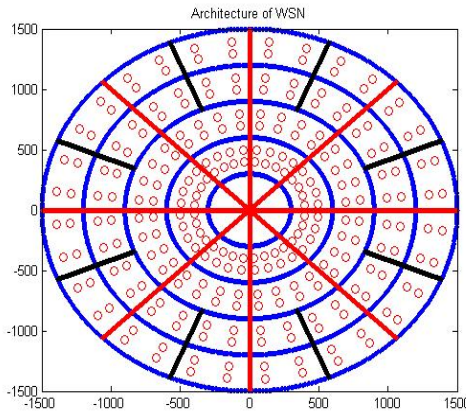
### 3.1 Proposed Architecture

#### 3.1.1 Nodes Deployment

The network is in the form of levels and sectors. The concentric circles in the figure are the levels. Each level has equiangular sectors. Level 1 is towards the sink which is the centre of all the circles. Level 3 sectors are quite large in comparison to the sectors in level 1 and 2. If the node has to forward data in these sectors from one end to end it is quite a large distance [5]. This approach is not efficient. Instead of forwarding the packet to the CH in this case the node should itself forward the packet to its nearest neighbour in the next level neighbouring sector. To identify the distance between the CH and the communicating node all the sectors have specific parameter name 'part'. If the sector size is large enough then the sector is divided into two parts. When the

communicating node is in part one and CH is in the other part of the same sector then the node will not forward the data to the CH. Here it will forward data directly to the nearest neighbour in the next level. The algorithmic detail of this approach is given in the following pseudo code.

The figure shown below represents layered architecture of wireless sensor network. The blue concentric circles depicts levels, red lines in the form of spokes illustrates the sector boundaries and the black lines in figure are partitioning sectors in level 3 and 4 into 2 parts. This has been done to approximately equalise the area of all the sectors. This concept has been termed as two part sectoring.



**Fig. 1.** Architecture of proposed routing mechanism

In this figure the blue concentric circles are for levels. Red lines in the form of spokes depict the sector boundaries and black lines in figure are partitioning sectors in level 3 and 4 into 2 parts. This has been done to approximately equalise the area of all the sectors. This concept has been termed as two part sectoring.

The proposed algorithm reuses the architecture based which included sector n level based approach [5]. Algorithm followed by partition phase, level and sector set up phase then after CH election stage which includes near neighbour, next head selection discovery, multi-hop routing phase and finally with video transmission stage. The network progresses in terms of rounds. One round is equal to the time until which all the communicating nodes send one video frame.

1. Initialize the network and deploy WSN nodes in sector based architecture.
2. Use adaptive size equalization of all the sectors.
3. BS selects a cluster head (CH) per sector randomly, since initially all the nodes have same energy.
4. BS forwards distance adjacency matrix of all the nodes to them.
5. All the nodes identify their nearest neighbors in the next level towards BS

6. BS initializes all the state and monitoring variables for network monitoring such as the no of nodes, no of dead nodes, first node dead time, residual average node energy, standard deviation, no of frames transferred, round status, round progress, etc.
7. The nodes having video in buffer, start forwarding the video frames one by one to the sink.
8. Repeat steps 8 to for frames number 1 to nFrames
  - a) For all concurrent communicating nodes 1 to concurrent
  - b) Calculate total number of dead nodes for this round
  - c) Record first node dead time
9. Keeping track of progress of the round at stage four since CH selection is to be done at 25, 50 75 and 100% progress of the round with adaptive CH selection more in the levels near to the BS.
10. Nodes having Video traffic creates packet of the video frame
11. It asks for part variable of the CH.
12. If  $\text{node.part} < \text{CH.part}$  then the communicating node don't forward the packet to the CH. It looks for its nearest neighbor in next level and forward packet to it. By this temporarily it acts as its own CH.
13. Else forward the packet to CH
14. CH or next node in the forwarding path checks whether it is in 1<sup>st</sup> level, if true then forward the packet directly to the BS and jump to step 8 else it searches for next level neighbor which repeat steps 11 until the packet reaches BS.
15. If the progress of the round is 25, 50, 75 or 100% then perform the following:
  - a)  $\text{stage} = \text{stage} - 1$
  - b) CH selection up to no levels-stage levels every time depending on the residual energy
  - c) Move to step 7 for next frame communication from all the communicating sensor nodes
16. If some packet is recorded at the BS It will record its sender in sink log and gathers the frame to video buffer corresponding to the node id.
17. Here quality analysis of the frames are done, improper frames may be discarded and called again to resend
18. Cluster head selection for next round depending on the residual energy move to step 7 until whole video traffic is not transmitted.

## 4 Simulation and Results

The simulated WSNs consist of 256 nodes deployed within an area which is divided into circular levels at fix distances. Each of the level is divided into some sectors. As the distance of any level becomes greater than a threshold the sector size also becomes adaptive. Size of all the sectors is not same. The base station is located at the

center of the field. The most of the parameters and the energy dissipation model of radio and hardware are similar [1]. The simulations are conducted for 4-levels and 8-sectors, the number of frames transmitted are varied. The simulation parameters are summarized in Table 1. Following table gives the details of the simulation results of the implementation of both SBMC and our proposed method.

**Table 1.** Simulation Parameter

Parameters	Value
Eo-Initial Energy	15,20,25(J) (variable)
Eelec Energy dissipation per bit for the transceiver circuit	50nJ/bit
Total No. of nodes	256
Video Communication Node	8
$E_{fs}$	10pJ/bit/m <sup>2</sup>
$E_{amp}$	0.0013pJ/bit/m <sup>4</sup>
Video Format	MPEG
Number of Frames	200 (variable)

**Table 2.** Simulation Results

Protocol	Initial Energy (Eo)	FirstNode Dead (FND) Round	FND Time	Avg. PSNR
Proposed Method	15	134	1.42	29.4
	20	185	1.47	40.96
	25	Node Still Alive	Node Still Alive	42.0
SBMC	15	83	6.24	24.61
	20	117	8.24	35.27
	25	150	1.5	42.11

The metric for First Node Dead is used to indicate the round and time that when the first node in the network got dead. This phase is required to be delayed in the network as much as possible. The FND round and time are much later in our proposed method than SBMC. At last the peak signal noise ratio between the sent video and received video frames is calculated to evaluate quality of video traffic.

Following analytical data represents the details of the simulation in graphical manner for comparison between the proposed and the referred work. The bars are always high in the proposed algorithm. In the proposed work the network life time is

measured using metrics like first node dead time First node dead time is illustrated in figure: 2 which shows that proposed routing mechanism performs better than SBMC and leads to efficient power utilization. This is because the efficient routing mechanism dissipates energy evenly in all the sensor nodes which means communication between intermediate nodes in the network is carried without any interruption. In SBMC algorithm first node dies after the 117th round whereas proposed algorithm dead time of the first node is 185th round when initial energy is 20J.

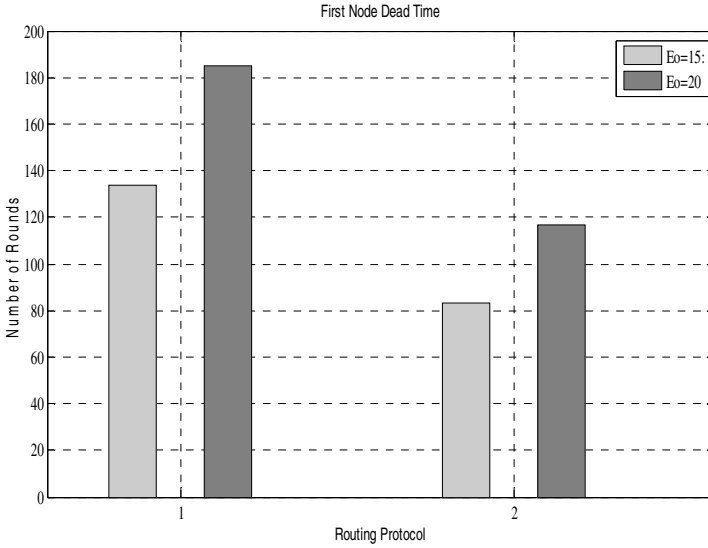


Fig. 2. Graph representing first node dead time

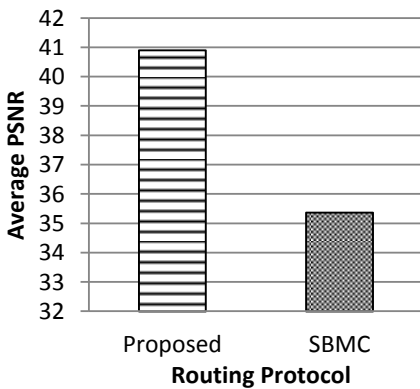


Fig. 3. Average PSNR

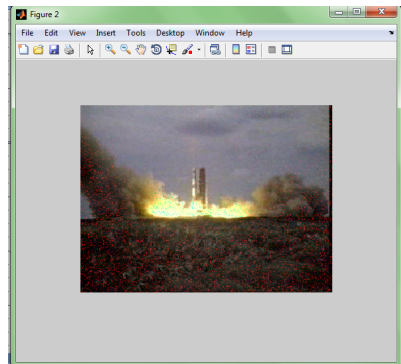


Fig. 4. Video Input

The peak signal to noise ratio in figure:3 indicates quality of video traffic at the sink is 40.9 in proposed routing mechanism and 35.27 in SBMC when initial energy is 20J.Hence better quality of video is obtained for the proposed mechanism.

## 5 Conclusion and Future Scope

This paper concludes a multipath routing algorithm for energy efficient video transmission from the source nodes to the sink. The sink gathers the video in the buffer. The scheme adds advantages by adaptively changing the sector size to equalize areas of all sectors. It also avoids the extra communication occurring in method from data node to distant CH. To avoid this, the data node acts as temporary CH for that communication. So packet is to be transmitted to a comparatively less distance. Concept of two part sector has been introduced and tested for performance. The results are shown that network life time and quality of video in proposed work increase in the wireless sensor networks. Future work in this protocol is required to make the packets available to the sink node in real time. This will increase the applicability of this algorithm in the WSN for video surveillance.

## References

1. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the Hawaii International Conference on System Sciences, Maui, Hawaii (2000)
2. Bandyopadhyay, S., Coyle, E.: An energy efficient hierarchical clustering algorithm for wireless sensor networks. In: Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), San Francisco, California (2003)
3. Gou, H., Yoo, Y., Zeng, H.: A partition based LEACH algorithm. In: IEEE Ninth International Conference on Computer and Information Technology, pp. 40–45 (2009)
4. Mirza, M.A., Garimella, R.M.: PASCAL: Power Aware Sectoring Based Clustering Algorithm for Wireless Sensor Networks. In: Proceeding of International Conference on Information Networking (ICOIN), pp. 1–6 (2009)
5. Bore Gowda, S.B., Puttamadappa, C., Mruthyunjaya, H.S., et al.: Sector based Multi-hop Clustering Protocol for Wireless Sensor Networks. International Journal of Computer Applications 43(13) (2012) (0975 – 8887)
6. Li-Qing, G., Xie, Y., Yang, C.H., et al.: Improvement on leach by combining adaptive cluster head election and two-hop transmission. In: Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao (2010)
7. Gong, B., Li, L., Wang, S., Zhou, X.: Multi-hop Routing Protocol with Unequal Clustering for Wireless Sensor Networks. In: Proceeding of IEEE CCCM, pp. 552–556 (2008)
8. Al-Karaki, J.N., Kamal, A.E.: Routing Techniques in Wireless Sensor Networks: A Survey. In: IEEE Wireless Communications, pp. 6–28 (2004)
9. Ahmed, K., Gregory, M.: A.: Wireless Sensor Network Data Centric Storage routing using Castalia. In: Telecommunication Networks and Applications Conference (ATNAC), Australasian, pp. 1–8 (2012)

10. Rosario, D., Costa, R., Paraense, H., et al.: A smart multi-hop hierarchical routing protocol for efficient video communication over wireless multimedia sensor networks. In: 2nd IEEE International Workshop on Smart Communication Protocols and Algorithms (2012)
11. Wang, X., Qian, L., Wu, J.: An energy and Distance Based Clustering Protocol for wireless sensor Networks. In: Novel Algorithm and Techniques in Telecommunications and Networking, pp. 409–412. Springer, Netherland (2010)
12. Farooq, M.O., Dogar, A.B., Shah, G.A.: MR-LEACH: Multi-hop Routing with Low Energy Adaptive Clustering Hierarchy. In: Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM), pp. 262–268 (2010)

# A Novel Pairwise Key Establishment and Management in Hierarchical Wireless Sensor Networks (HWSN) Using Matrix

B. Premamayudu<sup>1</sup>, K. Venkata Rao<sup>2</sup>, and P. Suresh Varma<sup>3</sup>

<sup>1</sup> Department of Information Technology, Vignan University, Vadlamudi, Guntur, Andhra Pradesh India

<sup>2</sup> Department of CSE, Vignan's Institute of Information Technology, Visakhapatnam (Dt), Andhra Pradesh, India

<sup>3</sup> Adikavi Nannaya University, Rajahmundry, W. Godavari (Dt), Andhra Pradesh, India  
{premayayudu, vrkoduganti}@gmail.com, vermaps@yahoo.com

**Abstract.** Key Management is an important challenging issue in wireless sensor networks to provide the confidentiality for the sensitive data which is sensed by the sensors from various fields. There are so many resource constraints for wireless sensor networks, implementing key management scheme in WSN. In general many key management schemes are proposed for Computer Networks like ECC, Diffie-Hellman, RSA and public key based schemes, but all are not suitable for wireless sensor networks. WSNS are suitable for many applications like Smart homes, Automation, Vehicular traffic management, Habitat Monitoring, Precision agriculture, Disaster detection and Surveillance. All the applications need secure communication and must protect from the eavesdropping, key compromising, sensor node capture attack, and message authentication. In this paper, we propose A Novel Pairwise Key Establishment and Management in Hierarchical Wireless Sensor Networks (HWSN) Using Matrix to prevent all the above attack and to achieve network connectivity, scalability, and resilience for the network. This scheme is used for the heterogeneous sensor networks. In this scheme, Base station prepares the key matrices for Cluster heads and Sensor nodes and generates the key chains for both cluster head and sensor nodes and preloaded before deployment. Our scheme prevents the following attacks: sensor capture attacks, key compromising attacks, malicious node attacks.

**Keywords:** Wireless sensor networks, Cluster, Base station.

## 1 Introduction

Wireless sensor networks (WSNs) increasingly become viable solutions to many challenging problems for both military and civilian applications, including target tracking, battlefield surveillance, intruder detection and scientific exploration. However, deploying sensors without security in mind has often proved to be dangerous in hostile environments. In wireless communication environments an



adversary not only can eavesdrop the radio traffic in a network, but also can intercept the exchanged data. To prevent the malicious node impersonating good nodes for spreading misleading data intentionally, secret keys should be used to achieve data confidentiality, integrity and authentication between communicating parties. Additionally, wireless sensors are not tamper resistant due to their low cost. Thus, the adversary may physically capture some sensors to compromise their stored sensitive data and communication keys. This serious attack is known as node capture attack, which makes the node's operation become under the control of the adversary. Hence key protection should be paid more attention in sensor networks.

## 2 Related Work

Eschenauer and Glor [1] proposed a random key pre-distribution scheme that focuses on symmetric encryption and decryption. To build the initial encrypting and decrypting key between sensor nodes, the system first generates a huge key pool. The sensor nodes then randomly choose several keys from the key pool and load them before deployment. The sensor nodes use these preloaded keys to generate pairwise key, which create safe communication channels between neighboring nodes. This communication channel is called a key path, and it allowed sensor nodes to connect with other nodes in the environment. To protect the confidentiality of the key path, each key corresponds to only one index value. However, when an attacker finds the key, the sensor nodes immediately change the index value to update the key and select a new pairwise key.

Chan and Perrig [2] proposed a protocol called peer intermediaries for key establishment. This approach, sensor nodes are trusted third parties and manage the key. Guorui et al. [4] proposed a group-based dynamic key management scheme. This system can update and change the key independently of the base station or cluster head. Cheng and Agrawal [5] proposed an effective method to build and manage the pairwise key. In this scheme, the system generates a two-dimensional key matrix, and each sensor node randomly stores one column and row of the key array from the matrix before deployment. After the sensor nodes are deployed, two adjacent sensor nodes can generate the pairwise key of each other.

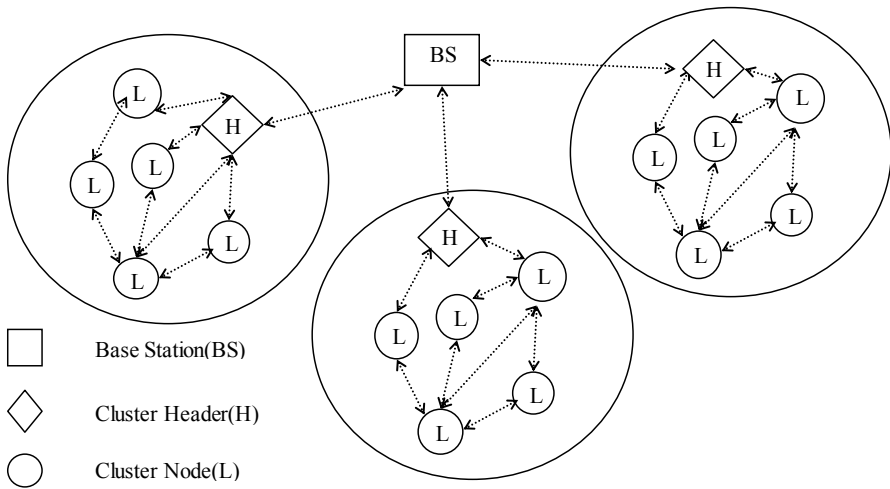
Kausar et al. [6] proposed a hierarchical sensor network consisting of a small number of high-end sensor(H-sensor node) and a large number of low-end sensor(L-Sensor node). The scheme is a scalable protocol for key management in the sensor networks to address the sensor nodes resource constraints, including computation, storage, and communication.

## 3 The Proposed Scheme

### 3.1 The Network Model

Based on the architectural consideration, wireless sensor networks may be broadly classified into two categories viz. (i) Hierarchical Wireless Sensor Networks(HWSN)

and (ii) Distributed Wireless Sensor Networks (DWSN). A Distributed WSN is easier for deployment, and there is no fixed infrastructure, and the network topology is not known prior to deployment. Sensor nodes are usually randomly scattered all over the target area. While a hierarchical WSN provides simpler network management, and can help further reduce transmissions. There is a pre-defined hierarchy among the participating nodes and three types of nodes in the descending order of capabilities: (a) base stations (BS), (b) cluster heads (H-sensors), and (c) sensor nodes (L-sensors). Base stations are many orders of magnitude more powerful than sensor nodes and cluster heads. Nodes with better resources, named as cluster heads (H-Sensor), may be used to collect and merge local data from sensor nodes and send it to base station. Sensor nodes (L-Sensor) are deployed around the neighborhood of the cluster heads. H-Sensors are implemented with the temper-resistant hardware and more processing capability and have more storage space. L-sensors are real working sensors in the deployment field that are very limited in terms of infrastructure like processing capability, memory and power. H-sensor are connected directory with Base station (BS) and can communicate. L-sensors can communicate with Base station via the H-sensor.



**Fig. 1.** Hierarchical wireless sensor network architecture

In this paper, we focus on three-tier hierarchical architecture, which is similar to [3], as shown in Fig. 1. Where sensor nodes are deployed, the clusters can be formed based on various criteria such as capabilities, location, signal strength etc. Each cluster has a cluster head and a set of sensor nodes. For ease of presentation, we assume there are at most  $m$  sensor nodes in each cluster and  $n$  cluster heads in the WSN, each cluster has a cluster head and  $m$  sensor nodes inside. We assume that sensor nodes are static once they are deployed. To reduce the energy consumption and the redundant traffic loads in the network, usually sensor nodes only communicate with its neighbors in the same cluster.

The Proposed System considers the following five communication rules

- (1) H-sensors can directly communicate with the BS.
- (2) The base station exchanges messages with L-sensors through H-sensors and vice versa.
- (3) H-sensors can send messages to specific L-sensors in the cluster.
- (4) H-sensors can broadcast messages to all L-sensors in the cluster.
- (5) L-sensors must exchange the messages with each other through an H-sensor. In other words, L-sensors cannot directly exchange messages with each other. Hence, a compromised L-sensor cannot affect the other L-sensor in the cluster.

### 3.2 Proposed Pre-distribution Key Management Scheme

Our key pre-distribution technique contains four phases: Matrices generation phase, key pre-distribution phase, pairwise key establishment phase, key ring establishment phase.

#### Phase I. Matrices Generation Phase

In our scheme, two different types of key matrices are used, one is  $M_{CH}$ , which is known by base station (BS) and all Cluster Heads (Hi). Second one is  $M_{Ci}$ , shared by sensor nodes(L-Sensors)in cluster  $i$  and their cluster head. The size of  $M_{CH}$  and  $M_{Ci}$  are  $x * x$  and  $y * y$  respectively,  $x = \sqrt{m}$  and  $y = \sqrt{n}$ . Where  $m$  is number of nodes in the cluster  $i$  and  $n$  is the number of cluster in the network.

**Step1.** The base station generates a pool of key  $p$  with more than  $2^{20}$  distinct keys. The keys are selected from a finite field  $FG(q)$  to create a matrix( $M_{CH}$ ). Where  $q$  is the prime number.

**Setp2.** The baset station computes  $m$  matrices  $M_{Ci}(i=1,2,3,...m)$ , for  $m$  clusters and one matrix  $M_{CH}$  for the cluster heads. The keys in the matrix must be distinct. So that the secret keys between any pair of nodes are unique. Base station prepares the different key chains from the generated matrices  $M_{CH}$ ,  $M_{Ci}$  by randomly selecting a row and a column from the respective matrices.

ID	1	2	3	4	.....	$x$
1	$K_{1,1}$	$K_{1,2}$	$K_{1,3}$	$K_{1,4}$	.....	$K_{1,x}$
2	$K_{2,1}$	$K_{2,2}$	$K_{2,3}$	$K_{2,4}$	.....	$K_{2,x}$
3	$K_{3,1}$	$K_{3,2}$	$K_{3,3}$	$K_{3,4}$	.....	$K_{3,x}$
4	$K_{4,1}$	$K_{4,2}$	$K_{4,3}$	$K_{4,4}$	.....	$K_{4,x}$
.	.	.	.	.	.....	.
.	.	.	.	.	.....	.
.	.	.	.	.	.....	.
$x$	$K_{x,1}$	$K_{x,2}$	$K_{x,3}$	$K_{x,4}$	.....	$K_{x,x}$

Fig. 2. An Example of constructed key matrix  $M_{CH}$

Fig. 2, illustrates an example of constructed key matrix  $M_{CH}$ , where each key has a unique two dimensional id denoted as  $K_{ij}(i,j=1,2,3,\dots x)$ . base station construct some key chains by randomly selecting row and a column from key matrix  $M_{CH}$ . We uses  $KC_{ij}(i,j=1,2,3,\dots x)$  to represent the key chains which is composed by  $i^{th}$  row and  $j^{th}$  column of the key matrix  $M_{CH}$ . The total number of constructed key chains are  $x$ , which is the same as number of clusters in the WSN.

Table-1 lists the constructed key chains. Any two key chains will hare exactly two common keys. When rows and columns are equal in two key chains, they will share all keys as common. All n clusters have the same setup of key chains prepared by base station.

**Table 1.** keys in key chain

Key chain ID	Keys in the key chain of $i^{th}$ cluster
CiK1,1	{ $K_{1,1}, K_{1,2}, \dots, K_{1,x}, K_{2,1}, \dots, K_{x,1},$ Session key }
.	.
.	.
CiKi,j	{ $K_{i,1}, K_{i,2}, \dots, K_{i,x}, K_{1,j}, \dots, K_{x,j},$ Session key }
.	.
.	.
CiK $x, x$	{ $K_{x,1}, K_{x,2}, \dots, K_{x,x}, K_{1,x}, \dots, K_{x-1,x},$ Session key }

The same procedure will be applied for the L-Sensor in each cluster using  $M_{Ci}(i=1,2,\dots,y)$  matrices.

**Phase 2. Key Pre-distribution Phase**

Base station: To establish a secure communication and authentication between the base station and cluster header nodes, base station stores the one key chain randomly, which is generated from the  $M_{CH}$  matrix and all other key chains are in its memory. The selected key chain is used to establish a pairwise key with the cluster head nodes.

**Cluster head(Hi).** Cluster head needs to communicate with base station, its neighboring cluster heads and also with their cluster members. That means, it required to store the two key chains one from the  $M_{CH}$  matrix and other from the  $M_{Ci}$  matrix.

**Sensor node ( $L_{ij}$ ).** To reduce the storage overhead of sensors, each sensor node in the Ci needs to store another key chain from  $M_{Ci}$  matrix. This key chain will be used to establish pairwise key with its cluster members and cluster header node. Each L-sensor can store only one key chain, it contains very limited number of keys, which reduce the storage over head than other existing methods discussed in the section II.

**Phase 3. Pairwise Key Establishment Phase**

After key pre-distribution, each cluster member node needs to establish a pairwise key with its neighbors and its cluster head to secure the communication inside the cluster.

To establish a pairwise key between sensor node  $L_{ij}$  and  $L_{ik}$ , the following steps are need to be done.

1. Sensor  $L_{ij}$  sends its key chain ID  $C_iK_{i,m}$  to sensor  $L_{ik}$ .
2. Upon receiving key chain ID from sensor  $L_{ij}$ , sensor  $L_{ik}$  can identify which keys they shared in common, computes the pairwise key  $KC_{i-jk}$  and replies with  $C_iK_{i,j}$ ,  $H(KC_{i-jk})$
3. Upon receiving key chain ID  $C_iK_{i,j}$  from  $L_{ik}$ ,  $L_{ij}$  computes pairwise key  $KC_{i-kj}$  and checks if  $H(KC_{i-jk})=H(KC_{i-kj})$ .
4. If sensor  $L_{ij}$  verifies  $H(KC_{i-jk})=H(KC_{i-kj})$ , that is to say  $KC_{i-kj}=KC_{i-jk}$ . The sensor  $L_{ij}$  generates the message with pairwise key  $KC_{i-kj}$   $\{ok, H(KC_{i-kj})\}$  and replays to sensor  $L_{ik}$ . Otherwise,  $L_{ij}$  generates the error message  $\{err, L_{ij}\}$  and broadcast to its H-sensor and  $L_{ik}$ . The error message is validated and managed using  $\mu$ TESLA.

The sensor  $L_{ij}$  and  $L_{ik}$  compute their pairwise key using the following equation 1

$$KC_{i-kj}=KC_{i-jk}=C_iK_{i,j}+L_{ij}+L_{ik}+C_iK_{i,m} \quad (1)$$

Hence the calculated pairwise key is unique in the whole network and shared between them. This unique key is not possible to computer any other guess. It is used for secure link between sensors.

#### Phase 4. Key Ring Establishment

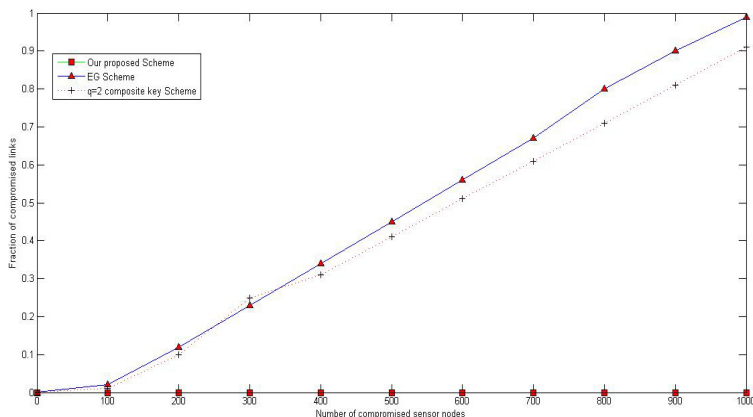
When a sensor computes all the pairwise keys with its neighbors, it deletes all the pre-loaded keys information from its memory to avoid the node capture attack and the possibility of compromising. Only the computed pairwise keys are stored in the memory. With this information, it is not possible to know the other sensor keys and key chain information. The key compromise and sensor capture attack may require some time after deployment of the network. At same that time, sensor also deletes key chains of sensor, plain messages and sensitive information of sensor. Hence proposed method completely prevents the node capture attack and key compromising.

## 4 Security Analysis

Node compromise attack is the main problem for wireless sensor networks. This paper, addresses the node compromise attack and node capture attack. We study the network resilience property of the proposed scheme based on the compromised sensor nodes. From the assumptions of the wireless sensor network, cluster header nodes(H-sensors) and Base Station(BS) are using temper resistant hardware. If the adversary capture the sensor nodes(L-sensor or cluster members), she may get the pairwise keys of its neighbors and its cluster head. But pairwise key cannot give any original keys in network and guessing information of keys to her. Even all the sensor nodes are compromised, it is not possible to know the pairwise keys of non-compromised sensor nodes.

We compared our proposed scheme with EG Scheme [1], q-composite scheme [3], as the number of compromised nodes increases, the fraction of affected communication nodes decreases. Our proposed key pre-distribution scheme, as the compromised nodes increases, the fraction of affected communication links are

almost zero. Because each pair wise key is distinct from others. Even 90% of the sensor nodes in the network are compromised, that will not effect the non-compromised node communication links. Figure 3 shows the network resilience against the sensor node compromise attack.



**Fig. 3.** Number of compromised sensor nodes vs. Number of compromised links

## 5 Conclusion

In this paper, we proposed a key establishment and management scheme for hierarchical sensor network. We analyzed the scheme based on the node capture and compromise attack. The results of our scheme is completely achieved the node capture attack in hierarchical sensor networks. It provides a tuff security to the communication links and indirectly addresses all the attacks like eavesdropping, replay, guessing attacks. We extend this works for the analysis on scalability, communication overhead, and storage capacity in future work.

## References

1. Eschenauer, L., Gligor, V.D.: A key-management scheme for distributed sensor networks. In: Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002), pp. 41–47 (November 2002)
2. Chan, H., Perrig, A.: Pike: peer intermediaries for key establishment in sensor networks. In: Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005), pp. 524–535 (2005)
3. Chan, H., Perrig, A., Song, D.: Random key pre-distribution schemes for sensor networks. In: IEEE Symposium on Security and Privacy, Berkeley, California, May 11-14, pp. 197–213 (2003)

4. Guorui, L., Jingsha, H., Yingfang, F.: A group-based dynamic key management scheme in wireless sensor networks. In: Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops/Symposia (AINAW 2007), pp. 127–132 (May 2007)
5. Cheng, Y., Agrawal, D.P.: Efficient pairwise key establishment and management in static wireless sensor networks. In: Proceedings of the 2nd IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2005), pp. 544–550 (November 2005)
6. Kausar, F., Hussain, S., Yang, L.T., Masood, A.: Scalable and efficient keymanagement for heterogeneous sensor networks. *Journal of Supercomputing* 45(1), 44–65 (2008)
7. Wen, M., Zheng, Y., Li, H., Chen, K.: A Hierarchical Composition of LU Matrix-Based Key Distribution Scheme for Sensor Networks. In: Washio, T., et al. (eds.) PAKDD 2007. LNCS (LNAI), vol. 4819, pp. 608–620. Springer, Heidelberg (2007)
8. Hunag, J.-Y., Liao, I.-E., Tang, H.-W.: A Forward Authentication Key Management Scheme for Heterogeneous Sensor Networks. *EURASIP Journal on Wireless Communications and Networking* 2011, Article ID 296704, 10 pages (October 2011)

# An Agent-Based Negotiating System with Multiple Trust Parameter Evaluation across Networks

Mohammed Tajuddin<sup>1</sup> and C. Nandini<sup>2</sup>

<sup>1</sup> Department of CSE, Dayananda Sagar College of Engineering

<sup>2</sup> Department of CSE, Dayananda Sagar Academy of Technology  
{tajdsce, laasyanandini}@gmail.com

**Abstract.** Mobile software agents can play a vital role to establish a strong relation between organizations. Gaining trust between software agents and organization is one of the key challenges. For the transactions to be successful, it is important to rely upon the trust exchange formalism, trust acknowledge and recording system for every major transaction. This paper presents the mechanism established, trust validity, trust verification and authentication strategies for achieving successful transactions in networks.

**Keywords:** Mobile Software Agents, E-commerce and E-business, Trust exchange Formalism, Validation and verification.

## 1 Introduction

The term software agent has become a marketing buzzword to describe everything from spreadsheet macro functions to complex mobile code that can roam across the network to meet our requirements. Beyond the publicity, it is necessary to understanding of the nature of agent software that can help in solving business and scientific problems. The mobile agent comprises of two separate and distinct concepts called as mobility and agent. Mobile agents refer to independent and identifiable computer programs that can move across the network and act on behalf of the user. The idea of a self-controlled program execution near the data source has been proposed as the next sign to replace the client-server model as a better, more efficient and flexible mode of communication. The mobile agent model as reported in the survey currently has two general goals: reduction of network traffic and asynchronous interaction. The goals of mobility are surprisingly close to those of the well-researched process migration concept, although at a different level of construct. These are reduction of network traffic, load balancing, fault tolerance, asynchronous interaction, and data access within the network. Once the network is free from traffic the transaction will become fast. Hence, there is no chance of losing any packets across the network. Mobile agents are used to implement network management by designation and deliver of network services. The mobile agent model proposes to consider network as multiple agent-environment. It further considers the agents as programmatic entity that move from one location to other location to perform the user



need. Agents can either function independently or work together with other agents to solve user problems [1].

Some of the issues addressed by mobile agents are agent transfer mechanisms, control of the mobile agent, naming, addressing, and locating a mobile agent. It also includes mobile agent data transfer, transparent communication, security, stability, performance, scalability, portability, resource management and data discovery [1].

Agent transfer mechanism is responsible for all the transaction in a network. Mobile agents will control the flow of data across the network. Agents are also responsible to name and locate packets in a network. Smart agents are responsible for transfer mechanism between two entities. For example a transaction is carried out in bank [1, 3].

When any transaction takes place across the network, the mobile agents are responsible to make the transaction between the source and the destination places, the software agents ensure that the transaction completed successful and agents also records transaction if it is necessary [5].

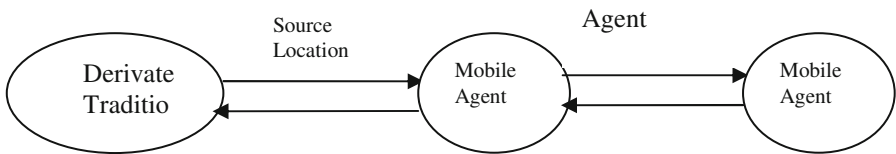


Fig. 1. Sending message from source to destination location

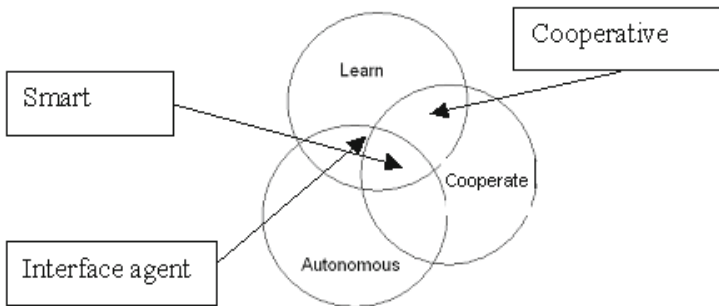


Fig. 2. Agent's overview

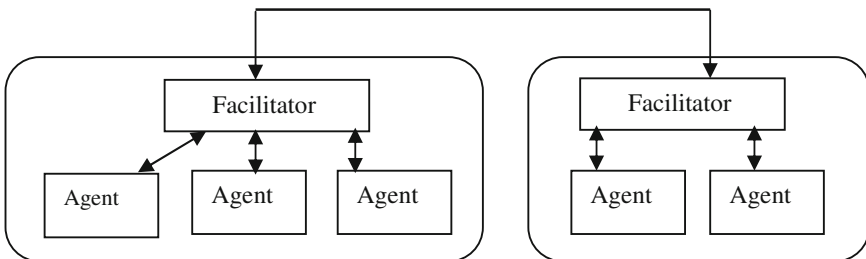
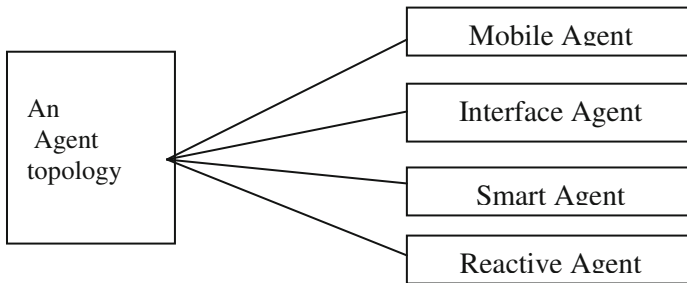


Fig. 3. Communication of agent to agent in a network system

Software agents are playing a vital role in the network, these agents are crucial for the transactions. Mobile agents are programs that bind data and the code, may be transfer from a source computer and transported to a remote computer for execution. When large quantities of data are stored at a remote computer, moving the computations to the information to is more practical and feasible approach, compared to allowing migrating data to the site of computations. Instead of gathering information from remote places at a centralized place directly. Users can send the mobile agents to a destination to perform information gathering and filtering, and to return to the source. Thus, the information transmitted over the network is minimized. Especially important when the network have low-bandwidth network. As shown in figure 3. when large quantity of data is transmitted over the network, it will be fragmented, depends on the bandwidth of the network and all packets are defragmented at the destination place, each fragment packets will have a sequence number, by using these fragment number at the destination used to defragment the entire original message. The facilitator (channel) transmits the data to agent in a network as the message comes to the facilitator.



**Fig. 4.** Types of Agents

## 2 Trust

Trust is a word, oral statement, promise or a written statement of an individual or a group that can relied upon. When a useful transaction occurs, the users and the end user should trust it. In the perspective of software agents trust means letting the process act on one’s behalf and accepting the risk due to that. Trust should exist between the two organizations when they are using the shared data or shared server. Trust factors will play an important role in trust management for successful transaction.

### 2.1 Trust Factors

The main trust factor is the ability of the user to trust the agent when there are multiple users and multiple transactions in a single network. The users must initially have a positive attitude when they approach any trust situation. Thus, users have to trust other users for any successful transaction. It demands the users to choose the

suitable software agent. For the transaction over the internet, the users are mainly concerned about the privacy and protection aspects of data, even with the existence of such measures. The following are the trust factors.

### **Experience**

An experienced agent will provide a better service than a learner or beginner agent. An experienced software agent gives complete information to the users. It includes the agent's working operations over the network during the processing so that the users feel secured of their transactions.

### **Communication**

Communication between agents and the users also determine the level of trust. Normally, agents are the responsible for the trust system. Hence, the agent is responsible for any error occurrences during the communication. The agent's feedback to the transactions will play a major role in communication system. Thus, feedback enables the user to trust the agent. Nevertheless, the agents should not harm the users because of failure transactions.

### **Interface Design**

Interface design of application controls the agent and flow of data over the network. Interface design includes appearance, behaviour, functionality, response, usage and operation. The interface should be user friendly and acknowledge proper message to the user. It further should provide options for system to recall the previous transaction or undo an action. Thus, it plays an important role towards trust management.

### **Predictable Performance**

Users trust the system for predictable, consistent and reliable transactions. Hence, predictable performance is one of the trust factors that should exist between users and agents in network system.

### **Trust Factor Estimation Model**

Trust factor estimation model verifies all trust factors before transaction occurs. In this model, from the user, transaction parameter passes to the application agent. The application agent will forward these transaction parameters to the trust table where they are evaluated according to the trust factors. An approval of verification for all the parameters goes back to the agents. However, upon completion of evaluation of nearly 80% of the trust parameters, the table will pass the transaction to the next phase. Figure 5. Depicts the model of estimating trust values before the start of transaction.

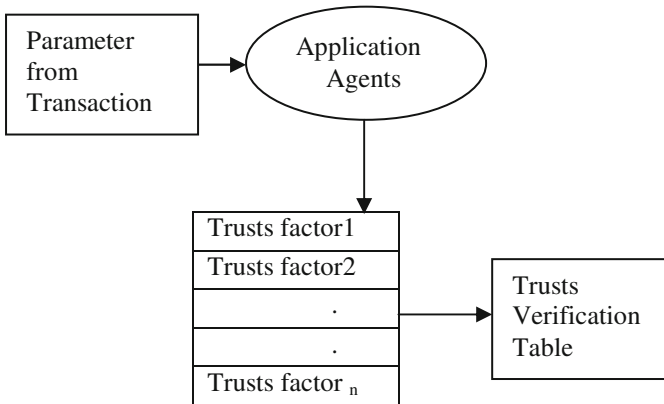


Fig. 5. The trust verification for all the transaction between the users

### 3 Mobile Electronic-Commerce

Mobile Electronic-commerce defined as all activities related to a commercial transaction conducted through communication networks that interface with wireless (or mobile) devices.

Trust management is an important issue in e-commerce. The traders never met and know nothing about each other, but still they carry our transaction. The wrong information of reputed traders may cause the mistrust, which effect the E – commerce economic efficiently. There has been enormous increase in transaction and the cooperative computing services on the Internet. Globally transaction and services over the Internet where the user in known or unknown to the service provider. Hence, it is very important to the service provider to trust the user based on the security policy made by the service provider before the transaction takes place. The service provider and the user have to make a contract. The contract list out the rules and confidential between the user and the service provider. The user will use the resource without violating the policy that prevail the across the boundary. The efficiency and usefulness of electronic commerce are well known. Internet has become an important medium for information exchange globally, for doing electronic commerce and electronic business. The available technologies are helps the traders to place their resources and workflow systems over the Internet.

The users can access, interact and transfer the information between the people and businesses or marketing. Electronic commerce is widely used over the Internet, hence lack of trust may creates large scale problem in electronic commerce transaction. The absence of trust among unfamiliar people, calls for the necessity to rely on various factors, for making successful transactions. Every service provider has a policy, which enables the availability trust to the users. Trust is the extent to which one party is willing to depend on somebody, in a given state with a feeling of relative security, even though negative factors are possible. It is a persistent notion and, as such, it has been studied thoroughly in a variety of different fields, including the social sciences,

economics and philosophy. It is an important observation from all these sources is that trust that in a sense is “one individual’s opinion of another” is a subjective believes, and every individual decides whether to trust another one or not based on the evidence available for personal evaluation. Trust is not symmetric that means the two individuals do not need to have similar levels of trust. Even if two entities get the same trust confirmation, they might not necessarily interpret this information in the same way. It is self-preserving and self-amplifying, it increases through periodic successful interactions, and degrades through disuse or misuse.

Trust is a mechanism for carrying the trade forward. Individuals write trade contract, which helps them as a legal document if some dispute among them may arises future. In an ideal world, the parties would write a group contract specifying exactly all deliverables in each state. However, if the number of states were very large, such a contract would be prohibitively expensive. In such situations, the parties will write an incomplete contract. When the state of nature is clear, they will renegotiate the contract, since at this stage they know what kind of goods are to be traded or each one’s responsibilities and accountabilities are clear. Success of one stage increases the trust for the next level. When the setup of trade involves complex information and trust. In such situations, such incomplete contract is the only efficient way for the transaction to succeed and is referred to as optimal contract. Thus, any trust management should support at least specification of contracts that are used in practice traditionally if not more. It is becoming increasingly important for the electronic community to have means and methods for tackling trust related issues for electronic transactions. We feel that trust in electronic commerce enhanced by the following factors. Establish trust, enable trust and enforced trust.

The users over the Internet will trust the reputed traders while buying the product or selling the product. The reputed traders establish that they the trust based on their quality policy and the service provides. Once the customer gets satisfy with a product and its service then a trust will established between the users, traders and the service provides.

There are several technological issues existing in trust management system and electronic commerce. In domain of trust management for electronic transactions, it is possible to reduce have lot human involvement at all stages of the trust management. The major question is to see how the combination of technology and human involvement enhances the trust of both. Some of the characteristic features of any transactions that the software agents may can include the following issues.

- Identification and authentication.
- Message confidentiality.
- Message integrity.
- Transparency and accountability.

When any transaction is made by the agents, It has authenticate and account the transaction. When a task is executing in network system, the above four characteristics are used to ensure the transaction is traceable and accountable towards the success of transaction.

Primary key infrastructure provides a framework to generate, distribute and maintain the cryptographic keys for authentication and verification and data integrity to establish communication over the Internet. The agent uses the key to transmit the data between the two organizations, this framework uses a data structure known as digital certificate. The framework having the digital certificate is very useful framework for trust establishment, trust enhancement and trust enforcement. The agent can use the cryptographic key to generate the original message at the destination. PKI can be used for trust agent transaction also ensuring successful transmission trust worthy data over the Internet as shown in figure (1).

## 4 Conclusion

The use of software agents in real-world applications offers opportunities for business and scientific organizations to realize trust in transaction over the network. Trust in cooperative computing services and technologies have become an important issue in today's communication over the Internet or a transaction between the two users. The trust among the business, consumer and the traders are crucial for the electronic business. There is urgent need to develop an environment, which becomes scalable in trusts management for high values over the Internet. The software agents can play a crucial role for trust worthy data exchange between the agents, traders and the users. Electronic communication totally depends on the trust management system to be maintained by the agents and traders. Trust is the key factor for continues growth and success of electronic commerce and hence software agents with multiple parameters governing trust mechanism is an inevitable field of research.

## References

- [1] Luck, M., McBurney, P., Preist, C.: Agent Technology: Enabling Next Generation Computing. AgentLink (2003) ISBN 0854 327886
- [2] McKnight, C.: Electronic Journals: What do Users think of them? In: Sugimoto, S. (ed.) Proceedings of International Symposium on Research, Development and Practice in Digital Libraries, ISDL 1997. University of Library and Information Science, Tsukuba (1997)
- [3] Luther, J.: White Paper on Electronic Journal Usage Statistics. The Journal of Electronic Publishing 6(3) (2000, 2001) ISSN 1080-2711
- [4] Tenopir, C., Hitchcock, B., Pillow, A.: Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies. Council on Library and Information Resources (2003)
- [5] Brazier, F.M.T., Oskamp, A., Prins, J.E.J., Schellekens, M.H.M., Schreuders, E., Wijngaards, N.J.E., Apistola, M., Voulon, M.B., Kubbe, O.: ALIAS: Analysing Legal Implications and Agent Information Systems. Technical Report no. IR-CS-004, Computer Science, Faculty of Sciences, Vrije Universiteit Amsterdam (2003)
- [6] Adam, N., et al.: Electronic Commerce: Technical, Business, and Legal Issues. Prentice Hall, Upper Saddle River (1998)

- [7] Amgoud, L., Maudet, N., Parsons, S.: Modelling dialogues using argumentation. In: Durfee, E. (ed.) Proceedings of the 4th International Conference on Multi-Agent Systems (ICMAS 1998), pp. 31–38. IEEE Press, Boston (1998)
- [8] Bickmore, T., Cassell, J.: Relational Agents: A Model and Implementation of Building Trust. ACM Press, New York (2001)
- [9] Youll, J.: Agent Based Electronic Commerce: Opportunities and Challenges. IEEE CS Press, Los Alamitos (2001)
- [10] Ericson, T.: Designing Agents as if People mattered Software Agents. J.M. Bradshaw AAI Press, Menlo Park (1997)

# Enabling Self-organizing Behavior in MANETs: An Experimental Study

Annapurna P. Patil, K. Rajanikant, Sabarish, Madan, and Surabi

Department of Computer Science and Engineering, M.S. Ramaiah Institute of Technology,  
Bangalore 560054, India  
annapurnap2@yahoo.com, sabarishchandramouli@gmail.com

**Abstract.** Mobile Ad Hoc Networks (MANETs) require special management when compared to wired networks because of their hardware and energy limitations. A key concept that could help provide this management is Self-organization. Here the concentration is on typical MANET scenarios that depend heavily on the reliability of the network. Reliability is a direct cost of the overall network lifetime, i.e., energy of the nodes. In this paper a Self-Organizing AODV protocol (SO\_AODV) has been developed by integrating self-organizing concepts found in Efficient Self-Organization Algorithm (EESOA) and Ad hoc On-demand Distance Vector (AODV) algorithm. SO\_AODV has been compared with Traditional AODV and basic EESOA using the following metrics: network lifetime, average end to end delay, delivery ratio and convergence time. In order to appreciate the performance of the proposed SO\_AODV algorithm, an implementation in the NS2 simulator is done. The observations from the simulation are: SO\_AODV has better network lifetime and end to end delay than Traditional AODV and has better values of convergence time when compared to EESOA.

## 1 Introduction

Mobile Ad Hoc Networks (MANETs) are a type of wireless networks. They have no fixed routers, every node could be router. All nodes are capable of movement and can be connected dynamically in an arbitrary manner. The responsibilities for organizing and controlling the network are distributed among the terminals themselves. In this type of network, some pairs of terminals may not be able to communicate directly with each other and have to rely on other terminals so that the messages are delivered to their destinations. Such networks are often referred to as multi-hop or store-and-forward networks. The nodes of these networks which are functioning as routers discover and maintain routes to other nodes in the networks. Mobile Ad-hoc Networks are having usage in wide application areas [1]. Self-organizing an ad hoc network allows the creation of a different view of radio topology called virtual structure, introducing one or more hierarchy levels and it facilitates the establishment of the services necessary for expected network operations such as routing. The functions that self-organization should accomplish in such networks are: Resource



sharing, structure formation and maintenance, helping the deployment of communication protocols, and resource management. Energy is a scarce resource in ad hoc wireless networks [2]. Thus it is of paramount importance to use energy efficiently when establishing communication patterns.

This paper concentrates on developing energy efficient, adaptive and self-organizing AODV routing protocol for MANETs. The SO\_AODV protocol has been developed by integrating self-organizing concepts such as those found within EESOA with concepts used by AODV.

The rest of the paper is organized as follows. In section 2, we briefly discuss the literature related to our paper. In section 3, we briefly discuss the related work. Section 4 provides a detailed description of the design and implementation. Section 5 contains the results. The conclusion is provided in Section 6. Section 7 suggests the future work that can be taken up.

## **2 Literature Survey**

### **2.1 Overview**

The requirements of MANETS represent a broad spectrum of network challenges. During the last few years, almost every aspect of MANETs has been explored to some level of detail. Yet, more questions have arisen than been answered. Many major open problems such as MANETs being autonomous, their dynamic topology, device discovery task, bandwidth optimization requirement, limited resources available, scalability, limited physical security, infrastructure-less and self-operated, poor transmission quality, ad hoc addressing, network configuration, etc. have to be addressed [5].

### **2.2 Self-organization**

Self-organization is a very important concept for building scalable systems consisting of huge numbers of subsystems. The primary objectives here can be addressed as coordination and collaboration for a global goal. The concept of Self-organization mechanisms can be found in many of our day-to-day activities. From an academic point of view, self-organization was first analyzed in biological systems. This research was soon extended to technical systems and engineering in general. Self-organization can be defined as the interaction of multiple components for achieving a common global objective. This collaborative work is done without central control. Instead, the interaction is done using a local context, e.g. the direct environment that can be changed and adapted by each individual and, therefore, affects the behavior of other individuals. The primary objectives of self-organization are scalability, reliability, and availability of systems composed of a huge number of subsystems [6]. The properties and constraints on mechanisms that are required for a system to be described as self-organizing with emergent properties have been, clearly defined in the field of biology by Camazine et al. [2001] as “a process in which a pattern at the global level of a system emerges from the numerous interactions among lower-level

components of the system". This definition captures important aspects of self-organization such as autonomous components that take decisions using only local information, and how interactions between components cause system properties to emerge [7].

### 3 Related Work

The proposed work is aimed at developing an energy efficient routing protocol using self-organization. This section documents in detail the Energy Efficient Self-Organization Algorithm (EESOA) developed by researchers in the field. EESOA is fully distributed and no node has the whole knowledge of the virtual backbone, i.e., each node is only aware of its neighbours within one hop [8]. A way to overcome the problem of lack of global knowledge is the use of group-based algorithms [9].

#### 3.1 Network Model

The wireless network is described by an undirected communication graph  $G(V, E)$ , where the set  $V$  represents the wireless devices in the network and the set of edges  $E$  represents bidirectional communication channels operating between neighbour nodes. Between any two nodes  $u$  and  $v$ , there will be an edge  $(u, v)$  if the transmission from node  $u$  is received by the node  $v$  with a signal strength greater than  $RX_{threshold}$ .

In classical approaches for choosing the leader in the group, the lowest identifier is often used. However, such a scheme is inefficient since this criterion cannot reflect the aptitude of a node to act as a backbone member for a long time and create a stable structure. Different metrics can be used to determine the more suitable nodes. In this scheme, the characteristics of nodes are taken into account when constructing a backbone, where the characteristics could be the node's degree, the remaining energy, the link quality, the communication capacity, the processing power etc. In this case for simplicity's sake, only the node's degree and the remaining energy will be used to calculate the weight of the node.

Under this model, the following assumptions are made: Each node has a unique identifier ID e.g. its IP address, Each node only knows the information of neighbours to one-hop, The nodes can move, arrive, or leave the network, Each node can adjust its transmission power, The agent can use overhearing to obtain important information for reducing the message transmission, Nodes keep a neighbours table and a weight that is equal to the product between the number of neighbours and the residual energy units, The nodes do not know their geographical positions, The wireless nodes are placed in the 2-dimensional Euclidean plane (it also works correctly if agents are located in three-dimensional space).

#### 3.2 Radio Propagation Model

There are factors that influence the communication in a network. For example, the background noise produced by signals naturally present in the environment. In order to make a more realistic environment, the two-ray ground reflection and free space

propagation models are used [10]. However, the two-ray model does not give a good result for a short distance due to the oscillation caused by the constructive and destructive combination of the two rays. Instead, the free space model is used when distance is small [11].

### 3.3 Algorithm Description

In the proposed algorithm, the generated structure is based on groups; it is composed of four different roles (leader, gateway, member and bridge); there is only one leader per group. Leaders are in charge of doing all the communication inside the group they coordinate. Thus, it is necessary to choose the most suitable nodes to play the leader role. The gateway node makes the communication happen between the leaders; it can link to more than two leaders. In this way the virtual backbone is formed by leaders and gateways. The communication is then carried out by means of these nodes. The member node only takes care of its own tasks, there can be zero or more members connected to one leader. Because the algorithm is localized, it is impossible to determine a whole connection by using only leaders and gateways in the backbone. A hello message sent by a leader will inhibit all nodes that receive it. Whenever a new node arrives to the network, if it receives a hello message from a leader, the node will be inhibited and become a member. If no hello message from leader is received, the node starts to discover the neighbourhood and follows the algorithm. A node that is not inhibited by a neighbour will get the leader role as long as it has the greatest weight in the group, with ties broken by the node id. A node will wait if it has an uninhibited neighbour either with a greater weight or with the same weight but with a greater id. The segmentation problem can be solved by using bridge nodes [8].

### 3.4 Table Management

Neighbourhood management essentially has three operations: insertion, erasing, and updating. Each node broadcast hello messages to discover the neighbourhood. Since all these nodes wake up at the same time, EESOA [8] attempts to schedule the times at which nodes send their broadcast messages, so that not all nodes send their messages at once.

## 4 Design and Implementation

The algorithm which we propose integrates the leader selection process of EESOA into AODV in an efficient way. Each node assumes one of three roles: leader, member or bridge. The node with the highest weight in a neighbourhood is selected as the leader. *Notations:* Meaning: -  $W_i$ : Weight of node  $v_i$ ,  $D_i$  : Degree of node  $v_i$ ,  $E_i(t)$  : Remaining energy of node  $v_i$  at time  $t$ , Degree: The degree of a node is the number of neighbours that it has. A node is a neighbor to another node if it is within the transmission range of the other. Remaining Energy Capacity: This metric makes the fairness of energy consumption the main focus. The weight of the node is calculated as follows:  $W_i = D_i * E_i(t)$

**Algorithm Description:**

```

Algorithm for Role selection of a node vi
Lead←1
Inhibitcount←0
for every node vj that is in the neighbour list of node
    vi
        if vj is a leader
            Inhibitcount←Inhibitcount+1
        if Wj>Wi and vj is not inhibited
            set node vi as inhibited
Lead←0
    if lead=1
        set role of node vi to leader
        set node vi as uninhibited
    else if Inhibitcount=1
        set role of node vi to member
    else if Inhibitcount>1
        set role of node vi to bridge
    
```

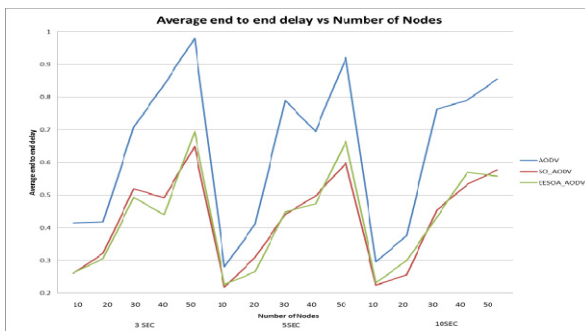
**Algorithm for flexible routing at a node vi**

```

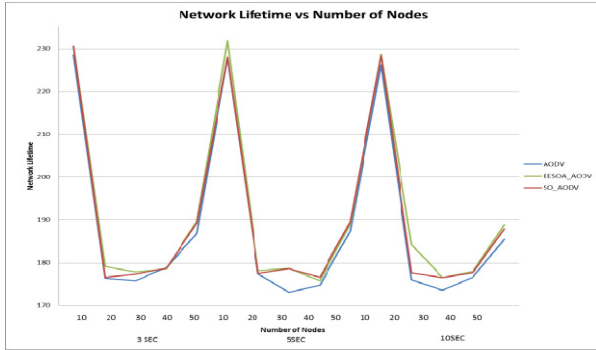
Switch←0
if node vi is a member
    for every node vj that is in the neighbour list of
        node vi
            if node vj is a leader
                set destination address to node vj
                Switch←1
                break
if Switch=0
    set destination address to broadcast
    
```

**5 Testing and Results**

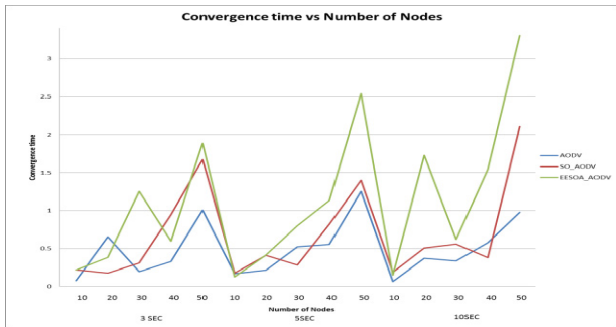
The four performance metrics are measured for SO\_AODV, EESOA and Traditional AODV Protocols. Protocols are tested with four conditions: Low density and low mobility, Low density and high mobility, High density and low mobility, High density



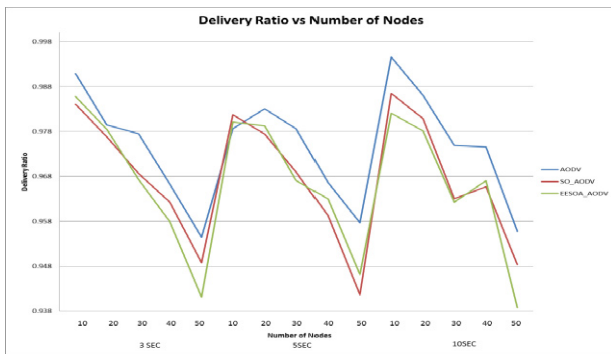
**Fig. 1.** Average End to End Delay Vs Number of Nodes



**Fig. 2.** Network Lifetime Vs Number of Nodes



**Fig. 3.** Convergence Time Vs Number of Nodes



**Fig. 4.** Packet Delivery Ratio Vs Number of Nodes

and high mobility. Simulations are done for: Number of nodes varying from [10,50] in steps of 10, Pause Time varied with values 3, 5 and 10 and with nodes given random initial energy values from [1,10].

## 6 Conclusion

In this work, SO\_AODV was developed by integrating self-organizing concepts such as those found within EESOA in AODV. The main object of the work was to enable self-organizing behaviour in MANETs by proposing an algorithm which groups the node elements in clusters. This was achieved by means of a self-organization strategy based on three possible roles for each node in the network: leader, member and bridge. The SO\_AODV routing protocol was compared with Traditional AODV and the basic EESOA routing protocol. It was found that SO\_AODV had better values for network lifetime and end to end delay than Traditional AODV whereas Traditional AODV had better values for delivery ratio than SO\_AODV. SO\_AODV had better values of convergence time than EESOA and both SO\_AODV and EESOA had the same end to end delay, network lifetime and packet delivery ratio. The application of self-organizing techniques indeed improved the performance of AODV with respect to the metrics mentioned. Thus the concept could lead to promising benefits even when MANETs are deployed in real world environments.

## 7 Future Work

In this work the topology of the network and the connection pattern were designed. Certain options such as traffic pattern, initial energy and packet queue length were configured for each mobile node. The same options were chosen for all the nodes. An algorithm can be developed to handle heterogeneous nodes with varying communication capacities links with differing quality. This would make the simulation closer to a realistic situation. During the simulation tests it was found that Traditional AODV had better values for throughput, convergence time and delivery ratio than SO\_AODV. Methods to improve the QoS parameters of SO\_AODV such as packet delivery ratio, (which was reduced due to congestion at the leader node) can be considered for future work, also other parameters as fault tolerance, delay can be tested as a part of future work.

**Acknowledgments.** We acknowledge the students Madhuri and Sowmya for their contributions in implementing the proposed concept.

## References

1. Han, L.: Wireless Ad-hoc Networks (October 8, 2004)
2. Nikolaidis, I., Barbeau, M., Kranakis, E.: Ad-Hoc, Mobile, and Wireless Networks. In: Third International Conference, ADHOC\_NOW (2004)
3. Cao, L., Dahlberg, T., Wang, Y.: Performance Evaluation of Energy Efficient Ad Hoc Routing Protocols. IEEE (2007)
4. Pushpalatha, M., Venkataraman, R., Ramarao, T.: Trust Based Energy Aware Reliable Reactive Protocol in Mobile Ad Hoc Networks. World Academy of Science, Engineering and Technology (2009)

5. Taneja, K., Patel, R.B.: An Overview of Mobile Ad hoc Networks: Challenges and Future. In: Proceedings of National Conference on Challenges & Opportunities in Information Technology (COIT 2007). RIMT-IET, Mandi Gobindgarh (March 23, 2007)
6. Dressler, F.: Self-Organization in Ad Hoc Networks: Overview and Classification (2006)
7. Biskupski, B., Dowling, J., Sacha, J.: Properties and mechanisms of self-organizing MANET and P2P systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 2(1(1)) (March 2007)
8. Olascuaga-Cabrera, J.G., Mendez-Vazquez, A., López-Mellado, E.: A Novel Distributed Energy-Efficient Self-Organized Algorithm for Wireless Ad-hoc Networks. In: Eighth International Conference on Intelligent Environments (2012)
9. Olascuaga-Cabrera, J., López-Mellado, E., Mendez-Vazquez, A., Ramos-Corchado, F.: A self-organization algorithm for robust networking of wireless devices. *IEEE Sensors Journal* 11(3), 771–780 (2011)
10. Li, N., Hou, J.C., Sha, L.: Design and analysis of an mst-based topology control algorithm. *IEEE Transactions on Wireless Communications* 4(3), 1195–1206 (2005)
11. Issariyakul, T., Hossain, E.: Introduction to Network Simulator NS2, 1st edn. Springer Publishing Company, Incorporated (2008)
12. Correia, L.H., Macedo, D.F., dos Santos, A.L., Loureiro, A.A., Nogueira, J.M.S.: Transmission power control techniques for wireless sensor networks. *Computer Networks* 51(17), 4765–4779 (2007)

# Secure Hybrid Routing for MANET Resilient to Internal and External Attacks

Niroj Kumar Pani and Sarojananda Mishra

Department of Computer Science, Engineering and Application,  
Indira Gandhi Institute of Technology, Sarang, India  
{nirojpani, sarose.mishra}@gmail.com

**Abstract.** An ad hoc network is the cooperative engagement of a collection of mobile nodes without the required intervention of any centralized access point or existing infrastructure. Although there is an increasing trend to adopt ad hoc networking for commercial uses, their main applications lie in military, tactical and other security-sensitive operations. In these and other applications of ad hoc networks, secure routing is an important issue. Quite a good number of protocols have been suggested in this area, but most of them are either proactive or reactive in approach. Studies reveal that, either a pure proactive or a pure reactive approach of routing performs well in a limited region of network setting. However, in diverse applications of ad hoc networks the performance of either class degrades dramatically. In this paper, we have presented a Secure Hybrid Routing Protocol (SHRP) for MANET, which aims at addressing the above limitations by combining the best properties of both proactive and reactive approaches. The protocol is based on the design of zone routing protocol (ZRP). The paper details the design of the protocol and analyses its robustness in the presence of multiple possible security attacks against ad hoc routing caused either by an internal compromised node or an external adversary.

## 1 Introduction

An ad hoc network is a collection of wireless computers (nodes), communicating among themselves over possibly multi-hop paths, without the help of any infrastructure such as base stations or access points [1, 2, 3, 4]. Unlike traditional mobile networks, ad hoc networks have no fixed infrastructure. Mobile nodes that are within each other's radio range communicate directly via wireless links, while those far apart rely on other nodes to relay messages as routers. In ad hoc network each node acts both as a host and a router which forwards the data intended for some other node. Applications of ad hoc network range from military operations and emergency disaster relief, to commercial uses such as community networking. In these and other applications of ad hoc networks, secure routing is an important issue.

Ad hoc network routing protocols [5, 7, 8, 9, 10] are challenging to design and secure ones are even more, due to the unique characteristics of ad hoc networks such as, lack of central authority, rapid node mobility, frequent topology changes, shared radio channel and limited availability of resources. A number of protocols have been



proposed in the literature for secure routing. A survey of the protocols is given in [1, 2, 15]. Most of these protocols are either proactive or reactive in approach. However, both the approaches have their own limitations [10, 11]. For example, the proactive protocols use excess bandwidth in maintaining the routing information while, the reactive ones have long route request delay. Reactive routing also inefficiently floods the entire network for route determination.

In this paper, we have presented the design and analysis a secure hybrid routing protocol (SHRP) for MANET, which combines the best properties of both proactive and reactive approaches. It ensures end to end authentication, message integrity, non-repudiation and data confidentiality by employing inexpensive cryptographic primitives. We have described the possible attacks against ad hoc routing and analyze how this protocol successfully detects and protects against all identified threats caused by both external and internal compromised nodes.

Rest of the paper is organized as follows: Section 2 presents the security exploits possible in ad hoc routing and the security requirements of any ad hoc network. Section 3 presents the secure hybrid routing protocol SHRP. Section 4 shows the security analyses of the protocol and finally Section 5 offers concluding remarks.

## 2 Security Attacks and Needs

Attacks on ad hoc networks can be classified as passive or active [1, 4]. In passive attack the attacker attempts to discover the nodes information (e.g., IP addresses) by without disrupting the normal operation of the network. Detection of passive attack is very difficult since the operation of the network itself doesn't get affected. An active attack attempts to alter or destroy the data being exchanged in the network by disrupting the normal functioning of the network. This attack involves actions such as the impersonation, information discloser, modification, fabrication and replication.

In Impersonation attack, a malicious node misrepresents its identity in order to alter the vision of the network topology that a benign node can gather. The Black Hole and Wormhole attacks [1] fall in this category. An information disclosure attack can reveal something about the physical location of nodes or the structure of the network. Modification attack involves the alteration of routing messages. A possible Denial of Service (DoS) attacks can also be done by altering control message fields. The notation "fabrication" refers to attacks performed by generating false routing messages. Such attacks can be difficult to identify as they come as valid routing constructs.

A good secure routing protocol should prevent each of the exploits. All secure ad hoc routing protocols must satisfy the following requirements: (1) Fabricated routing messages cannot be injected into the network; (2) Routing messages cannot be altered in transit, except according to the normal functionality of the routing protocol; (3) Route signaling cannot be spoofed; (4) Routing loops cannot be formed through malicious action; (5) Routes cannot be redirected from the shortest path by malicious action; (6) Unauthorized nodes should be excluded from route computation and discovery. (7) The network topology must not be exposed either to adversaries or to authorize nodes by the routing messages as exposure of the network topology may be an advantage for adversaries trying to destroy or capture nodes.

### 3 The Secure Hybrid Routing Protocol

The Secure Hybrid Routing Protocol (SHRP) is based on Zone Routing Protocol (ZRP) [10, 11]. Like ZRP it performs Intra-zone [12] and Inter-zone [13] routing but, unlike ZRP where no security measures have been taken, SHRP is designed to address all measure security concerns like end to end authentication, message/packet integrity and data confidentiality. End to end authentication and message integrity is ensured by digital signature mechanism [16]. Data confidentiality is provided by an integrated approach of both symmetric and asymmetric key encryption [16].

The process of signing and encrypting requires keys. Each communicating node In order to perform secure routing (intra-zone and inter-zone) must have two pairs of private/public keys, one pair for signing and verifying and the other for encrypting and decrypting. For a node  $X$  the signing and verifying keys are  $SK_X$  and  $VK_X$  respectively where as encryption and decryption keys are  $EK_X$  and  $DK_X$  respectively. Among these keys  $SK_X$  and  $DK_X$  are private keys where  $VK_X$  and  $EK_X$  are public keys.

We assume that SHRP makes the use of public key certificates [15, 16] for key distribution and management. For the process of public key certification, SHRP assumes the presence of trusted certification servers called the *certification authorities* (CAs) in the network in addition to the communicating nodes which we call the *common nodes* (CNs). The public keys of the CAs are known to all valid CNs. Keys are generated apriory and exchanged through an existing, perhaps out of band, relationship between CA and each CN. Before entering the ad hoc network, each node requests a certificate from its nearest CA. Each node receives exactly one certificate after securely authenticating their identity to the CA. The methods for secure authentication to the certificate server are numerous and hence it is left to the developers; a significant list is provided by [16]. A common node  $X$  receives a certificate from its nearest CA as follows:

$$CA \rightarrow X: \text{cert}_X = [IP_X, VK_X, EK_X, t, e] | \text{sign}_{CA}$$

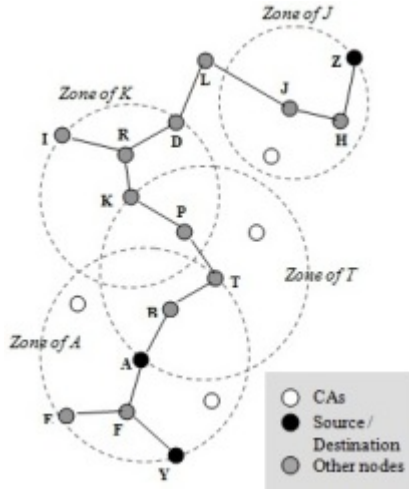
$$\text{where, } \text{sign}_{CA} = [IP_X, VK_X, EK_X, t, e] SK_{CA}$$

The certificate contains the IP address of  $X$ , the two public keys  $VK_X$  and  $EK_X$  of  $X$ , one for verifying the signature signed by  $X$  and other for encrypting a packet to be send to  $X$ , a timestamp ' $t$ ' of when the certificate was created, and a time ' $e$ ' at which the certificate expires, all appended by the signature  $\text{sign}_{CA}$  of CA. All nodes must maintain fresh certificates with their nearest CA. Once a node acquires the certificate it can perform secure intra-zone and inter-zone routing.

#### 3.1 Secure Intra-Zone Routing

Secure Intra-zone Routing (SIAR) is a limited depth proactive [5] link-state routing approach [6] with added security features. To perform intra-zone routing each node periodically computes the route to all intra-zone nodes (nodes that are within the routing zone of a node) and maintains this information in a data structure called SIAR routing table. For example in Fig.1 (for the illustration of intra-zone routing we consider node

A as the source and node Y as the destination in the network given in Fig.1), node A proactively computes the route to nodes B, T, E, F and Y and stores this information in its SIAR routing table. This process is called *proactive route computation*.



**Fig. 1.** Intra-Zone and Inter-Zone destinations of node A (zone radius  $\beta = 2$ )

In order to detect the neighbor nodes and possible link failures, SHRP relies on the *neighborhood discovery protocol (NDP)* [12] similar to that of ZRP. NDP does this by periodically transmitting a HELLO becon (a small packet) to the neighbors at each node and updating the *neighbor table* [12] on receiving similar HELLO becons from the neighbors. NDP gives the information about the neighbors to the node and also notifies the node when the *neighbor table* updates. We have assumed that NDP is implemented as a MAC layer protocol. A number of security mechanisms suggested in [4, 18] for MAC layer can be employed to secure NDP.

For *proactive route computation* each node within its routing zone periodically advertises a *link state packet (LSP)*. For example, node A advertises the LSP within the zone of A.

$$A \rightarrow \text{brdcast} : [LSP, IP_A, cert_A, \beta, TTL, SNo, neighbour[n], link\_metric[n]] \mid sign_A$$

Where,  $sign_A = [LSP, IP_A, cert_A, \beta, TTL, SNo, neighbour[n], link\_metric[n]] SK_A$

The packet contains a packet type identifier “LSP”, the IP address of the broadcasting node A, the certificate of A, the zone radius ‘ $\beta$ ’, a time-to-live (TTL) value, the sequence number *SNo* of the packet which is used to track the link state history of the source node A, the list of neighbors of A, and link metrics, all appended by the signature  $sign_A$  of A. The TTL field is used to control the scope of the packet which is initialized to  $\beta - 1$  hops by A. Upon receipt the packet, the TTL value is decremented and as long as the value is greater than 0, the LSP is rebroadcasted.

When a neighbor of A, receives the LSP, it verifies the authenticity of the packet using  $VK_A$  which it extract from A’s certificate in the LSP, add LSP’s information to

its *link-state table* [6], decrement the value of TTL field and again forwards this LSP as long as the value of TTL field is greater than 0 else the LSP is dropped. Because every node within the zone of  $A$  receives the same LSPs, all the nodes build the same link state table. A typical link state table contains at least the following fields:  $\langle \text{Source address, Zone radius, Neighbor ID, Insert time, route metrics} \rangle$ .

Once the link-state table is built, each node computes the route to every other node within its zone by applying the Dijkstra algorithm [6] to its link state table and stores this information in its SIAR routing table. A typical SIAR routing table maintained at a node contains the following fields:  $\langle \text{Dest\_Address, Routes, Route metrics} \rangle$  and has entries for all intra-zone nodes.

After proactive route computation the following steps are taken by node  $A$  to route the data packet to  $Y$ .

**Step 1:**  $A$  looks for the route to  $Y$  in SIAR routing table and finds it to be A-F-Y.

**Step 2:**  $A$  sends a Session Key Request (SKREQ) packet to  $Y$  along this route requesting a session key  $K_{AY}$  between  $A$  and  $Y$ .

$$A \rightarrow Y : [SKREQ, IP_Y, cert_A] \mid sign_A$$

where,  $sign_A = [SKREQ, IP_Y, cert_A] SK_A$

The SKREQ packet contains a packet type identifier “SKREQ”, the IP address of the destination  $Y$ , and  $A$ ’s certificate, all appended by the signature  $sign_A$  of  $A$  signed using  $SK_A$ .

**Step 3:**  $Y$  on receiving this request, verifies the signature using  $VK_A$ , which it extracts from  $A$ ’s certificate, creates the session key  $K_{AY}$ , encrypts it using  $EK_A$  and sends it to  $A$  as Session Key Reply packet (SKREP) packet along the reverse route Y-F-A.

$$Y \rightarrow A : [SKREP, IP_A, cert_Y, \{K_{AY}\}EK_A] \mid sign_Y$$

where,  $sign_Y = [SKREP, IP_A, cert_Y, \{K_{AY}\}EK_A] SK_Y$

The packet contains a packet type identifier “SKREP”, the IP address of  $A$ , the certificate of  $Y$  and the session key  $K_{AY}$  encrypted using  $EK_A$ , all appended by the signature  $sign_Y$  of  $Y$  signed using  $SK_Y$ .

**Step 4:**  $A$  on receiving the SKREP packet, verifies it using  $VK_Y$ , conforms the authenticity of the packet, decrypts it using  $DK_A$  and extracts the session key  $K_{AY}$ .

Once  $A$  gets the session key, it can encrypt the data packet using  $K_{AY}$  and send it to  $Y$  along the same route A-F-Y. All further communication between  $A$  and  $Y$  takes place similarly, using this session key.

### 3.2 Secure Inter-Zone Routing

Secure IntEr-zone Routing (SIER) is a reactive routing protocol [5] with added security features like ARAN [17]. SIER offers on demand secure route discovery and route maintenance services based on local connectivity information. Inter-zone routing is presented in this section and the route maintenance services offered by SIER are discussed in Section 3.3.

To minimize the delay during inter-zone route discovery, SIER uses *border-casting technique* [14] similar to ZRP, but with little modification. The border-casting technique adopted here not only forwards secure route discovery packets (discussed later) to the peripheral nodes of the border-casting node but also sets up a reverse path back to the neighbor by recording its IP address. It uses the SIAR routing table to guide these route queries.

Secure inter-zone routing is initiated when a node searches the path to another node in its SIAR routing table and fails to find it (because the destination is outside the zone of the source node). In our case when node  $A$  wants to send a packet to node  $Z$  (for the illustration of intra-zone routing we consider  $A$  as the source and  $Z$  as the destination in the network given in Fig.1),  $A$  looks in its SIAR routing table for a valid route to  $Z$ . Since  $Z$  is not within the zone of  $A$ ,  $A$  fails to find the route.

In this case, the following steps are taken by  $A$  to route the data packet to  $Z$ :

**Step 1:**  $A$  begins a *secure route discovery* process to  $Z$  by border-casting to its peripheral nodes  $T$ ,  $E$  and  $Y$ , a Secure Route Discovery (SRD) packet.

$$A \rightarrow \text{bordercast} : [SRD, IP_Z, cert_A, \beta, N_A, t] \mid sign_A \\ \text{where, } sign_A = [SRD, IP_Z, cert_A, \beta, N_A, t] SK_A$$

The packet contains a packet type identifier “SRD”, the IP address of the destination  $Z$ ,  $A$ ’s certificate, the zone radius ‘ $\beta$ ’, a nonce  $N_A$  created by  $A$  and the current time  $t$ , all appended by the signature  $sign_A$  of  $A$ . The nonce  $N_A$  is monotonically increased every time  $A$  performs route discovery.  $N_A$  and  $t$  together with the IP address of  $A$  ( $IP_A$ ) uniquely identify the SRD which prevents the replay attack.  $N_A$  is made large enough such that, it will not need to be recycled within the probable clock skew between receivers. If a nonce later reappears in a valid packet that has a later timestamp, the nonce is assumed to have wrapped around, and is therefore accepted. It should be noticed that a hop count is not included with the message.

**Step 2:** When a peripheral node of  $A$  ( $T$ ,  $E$  or  $Y$ ), receives the SRD, it checks the  $(IP_A, N_A, t)$  tuple to verify that it has not already processed this SRD. Nodes do process packets for which they have already seen this tuple. The receiving node uses  $A$ ’s public key, which it extracts from  $A$ ’s certificate, to validate the signature and verify that  $A$ ’s certificate has not expired. If the packet is found to be authentic, it sets up a reverse path back to the source  $A$  by recording the neighbor from which it received the SRD, for example when the peripheral node  $T$  receives the SRD it sets up a reverse path back to  $A$  by recording the neighbor  $B$  from which it received the SRD ( $B$  sets up a reverse path to  $A$  during border-casting. Now,  $T$  sets up the reverse path to  $B$ . So a reverse path from  $T$  to  $A$  is set).

The peripheral node then signs the contents of the message originally border-cast by  $A$  and appends this signature and its own certificate to the SRD. It checks in its SIAR routing table whether it has a valid path to the destination  $Z$ . If it has ( $Z$  is within the zone of the node), it forwards the SRD directly to  $Z$  along this route, otherwise it rebroadcasts the packet to its peripheral nodes. In the present case since none of the peripheral nodes  $T$ ,  $E$  and  $Y$  has the route to  $Z$  ( $Z$  is not within the zone of

$T$ ,  $E$  or  $Y$ ), all rebroadcasts the SRD to their peripheral nodes, for example,  $T$  rebroadcasts the SRD to  $K$ .

$$T \rightarrow \text{bordercast} : [[SRD, IP_Z, cert_A, \beta, N_A, t] | sign_A] | sign_T, cert_T \\ \text{where, } sign_T = [[SRD, IP_Z, cert_A, \beta, N_A, t] | sign_A] SK_T$$

**Step 3:** Upon receiving the SRD,  $T$ 's peripheral node  $K$  checks the  $(IP_A, N_A, t)$  tuple, validates  $T$ 's signature and sets up the reverse path to  $T$  (if the signature is authentic).  $K$  then removes  $T$ 's certificate and signature, signs the contents of the message originally broadcast by  $A$  and appends this sign along with its own certificate to the SRD. It checks in its SIAR routing table whether it has a valid path to  $Z$ . Since it doesn't, it again rebroadcasts the packet to its peripheral nodes  $I$  and  $D$ .

$$K \rightarrow \text{bordercast} : [[SRD, IP_Z, cert_A, \beta, N_A, t] | sign_A] | sign_K, cert_K \\ \text{where, } sign_K = [[SRD, IP_Z, cert_A, \beta, N_A, t] | sign_A] SK_K$$

Each node along the path repeats these steps of validating the previous node's signature, recording the previous node's IP address for setting up the reverse path, removing the previous node's certificate and signature, signing the original contents of the message, appending its own certificate and rebroadcasting the message, until the SRD reaches a node, that has a valid route to the destination  $Z$  ( $Z$  is within the zone of the node). In this case the node instead of rebroadcasting the SRD, directly forwards it to  $Z$ . For example, when the SDR reaches  $J$ , it validates the packet, sets up the reverse path to the bordercasting node  $D$ , removes  $D$ 's certificate and signature, signs the contents of the message originally broadcast by  $A$ , appends this signature and its certificate and forwards the SRD to  $Z$ .

$$J \rightarrow Z : [[SRD, IP_Z, cert_A, \beta, N_A, t] | sign_A] | sign_J, cert_J \\ \text{where, } sign_H = [[SRD, IP_Z, cert_A, \beta, N_A, t] | sign_A] SK_J$$

**Step 4:** Finally, the SRD arrives at destination  $Z$ , which replies to the first SRD that it receives for a source and a given nonce. There is no guarantee that the first SRD received traveled along the shortest path from the source. A SRD that travels along the shortest path may be prevented from reaching the destination first if it encounters congestion or network delay, either legitimately or maliciously manifested. In this case, however, a non-congested, non-shortest path is likely to be preferred to a congested shortest path because of the reduction in delay. Because SRDs do not contain a hop count or specific recorded source route, and because messages are signed at each hop, malicious nodes have no opportunity to redirect traffic.

$Z$  on getting this SRD packet verifies it using both  $VK_J$  and  $VK_A$ , confirms its authenticity and extracts  $EK_A$ .  $Z$  creates a *secure route reply* (SRR) packet and *unicasts* it back to the source along the reverse path. The first node that receives the SRR sent by  $Z$  is  $H$ .

$$Z \rightarrow H : [SRR, IP_A, cert_Z, N_A, t, \{K_{AZ}\}EK_A] | sign_Z \\ \text{where, } sign_Z = [SRR, IP_A, cert_Z, N_A, t, \{K_{AZ}\}EK_A] SK_Z$$

The SRR includes a packet type identifier ‘‘SRR’’, the IP address of  $A$ , the certificate of  $Z$ , the nonce  $N_A$ , the associated time stamp  $t$  sent by  $A$  and a session key  $K_{AZ}$  between  $A$  and  $Z$  encrypted with  $EK_A$ , all appended by the signature  $sign_Z$  of  $Z$ . Nodes that receive the SRR forward the packet back to the predecessor from which they received the original SRD. Each node along the reverse path back to the source signs the SRR and appends its own certificate before forwarding the SRR to the next hop. Since,  $J$  is the next hop node to the source  $A$  after  $H$ :

$$H \rightarrow J : [[SRR, IP_A, cert_Z, N_A, t, \{K_{AZ}\}EK_A] | sign_Z] | sign_H, cert_H$$

Where,  $sign_H = [[SRR, IP_A, cert_Z, N_A, t, \{K_{AZ}\}EK_A] | sign_Z] SK_H$ .

$J$  on getting the SRR validates  $H$ 's signature on it, removes  $H$ 's signature and certificate, signs the contents of the message and appends this signature and its own certificate before unicasting the SRR to its neighbour  $L$ .

$$J \rightarrow L : [[SRR, IP_A, cert_Z, N_A, t, \{K_{AZ}\}EK_A] | sign_Z] | sign_J, cert_J$$

Where,  $sign_J = [[SRR, IP_A, cert_Z, N_A, t, \{K_{AZ}\}EK_A] | sign_Z] SK_J$

Each node checks the nonce and signature of the previous hop as the SRR is returned to the source. This avoids attacks involving impersonation and replay of the message. Eventually the source  $A$  receives the SRR.

**Step 5:** On getting the SRR,  $A$  verifies  $Z$ 's signature and the nonce returned by  $Z$  to conform its authenticity. It then extracts the session key  $K_{AZ}$ .  $A$  now encrypts the data packet using  $K_{AZ}$  and sends it to  $Z$  along the same route.

### 3.3 Route Maintenance

For route maintenance, SIER at each node keeps track of routes whether they are active or not. When there is no flow of traffic on an existing route for that route's lifetime or the link of an active route is broken due to node mobility or some other reasons, the route is deactivated by the node. Data received on an inactive route causes nodes to generate an Error (ERR) message. The node sends the ERR message to the source along the reverse path. All ERR messages must be signed to check the authenticity of the sender as well as the message. For a route between source  $A$  and destination  $X$ , a node  $M$  generates the ERR message for its neighbor  $N$  as follows:

$$M \rightarrow N : [ERR, IP_A, IP_X, cert_M, N_M, t] | sign_M$$

Where,  $sign_M = [ERR, IP_A, IP_X, cert_M, N_M, t] SK_M$

This message is forwarded along the path to the source without modification. A nonce and timestamp ensure that the ERR message is fresh. Since the ERR messages are signed, malicious nodes cannot generate ERR messages for other nodes. The non-repudiation provided by the signed ERR message allows a node to be verified as the source of each ERR message that it sends. The source node drops the duplicate ERR message with same nonce and time stamp.

## 4 Security Analysis of SHRP

In this section, we analyze the security aspects of SHRP by evaluating its robustness in the presence of multiple security attacks as mentioned in section 3 caused both by an external adversary and an internal compromised node within the network.

**Attacks involving impersonation:** SHRP participants, accept only those packets that have been signed with a certified key issued by a CA. In intra-zone routing since the SKREQs and SKREPs can only be signed by an authenticated source with its own private signature key, nodes can't impersonate (spoof) other nodes. Inter-zone routing follows hop-by-hop authentication during route discovery and end-to-end authentication during the route reply phase. So it is impossible for an external node or an internal compromised node to impersonate an intermediate node during inter-zone routing. Further since the SRD packet is signed by the source node using its private key, it guarantees that only the source can initiate a route discovery process. Similarly, the SRR packets include the destination's certificate and signature, ensuring that only the destination can respond to the route discovery. This prevents attacks where the source, the destination or any intermediate nodes are spoofed e.g. black-hole, wormhole or DoS attacks.

**Prevention from Information Disclosure:** No hop count information is present in the SRD or SRR packets. This prevents an external adversary or an internal compromised node from getting any kind of information about the network topology. Topology information is restricted to nodes within a zone. This is harmless as nodes accept packets only after verifying the sender's signature. Further all the data packets and the control packets that contain the session key are encrypted which ensures the confidentiality of information.

**Routing message Modification:** SHRP specifies that all fields of LSPs, SKREQ, SKREP, SRD and SRR packets remain unchanged between the source and the destination. Since all packets are signed by the initiating node, any alterations in transit would be immediately detected by intermediary nodes along the path, and the altered packet would be subsequently discarded. Repeated instances of altering packets could cause other nodes to exclude the errant node from routing. Thus, modification attacks like routing table poisoning are prevented.

**Fabrication of routing messages:** Messages can be fabricated only by the internal compromised nodes with certificates. In that case, SHRP does not prevent fabrication of routing messages, but it does offer a deterrent by ensuring non-repudiation. A node that continues to inject false messages into the network may be excluded from future route computation.

**Replay Attacks:** Replay attacks like wormhole attack are prevented by including a nonce and a timestamp with routing messages.

## 5 Conclusion

In this paper, we have presented the design and analysis of a new secure hybrid routing protocol for MANET called SHRP based on the design of ZRP. In designing SHRP, we carefully fit the inexpensive cryptographic primitives to each part of the protocol functionality to create an efficient and practical protocol that is robust against multiple



active attackers or compromised nodes in the network. SHRP gives a better solution towards achieving the security goals like message integrity, data confidentiality and authentication, by taking an integrated approach of digital signature and both symmetric and asymmetric key encryption. Together with existing approaches for securing the physical layer and MAC layer within the network protocol stack, the SHRP provides a foundation for the secure operation of an ad hoc network.

## References

1. Murthy, C.S.R., Manoj, B.S.: *Ad Hoc Wireless Networks, Architecture and Protocols*. Prentice Hall PTR (2004)
2. Basagni, S., Conti, M., Giordano, S., Stojmenovic, I.: *Mobile Ad Hoc Networks*. IEEE Press, A John Wiley & Sons, Inc., Publication (2003)
3. Aggelou, G.: *Mobile Ad Hoc Networks*, 2nd edn. Mc Graw Hill Professional Engineering (2004)
4. Chlamtac, I., Conti, M., Liu, J.J.-N.: *Mobile Ad Hoc Networking: Imperatives and Challenges*. Elsevier Network Magazine 13, 13–64 (2003)
5. Belding-Royer, E.M., Toh, C.-K.: A review of current routing protocols for ad-hoc mobile wireless networks. *IEEE Personal Communications Magazine*, 46–55 (April 1999)
6. Forouzan, B.A.: *Data communication and Networking*, 2nd edn. Tata McHill Publication (2001)
7. Johnson, D.B., Maltz, D.A.: Dynamic source routing in adhoc wireless networks. In: Imielinski, T., Korth, H. (eds.) *Mobile Computing*, pp. 153–181. Kluwer Academic Publishers (1996)
8. Perkins, C.E., Royer, E.M.: Ad hoc on-demand distance vector routing. In: *IEEE Workshop on Mobile Computing Systems and Applications*, pp. 90–100 (February 1999)
9. Jacquet, P., Muhlethaler, P., Qayyum, A.: *Optimized Link State Routing Protocol*. Internet Draft, draft-ietf-manetolsr-00.txt (November 1998)
10. Haas, Z.J., Pearlman, M.R., Samar, P.: *The Zone Routing Protocol (ZRP)*. IETF Internet Draft, draft-ietf-manet-zone-zrp-04.txt (July 2002)
11. Schaumann, J.: *Analysis of Zone Routing Protocol*. Course CS765, Stevens Institute of Technology Hoboken, New Jersey, USA (December 8, 2002)
12. Haas, Z.J., Pearlman, M.R., Samar, P.: *Intrazone Routing Protocol (IARP)*. IETF Internet Draft, draft-ietf-manet-iarp-01.txt (June 2001)
13. Haas, Z.J., Pearlman, M.R., Samar, P.: *Interzone Routing Protocol (IERP)*. IETF Internet Draft, draft-ietf-manet-ierp-01.txt (June 2001)
14. Haas, Z.J., Pearlman, M.R., Samar, P.: *The Bordercast Resolution Protocol (BRP) for Ad Hoc Networks*. IETF Internet Draft, draft-ietf-manet-brp-01.txt (June 2001)
15. Zhou, L., Haas, Z.J.: *Securing Ad Hoc networks*. *IEEE Network Magazine* 13(6) (December 1999)
16. Forouzan, B.A.: *Cryptography and Network Security, special Indian edn*. Tata McHill Publication (2007)
17. Sanzgir, K., Dahill, B.: A secure routing protocol for ad hoc networks. In: *Proceeding of the 10th IEEE International Conference on Network Protocols*, pp. 1–10 (2002)
18. Michiardi, P., Molva, R.: Ad hoc networks security. In: Basagni, S., Conti, M., Giordano, S., Stojmenovic, I. (eds.) *Ad Hoc Networking*. IEEE Press Wiley, New York (2003)

# Compact UWB/BLUETOOTH Integrated Uniplanar Antenna with WLAN Notch Property

Yashwant Kumar Soni and Navneet Kumar Agrawal

Department of Electronics and Communication,  
College of Technology and Engineering,  
Maharana Pratap University Agriculture and Technology,  
Udaipur, India  
{yash.swarnkar,navneetctae}@gmail.com

**Abstract.** A low-profile and planar dual band integrated antenna is presented and discussed, which possesses notched frequency band to eliminate interference signal. In this paper the designed antenna has got speciality having dual frequency band operation covering Bluetooth (2.4–2.484 GHz) and UWB (3.1–10.6 GHz) and above all single band 5.15-5.825GHz (WLAN) notch features. Designed antenna is fed by coplanar waveguide (CPW) and built on the low cost (FR-4) substrate with height of 0.76 mm and 26×20 mm<sup>2</sup> plane area. The antenna uses a new technique comprising of cutting slots at the periphery of the rectangular stub and tapered shape in the vicinity of the feed-line to improve its VSWR characteristics over the UWB and Bluetooth band. An effective half-wavelength tuning element is used to generate the necessary band-notch at the WLAN interfering band. Full wave analysis of the proposed antenna design in frequency domain and optimize design is obtained by using the commercial simulation tool ANSOFT HFSS\_v12.1 which is based on finite element analysis. The antenna exhibits stable omnidirectional radiation patterns and acceptable gain flatness across the entire UWB and Bluetooth band.

**Keywords:** Bluetooth, Integrated antenna, Ultrawideband, Band-notch characteristics.

## 1 Introduction

Ultra-wideband (UWB) system design and application have become the objective of the wireless communication since Federal Communications Commission (FCC) released the frequency band of 3.1-10.6 GHz for commercial applications in 2002 [1]. By April of that year, the FCC gave formal approval for the unlicensed use of the technology in this band [1]. Because of the rapid development of UWB systems in recent years, UWB antenna designing has become a very competitive and challenging topic for research and development in telecommunications industry and academia. Challenges involved with the development of a feasible UWB antenna design for consumer electronics applications include compact size and low manufacturing cost as well as impedance matching and radiation stability over the entire UWB.

UWB antennas have a double layer structure and normally excited by a microstrip feed or a probe feed with a large ground plane. Uniplanar structure gained attention due to advantages like single metallic layer structure and compatible to MMICs. CPW is the widely used uniplanar feeding technique. In this paper a compact uniplanar CPW fed printed antenna is presented and discussed.

The nominated bandwidth of UWB system will cause electromagnetic interference with existing narrowband wireless communication technology, for instance, the wireless local area network (IEEE 802.11a) operating at 5.15-5.825 GHz. One band stop filter can be used with the UWB antenna to reject this band. However, this increases the size and complexity of the system. So it is desirable to design the UWB antenna with single notched frequency band at 5–6 GHz to minimize the potential interferences between UWB and narrowband systems.

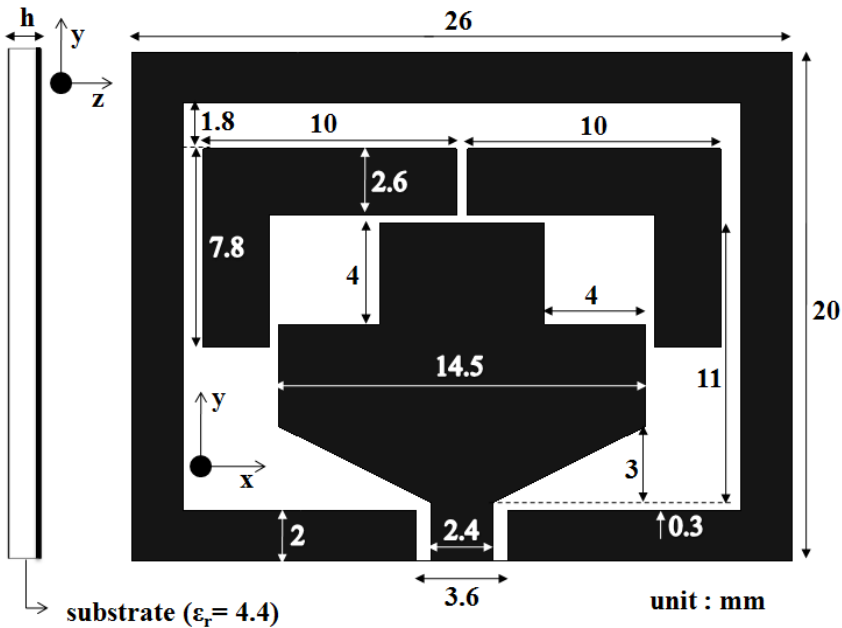
In the literature, several techniques have been reported to design printed UWB antennas with single band notched property. The normally used techniques to realize a band-notched characteristic include etching slots or slits on the radiator [2], placing parasitic strips in close proximity to the antenna [3]. To design slot antennas with band-notched property, one simple and effective way is to incorporate slots in the antenna's tuning stub, such as a U-shaped [4], V-shaped [3], C-shaped slots [4], etc. Furthermore, Yi-Cheng Lin et al. discussed the designs of three, advanced band-notched (5–6 GHz) UWB rectangular aperture antennas [5]. The antenna structure is quite simple and the aperture size is compact. Broad impedance bandwidth and stable radiation patterns are obtained, but the ground plane dimension is a bit large. In practice, when integrated with the system board of different size of ground plane, the antenna might need a retuning for the optimized dimensions. Wang-Sang Lee et al. proposed the design of wideband planar monopole antennas with dual band-notched characteristics [6], which is suitable for creating UWB antenna with narrow frequency notches or for creating multiband antennas. However, this antenna is not suitable for integration with compact systems, because its ground plane is very large and it is perpendicular to the radiator, which limits its applications in compact UWB systems. Furthermore, the bandwidth performance of this antenna is from 2 GHz to 6 GHz, which can not satisfy the UWB requirement. In UWB systems, printed slot (aperture) antennas with different structures and performances have been developed [2, 3, 7 and 8]. From the literature survey, it has been observed that the slot antenna [9-10] is one of the promising candidates for UWB applications. CPW fed square slot antennas have been presented for wide band characteristics [11, 12]. In general, the wideband slot antenna can be developed by tuning their impedance. One of the impedance tuning techniques is to vary the slot dimensions, which has been carried out with various slot geometries like bowtie [13], wide rectangular [14, 15], circular [7, 9] and hexagonal slots [16].

On the other hand, the ambitious goal of wireless connectivity for “everything for everybody at any place and any time” requires a comprehensive integration of existing narrowband (Bluetooth) and new UWB systems that link devices as diverse as fixed and portable appliances, PCs, and entertainment equipment. However, the design of those proposed previous works [2-16] was not integrated ultra wideband and Bluetooth and quite complex and the approach in reference [17] achieves the

integration of Bluetooth and UWB antenna but it is not compatible to suppress the interference with other narrowband service like WLAN. Hence, this has motivated us to design up a compact, less complex and low cost antenna to cover entire Bluetooth as well as UWB band with WLAN notched band characteristics.

## 2 Antenna Design Parameters and Methodology

Fig. 1 shows the geometry and configuration of the proposed CPW-fed integrated uniplanar antenna with band-notched characteristics. The antenna consists of a rectangular aperture etched in the ground-plane of a PCB and is excited using a rectangular stub. Reduction in the antenna's rectangular aperture area, affects its input impedance match over the operating band, especially at lower frequencies. The aperture area of  $22 \times 16 \text{ mm}^2$  was determined to satisfying the antenna's impedance match across the UWB and Bluetooth band. The CPW-fed line is designed to have a characteristic impedance of  $50 \Omega$ . The advantage of this configuration is twofold; one is ease in manufacturing and the other is low cost. The CPW line allows the antenna to be easily integrated with microwave/radio-frequency circuitry on the system board.



**Fig. 1.** Geometry and configuration of the proposed integrated uniplanar antenna

To mitigate the interferences with the WLAN (IEEE802.11a) systems operating in the 5.15-5.825GHz band the need for a notch-band is highly desirable in the UWB system. In this paper, the band-notched characteristic is realized by introduction of an

additional resonant structure to the radiating structure. The pair of L-shaped tuning stubs on aperture as a resonant structure has been used independently to obtain 5-6GHz stop band as shown in Fig. 1. Usually, the length of the resonant structure is made approximately equal to the half of the guided wavelength at the notch frequency, which is given by

$$L_{\text{WLAN}} = \frac{\lambda_g}{2} = \frac{c}{2f_{\text{notch\_WLAN}} \sqrt{\frac{\epsilon_r + 1}{2}}} \quad (1)$$

where  $L_{\text{WLAN}}$  is the length of the L-shaped tuning stub. For the notch frequency  $f_{\text{notch\_WLAN}} = 5.15$  GHz the length of the tuning stub is calculated as 17.72 mm at the beginning of the design. Finally the lengths are adjusted to 17.8mm by using simulator for obtaining the desired results. The optimum width of tuning stub is found to be 2.6mm after the exhaustive simulation studies. The appearance of tuning stubs will change the surface current and electric field distribution on the antenna and does not radiate noticeably, which reduces the input impedance of the antenna and causes significant mismatch at the desired WLAN band. The simulated VSWR characteristics for these configurations of tuning stubs are shown in Fig. 2. Due to the resonant structure inside the aperture, maximum current flows back to the feeding part and degenerates radiation around 5.1 GHz to 6.2 GHz. All the optimized parameters of the tuning stubs configuration are provided in Fig. 1.

In addition to this, cutting slots at the periphery of the rectangular stub can be used to act as an impedance matching network to control the impedance bandwidth of the proposed antenna. The cutting slots create a capacitive load that neutralizes the inductive behavior of the stub antenna to produce nearly-pure resistive input impedance. By creating tapered shape in the rectangular stub near the fed line, the antenna's VSWR characteristics can be improved. For the simulation model, the conductive parts (antenna metallization, stubs and feed line) are modeled as single-sheet perfect electric conductor referred to as the simplest case. This printed antenna provides a cheap solution compatibility with standard PCB manufacturing that can be easily connected to electronic sensors and circuitry.

### 3 Results and Discussions

The simulated VSWR characteristics obtained from Ansoft HFSS of the proposed antenna is shown in Fig.2. The simulated curve obtained shows that antenna has two resonances at Bluetooth and UWB frequency band respectively. The influence of L-shaped tuning stubs on antenna performance is also studied; the WLAN stop band is due to the L- shaped tuning stubs.

The E-plane and H-plane radiation patterns at three resonant frequencies 2.4, 5.6, and 8.8 GHz have been found by simulation as given in Fig.3. The antenna exhibits

the uniform omnidirectional radiation behavior in the H- plane and dipole like nature in the E-plane across the Bluetooth and UWB frequency bands. However the E-plane and H-plane patterns are stable in entire UWB and Bluetooth region. Thus it may be considered as good antenna characteristics for UWB/Bluetooth applications.

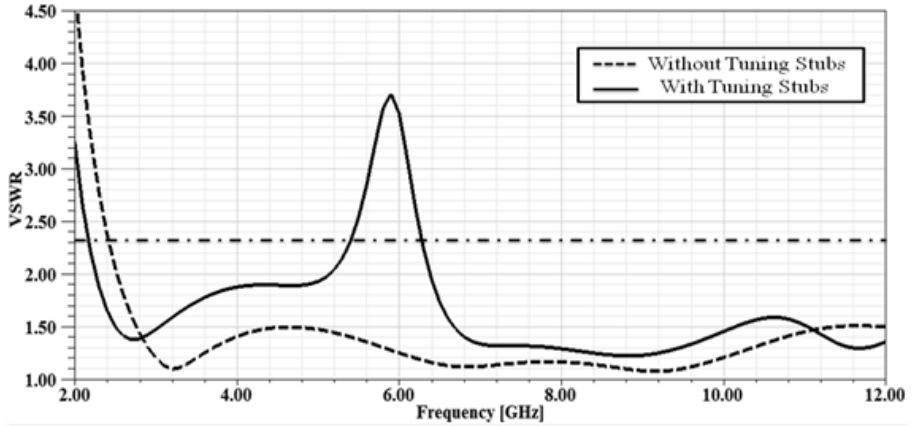


Fig. 2. Simulated VSWR with and without L-shaped tuning stubs

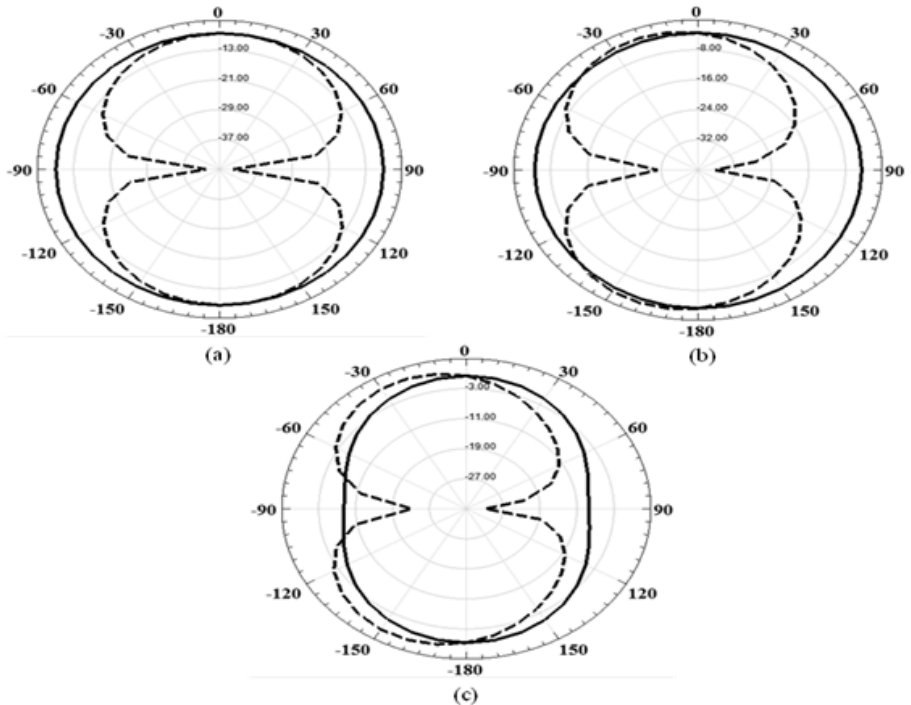


Fig. 3. E-plane (dashed line) and H- plane (solid line) radiation pattern at (a) 2.4GHz, (b) 5.6GHz and (c) 8.8GHz

In the notched frequency band most of the power fed into the antenna is reflected back which leads to a decrease of the antenna efficiency and hence the antenna gain. The simulated gain and radiation efficiency of the proposed antenna is shown in Figs. 4 and 5 respectively. This is a sharp decrease in gain and efficiency around 5.7 GHz in notch band. As observed in the figure the gain in WLAN rejection band is expected to be sharply reduced as low as -8.2 dBi.

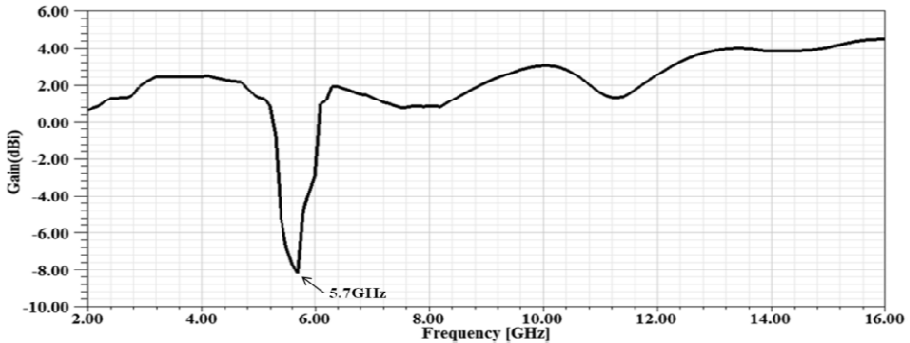


Fig. 4. Simulated peak gain of the proposed antenna as a function of frequency

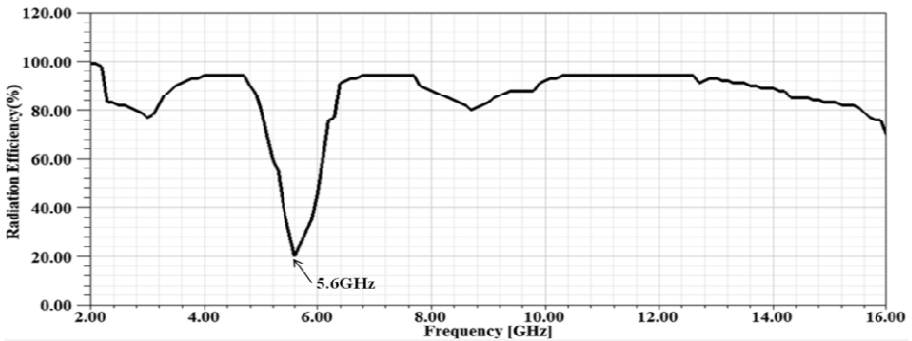
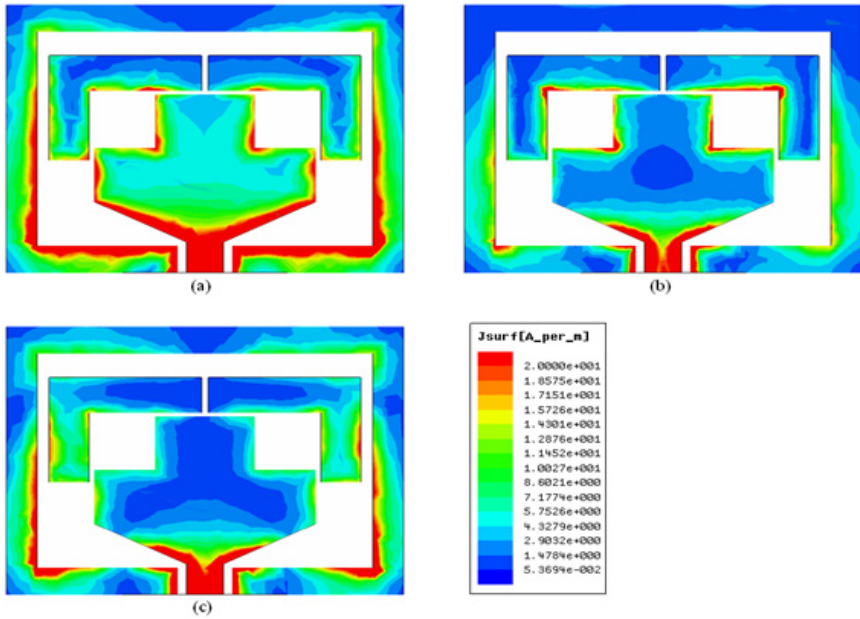


Fig. 5. Frequency versus antenna radiation efficiency plot

The proposed antenna provides more than 77% efficiency and the gain varies from 1.2 dBi to 3 dBi over the UWB/Bluetooth frequency range except in notch band. For other frequencies outside the rejected band, the gains remain good and stable in performance. These characteristics can make sure the ability of the proposed antenna to reject unwanted band effectively. For better understanding of the proposed antenna behavior, the simulated current distribution on the antenna at the frequencies of 2.4 GHz, 5.6GHz and 8.8 GHz is presented in Fig. 6. At Bluetooth frequency band of 2.4 GHz the distribution of surface current is almost uniform as is evident in Fig. 6(a). According to Fig. 6(b), L-shaped tuning stubs in the rectangular aperture area are responsible for WLAN stop band characteristics. In Fig. 6(c) the current distribution is concentrated near the feed line at frequency 8.8GHz.



**Fig. 6.** Surface current distributions on the antenna at (a) a Bluetooth frequency of 2.4 GHz, (b) WLAN notched band at 5.6 GHz (c) pass band frequency of 8.8 GHz

#### 4 Conclusion and Future Work

A CPW fed Bluetooth/UWB integrated planar aperture antenna with single notched band has been proposed and discussed. Single stop band is realized by pairs of half-wavelength L-shaped tuning stub. The single stop band of the antenna can significantly suppress potential EM interference in the UWB region. The VSWR performance of this antenna has been obtained by simulation. The omnidirectional radiation patterns are very stable across the operating band of Bluetooth and UWB. The UWB response fully encloses the 3.1–10.6 GHz band and is not affected by the Bluetooth resonance. The antenna exhibits good impedance matching and uniform gain behavior. Therefore the proposed antenna is expected to be a good candidate in various applications. Authors have developed the three separate drawing of the antenna on PCB as a part of hardware implementation. The antennas are in the laboratory testing stage. The hardware design will be further refined to correlate the simulated results with the field testing results.

**Acknowledgment.** Authors would like to thank technical team of Techno INJR institute of technology, CTAE and GITS for supporting and providing laboratory facility to carry out the experimental analysis and testing of the antenna.



## References

1. Federal Communications Commission, First report and order, revision of Part 15 of Commission's rule regarding ultra-wideband transmission system FCC 02-48 (2002)
2. Liao, X.-J., Yang, H.-C., Han, N., Li, Y.: Aperture UWB antenna with triple band-notched characteristics. *Electronic Letters* 47(2) (2011)
3. Mehdipour, A., Parsa, A., Sebak, A.R., Trueman, C.W.: Miniaturised coplanar waveguide-fed antenna and band-notched design for ultra-wideband applications. *IET Microwave Antennas Propagation* 3(6), 974–986 (2009)
4. Zhou, H.-J., Sun, B.-H., Liu, Q.-Z., Deng, J.-Y.: Implementation and investigation of U-shaped aperture UWB antenna with dual band notched characteristics. *Electron Letters* 44(24) (2008)
5. Cho, Y.-J., Kim, K.-H., Choi, D.-H., Lee, S.-S., Park, S.-O.: A miniature UWB planar monopole antenna with 5-GHz band-rejection filter and the time-domain characteristics. *IEEE Trans. Antennas Propagation* 54, 1453–1460 (2006)
6. Lee, W.S., Kim, D.Z., Kim, K.J., Yu, J.W.: Wideband planar monopole antennas with dual band-notched characteristics. *IEEE Transaction Microwave Theory Tech.* 54, 2800–2806 (2006)
7. Elboushi, A., Ahmed, O.M.H., Sebak, A.R., Denidni, T.A.: Study of elliptical slot UWB antennas with 5.0-5.6 GHz band-notched capability. *Progress In Electromagnetic Research C* 16, 207–222 (2010)
8. Lui, W.-J., Cheng, C.-H., Zhu, H.-B.: Improved frequency notched ultra-wide band slot antenna using square ring resonator. *IEEE Trans. on Antennas and Propagation* 55(9), 2445–2449 (2007)
9. Gao, G.-P., Mei, Z.-L., Li, B.-N.: Novel circular slot UWB antenna with dual band notched characteristic. *Progress In Electromagnetics Research C* 15, 49–63 (2010)
10. Behdad, N., Sarabandi, K.: A multiresonant single-element wideband slot antenna. *IEEE Antennas and Wireless Propagation Letters* 3, 5–8 (2004)
11. Chiou, J.Y., Sze, J.Y., Wong, K.L.: A broad-band CPW-fed strip-loaded square slot antenna. *IEEE Trans. on Antennas and Propagation* 51(4), 719–721 (2003)
12. Chen, H.D.: Broadband CPW-fed square slot antenna with a widened tuning stub. *IEEE Transaction on Antennas and Propagation* 51(8), 1982–1986 (2003)
13. Marantis, L., Brennan, P.: A CPW-fed bow-tie slot antenna with tuning stub. In: *Proc. 2008 Loughborough Antennas & Propagation Conference*, pp. 389–393 (2008)
14. Eldek, A., Elsherbeni, A.Z., Smith, C.E.: Rectangular slot antenna with patch stub for ultra-wide band applications and phased array systems. *Progress in Electromagnetic Research* 53, 227–237 (2005)
15. Chen, W.Q., Ding, G.F., et al.: Design and simulation of broad-band unidirectional CPW-fed rectangular slot antennas. In: *IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, pp. 632–635 (2007)
16. Kraisor, S.-K., Vivek, V., Akkaraekthalin, P.: A broadband CPW-fed equilateral hexagonal slot antenna. In: *IEEE ICSIT Proceedings*, pp. 783–786 (2006)
17. Yildirim, B.S., Cetiner, B.A., Roqueta, G., Jofre, L.: Integrated Bluetooth and UWB Antenna. *IEEE Antennas Propag. Letters* 8 (2009)

# Open Security System for Cloud Architecture

S. Koushik and Annapurna P. Patil

Department of ISE, Department of CSE,  
M.S. Ramaiah Institute of Technology,  
Bangalore -560054

koushik.85@gmail.com, annapurnap2@yahoo.com

**Abstract.** Cloud computing is a computing platform that delivers computing resources as a service over a network. Infrastructure, data, software, platform and many more such computing resources are provided by different vendors for different purposes. This enables the capability to hold control on the resources that the vendors are providing to its users. This paper focus on security aspects of cloud computing. Any operations over a network are vulnerable to attacks. The data present on cloud server is viable to risks such as theft or loss of data. In this context, securing data becomes top priority. Currently the cloud providers are providing their own security mechanisms. Building the security around the cloud may prove costly in terms of cost and time for a cloud provider. This paper focus on providing an open security mechanism that can be used by all cloud providers, thus achieving high security and manageability at affordable cost.

**Keywords:** Cloud, Open Security, Public Cloud, Private Cloud, Security Architecture, Cloud Security Issues.

## 1 Introduction

Cloud as we know is one of the most exciting platforms in the field of computing. There is a huge impact of technological change that is built around the cloud. Amazon, Google, Apple, Microsoft and many more companies are focusing on cloud technology and are offering different services on cloud. Not just corporations, even government bodies are also keen using this technology for their purposes. Government operated clouds like U. S (Apps.gov), U. K (G-Cloud), and Canada (Canadian Government Cloud) are also trying to move to this technology. Cloud applications are accessed using a web browser, the control of the software application remains with the company itself which is easier to manage. The applications provided on cloud are provided as a service. Cloud is basically modelled as XaaS i.e. X-as-a-service where 'X' stands for different services viz. Software, Platform, Infrastructure, Database, Network, Storage and many more [14, 12].

Many mid-size and small companies are not exploiting the cloud technology to the fullest because of the drawbacks in current security policies adopted by cloud. When one gets assurance about the secure data that is being hosted on cloud is as safe as they have a local copy then, moving to cloud would become easy.

The technology behind cloud is virtualization. Virtualization has its own advantages and disadvantages, one of the major problems being security. Operation, administration and maintenance cost is more if the system is not local. Remote monitoring is used to fix the issues on remote sites, network failures results in more overhead. Different cloud providers have different SLAs (Service Level Agreements), which makes building a secure cloud difficult [7].

Security has many paradigms. Various researchers are discussing about the cloud security from their own viewpoints. One of the major concerns of security in cloud is when the data is posted on the public cloud. Private cloud owners have their own security mechanisms and due to the data being local in the private cloud it becomes easier for organizations to secure the data [7].

## **2 Types of Cloud**

### **2.1 Private Cloud**

Private cloud is commonly called as internal cloud or corporate cloud. Private clouds provide services to limited number of users and have a private firewall. The services are provided to people behind the firewall. This private cloud is owned by any corporate company to cater the services internal to the organization. The cloud can either be built by the organizations or bought by a third-party [12, 13].

### **2.2 Public Cloud**

A public cloud may be established where several organizations have similar requirements and seek to share infrastructure so as to realize some of the benefits of cloud computing. The costs are spread over fewer users than a public cloud (but more than a single tenant) [10]. This option offers a higher level of privacy, security, and/or policy compliance. In addition, it can be economically attractive as the resources (storage, workstations) utilized and shared in the community are already exploited.

Few other cloud mechanisms that are existent are: Community cloud, Hybrid cloud and inter-cloud. These cloud computing platforms are useful in their own rights [12].

## **3 Security in Cloud**

Security controls in cloud computing are, for the most part, no different than security controls in any IT environment. However, because of the cloud service models employed, the operational models, and the technologies used to enable cloud services, cloud computing may present different risks to an organization than traditional IT solutions [7].

One of the attractions of cloud computing is the cost efficiencies afforded by economies of scale, reuse, and standardization [8]. To bring these efficiencies to bear, cloud providers have to provide services that are flexible enough to serve the largest customer base possible, maximizing their addressable market. Unfortunately, integrating security into these solutions is often perceived as making them more rigid [6].

## 4 Open Issues in Cloud Security

From an architectural perspective, there is much confusion surrounding how Cloud is both similar and differs from existing models and how these similarities and differences might impact the organizational, operational and technological approaches to Cloud adoption as it relates to traditional network and information security practices. There are those who say Cloud is a novel sea-change and technical revolution while others suggest it is a natural evolution and coalescence of technology, economy, and culture. The truth is somewhere in between [4].

There are many models available today which attempt to address Cloud from the perspective of academicians, architects, engineers, developers, managers and even consumers. We will focus on a model and methodology that is specifically tailored to the unique perspectives of IT network and security professionals [6].

The keys to understanding how Cloud architecture impacts security architecture are a common and concise lexicon coupled with a consistent taxonomy of offerings by which Cloud services and architecture can be deconstructed, mapped to a model of compensating security and operational controls, risk assessment and management frameworks and in turn, compliance standards [2, 3, 4].

## 5 Open Security Architecture

The proposed open security architecture is based on the control catalogue in OSA (Open Security Architecture). The OSA is based on National Institute of Standards and Technology (NIST) model. This model provides one among the best catalogue for the IT industry. The catalogue is built on open standard; the usage of this standard is free and can be used without restriction. The same standard is also available in ISO17799 [6].

By taking a single control catalogue we allow you to clearly establish how you can meet the objectives of many standards, without having to repeatedly work out what controls are needed and how they can be implemented. In addition we map against threats and supply tests, so you can quickly establish whether a particular control is relevant for your situation, and can check it's working correctly.

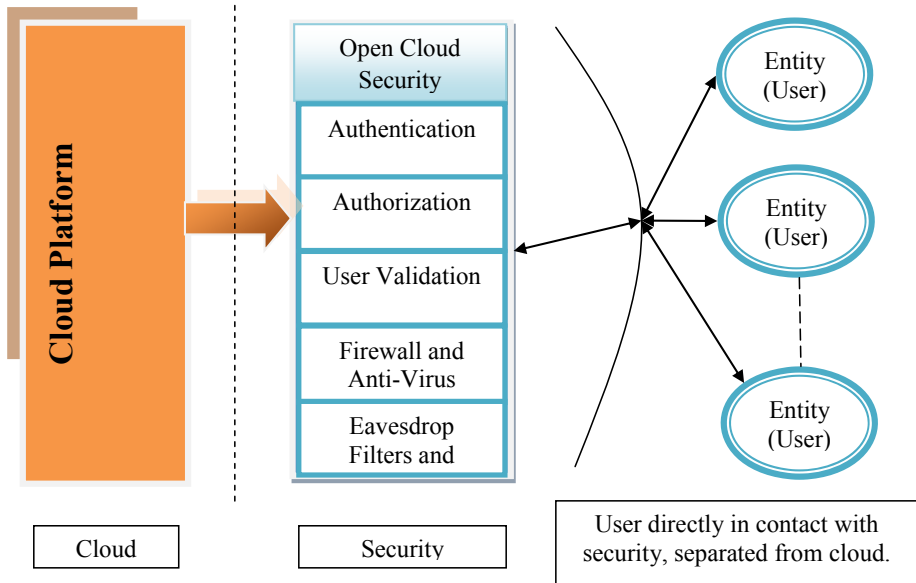
### 5.1 Proposed Architecture for Open security Architecture

In this paper we propose open security model for cloud. This open security model can be used by different cloud providers at the same time. Each cloud provider has its own security mechanism, but this model enables to share a security mechanism between different cloud vendors and concentrate on just the other services that can be provided on cloud.

Open security is useful when a cloud app is built for multiple cloud vendors and migration of data from one cloud to another is required. For example let us assume that the data and application is shared among different clouds. Each cloud provider has its own security mechanisms in usage. So understanding different security

mechanisms takes some time to access data on different cloud providers becomes difficult and time consuming. Having a common security model or open security model solves this problem. Implementing a security model where all the cloud providers can access this security mechanism and then provide to the user makes managing and securing data more easy.

In the figure 1 we can see the architecture diagram for open security between multiple clouds.



**Fig. 1.** Proposed Open Security Architecture

Figure 1 suggests that having an open security mechanism improves the performance of security as the service concentrates only on security part of the cloud service. When this security mechanism can be made as a service like Security-as-a-service then the user will first login to security mechanism and authenticate. Once the authentication process is complete the security mechanism allows the user to login to access the data and work on it.

**5.2 Advantages of Open Security**

Public cloud is very vulnerable to threats and eavesdropping. Advantage of having this model is that the user login and authentication is controlled by a common registrar. The user is registered with the security registrar instead of the cloud vendor itself. This registrar will be responsible for authenticating and authorizing the user before accessing the cloud services.

The open security registrar consists of a firewall and anti-virus that takes care of external threats and other types of data misappropriation. Instead of directly allowing access to the data and application on the cloud the open security acts as a filter that resides outside cloud that takes care of security policies, this mechanism separates users and data that makes data more secure and hidden [6].

This also provides more advantages of having control over data by organizations than posting data over cloud [7, 8].

## 6 Conclusion

The proposed open security policy for cloud can be used by multiple cloud providers for securing their private as well as public cloud. Choice of encryption algorithm and policies should depend on the providers using the security. This enables the interoperability between clouds which make cloud computing more flexible and reliable. The availability of data and applications can be made easier using this security.

## References

- [1] Bass, L., Clements, P., Kazman, R.: Software Architecture in Practice, 2nd edn. Addison Wesley (2003)
- [2] Garlan, D., Perry, D.: guest editorial to the IEEE Transactions on Software Engineering (April 1995)
- [3] Perry, D.E., Wolf, A.L.: Foundations for the Study of Software Architecture. ACM SIGSOFT Software Engineering Notes 17(4) (October 1992)
- [4] Cloud Security, Rational Survivability homepage on Cloud Security Architectural Framework (2009),  
<http://www.rationalsurvivability.com/blog/?p=1150>
- [5] Gerber, A., van der Merwe, A., Barnard, A.: A Functional Semantic Web Architecture (2009)
- [6] Open Security Architecture, homepage on open security architecture framework (2011),  
<http://www.opensecurityarchitecture.org/cms/index.php>
- [7] Almorsy, M., Grundy, J., Ibrahim, A.S.: Collaboration-Based Cloud Computing Security Management Framework (2011)
- [8] Moshovos, A.: Advanced Computer Architecture. Fall (2005)
- [9] Taylor, R.N.: Software Architectures (2008)
- [10] Web Services Architecture. W3C Working Group Documentation (2010)
- [11] Wood, J., Brodlie, K., Seo, J., Duke, D., Walton, J.: A Web Services Architecture for Visualization (2008)
- [12] Petcu, D., Craciun, C., Neagul, M.: Silviu, P., Architecturing A Sky. Computing Platform (2011)
- [13] Riteau, P.: Large Scale Sky Computing Applications with Nimbus
- [14] Cloud Computing, Andy Bechtolsheim (2008)
- [15] Security Guidance for Critical Areas of Focus in Cloud Computing V2.1. Cloud Security Alliance (December 2009)

# Client-Side Encryption in Cloud Storage Using Lagrange Interpolation and Pairing Based Cryptography

R. Siva Ranjani, D. Lalitha Bhaskari, and P.S. Avadhani

Dept.of CS&SE, Andhra University, Visakhapatnam, Andhra Pradesh, India  
rsivaranjani552008@gmail.com, lalithabhaskari@yahoo.co.in,  
psavadhani@yahoo.com

**Abstract.** The rapid growth in the availability and popularity of cloud services allows on demand remote storage and the computation. Security and the privacy are the main concern to establish the trust cloud service; a solution is needed to achieve confidentiality and integrity of the user data. Many modern cryptographic techniques are available to protect the user's data in the cloud storage. In this research, we proposed an asymmetric cryptographic technique for securing the data which encrypts the original message before placing it in the cloud storage. The proposed algorithm is providing the following securities: security in transit and security at rest. Our proposed protocol equipped with the security parameters like confidentiality, authentication.

**Keywords:** Client side Encryption, Cloud storage, bilinear map, broadcast encryption, Lagrange interpolation, cloud data storage.

## 1 Introduction

Nowadays many organizations are providing the cloud services. The recent rise in the availability of the cloud services makes them enamoring and economically sensible for the clients with limited computing or the storage resources that are unable to procure or maintain their own resources for computing. Clients can easily utilize large amounts of data and computation to remote locations and also run the application directly from the cloud. Hence, there are two main issues to be riveting in cloud computing are: capable of data storage and data security.

Data security is the raising thrust area in the cloud computing, because organizations critical data is proceeded to geographically through the cloud platforms. Many security models [1][2][3][4] were proposed under different systems. Securing the data [5][6][7][8][9] in the cloud can be done at the server side or client side. In server side security, server decides the algorithm for encrypting the data. Where as in client side security, user uses own algorithm to encrypt the data before playing it on the cloud. Client side encryption is a security and privacy measure that involves encrypting data on the user's computer. The main benefits to this approach are increased security and privacy. Since the service provider does not have access to the key, not even an employee of the company can access the data.

## 1.1 Motivation and Contribution

The existed protocols are secured only against the passive adversaries. But, in the real world attackers are usually active. The active attackers can be able to do man-in-middle attack. Hence it is essential to secure the data against the active adversaries. Amazon S3 are providing to encryption methodologies: 1:server side encryption and 2: client side encryption. In, server side encryption the server will encrypt the user data and store it on a server. In client side encryption the user encrypts the data on the client side before uploading it on Amazon s3 server. Currently the only AWS SDK for Java supports client side encryption, uses a process called envelope encryption. In envelope encryption, the client generates a one-time-use symmetric key that the client uses to encrypt your data.

1. The client encrypts the envelope symmetric key using your private encryption key.
2. The client then uploads the encrypted envelope key along with your encrypted data to Amazon Simple Storage Service (S3).

Retrieving and decrypting the client-side encrypted data from Amazon S3 is the reverse of the encryption flow above:

1. The client retrieves your encrypted data from Amazon S3 along with the encrypted envelope key.
2. The client then decrypts the encrypted envelope key using your private encryption key. The client decrypts your data using the envelope key.

The main drawback with this technique is client is sending the symmetric key through the envelope, then there is a scope to access and extract the key during the transmission over the network. In this paper our aim is to secure the data and key from both passive attacks and active attacks by using Asymmetric Authenticated Group key[14] algorithm derived using pairing based cryptographic concepts. The proposed one is an asymmetric key encryption algorithm. In normal asymmetric key algorithms, each user is having their own public and private key. But in the proposed one, there is only one common group public key derived by all the users having the permission to use the data storage and a separate private key to each user. Here the user need not send any key to other users, whenever he extracts data from data storage, decrypts the data using his own key.

## 1.2 Related Work

Using Amazon S3 [11] as an example, cloud storage providers tend to offer a plain file - system like interfaces to the end users without exposing them to the complicated management of physical servers After a client moves its data to the cloud, the client relinquishes its ultimate control over the data, which is now entirely managed by the cloud service provider. Thus, it is essential for the client to be able to verify that her data is still available on the cloud in its original form and is ready for retrieval when necessary. For instance, the client might want to make sure that her data has not been



corrupted (deleted or modified) or moved to an off-line unavailable storage medium, that could be caused by either an attempt of a dishonest provider to save storage costs or by outages and security breaches within the cloud services themselves. Juels and Kaliski [10] proposed the basic scheme, given the solution to the cloud security by implementing a proof of Retrievability(POR). In this method, the client first calculates the error correcting code to the original file, then encrypts the file and finally adds some random sentinel points at different locations before storing on the cloud server.

Many modern symmetric key encryption techniques are existed for encrypting the data, before storing the data on the cloud server. There is a wide range of encryption algorithms e.g. AES, DES, TripleDES, Blow fish are available, which are secured enough for converting the plaintext data into cipher text format. For checking the correctness of the received data the user can choose any hash coding techniques e.g. MD5,SHA-1. Thus, the combination of the encryption algorithm and hash coding is used for achieving the confidentiality and integrity. Preferring to use those cryptographic techniques leads to computational overhead and the communication overhead, this affects the performance of the cloud system.

## 2 Preliminaries

In this section, we put forward the notations, definitions that we used in the discussion of the forthcoming sections.

### 2.1 Bilinear Maps

We review the basic notations of the bilinear maps [12,13] under our proposal, Let  $(G_1,+)$ ,  $(G_2,+)$  and  $(G_T, \cdot)$  are the three multiplicative group of prime order  $q > 2k$  for a security parameter  $k < N$ . We say that, there exists a bilinear map  $\hat{e} : G_1 \times G_2 \rightarrow G_T$  satisfies the following properties:

1. Bilinearity:  $\hat{e}(aP, bQ) = \hat{e}(P, Q)^{ab}$ ,  $\forall P, Q \in G_1, \forall a, b \in \mathbb{Z}$ .
2. Non-degeneracy:  $\forall P \in G_1$ , and  $\forall Q \in G_2$  then  $\hat{e}(P, Q) = 1$  if  $P = 1 \in G_1$ .
3. Computability:  $\forall P \in G_1$  and  $\forall Q \in G_2$  then  $\hat{e}(P, Q)$  is efficiently computable.

if  $G_1=G_2$  and  $P=Q$  then it is called as symmetric bilinear map groups. A bilinear map is defined as a probabilistic polynomial time algorithm  $(E)$  that takes a security parameter  $k$  and returns a uniformly random tuple  $(G_1, G_2, G_T, \hat{e}, g, q)$  of bilinear parameters, where  $g$  is the generator of  $G_1$  and  $\hat{e}$ : is the bilinear map.

## 3 Protocol

This section summarizes the description of the proposed protocol. In this protocol, the organization should mention the users list(who will access the data stored on the cloud) to the service provider. The cloud service provider collects the information

from all the users who access the data storage, and then broadcast the secret information to the users as the response. Each user in the group collects the secret information to derive the group master public key is shown in figure 1.

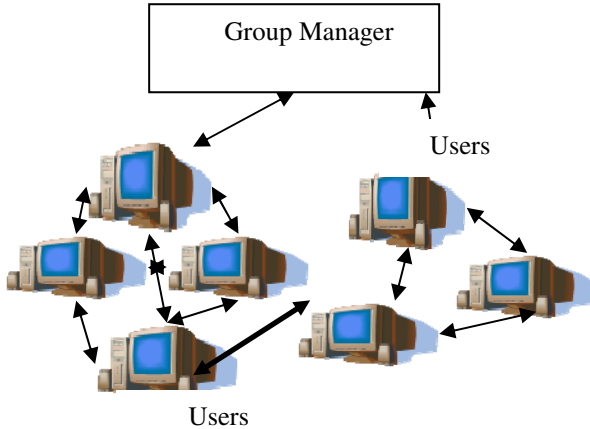


Fig. 1. Group key Derivation

The proposed protocol can be organized as the combination of following algorithms.

**Step 1: System setup:**

In this step, the user  $U_i$  in the organization list, has to get the permission from the cloud service provider ( $U_0$ ) by sending their contribution ( $m_i$ ) to  $U_0$ , who collects all  $U_i$  contributions to computing the master secret value. This is done by performing the following operations.

**a. User  $U_i$  contribution Preparation and signature generation**

Each user  $U_i$  chooses a random message  $(m_i) \in Z_q^*$  and takes the current time stamp  $T$ , formulates  $(ID_i; ID_0; m_i; T)$  and then encrypted using  $U_0$ 's public key, finally sent it to service provider  $U_0$ , i.e

$$e = Pu_0\{ ID_i; ID_0; m_i; T \}$$

Also calculates the signature  $sign_i$  of  $(ID_i; ID_0; m_i; T)$  using its private key (Pri)

$$sign_i = Pri\{ ID_i; ID_0; m_i; T \}$$

Each user  $U_i$  sends  $\{e, sign_i\}$  to the Cloud Service provider ( $U_0$ );  $U_i \rightarrow U_0: \{e, sign_i\}$

**b.  $U_i$  message reception and Verification by  $U_0$**

The  $U_0$  receives all the messages from all the users and decrypts them.  $U_0$  then verifies the signatures of corresponding user  $U_i$ 's. It also checks for the validity of timestamp.  $U_0$  accepts the  $U_i$ 's message if their signature and timestamp are valid.

**c. master secret key generation**

The  $U_0$  collects all the  $x_i$  contributions from the cloud users and then computes a master secret key  $K=XOR(m_1,m_2,-----,m_n)$ . On the input security parameter  $l$ , the  $U_0$  generates a random tuple  $(G_1, G_T, \hat{e}, g, q)$ . The two multiplicative groups  $G_1$  and  $G_T$  with the prime order  $q$ , where  $G_1$  is generated by  $g$  and  $\hat{e}: G_1 \times G_1 \rightarrow G_T$ . The  $U_0$  also selects  $K_1, K_2, -----, K_n \in G_1$ , where  $n$  is number of users are involved in that session for cloud access, a master secret key  $K \in Z_q^*$  and sets  $g=g_1^k$  and also chooses hash functions  $H_i\{0,1\}^* \rightarrow G_1$ . Finally,  $U_0$  publishes the system parameters list  $\pi =(G_1, G_T, \hat{e}, g, q, K, K_1, K_2, ---- K_n, H_1)$ . Where  $n$  is the largest cloud users count that the system can support.

**Step 2: Private Key Extraction:**

In this step, each user  $U_i$  will collect the master secret key  $K$  coming from the  $U_0$  and used it in a private key generation. The user  $U_i$  chooses a random number  $S_i$  and then computes the private key for the group communication  $PR_i= S_i^K$ .

**Step 3: Key Establishment:**

In this stage, the user generates and publishes the messages which will be used in generation of used in generation of group encryption and decryption keys. Without loss of generality, all the participants  $U_1, U_2, --- U_n, n < P$ , A user  $U_i$  holding her private key  $PR_i$  performs the following steps:

- (i) Randomly chooses  $h_i \in G_1, r_i \in Z_q$  and compute  $x_i= g^{-r_i}$  and  $A_i= \hat{e}(h_i, g)$ .
- (ii) For all the users in the group  $1 \leq j \leq n$  he computer  $\beta_{i,j}=h_i * K_j^{r_i}$
- (iii) Generate a signature  $\delta_i$  on  $M_i$  using the private key  $PR_i$ . where
  - a.  $\delta_i= E_{PR_i} \{ M_i \}$  and  $M_i= \{ \Delta_{i,1}, \Delta_{i,2}, -----, \Delta_{i,i-1}, \epsilon, \Delta_{i,i+1}, -----, \Delta_{i,n}, x_i, A_i, ID_i \}$
- (iv) Finally, publish the parameters
  - {  $\Delta_{i,1}, \Delta_{i,2}, -----, \Delta_{i,i-1}, \epsilon, \Delta_{i,i+1}, -----, \Delta_{i,n}, x_i, A_i, \delta_i, ID_i \}$

After completion of this stage, each participant can get the message as shown in the table 1.

In the table 1  $\beta_{i,i}=h_i * g_i^{r_i}$ , only known to  $U_i$  and will not be published to other users in the group.

**Step 4: Common encryption key Derivation:**

Any user in the group can compute the common encryption key  $(X,R)$  using the formulae

$$X = \prod_{i=1}^n x_i ; R = \prod_{i=1}^n A_i = \prod_{i=1}^n \hat{e}(h_i, g) = \hat{e}(\prod_{i=1}^n h_i , g)$$

The common encryption key  $(X,R)$  is accepted only after checking the validity of all users signatures  $\delta_1, \delta_2, \dots, \delta_n$ .

**Table 1.** Messages received by various participants

User	$U_1$	$U_2$	$U_3$	----	$U_n$	All
$U_1 \Rightarrow$	$\beta_{1,1}=\epsilon$	$\beta_{1,2}$	$\beta_{1,3}$	----	$\beta_{1,n}$	$(x_1, A_1, \delta_1, ID_1)$
$U_2 \Rightarrow$	$\beta_{2,1}$	$\beta_{2,2}=\epsilon$	$\beta_{2,3}$	----	$\beta_{2,n}$	$(x_2, A_2, \delta_2, ID_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
$U_n \Rightarrow$	$\beta_{n,1}$	$\beta_{n,2}$	$\beta_{n,3}$	----	$\beta_{n,n}=\epsilon$	$(x_n, A_n, \delta_n, ID_n)$
Key	$d_1$	$d_2$	$d$	----	$d_n$	$(X,R)$

**Step 5: Individual decryption key Derivation:**

All the user  $U_i$ 's can calculate their decryption key( $d_i$ )  $d_i = \prod_{j=1}^n \beta_{i,j}$ , accepts the  $d_i$  if all users signatures  $\delta_1, \delta_2, \dots, \delta_n$  are valid. The attacker cannot compute  $d_i$  since  $\beta_{i,i}$  is not published. After successful completion of this instance last row (decryption key) is generated as output shown in the table. Expand the decryption key corresponding to the group key encryption key, we get  $d_i = \prod_{j=1}^n \beta_{j,i} = \prod_{j=1}^n h_j K_i^{r_j} = \prod_{j=1}^n h_j K_i^{\sum_{j=1}^n r_j}$

**Step 6: Encryption:**

Any selected user wants to store data in the cloud first he has to convert data into polynomial function, and then calculate the interpolation points by giving different x values to the polynomial. All the coordinate points can be encrypted by using the public key(X,R). The cipher text will be stored on the server. So, the data which are passing through the cloud is in cipher format and stored in the server in the encrypted format. Hence, data in storage and data at transit security is achieved. The interpolation point  $m \in G_T$  can be encrypted by the following steps:

- I. Select a random number t
- II. Then compute  $C_1 = g^t, C_2 = X^t, C_3 = m.R^t$  and  $C_4 = H_1(C_1, C_2, m)$
- III. Transmit the ciphertext  $C = (C_1, C_2, C_3, C_4)$  onto the Data storage in the cloud.

**Step 7: Decryption:**

When any user wants to access the data from the server, first, he read the cipher data from the cloud. By using his private key, the cipher text is decrypted to access the original interpolation points. So, during the data access from the server, it is in a cipher text format, an unauthorized person can extract the original until he knows the private key of the user. After finding the interpolation points, polynomial function is formulated by using the Lagrange interpolation [15]. From the polynomial function the group user can finally retrieve the data for further utilization by the user. After extracting the cipher text from the data storage the user can do the following steps to extract the interpolation points.

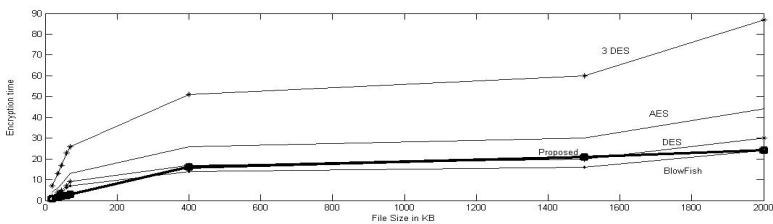
$$\text{Calculate } m = \frac{C_3}{\hat{e}(d_i, C_1) \hat{e}(K, C_2)}$$

## 4 Performance Evaluation and Comparisons

This section compares the performance of different symmetric key cryptographic algorithms with the proposed scheme. In the simulation, we used the pairing based cryptography library run on the laptop using intel® core™ i5-2400 CPU @ 3.10GHz and 2.91 GB of RAM. The security parameter  $L$  is set to be 160, the length of the group element in  $G_1$  and  $G_2$  is 171 bits and 1024 bits respectively. The computation time of encryption is shown in figure 2. Experimentation was done for different file sizes and then measured the time required for encryption. It is observed that, the proposed technique is taking very less time for encryption and decryption. Table 2 shows the security Comparison among various symmetric encryption techniques and proposed algorithm, Based on factors like Key length, Speed, Attacks, Number of Rounds. We observed that the proposed technique is useful in group communications as the computation speed is high when compared to others and allows increased in key length.

**Table 2.** Comparison of various encryption techniques

Parameter	AES	DES	Triple DES	Blow fish	Our Scheme
Algorithm Type	Symmetric	Symmetric	Symmetric	Symmetric	Asymmetric
Key length(in bits)	128,192,256	56	168	28-442	160-1024
No of rounds	9,11,13	16	48	16	3
Attacks	Side channel	Brute force attack	Theoretically Possible	Theoretically Possible	Theoretically Possible
Speed	Fast	Slow	Very slow	Fast	Very fast
Speed depends on size	Yes	-	No	Yes	Yes
Type of key	Key distribution	Key distribution	Key distribution	Key distribution	Key agreement



**Fig. 2.** File Encryption Time

## 5 Conclusion

We have proposed an Authenticated asymmetric key protocol for securing the data, which is stored in the cloud. A client side cloud security is achieved by encrypting the

original document and then stored in the cloud, which provides security against active and passive attacks. Evaluation of our protocol has more computation overhead but provides more security than the existing symmetric cryptographic algorithms. The protocol is well suitable for dynamic cloud group, where the users utilizing the data store may change. For every updating in the users(join/leave), the cloud provider gives new parameters results, from them a new set of keys are generate, results secured cloud storage with dynamic client key pairs.

## References

- [1] Wang, Q., Wang, C., Li, J., Ren, K., Lou, W.: Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing. In: Backes, M., Ning, P. (eds.) ESORICS 2009. LNCS, vol. 5789, pp. 355–370. Springer, Heidelberg (2009)
- [2] Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., Song, D.: Provable Data Possession at Untrusted Stores. In: Proc. 14th ACM Conf. Computer and Comm. Security (CCS 2007), pp. 598–609 (2007)
- [3] Juels, A., Kaliski Jr., B.S.: Pors: Proofs of Retrievability for Large Files. In: Proc. 14th ACM Conf. Computer and Comm. Security (CCS 2007), pp. 584–597 (2007)
- [4] Shacham, H., Waters, B.: Compact Proofs of Retrievability. In: Pieprzyk, J. (ed.) ASIACRYPT 2008. LNCS, vol. 5350, pp. 90–107. Springer, Heidelberg (2008)
- [5] Bowers, K.D., Juels, A., Oprea, A.: Proofs of Retrievability: Theory and Implementation. Report 2008/175, Cryptology ePrint Archive (2008)
- [6] Naor, M., Rothblum, G.N.: The Complexity of Online Memory Checking. In: Proc. 46th Ann. IEEE Symp. Foundations of Computer Science (FOCS 2005), pp. 573–584 (2005)
- [7] Wang, Q., Ren, K., Lou, W., Zhang, Y.: Dependable and Secure Sensor Data Storage with Dynamic Integrity Assurance. In: Proc. IEEE INFOCOM, pp. 954–962 (April 2009)
- [8] Ateniese, G., Pietro, R.D., Mancini, L.V., Tsudik, G.: Scalable and Efficient Provable Data Possession. In: Proc. Fourth Int'l Conf. Security and Privacy in Comm. Networks (SecureComm 2008), pp. 1–10 (2008)
- [9] Wang, C., Wang, Q., Ren, K., Lou, W.: Ensuring Data Storage Security in Cloud Computing. In: Proc. 17th Int'l Workshop Quality of Service, IWQoS 2009 (2009)
- [10] Juels, A., Kaliski, B.: PORs: Proofs of retrievability for large files. In: Proceedings of the 14th ACM Conference on Computer and Communications Security, CCS 2007, pp. 584–597. ACM, New York (2007)
- [11] Amazon.com, Amazon web services (aws), <http://aws.amazon.com/> (accessed on July 12, 2013)
- [12] Zhang, L., Qin, B., Wu, Q., Zhang, F.: Efficient many-to-one authentication with certificateless aggregate signatures. *Computer Networks* 54(14), 2482–2491 (2010)
- [13] Boneh, D., Boyen, X., Goh, E.-J.: Hierarchical Identity Based Encryption with Constant Size Ciphertext. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 440–456. Springer, Heidelberg (2005)
- [14] Sivaranjani, R., Lalitha Bhaskari, D., Avadhani, P.S.: Current Trends in Group Key Management. *International Journal of Advanced Computer Science & Applications* 2, 82–86 (2011)
- [15] Sivaranjani, R., Lalitha Bhaskari, D., Avadhani, P.S.: Secure Message Transmission using Lagrange Polynomial Interpolation and Huffman Coding. *International Journal of Computer Applications* 55(1) (October 2012)

# Analysis of Multilevel Framework for Cloud Security

Vadlamani Nagalakshmi and Vijeyta Devi

Department of Computer Science,  
GITAM University, Andra Pradesh, India  
Vijeyta11aug@gmail.com, ram\_yadav536@yahoo.com

**Abstract.** This article proposes a novel and trusted multilevel security framework for securing cloud resources used in a collaborative deployment. A secure communication protocol is also proposed for communication among the cloud resources from different Cloud Service Users (CSU) and among the non trusted groups. Two level security frameworks are Domain level and cloud service provider (CSP) level. A Domain contains a number of CSUs for the same trusted group. The Domain level security ensures and evaluates the trustworthiness of its individual CSU. The CSP level security ensures and evaluates the trustworthiness of various domains. In the proposed framework, security agents are to be deployed both at the Domain level and at the CSP level to evaluate and maintain the trust To support collaboration user must allow inbound traffics to its own Virtual Machine (VM) from VMs of other (CSU). This can be used by the agents to control access between two VMs belonging to the same domain or in two different security groups. The proposed framework ensures that the trust-level of an entire domain does not fall due to malicious activities of only a small minority of members.

**Keywords:** Cloud service users, Virtual machine, Domain level, CSP Level, security Group Graph, security protocol.

## 1 Introduction

Cloud computing has emerged as a new computing paradigm that allows a programmatic access to the Internet based services. It covers a wide range of services including Infrastructure (IaaS), Platform (PaaS) and Software (SaaS) [1]. IaaS deals with the bare hardware requirements of the Cloud Service Users (CSU) while PaaS provides a development platform and SaaS provides a configurable application itself for the CSU so that he can configure it according to his needs. The deployed environment is self-healing, SLA-driven, multi-tenant, service-oriented, virtualized and scalable. Another important feature of cloud computing is the elasticity i.e. quick ale up and scale down according to the changes in the deployed environment. There are many Cloud Service Providers (CSP) e.g. Amazon, IBM, Microsoft etc. who provide those services in a pay-as-youuse model. As a result, a CSU need not spend large capital towards space, power and administrative knowledge. Also a number of CSUs can collaborate with their rented services to gain maximum benefit out of the cloud. However, none of these benefits of cloud deployment comes free of cost. The

inception of cloud technology has increased many different security threats and risks. The VMs along with the data and services loaded there are exposed to these risks. Traditional methodologies are not enough to adopt for protecting cloud resources as they become obsolete with respect to the ever evolving security threats as well as to avoid data losses in the cloud environment. Security and privacy even with traditional information security systems and networks has been difficult to satisfy and this is also a challenging job for cloud environment [4]. To alleviate those concerns the CSPs have already considered security at various levels. As for example, VM isolation for the hypervisor level whereas security groups for the application level. Again these considerations are not enough when the CSUs use the cloud in a collaborative deployment. This paper aims to propose a novel and trusted security framework for securing cloud resources used in a collaborative deployment considering the secure communication among the cloud resources from different CSUs.



**Fig. 1.** Cloud Computing

## 2 Background Work

Most of today's security models are based on traditional cryptographic approaches. However, we are more concerned about the security of accessing the cloud resources that may need thinking beyond the traditional cryptographic approaches. Some referred to dynamic security measures for a cloud environment [3] while other domain based applications discussed the growing security concerns of cloud infrastructure. In [2] domain trust concept is used to develop a secure cloud infrastructure. However not much work has been done considering the security of the cloud while used in collaborative way. Amazon has introduced the Security Groups (SG) [6] to implement a role based access control for inbound traffics allowed to the VMs. A SG can define access rule for a port to an individual IP address, a range of IP addresses (using CIDR notation), the entire Internet (0.0.0.0/0) or another SG (all VMs launched from this



SG are allowed to access) . A novel approach has been demonstrated in [6] to detect the mis-configurations of SGs using attack graphs and to find an overall vulnerability assessment of a multi-tier infrastructure. An agent based multilevel security approach has been introduced in [5] to control access to cloud resources based on identity and trust. But it doesn't consider the trust measurement within the members of a trusted group or among the members of not trusted groups. Also it does not cover the secure communication among the members of non-trusted groups.

### 3 Cloud Security Framework

A two level cloud security framework has been proposed considering the security at the Domain level and CSP level. A Domain contains a number of CSUs for the same trusted group. The trustworthiness of its individual CSU. The CSP level security ensures and evaluates the trustworthiness of various domains. In the proposed framework, security agents are to be deployed both at the Domain level and at the CSP level to evaluate and maintain the trust. The domain level also deploys a Proxy Server to ensure that all its CSUs must access the cloud resources from the CSP using this Proxy. The Proxy Server consults the corresponding security agent in the domain to check the trust level of the requesting CSU. If the trust value for the CSU is lower than a pre-set threshold, then the request is rejected. This makes sure that a malicious member in a domain is not exposed outside such that the trust value of the domain remains un-affected. The request from a trusted CSU is forwarded to the CSP. The CSP then consults its corresponding security agent to check the trust level of the corresponding Domain. The cloud resource is allowed to be accessed only when the domain has a trust value greater than a pre-defined threshold level. The CSP actively monitors the state of the resource during such access. If the resource is compromised, the CSP consults the corresponding security agents to re-evaluate the trust level of the Domain at the CSP level and the trust level of the CSU at the Domain level. Domain level security ensures and evaluates the

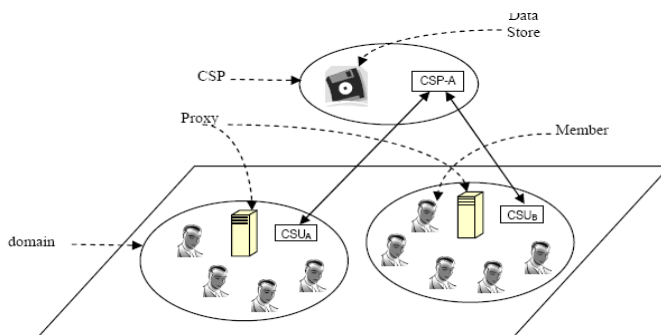
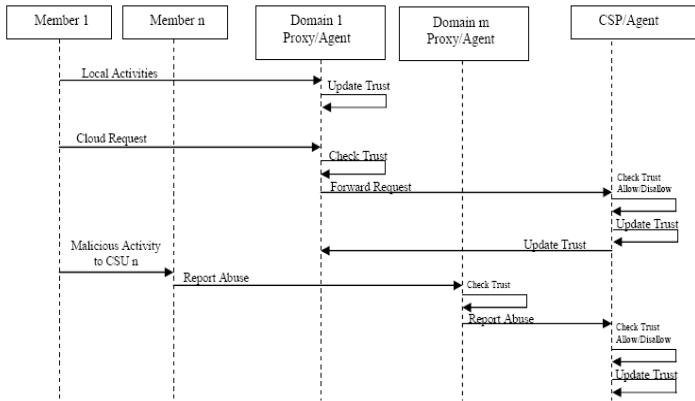
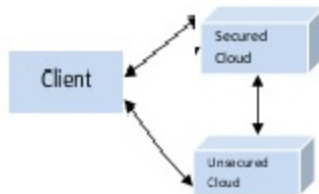


Fig. 2. Secure Cloud Framework



**Fig. 3.** Sequence Diagram of the Framework

In this 2-tier security framework, the trust values of the members in a domain are computed based on their behavior at both the levels. The CSU in a domain would exchange control packets with the domain members in promiscuous mode. The performance of the domain members is to be measured using parameters like packet delivery ratio, transmission delay, etc.



**Fig. 4.** 2-tier security framework

These may be used to compute an initial trust value for each domain member. When a trusted domain member is allowed to operate at the CSP level, the feedback from the agent at CSP to the CSU agent would be used to re-evaluate the trust value of the member node. This 2-tier trust assessment model ensures that the trust value of a node changes dynamically according to its behavior. A node which is found trusted at a time instance  $t_i$  may go down and its trust level may fall below the threshold at some instance. On the other hand, a node gains trust if its behavior is found to be good over a period of time by the corresponding CSU agent in the domain. Let's consider that collaboration exists between the two members  $M_A$  and  $M_B$  from two separate domains A and B. If one of these (say  $M_A$ ) experience a malicious behavior from the other user (say  $M_B$ ), the member  $m_A$  sends a report to its CSU agent, say CSUA, The CSUA looks at the current trust value for  $m_A$ . If the member A is trusted, then the report from  $m_A$  is forwarded to the agent at CSP, i.e., CSP\_A. In either case, the CSU agent for domain A, would start monitoring the behavior of the member  $m_A$

to detect if this act of reporting by mA itself is an attack. The novelty of the architecture is that a domain will not be blocked by a CSP even if the domain contains some non-trusted CSUs. This is because the Domain itself will block those CSUs before the CSP block the domain. A domain will only be blocked when it contains a large of not trusted CSUs.

#### 4 Secured Protocol

Let's consider that a source VM (say, VMS) wants to connect a destination VMS, say VMD through a common port, say P. Various CSPs introduce the notion of Security Groups (SG) that define the rules for inbound traffics allowed to the VMs. A SG can allow access rule for a port to an individual IP address, a range of IP addresses (using CIDR notation), the entire Internet (0.0.0.0/0) or another SG (all VMs launched from this SG are allowed to access) . An important feature of the SG is that they can redefine the rules dynamically. Along with the cloud security framework, we also propose a secure communication protocol that can be incorporated to the proposed framework. The CSP agent would maintain a SG graph that determines the reachability between the various VMs through different ports. However creating access rules across Domains may create some vulnerability without the knowledge of the owner of a Domain. For example, if a SG g1 allow the SG g2 and g2 allow the SG g3, then all VMs launched from g1 will be accesses by the VMs launched from g3. As long as g1 and g3 belongs to the same Domain, it will not create a problem but we have to be concerned if they belong to different Domain. For that reason our secure communication protocol will not allow the creation of such edges inside the security group graph rather it will allow such access on the fly considering the trust of the source. The secure communication protocol will be installed inside the security agents maintained by a CSP and the domain proxy.

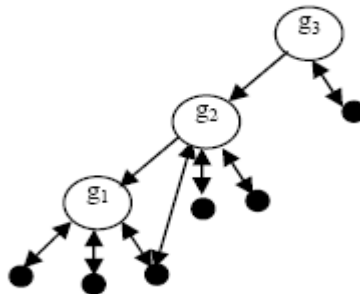


Fig. 5. A Security Group (SG) graph

**The access rule of the protocol is described below:**

**Rule 1:** If source VM and destination VM belongs to the same user allow access.

**Rule 2:** If source VM and destination VM belong to the different user but same domain, consult the security agent at that domain for the trust of the source User. Allow access if the trust value of the source VM is above the threshold.

**Rule 3:** If they belong to the different domain, consult the security agent maintained by the CSP for the trust of the source user. Find a directed path in the security group graph from the source VM to the destination VM if the trust value of the source VM is above the threshold. It may be considered here that accessing a port through VMs is non-transitive, i.e.,

$$VM_A \xrightarrow{P} VM_B \wedge VM_B \xrightarrow{P} VM_C \text{ is not equivalent to } VM_A \xrightarrow{P} VM_C \quad (1)$$

Thus in order to connect VMS with VMD on port P, the intermediate hops need to be SG nodes only. All these nodes including the VMD and VMS must allow access to the target communication port P. If found, allow the access otherwise access through the following

#### Algorithm:

**Step 1:** Authenticate source user by the destination domain.

**Step 2:** Allow the communication port for the source VM to one of the SG belongs to the destination VM (should be carried out by the destination domain).

**Step 3:** Insert an edge from the source VM to the SG for the corresponding port

**Rule 4:** When the trust level of a user goes below a predefined threshold, maintained by the security agent at the CSP, all the edges from the VMs running under the user to the SGs of a different domain are deleted. Also the corresponding permissions are revoked. One major advantage of the solution offered in this paper is that the existing methods for collaborative trust assessment may be reused with very little customization with the proposed security framework.

## 5 Conclusions

In this paper, a security framework for cloud environment is proposed. In applications, the virtual Machine (VMs) are highly exposed to the security breaches. In this paper, a secure communication protocol has also been proposed that tries to reduce these vulnerabilities on the fly. The decisions are taken in a distributive way the agents installed at the cloud service provider (CSP) level as well as the Domain level. In future this research may be extended towards standardizing the creation of the security group graph so that the agents can analyze the graph to find out the possible attack path at a particular instance of time.

## References

1. Ramgovind, S., Eloff, M.M., Smith, E.: The Management of Security in Cloud computing. In: Proc. Information Security for South Asia (ISSA 2010), pp. 1–7. IEEE Press (2010)
2. Wang, C., et al.: Toward Publicly Auditable Secure Cloud Data Storage Services. *IEEE Network* 24(4), 19–24 (2010)
3. Takabi, H., Joshi, J.B.D., Ahn, G.-J.: SecureCloud: Towards a Comprehensive Security Framework for Cloud Computing Environments. In: Proc. 2010 IEEE 34th Ann. Computer Software and Applications Conf. Workshops, pp. 393–398. IEEE Press (2010)
4. Zhou, M., et al.: Security and Privacy in Cloud Computing: A Survey. In: Proc. 6th Int'l Conf. Semantics, Knowledge and Grids, pp. 105–112. IEEE Press (2010)
5. Popovic, K., Hocenski, Z.: Cloud Computing Security Issues and Challenges. In: Proc. 33rd Int'l Convention on Information and Comm. Technology, Electronics and Microelectronics (MIPRO 2010), pp. 344–349. IEEE Press (2010)
6. Morsy, M.A., Grundy, J., Müller, I.: An Analysis of the Cloud Computing Security Problem. In: Proc. 17th Asia Pacific Software Eng. Conf. 2010 Cloud Workshop (APSEC 2010). IEEE Press (2010)
7. Grobauer, B., Walloschek, T., Stöcker, E.: Understanding Cloud-Computing Vulnerabilities. *IEEE Security and Privacy* 9(2), 50–57 (2011)
8. Lua, P., Yow, K.C.: Mitigating DDoS Attacks with Trans- parent and Intelligent Fast-Flux Swarm Network. *IEEE Network* 25(4), 28–33 (2011)
9. Pham, V.H., Dacier, M.: Honeypot Trace Forensics: The Observation Viewpoint Matters. *Future Generation Computer System—Int'l J. Grid Computing and E- Science* 27(5), 539–546 (2011)
10. Phillips, C., Swiler, L.P.: A graph-based system for network-vulnerability analysis. In: Proceedings of the 1998 Workshop on New Security Paradigms (NSPW 1998), pp. 71–79. ACM, New York (1998)
11. Bleikertz, S., Schunter, M., Probst, C.W., Pendarakis, D., Eriksson, K.: Security audits of multi-tier virtual infrastructures in public infrastructure clouds. In: Proceedings of the 2010 ACM Workshop on Cloud Computing Security. ACM, New York (2010) ISBN: 978-1-4503-0089-6

# Agent Based Negotiation Using Cloud – An Approach in E-Commerce

Amruta More<sup>1</sup>, Sheetal Vij<sup>1</sup>, and Debajyoti Mukhopadhyay<sup>2</sup>

<sup>1</sup> Department of Computer Engineering,  
Maharashtra Institute of Technology, Pune- 411038, India

<sup>2</sup> Department of Information Technology,  
Maharashtra Institute of Technology, Pune- 411038, India  
{moreamruta930, sheetal.sh,  
debajyoti.mukhopadhyay}@gmail.com

**Abstract.** 'Cloud computing' allows subscription based access to computing. It also allows storage services over Internet. Automated Negotiation is becoming an emerging, and important area in the field of Multi-Agent Systems in E-Commerce. Multi-Agent based negotiation system is necessary to increase the efficiency of E-negotiation process. Cloud computing provides security and privacy to the user data and low maintenance costs. We propose a Negotiation system using cloud. In this system, all product information and multiple agent details are stored on cloud. Both parties select their agents through cloud for negotiation. Agent acts as a negotiator. Agents have user's details and their requirements for a particular product. Using user's requirement, agents negotiate on some issues such as price, volume, duration, quality and so on. After completing negotiation process, agents give feedback to the user about whether negotiation is successful or not. This negotiation system is dynamic in nature and increases the agents with the increase in participating user.

**Keywords:** Cloud computing, negotiation, multi-agent, E-Commerce.

## 1 Introduction

In business negotiation two or more parties come together to find mutually agreeable contractual decision. For negotiation process both parties must show their interest, thus, negotiation can be complicated and lengthy process. When all parties have to come to final decision, negotiation will be stopped. In negotiation, each individual aim to achieve the best possible outcome for their organization.

Negotiator is an individual representing an organization which listens to all the parties' decision carefully and takes his own decision which gives profit to his organization. In negotiation process organization profit depends on organization's negotiator so that negotiator needs to understand situation and all other organization's negotiator. Negotiator must know how to negotiate well to successfully close deals, avoid conflicts, and establish better relations among the other organization's negotiators making the organization a better place to work. For successful negotiation individuals or negotiator must learn to compromise and stop finding faults in each other.

Today cloud computing is widely used and is becoming a popular technology. Cloud is a remote server, where the user can store their data and access the data remotely whenever it's required. Cloud computing provides security and privacy to the user data. User has no burden in maintaining huge amount of data stored on cloud. Cloud computing is sometimes referred to as "on-demand resources" and is usually based on pay-per-use basis. If business owner requires more space as compared to his previous space on cloud, owner can easily request for additional data storage on the cloud. Also owner can easily request for additional bandwidth, processing speed, and additional licenses. Using cloud computing, user can access information from any device like desktop, minicomputer, mobile etc. anywhere and anytime.

Amazon, Microsoft, Openstack, Google all these are the cloud providers. Google Apps is one example of cloud.

In Agent based negotiation system, we propose a system for negotiation between a provider and a consumer using cloud. In this system, all product information and multiple agent details are stored on the cloud. Both provider and consumer will select their agents through cloud for negotiation. An agent acts as a negotiator. Agent has user's details and their requirements for a particular product. Using user's requirement, agents negotiate on certain features or issues.

## 2 Features of Cloud for E-Negotiation

1. **Rapid Scalability:** Cloud computing has the ability of scaling the resources both ways for the consumers and as per the need. Cloud is infinite and one can buy the computing power as per the need. Negotiation system is dynamic, so that if more data storage is required, it can be easily made available by cloud.
2. **Security:** It is the core feature of cloud computing. Security is much stricter in cloud computing. Data is shared within a server hence, the provider must ensure that each account is secured, and only authorized users in one account can access it. Any product or negotiation process information is stored in a secure manner. Only authorized agents have access to the product information and negotiation process.
3. **No Need of Maintenance:** User can store all types of data on the cloud, and do not have to worry about maintenance of data. The product data can be stored on cloud, so that organizations do not require any server and maintenance of that server. Simultaneously maintenance cost is also reduced.
4. **No Need of Backup:** Business owner do not need to worry about the backup responsibilities, as the supplier has already taken effort to put up a great system for backup. Disk failure, server crash or system failure won't create much problem as the supplier can easily restore the latest backup from the cloud.
5. **Device and Location Independent:** User can access cloud from any device like mobile, mini-computer, and desktop etc. anywhere and anytime.
6. **Transparent Software Updates:** Softwares which are necessary with any e-commerce system, are updated, transparently and require minimum download time.

### 3 Technical Literature Survey

Li Pan [1] introduced a framework for automated service negotiation in cloud computing environments. In this framework, software agents negotiate with each other on behalf of service consumer and provider. This system also used a bilateral multi-step monotonic concession negotiation protocol for service negotiation in cloud computing environments. Service provider and consumer agents interact with each other due to the negotiation process and they make decisions according to the negotiation protocol.

Miguel A. Lopez-Carmona, Ivan Marsa-Maestre and Mark Klein [2] says that, consensus policy based mediation framework is used to perform multi-agent negotiation. This paper also proposed a mediation mechanism which is used to perform the exploration of negotiation space in the multiparty negotiation setting. The performance of mediator mechanism is under guidance of aggregation of agent performance and on the set of alternatives the mediator proposes in each negotiation round.

Mikoto Okumura, Katsuhide Fujita proposed [3], a collaborative park-design support system which is an example of collective collaboration support systems based on multi-agent systems. In this system, agents collect user information, many alternatives and reach optimal decision using automated negotiation protocol. Especially, in this paper, the attribute space and utility space of user in real world is decided. At the end of the system user gives feedback. According to the user's feedback, if most of the users agree on some alternative, then this alternative is final or optimal.

Amir Vahid Dastjerdi and Rajkumar Buyya [4] described SLA negotiation challenges in a cloud computing environment. This system also proposed time dependent negotiation which solves negotiation challenges. To increase the dependability of negotiation process, this system has included reliability assessment. Cloud providers can accommodate more requests and thus increase their profit by discriminating regarding the pattern of concession

Ivan Marsa-Maestre, Miguel A. Lopez-Carmona and Mark Klein [5] presented a framework for characterization and generation of negotiation process. Considering both the structural properties of agent utility functions, and the complexity due to relationships between utility functions of the different agents, a set of metrics to measure high-level scenario parameters is provided. Then a framework is presented to generate scenarios in a parametric and reproducible way. The basis of generator is the aggregation of hyper volumes which is used to generate utility functions. Generator is also based on the use of shared hyper volumes and nonlinear regression which is used to generate negotiation scenarios.

Bo An, Victor Lesser, David Irwin, Michael Zink [6] designed a system for dynamic resource allocation problem and implements a negotiation system. In negotiation model, multiple sellers and buyers are allowed to negotiate with each other concurrently. At the same time an agent is allowed to de-commit from an agreement at the cost of paying a penalty. This system also presents negotiation strategies for both seller and buyer.



Moustapha Tahir Ateib [7] has presented a fuzzy logic based negotiation modeling that can be used to overcome the complexity of automation negotiation processes. This system uses fuzzy logic to deal with ambiguity and uncertainties.

Yan Kong, Minjie Zhang [8] proposed a negotiation-based method which is used for task allocation under time constraints in an open, dynamic grid environment. In this environment, both consumers and provider agents can enter into or, exit the environment freely at any time. There is no central controller so that agents are negotiating with each other for task allocation based only on local views.

Hsin Rau, Chao-Wen Chen, and Wei-Jung Shiang [9] developed a negotiation model which is used for a supply chain with one supplier and one buyer. This model is useful to achieve coordination under incomplete information environment. To find an optimal solution, an objective programming approach is applied.

Liu Xiaowen, Yu Jin [10] introduced automated negotiation model for tourism industry. To improve the negotiation efficiency and success rate, this system proposed RBR and CBR. The model employs CBR method to support an automated negotiation by past successful negotiation cases used for those negotiation partners that have no contract rule existing in each other.

Mukhopadhyay et. al. recently proposed related solutions in negotiation over the Internet for efficient E-Commerce and negotiation prediction. [11] [12]

## **4 Proposed System**

### **4.1 Scope of the System**

Due to cloud computing, negotiation system will be secured, user can access it any time on any device like desktop, mobile etc. and organization maintenance cost is also reduced. Further we can use features like rule based reasoning and case based reasoning [10]. These two features improve the efficiency and success rate of the negotiation process.

### **4.2 Purpose**

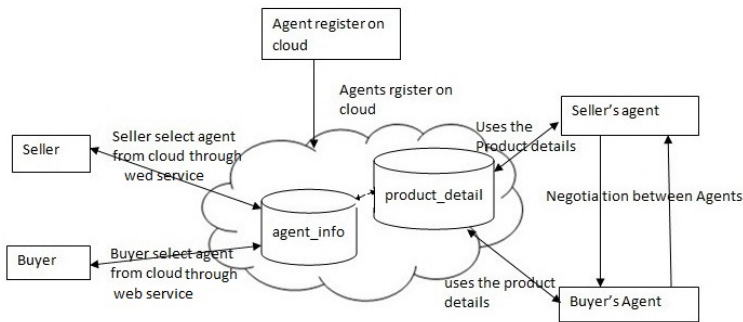
The objective of this system is to reduce maintenance cost of organizing data and provide security for data and negotiation process. So, it makes automated negotiation faster, flexible, secure and reliable.

### **4.3 Problem Definition**

In order to make some agreeable decision, two or more parties come together during the negotiation process. And there are organizations to maintain data of negotiation process and product data. But this maintenance is a very tedious job. In order to overcome this problem, all organizations' product data is stored on cloud. Hence, security and maintenance cost of organizations' data is reduced.

## 5 Proposed System Architecture

In agent based negotiation system, agents are the negotiator. The negotiation process is done by agents through cloud. Cloud is used to store all product details and agents information. Further we go on case based reasoning and rule based reasoning, we will add two more databases in this system.



**Fig. 1.** System Architecture of Agent Based System

If sometimes buyer has time for doing negotiation process but at the same time seller is busy in his/her work. In this situation buyer has to wait until seller is free. For this reason we can use agent based system. In this system, firstly, agents are registered on the cloud. All information related to agents (such as agent name, experience etc.) is stored on cloud database. Cloud database also has product details (for example product of company, price, features of product etc). Seller and Buyer of product selects agent using cloud database for negotiation. Agents have all requirement and details of seller and buyer respectively. Using these requirements agents negotiate. After completing negotiation, respective agents will give feedback to the seller and buyer through cloud.

For this system, we can use Amazon Simple Storage Service (Amazon S3). Amazon S3 is used as huge storage area for the internet. It is designed to make easy development of web-scale computing. Amazon S3 offers a simple web services interface which can be used to store, access and retrieve any amount of data at any time. On Amazon S3, user can write, read, and delete objects containing from 1 byte to 5 terabytes of data each. User can store unlimited number of objects.

### 5.1 System Components

For this system, we can use three modules. Using these components, system becomes easy to use and works efficiently.

1. **Store data on cloud and Agent registration:** For storing data, we can use Amazon S3 service. It makes our system fast, secure and highly reliable. In this component, agent detail and product information is stored on cloud in proper format. Only authorized agent can access product details.

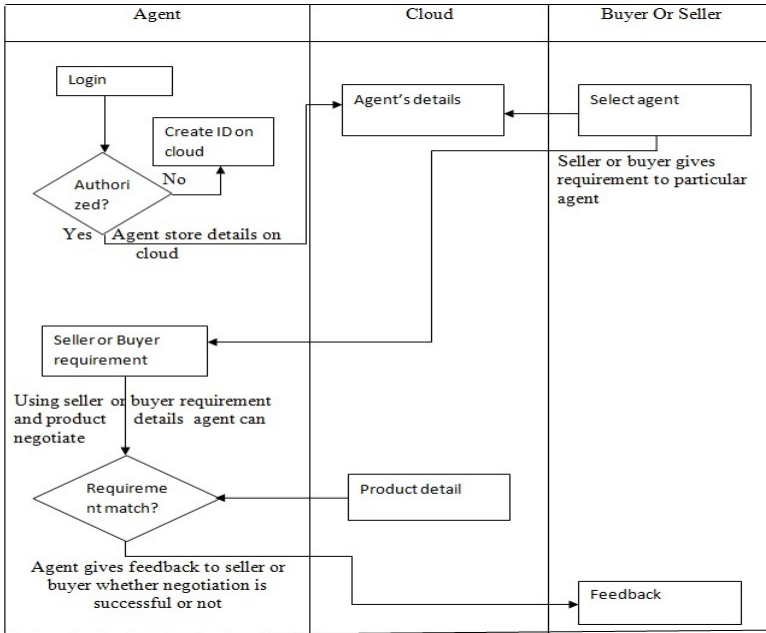


Fig. 2. Flowchart of the System

- Negotiation process:** For negotiation process, seller and buyer select their agents respectively. After that they can give their requirement to agent in encrypted format that is to generate the hash code of that requirement and encrypt that hash code using agent's public key. Agent's public key is known to sellers and buyers.

$$\text{Buyer or Seller Requirements} = E \{ H (m), A_{pk} \}. \tag{1}$$

Where, for generating hash function MD5 algorithm is used.  $A_{pk}$  is agent's public key. We can use encryption for security purpose, in this process, we can use digital signature concept same as seller. Buyer can do same process for generating hash code of his requirement. After getting requirement, agent decrypts the hash code using his private key. And calculates the hash code for checking whether this message comes from appropriate seller or buyer and whether it is modified or not.

$$\text{Agent Receive Requirements} = D \{ H (m), A_{pri} \}. \tag{2}$$

Where,  $A_{pri}$  is agent's private key. When an agent confirms that this message comes from appropriate seller or buyer and message is not modified, then negotiation process will be start.

For negotiation process, we can use The Bilateral Negotiation Model presented by Moustapha Tahir Ateib [7].

Let  $x$  ( $x \in \{x_1, x_2, \dots, x_m\}$ ) represents the buyer agent and  $y$  ( $y \in \{y_1, y_2, \dots, y_n\}$ ) be the supplier agent. And let then  $i$  ( $i \in \{i_1, i_2, \dots, i_n\}$ ) be the issues under negotiation, such as price, volume, duration, quality and so on. Each agent assigns to each issue  $i$ , weight  $W_i$  denoting the relative importance of that issue to the agent. Hence,  $W_i^x$  represents the importance of issue  $i$  to agent  $x$ , therefore the overall utility function of an offer  $O$  is:

$$U(O) = \frac{\sum_{i=1}^m W_i u_i(x_i)}{\sum_{i=1}^m w_i} \quad (3)$$

Where  $U(O)$  is the overall utility for the offer  $O$  ( $= [O_1, \dots, O_m]$ ) and  $u_i(x_i)$  is the individual utility function for issue  $i$  for  $u_i \in [0, 1]$  and the preference degree of an agent to an issue  $i$  is denoted as  $W_i \in [0, 9]$ . Each agent also specifies a minimum acceptable utility level  $[U_{\max}, U_{\min}]$  to determine if an offer is acceptable. Hence, for benefit-oriented, the utility function  $U_i(x_i)$  is computed as follows :

$$U_i(x_i) = \frac{x_i - x_{\{min\}}}{x_{\{max\}}^i - x_{\{min\}}^i} \quad (4)$$

For cost oriented however, the utility function can be written as:

$$U_i(x_i) = 1 - \frac{x_i - x_{\{min\}}}{x_{\{max\}}^i - x_{\{min\}}^i} \quad (5)$$

3. **Feedback:** When negotiation process is finished, agent gives feedback to his appropriate seller or buyer about negotiation whether it is successful or not. Then the actual E-commerce part would start which is not part of our current research.

## 6 Conclusion

Cloud computing provides various features such as security, scalability, reliability and low maintenance which are beneficial to negotiation process in E-Commerce. In this paper, we propose an agent based negotiation system, in which the agent uses cloud for storage of data and product information. Agents negotiate on some issues using the product information and seller's or buyer's requirement. After completing negotiation process, agents give feedback to user about whether the negotiation is successful or not. This negotiation system is dynamic, if number of users is increased, then number of agents also increases automatically. Our future work is to make a faster, secure and flexible E-negotiation agent using rule based reasoning and case based reasoning [10].

## References

1. Pan, L.: Towards A Framework For Automated Service Negotiation. In: Cloud Computing. IEEE (2011) 61284-204-2/11
2. Lopez-Carmona, M.A., Marsa-Maestre, I., Klein, M.: Consensus Policy Based Multi-Agent Negotiation. In: ANAC (2011)
3. Okumura, M., Fujita, K.: Implementation of Collective Collaboration Support System based on Automated Multi-Agent Negotiation. In: ANAC (2011)
4. Dastjerdi, A.V., Buyya, R.: An Autonomous Reliability-aware Negotiation Strategy for Cloud Computing Environments. IEEE (2012) 978-0-7695-4691-9/12
5. Marsa-Maestre, I., Lopez-Carmona, M.A., Klein, M.: A Scenario Generation Framework for Consistent Comparison of Negotiation Approaches. In: ANAC (2011)
6. An, B., Lesser, V., Irwin, D., Zink, M.: Automated Negotiation with Decommitment for Dynamic Resource Allocation in Cloud Computing. In: Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems AAMAS (2010)
7. Ateib, M.T.: Agent Based Negotiation in E-commerce. IEEE (2010) 978-1-4244-6716-7/101
8. Kong, Y., Zhang, M., Luo, X., Ye, D.: A Negotiation Method for Task Allocation with Time Constraints in Open Grid Environments. In: ANAC (2013)
9. Rau, H., Chen, C.-W., Shiang, W.-J.: Development of an Agent-based Negotiation Model for Buyer-supplier Relationship with Multiple Deliveries. IEEE (2009) 978-1-4244-3492-3/09
10. Xiaowen, L., Jin, Y.: Hybrid Approach Using RBR and CBR to Design an Automated Negotiation Model for Tourism Companies. IEEE (2012) 978-0-7695-4853-1/12
11. Deochake, S., Kanth, S., Chakraborty, S., Sarode, S., Mukhopadhyay, V.P.D.: HENRI: High Efficiency Negotiation-based Robust Interface for Multi-party Multi-issue Negotiation over the Internet, pp. 647–652. ACM Digital Library, USA (2012) ISBN 978-1-4503-1185-4
12. Mukhopadhyay, D., Vij, S., Tasare, S.: NAAS: Negotiation Automation Architecture with Buyer's Behavior Pattern Prediction Component, pp. 425–434. Springer, Germany (2012) ISSN 1867-5662, ISBN 978-3-642-31513-8
13. Zheng, X., Martin, P., Powley, W., Brohman, K.: Applying Bargaining Game Theory to Web Services Negotiation. In: IEEE International Conference on Services Computing (2010)

# Improve Security with RSA and Cloud Oracle 10g

Vadlamani Nagalakshmi and Vijeyta Devi

Department of Computer Science,  
GITAM University, Andra Pradesh, India  
Vijeyta11aug@gmail.com, ram\_yadav536@yahoo.com

**Abstract.** This paper is the extended part of “A PROSPECTIVE APPROACH ON SECURITY WITH RSA ALGORITHM AND CLOUD SQL IN CLOUD COMPUTING” in this paper research work is done on RSA Security with cloud Oracle 10g it helps organizations protect private information and manage the identities of people and applications accessing and exchanging that information. cloud Oracle 10g Identity Management solution consists of the Oracle Internet Directory as well as additional security components and services provided by Oracle Application Server 10g, including provisioning, authentication and single sign-on (SSO) to Oracle and RSA are designed to provide the most seamless e-security experience in the market, To satisfy the customer needs from anywhere the information posted by the customer is not maintained in a single site or computer, rather maintained in number of trusted nodes. Simultaneous and faster access by different users from different places is also supported. To get high reliability and availability the data processed by the customer is stored and updated in multiple machines. If any one node gets failed, the other one provides the service. It reduces the costs associated with computing, dynamic resource pools, virtualization, increases the efficiency of computing and high availability.

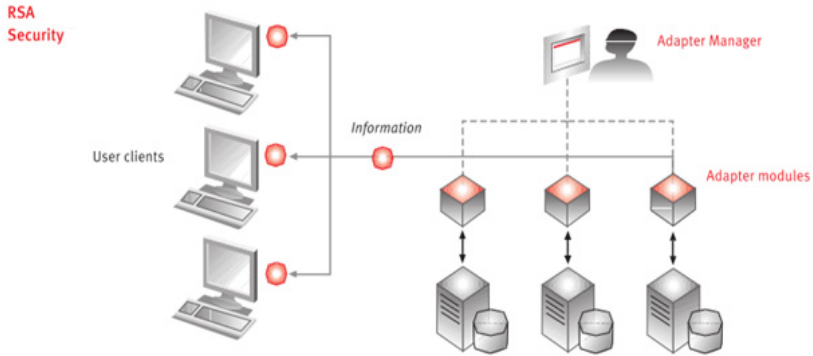
**Keywords:** Single Sign On(SSO), RSA, Oracle Application Server 10g, Internet directory, Virtualization.

## 1 Introduction

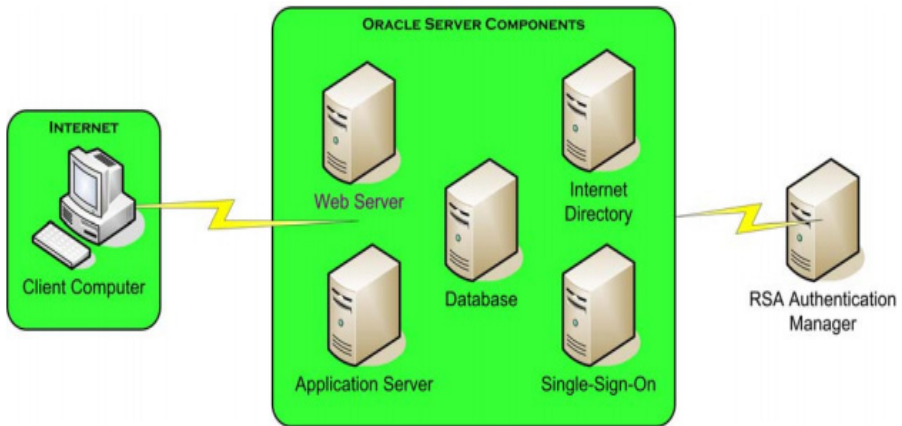
Cloud Computing is the key driving force in many small, medium and large sized companies and as many cloud users seek the services of cloud computing, the major concern is the security of their data in the cloud. Securing data is always of Cloud computing proposes new model for computing and related issues like compute, storage, software. It provides development environment, allocation and reallocation of resources when needed, storage and networking facility virtually. It satisfies the on-demand needs of the user. It facilitates the sharable resources “asa-service” model. For the organization, the cloud offers data centers to move their data globally. It eliminates the responsibility of local nodes for maintaining their data and also

cloud supports customizable resources on the web. Cloud Service Providers maintains computing resources and data automatically via software. Data security is an important aspect of quality of service. As a result, security must be imposed on data by using encryption strategies to achieve secured data storage and access. Because of opaqueness nature of cloud, it is still having security issues. The cloud infrastructure even more reliable and powerful than personal computing, but wide range of internal, external threats for data stored on the cloud. Since the data are not stored in client area, implementing security measures cannot be applied directly. In this work, we implement RSA algorithm before storing the sensitive data in cloud. When the authorized user request the data for usage then data decrypted and provided to the user. Oracle Cloud supports multi-tenant infrastructure in which, contents can be pushed in a short iteration cycle., Whenever new features introduced then automatically reflected in the browser by refreshing it. Additional functionalities released in small sized chunks, this leads to reduce the change management hurdles. Oracle provides support for cloud computing and it has been updated periodically in order to meet the customers current needs after getting feedback and usage statistics from millions of customers. In order to satisfy the customer needs from anywhere the information posted by the customer is not maintained in a single site or computer, rather maintained in number of trusted nodes. Simultaneous and faster access by different users from different places is also supported by oracle. To get high reliability and availability the data processed by the customer is stored and updated in multiple machines. If any one node gets failed, the other one provides the service. cloud oracle 10g is very easy to use and not requiring any other software

In this, we propose a way of implementing RSA with cloud oracle 10g to guaranty the data storage security in cloud. To achieve cloud Oracle 10g, the RSA Authentication Agent for the Web is installed on the Oracle web server, and identical usernames are added to both the infrastructure and portal user repositories. The agent is then configured to protect all Oracle resources. Oracle 10g Single-SignOn server establishes the identity of the user by using the RSA Cookie API to parse the RSA Authentication Agent Web cookie to serve personalized content. This allows strong 2-factor authentication to protect both Oracle and other third party resources, while requiring users to re-authenticate. This approach can be either implemented by the party who stores his data or by the service provider. Hence forth, concerns regarding data privacy and security are proving to be a barrier to the broader uptake of cloud computing services. Every cloud service(s) seeker either an individual or a company should ask the right questions to the cloud provider before hosting their data or applications on the cloud. As many companies move their data to the cloud the data undergoes many changes and there are many challenges to overcome. To be effective, cloud data security depends on more than simply applying appropriate data security procedures and countermeasures. Computer based security measures mostly capitalizes on user authorization and authentication.



**Fig. 1.** RSA Security Manager



**Fig. 2.** RSA with Oracle Authentication

## 2 Problems and Challenge of RSA and Oracle 10g

As enterprises put more applications on the Internet, they are facing increased risk of security breaches more users and more applications are the primary cause of the escalation of security breaches experienced by customers. Furthermore, customers are concerned about the cost of securing this growing problem. Consider the following:

- 16 minutes per day are spent by the average employee authenticating to the various applications needed. This alone is a 3 – 4% productivity loss per employee.
- A single password reset could cost between US\$14 and US\$25.
- The creation of one account for a new employee takes anywhere between 5 and 24 hours.



### 3 Solution Cloud Oracle and RSA

Cloud Oracle and Cloud RSA Security together offer multiple integration points designed to further tighten security and reduce cost of managing users and their access rights.

**Step1** • RSA ClearTrust software is certified with Cloud Oracle10g Application Server for Web-based access management.

**Step 2** • cloud Oracle Internet Directory, compliant directory or cloud Oracle10g Application Server, has been certified as a secure repository for digital certificates created and managed by RSA Keon software.

**Step3** • cloud Oracle Database securely stores digital certificates and works hand-in-hand with Oracle Internet Directory to ensure 100% secure environment for your authentication.

**Step4** • Cloud Oracle and RSA Security integration is optimized to work efficiently in customer environment.

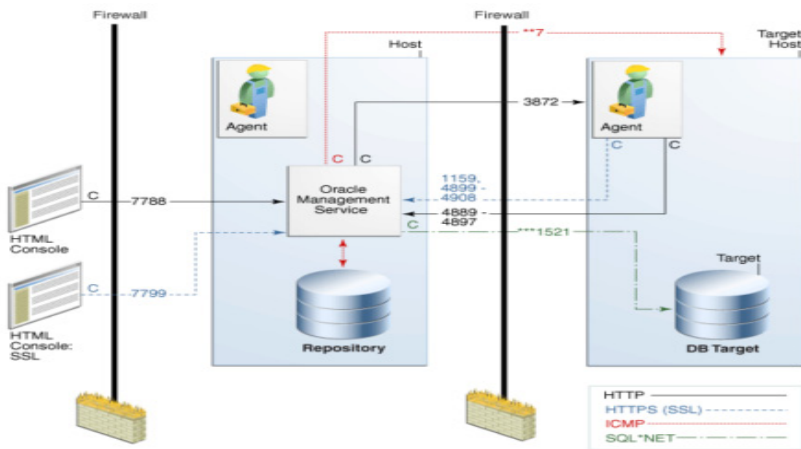


Fig. 3. Oracle Services & Protocols

## 4 User Interface by Using J2ee and Oracle 10g in Apache 1.3 Steps Are

### 4.1 RSA Authentication Agent Steps

**Step1.** The RSA Authentication Agent for Apache 1.3 must be installed to protect the Oracle web servers. Prior to beginning the installation, stop all Oracle infrastructure services by running the command “runstartupconsole.sh stop all” or use enterprise manager to stop the services Then proceed with the installation of the RSA Authentication Agent.

**Step2.** After completing the installation, copy the jar files included in the RSA SecurIDIntegrationModuletothe<ORACLE\_HOME>/j2ee/OC4J\_SECURITY/applications /sso/web/WEB-INF/lib directory. If you're using RedHat, copy the librsacookieapi.so file to <ORACLE\_HOME>/lib.

**Step3.** Edit the <ORACLE\_HOME>/sso/conf/policy.properties. Change the appropriate authentication plug-in to SSO SecurIDAuth.

**Step4.** Restart the Oracle infrastructure servers using “runstartupconsole.sh starts all” or the enterprise manager.

**Step5.** Run config from the RSA Authentication Agent for Web installation directory. During configuration, be sure to protect the URI for the logon area of the desired areas. For example, protect the logon button for the portal server or the /pls/orasso area of the infrastructure server.

## 4.2 Cloud Oracle 10g Via RSA Securid Authentication Agent Steps

**Step 1.** Once the resource has been protected by the web agent, users attempting to access that resource will be challenged to authenticate via RSA SecurID Authentication. For example, an Oracle user attempting to access a restricted intranet portal would see the following page.

### CASE Tool.(Sample)



Fig. 4. New User Authentication

### CASE Tool (Sample)

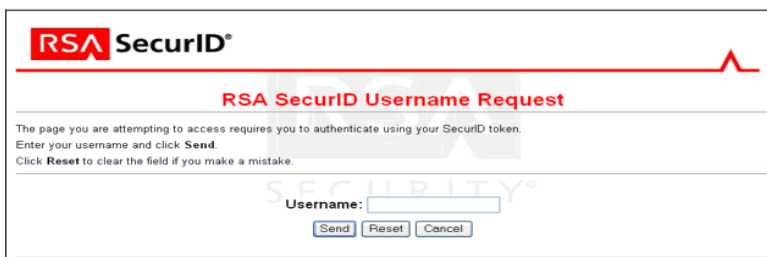
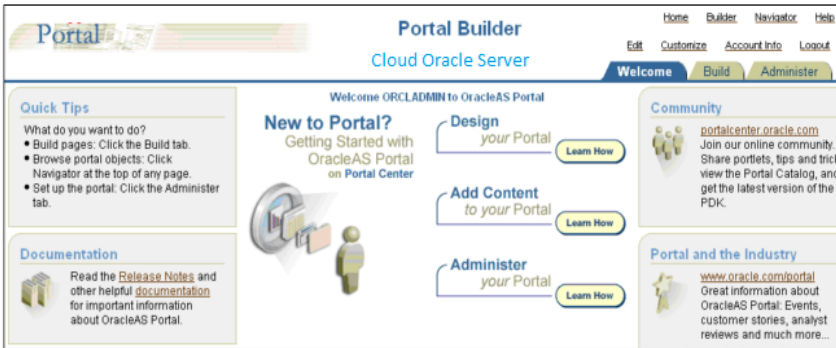


Fig. 5. Output page Restricted intranet portal

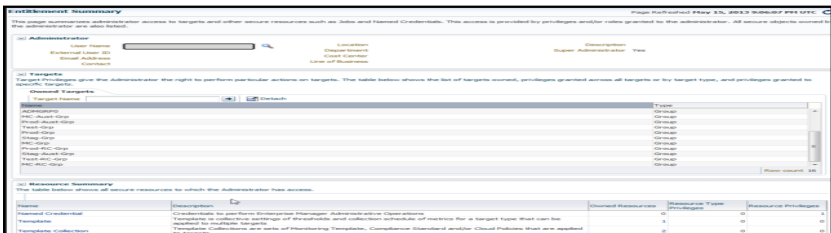
**Step 2.** After the user successfully authenticates, they are passed on to their destination. The Cloud Oracle SSO server will then notice the agent’s cookie and use it to identify and authenticate the user. The user is then redirected to his/her destination.

**CASE TOOL(SAMPLE)**



**Fig. 6.** Output Cloud Oracle Portal page

**CASE TOOLS (Sample)**



**Fig. 7.** Administrators Entitlement Page

**5 Conclusion**

In this we proposed two method for security and and cost saving 1. **RSA Authentication Agent** and 2. **cloud oracle 10g via RSA SecurID authentication agent** its provides the cloud lifecycle management solution. Oracle creates business value from IT by leveraging the built-in management capabilities of the Oracle stack for traditional and cloud environments, allowing customers to achieve unprecedented efficiency gains while dramatically increasing service levels. RSA & cloud oracle given its broad footprint in the data center. Oracle Cloud provides a robust set of security features and capabilities starting with secure framework level communication to a secure user model. Wherever possible, we attempt to incorporate best practices learned in the field into the product itself while balancing ultimate security and usability.

## References

1. Ramgovind, S., Eloff, M.M., Smith, E.: The Management of Security in Cloud computing. In: Proc. Information Security for South Asia (ISSA 2010), pp. 1–7. IEEE Press (2010)
2. Wang, C., et al.: Toward Publicly Auditable Secure Cloud Data Storage Services. *IEEE Network* 24(4), 19–24 (2010)
3. Takabi, H., Joshi, J.B.D., Ahn, G.-J.: SecureCloud: Towards a Comprehensive Security Framework for Cloud Computing Environments. In: Proc. 2010 IEEE 34th Ann. Computer Software and Applications Conf. Workshops, pp. 393–398. IEEE Press (2010)
4. Zhou, M., et al.: Security and Privacy in Cloud Computing: A Survey. In: Proc. 6th Int'l Conf. Semantics, Knowledge and Grids, pp. 105–112. IEEE Press (2010)
5. Popovic, K., Hocenski, Z.: Cloud Computing Security Issues and Challenges. In: Proc. 33rd Int'l Convention on Information and Comm. Technology, Electronics and Microelectronics (MIPRO 2010), pp. 344–349. IEEE Press (2010)
6. Morsy, M.A., Grundy, J., Müller, I.: An Analysis of the Cloud Computing Security Problem. In: Proc. 17th Asia Pacific Software Eng. Conf. 2010 Cloud Workshop (APSEC 2010). IEEE Press (2010)
7. Grobauer, B., Walloschek, T., Stöcker, E.: Understanding Cloud-Computing Vulnerabilities. *IEEE Security and Privacy* 9(2), 50–57 (2011)
8. Lua, P., Yow, K.C.: Mitigating DDoS Attacks with Transparent and Intelligent Fast-Flux Swarm Network. *IEEE Network* 25(4), 28–33 (2011)
9. Pham, V.H., Dacier, M.: Honeypot Trace Forensics: The Observation Viewpoint Matters. *Future Generation Computer System—Int'l J. Grid Computing and E- science* 27(5), 539–546 (2011)
10. Phillips, C., Swiler, L.P.: A graph-based system for network-vulnerability analysis. In: Proceedings of the 1998 Workshop on New Security Paradigms (NSPW 1998), pp. 71–79. ACM, New York (1998)

# Negotiation Life Cycle: An Approach in E-Negotiation with Prediction

Mohammad Irfan Bala<sup>1</sup>, Sheetal Vij<sup>1</sup>, and Debajyoti Mukhopadhyay<sup>2</sup>

<sup>1</sup> Department of Computer Engineering,  
Maharashtra Institute of Technology, Pune 411038, India

<sup>2</sup> Department of Information Technology,  
Maharashtra Institute of Technology, Pune 411038, India  
{mirfan508, sheetal.sh, debajyoti.mukhopadhyay}@gmail.com

**Abstract.** With the exponential increase in the use of web services it has become more and more important to make the traditional negotiation process automated and intelligent. Various tactics have been given till date which determines the behavior of the software agents in the negotiation process. Here we have given lifecycle of the negotiation process and presented a custom scenario to understand it better. Recently the active area of research has been prediction of partner's behavior which enables a negotiator to improve the utility gain for the adaptive negotiation agent and also achieve the agreement much quicker or look after much higher benefits. In this paper we review the various negotiation methods and the existing architecture. Although negotiation is practically very complex activity to automate without human intervention we have proposed architecture for predicting the opponents behavior which will take into consideration various factors which affect the process of negotiation. The basic concept is that the information about negotiators, their individual actions and dynamics can be used by software agents equipped with adaptive capabilities to learn from past negotiations and assist in selecting appropriate negotiation tactics.

**Keywords:** Electronic negotiation, decision functions, agent negotiation, neural networks.

## 1 Introduction

Negotiation is a form of interaction in which a group of agents, with conflicting interests and a desire to cooperate try to come to a mutually acceptable agreement on the division of scarce resources. These resources do not only refer to money but also include other parameters like product quality features, guarantee features, way of payment, etc. The tremendous successes of online auctions show that the dynamic trade based on e-negotiation will gradually become the core of e-commerce. Whether it is a case of B2B purchase or a case of online shopping [11], it is important to make the traditional negotiation pricing mechanism automated and intelligent. The

automation saves human negotiation time and computational agents are better at finding deals in combinatorial and strategically complex settings.

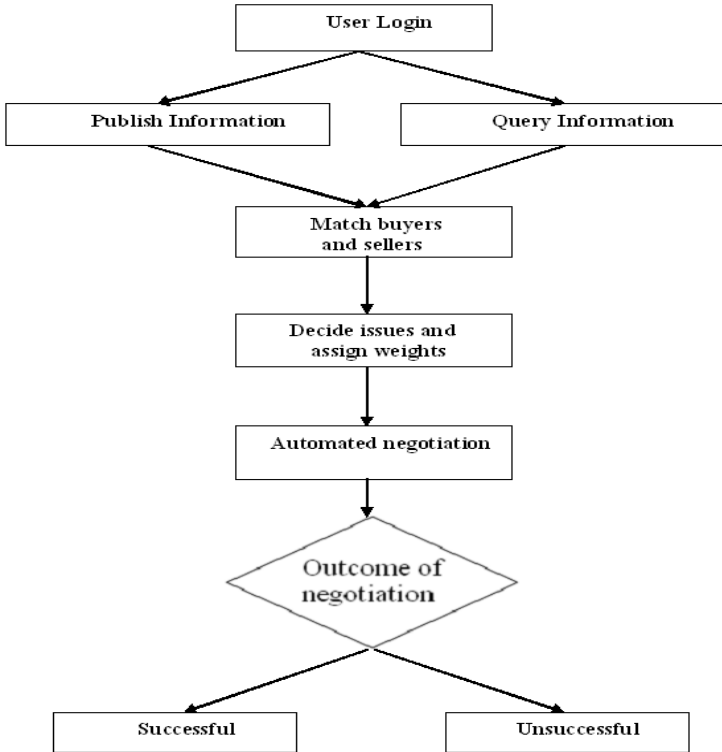
Traditionally e-negotiation processes have been carried out by humans registering at certain web pages, placing bids, making offers and receiving counter offers from other participants. One major disadvantage with this way of e-negotiation is that the knowledge and experience is kept within the human minds [11]. Agent mediated negotiations have received considerable attention in the field of automated trading. However various problems are faced by the negotiation agents such as limited and uncertain knowledge and conflicting preferences. Also agents may have inconsistent deadline and partial overlaps of zones of acceptance [13]. Moreover, multilateral negotiations are more complicated and time consuming than bilateral negotiations. These factors make it difficult to reach consensus.

The need is that the agents should be equipped with a decision making mechanism which allows them to adapt to the behavior of the negotiation partner [3]. Intelligent systems for negotiation aim at increasing the negotiators abilities to understand the opponent's needs and limitations. This ability helps to predict the opponent's moves which can be a valuable tool to be used in negotiation tasks. Various negotiation strategies have been proposed which are capable of predicting the opponent's behavior. The research presented here focuses on the online prediction of the other agent's tactic in order to reach better deals in negotiation. While the extensive coverage of all the prediction methods employed in negotiation is beyond the scope of the current work, it is useful to mention several key studies. In this paper we are also proposing a new architecture for prediction of opponent's behavior.

## 1.1 Negotiation Life Cycle

A negotiation model consists of three main elements: Negotiation protocols, negotiation objects and agent's decision making model. The relative importance of these elements may vary according to the negotiation and environmental context. Once the negotiation model is complete we need to decide what the agents will exchange with each other in the course of negotiation. For buyer and seller model, the objects of negotiation are offers and counter offers over a set of issues.

Figure 1 shows the flow of the negotiation system. The seller registers itself with a well known registration center and advertises itself making it visible to buyers. Interested buyers will look for the sellers of their interest. Once the buyers and sellers are matched, their respective agents should mutually decide the issues of conflict over which negotiation will take place. Issues are rarely viewed as equally important and the difference in their importance can be realized by assigning weights to each issue. Higher the importance of issue higher is its weight and the sum of weights of all the issues should be 100. Also a limit for the value of each issue is assigned. This value acts as a deadline and the negotiation will terminate if this deadline is exceeded for any of the issues.



**Fig. 1.** Life cycle of our E- negotiation system

Negotiation model also requires a utility function which will evaluate the offers and assign ratings to each offer. This rating is used to measure the improvement in current offer as compared to previous offer. Then we begin with the actual negotiation where the agents will exchange offers and counter offers. Each offer should contain a value for all the issues within the specified limits and we should ensure that the rating for any offer should be better than the previous offer. This exchange of offers and counter offers continues till agreement is reached or deadline of any of the issues is exceeded resulting in successful or unsuccessful negotiation.

## 2 Related Work

Predicting the agent's behavior and using those prediction results to maximize agents own benefits is one of the crucial issues in the negotiation process. It is necessary for an agent to produce offers based on his own criteria because an agent has limited computational power and incomplete knowledge about opponents. Various approaches [1,2,10,15,16,18] have been proposed for predicting the opponent's negotiation behavior. We reviewed some of the approaches to come up with certain conclusions regarding the efficiency of each approach and their short comings.

Initially game theory was used in the negotiation process. It treats negotiation as a game and the negotiation agents are treated as players of the game. Zeng and Sycara [9] used game-theoretic approach with Bayesian belief revision to model a negotiation counterpart. However game theory has two main drawbacks [1] which make it unsuitable for use in the negotiation process. First is that it assumes the agent has infinite computational power and secondly it assumes all the agents have common knowledge. These limitations of the game theory were overcome by the decision functions.

Faratin [18] proposed a bilateral negotiation model in which the two parties negotiate on an issue like price, delivery time, quality etc. The two parties adopt opposite roles (buyer and seller) and use one of the three families of negotiation tactics namely: Time dependant tactics, Resource dependant tactics and behavior dependant tactics. The offers exchanged between the agents are represented as  $x_{a \rightarrow b}^t$ . This is the offer generated by agent 'a' for agent 'b' at time 't'. All the offers are restricted in between  $\min^a$  and  $\max^a$  which specifies the range of all possible offers of 'a'. Each agent has a scoring function  $V^a$  which assigns a score to each offer produced. A sequence of alternating offers and counter offers by the agents is called negotiation thread. An agent may respond to the offer by any of the three ways: withdraw, accept or offer

$$response^a(t^n, x_{b \rightarrow a}^{t_{n-1}}) = \begin{cases} withdraw(a, b) & \text{if } t^n > t_{max}^a \\ accept(a, b, x_{b \rightarrow a}^{t_{n-1}}) & \text{if } V^a(x_{b \rightarrow a}^{t_{n-1}}) > V^a(x_{a \rightarrow b}^{t_n}) \\ offer(a, b, x_{a \rightarrow b}^{t_n}) & \text{otherwise} \end{cases}$$

$x_{a \rightarrow b}^{t_n}$  is the counter offer generated by agent 'a' in response to the offer  $x_{b \rightarrow a}^{t_{n-1}}$  of agent 'b'.  $t_{max}^a$  is the deadline for agent 'a' by which the negotiation should be complete.

Offers generated use one of the three families of tactics [18]. In time dependant tactics time is the predominant factor and each offer generated depends on the amount of time remaining and amount of time already consumed. In resource dependant family of tactics offers depend on how a resource is being consumed. Offers become more and more cooperative as the quantity of the resource diminishes. In behavior dependant family of tactics agent imitates the behavior of the opponent. These tactics differ depending on the behavior of the opponent they imitate and to what degree.

Time dependant tactics are further divided into three types [12,18] depending on how quickly the agent starts to concede to the opponent's demands. In *boulware* an agent does not concede until near the deadline. In *conceder* an agent starts giving ground fairly quickly and in *linear* an agent concedes same amount in each round. Similarly behavior dependant family of tactics is also divided into three types: relative tit-for-tat, relative absolute tit-for-tat and average tit-for-tat. In relative tit-for-tat offers produced imitate the opponent's behavior in previous offers in terms of percentage. Random absolute tit-for-tat is similar to relative tit-for-tat except that the behavior is imitated in absolute terms. In average tit-for-tat concession offered is averaged over the previous offers of the opponent. The above given figure shows the various curves where each curve represents different tactics of time dependant family.



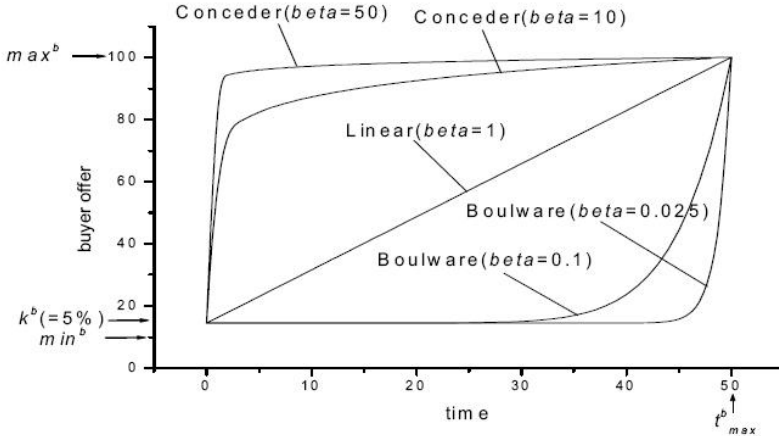


Fig. 2. Concession curves in time dependant family of tactics

Chongming Hou [1] proposed to use non linear regression approach for the prediction of the opponent’s tactics. It could predict the approximate value of opponent’s deadline and reservation values. The performance of the agent improved by using this approach as it reduced the number of negotiation breakdowns and caused early termination of unprofitable negotiations. But this approach is restricted for bilateral negotiations only and can be used only when the agent is sure that the opponent is using one of the above mentioned families of tactics for negotiation.

Many other prediction approaches have been proposed which are based on machine learning mechanism. Most of the work devoted to the learning approach is focused on learning from previous offers i.e. offline learning. They require training data and such agents need to be trained in advance. However this approach may not always work well for the agents whose behavior has been excluded from the training data. Also such data may not be always available.

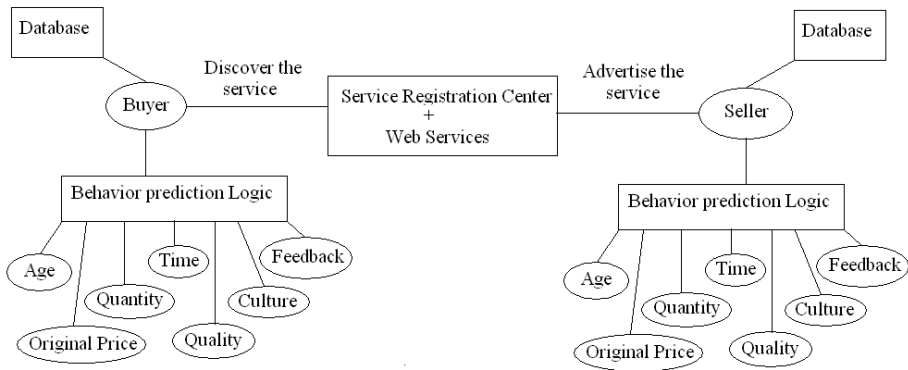
Brzostowski and Kowalczyk [10] presented a way to estimate partners’ behaviors by employing a classification method. They used a decision making mechanism which allows agents to mix time-dependant tactics with behavior dependant tactics using weights which can result in quite complex negotiation behavior. However this approach only works for the time dependent agent and the behavior-dependent agent, which limits its application domains. Gal and Pfeffer presented a machine learning approach based on a statistical method [14,17]. The limitation of this approach is the difficulty of training the system perfectly. Therefore, for some unknown kind of agents whose behaviors are excluded in the training data, the prediction result may not reach the acceptable accuracy requirements.

I. Roussaki, I. Papaioannou, M. Anagnostou [13] proposed an approach based on learning technique which has been employed by Client Agents and uses a feed-forward back-propagation neural network with a single output linear neuron and three hidden layer’s neurons. These neural networks require minimal computational and storage resources making it ideal for mobile agents. The agents use a fair relative tit-for-tat negotiation strategy and the results obtained were evaluated via numerous

experiments under various conditions. The experiments indicated an average increase of 34% in reaching agreements [13]. This approach has excellent performance when the acceptable interval of the negotiation issue overlaps irrespective of the concession rate. On the other hand if the acceptable intervals' overlap is limited and the deadline is quite high, this approach is likely to fail.

### 3 Proposed Architecture

We are proposing the architecture of behavior prediction module in the form of web services as depicted in Figure 3. It has already been established in [4] that providing negotiation as a service (NaaS) is a completely innovative application model of software which provides services through internet. Its benefits are: we can obtain stable visiting quantity, user need not concern about maintenance and upgrade of system as it is done on the server independently, saving human and material resources, automated negotiation system can make use of the existing basic facilities provided by e-commerce platform i.e. security, authentication, transaction management etc. ,saving costs of development.



**Fig. 3.** Proposed Architecture for behavior prediction

**Working:** The seller will advertise itself through a well known service registration center which will make it visible to all the interested buyers. All the available services at any point of time are stored in the service registration center. A buyer looking for some product will query the service registration center to discover the product of his interest. Once the preferences are matched, buyer and seller will directly communicate with each other and start negotiation. Each buyer and seller has its own module for behavior prediction. Complexity of the behavior prediction module may vary depending on the number of issues taken into consideration. Here we have taken seven issues into consideration during prediction although the number of issues may increase or decrease with corresponding increase or decrease in the complexity of the behavior prediction module. Here we assume that an agent can use one of the strategies from a set of pre-decided strategies and our prediction system will predict

the strategy used by opponent and then try to manipulate the offers so as to gain higher benefit. Also the system will initially work for only few real life situations which can be later extended depending on the success of the system.

The behavior prediction logic will use artificial neural networks which have been proved to be universal approximators when provided with sufficient hidden layer neurons and assuming that the activation function is bounded and non-constant. Neural networks also possess the abilities of being self adaptive and self learning. All the issues included in the behavior prediction logic are given as input to the prediction logic system and each issue is assigned some weight depending on its importance in the process of negotiation. Some of the issues like quality are subjective while some are continuous like age. Such issues should be categorized first to make the process of behavior prediction easy. Example: Instead of using the continuous values for 'age', it should be grouped as youth, middle-aged and old for age group of [10-25], [25-50], [50+] respectively. The architecture shown is for bilateral negotiations. However it can be extended to support multi lateral negotiations where each pair of agents has similar architecture in between them.

#### **4 Conclusion and Future Work**

This work reviews the various methods used for predicting the opponent's behavior and then proposes architecture for behavior prediction using artificial neural networks. It proposes the use of database for storing the results and suggests various issues that can be taken into consideration while predicting the opponent's behavior. The proposed intelligent agent based architecture is for bilateral negotiations and may be extended in future to multi lateral negotiations. The given architecture is for general use and may not produce optimal results in all situations. So a situation specific architecture is required in every case of negotiation, where the negotiation issues are selected accordingly. In future we would be making the system to simulate above architecture with the application of agent's behavior in web based negotiation. We plan to test it vigorously and do the necessary comparative study and analysis with above mentioned related systems which we have already studied as technical literature survey. We can also extend our research in the direction of multilateral negotiations after successful completion of bilateral system.

#### **References**

1. Hou, C.: Predicting agents' tactics in automated negotiation. In: Proc. IEEE/WIC/ACM Int'l Conf. Intelligent Agent Technology (IAT 2004), pp. 127–133 (2004)
2. Beheshti, R., Mozayani, N.: Predicting opponents offers in multi-agent negotiations using ARTMAP neural network. In: Second International Conference on Future Information Technology and Management Engineering, FITME 2009, pp. 600–603 (2009)
3. Carbonneau, R., Kersten, G.E., Vahidov, R.: Predicting opponent's moves in electronic negotiations using neural networks. *Expert Systems with Applications: An International Journal* 34(2) (2008)

4. Mukhopadhyay, D., Vij, S., Tasare, S.: NAAS: Negotiation Automation Architecture with Buyers Behavior Pattern Prediction Component. In: Meghanathan, N., Nagamalai, D., Chaki, N. (eds.) *Advances in Computing & Inform. Technology*. AISC, vol. 176, pp. 425–434. Springer, Heidelberg (2012)
5. Ren, F., Zhang, M.: Prediction of partners behaviors in agent negotiation under open and dynamic environments. In: *Proceedings of International Conferences on Web Intelligence and Intelligent Agent Technology*, pp. 379–382 (2007)
6. Rau, H., Chen, C.-W., Shiang, W.-J., Lin, C.J.: Develop an adapted coordination strategy for negotiation in a buyer-driven E-marketplace. In: *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, pp. 3224–3229 (2008)
7. Zulkernine, F.H., Martin, P.: An adaptive and intelligent SLA negotiation system for web services. *IEEE Transactions on Service Computing* 4, 31–43 (2011)
8. Haim, G., Kraus, S., Blumberg, Y.: Learning human negotiation behavior across cultures. In: *Second International Working Conference on Human Factors and Computational Models in Negotiation* (2010)
9. Zeng, D., Sycara, K.: Bayesian learning in negotiation. *International Journal of Human-Computer Studies* 48, 125–141 (1998)
10. Brzostowski, J., Kowalczyk, R.: Predicting partner's behaviour in agent negotiation. In: *Proc. Int'l Joint Conf. Autonomous Agents and Multiagent Systems*, pp. 355–361 (2006)
11. Mukun, C.: Multi-agent automated negotiation as a service. In: *7th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6 (2010)
12. Lin, R., Kraus, S.: *Magazine communications of the ACM*, vol. 53(1) (January 2010)
13. Roussaki, I., Papaioannou, I., Anagnostou, M.: Employing neural networks to assist negotiating intelligent agents. *2nd IET International Conference on Intelligent Environments* 1, 101–110 (2006)
14. Park, S., Yang, S.-B.: An automated system based on Incremental learning with applicability toward multilateral negotiations. In: *SICE-ICASE International Joint Conference*, pp. 6001–6006 (2006)
15. Liu, N., Zheng, D., Xiong, Y.: Multi-agent negotiation model based on RBF neural network learning mechanism. In: *International Symposium on Intelligent Information Technology Application Workshops*, pp. 133–136 (2008)
16. Jazayeriy, H., Azmi-Murad, M., Sulaiman, M.N., Udzir, N.I.: A review on soft computing techniques in automated negotiation. *Academic Journals for Scientific Research and Essays* 6(24), 5100–5106 (2011)
17. Li, B., Ma, Y.: An auction-based negotiation model in intelligent multi-agent system. In: *International Conference on Neural Networks and Brain*, vol. 1, pp. 178–182 (2005)
18. Faratin, P.: Automated service negotiation between autonomous compositional agents. PhD thesis, Queen Mary & Westfield college, University of London, UK (2000)
19. Mukhopadhyay, D., Vij, S., Bala, M.I.: Automated Negotiation And Behavior Prediction. *International Journal of Engineering Research & Technology* 2(6), 1832–1838 (2013)

# Dynamic Scheduling of Requests Based on Impacting Parameters in Cloud Based Architectures

R. Arokia Paul Rajan and F. Sagayaraj Francis

Pondicherry Engineering College, Pondicherry, India  
{paulraajan, fsfrancis}@gmail.com

**Abstract.** This paper focuses on the request assignment problem in distributed storage system with the limited resources which is a challenging issue. The pertinent constraints and variants to be considered for devising a solution are identified. Assignment of requests to the storage servers should be continuously monitored with the existing data and should be reconfigured when there is a change in the parameters that are observed. Thus assignment of users' requests – monitoring – reconfiguring of the data periodically gives the nature of agility for the storage servers. The study compares the performance of a few strategies that includes the constraints and variants fitting for cloud based architectures.

**Keywords:** Storage Networks, Request assignment, Monitoring, Constraints, Variants, Cloud.

## 1 Introduction

Voluminous growth of data usage in this information age throws a challenge to the corporate the way in which they looked the storage architectures in a different perspective [1]. Due to the data explosion there is a need for more and more storage networks which are made available to the larger group of network users. There are three major issues to be concentrated on any kind of heavy data storage centric applications which are processing huge users' requests. They are Data placement, data monitoring and data reconfiguration [2]. Storage Area Networks [3], Network Attached Storage are the most common examples of centrally managed enterprise storage systems which are the architectural deployment form of any storage centric applications. User requests' assignment problem is crucial to be addressed when we focus on Quality of Service in terms of performance tuning. The objective of the data assignment problem is to place the requests served by the appropriate storage node satisfying the constraints. This placement is done according to the nature of the users' demand restricted by influencing parameters.

Appropriate placement of data for the requests should be continuously monitored which is very crucial for further placement of requests as well as migration of data among the storage devices based on the factors like popularity of the data, criticality of the data and similar to that [4]. These factors are to be computed based on aggregate queries across the system which is comparatively easy on centralized managed storage systems but will be a very challenging task in decentralized systems.

The data request assignment problem can be comparable to the assignment of play stations for the users in a play station centre. Because the request management in storage network and play station centre both takes up the same kind of limiting constraints and varying parameters. In both the cases the objective is to maximize the profit earned by each service provider and to minimize the queue length for the service.

In this paper we enumerated some of the storage constraints and variants which are having high impact on the cloud based architectures in general [5]. These impacting parameters has to be considered when a request has to be assigned with a server as well as when the data has to be reconfigured among the storage nodes. Variants are changing in nature and it affects the storage constraints. These parameters are monitored continuously and whenever there is a raise of violation in the constraints, it is forcing the storage network to reconfigure by migrating data among the nodes.

This paper enumerates various storage constraints and variants which are very essential for inclusion in the constitution of larger scalable and virtualized storage environments which forms the basic characterization of cloud architectures in general. The rest of this paper is organized as follows: Section 2 summarizes the related work. Section 3 describes the problem and its simulated model. Section 4 introduces the strategies and its algorithms. Section 5 precisely presents the experimental model and qualitative results. Section 6 concludes the paper and gives direction for the possible future works that can be extended.

## 2 Related Work

Data request placement and assignment of servers is always a challenging area of research since then years. There is a need for adapting appropriate storage management policies as well as the technology in networked consumerism of data changes such as Cluster, Grid, Peer-to-Peer and Cloud computing [6]. There are researches in this area to fit storage policies for scalable distributed systems and applications which requires effective management of voluminous, distributed, and heterogeneous data. Some of these works are briefed below.

The dynamic request allocation problem proposed with a strategy, distributed dynamic rank-based request allocation and gi-FIFO scheduling scheme that aim to maximize the total profit charged by the cloud. Experimental results show that the scheme far outperforms the commonly deployed static allocation with either FIFO or Weighted Round Robin scheduling [7].

The centralized data placement problem analyzed and proposed with a Polynomial Time Approximation Scheme (PTAS). When the storage servers are reconfigured with migration round and in each migration, nodes are paired up. Paired nodes can exchange data during that round. The results are that the problem is NP-Hard and the authors present heuristics for the problem [8].

The request placement problem proposed with different scheduling strategies and introduced a batch scheduler. This scheduler implements techniques specific to queuing, scheduling, and optimization of data placement jobs, and provides a level of abstraction between the user applications and the underlying data transfer and storage resources [9].

### 3 Problem Description

In a storage network, the storage nodes are characterized by their load capacity. Load capacity specifies the number of requests that can be served at a time. Users' requests are consolidated periodically with fixed time slice. The data items are classified with a profit. For example, the data item which is having the maximum access frequency will be considered as the highest profitable one. Each request will have a stay period in the server which is uncertain to predict.

Each data item is having its own profit value which will be changing time to time. Profit value will be computed based on some principle like popularity of the item. Users' requests are consolidated at equal interval of time. At time  $t_1$ , the monitor collects the requests between time slice  $t_0$  to  $t_1$ . Monitor prepares the assignment table for the requests to the nodes based on the load capacity constraint. Requests are assigned to the nodes. Monitor started to record the requests which started at  $t_1$ . Monitor pre-assigns the nodes for the requests satisfying the constraint. The pre-assignment of requests with nodes is maintained by a queue. At time  $t_2$ , request queue of  $t_1$  is assigned. These processes are iterated till all the requests are served. The actual service is executed following this stream of configuration tables. The objective of the problem is to maximize the requests served by each node as well as to minimize the response time for each request.

In short, there is a collection of data items  $D(D_1, D_2, \dots, D_n)$  that are stored in a storage network consisting of  $N(N_1, N_2, \dots, N_j)$  nodes. Each node  $N_j$  is characterized by its load capacity  $L_j$  which indicates the maximum number of requests that it can serve. We are given a set of  $R(R_1, R_2, \dots, R_i)$  requests. Each request  $r_i$  seeks a particular data item  $d_n$  and has profit  $P_i$  associated with the data item that it seeks. The goal is to assign the requests to nodes with the objective of maximizing the total profit of requests served with minimized response time, subject to the capacity constraints of the storage nodes [8].

### 4 Identified Strategies with Algorithm

There are five strategies identified for the above said problem and all the algorithms constructed for the strategies follows the above notations: Let  $D_i$  be the data item;  $P_i$  be the profit associated with each data item;  $R_i$  be the request of each data item;  $N_j$  be the nodes;  $L_j$  be the load capacity of the nodes;  $k$  be the number of requests;  $T_k(R_i)$  be the given stay of time at random for each request.  $X_j(R_i)$  be a part of request that is assigned to a particular node  $N_j$ . The algorithm which describes the processes of placement, monitoring and reconfiguration as follows:

#### 4.1 Strategy 1: Request Assignment at Random

In this strategy, the requests are assigned to the storage servers based on their in-time and assignment is restricted by the load capacity constraint. The algorithm is presented below:

```

1  Algorithm Random_Assignment(L,D,P,R,T)
2  begin
3  Stay time of Requests  $T_k(R_i)$  are allotted in random.
4  // Allot requests to specific nodes based on random distribution.
5  Allot it randomly to any node  $N_j$  in any order as below:
6  repeat
7  {
8      Start with the unserved data item.
9      If  $R_i(D_i) \leq L_j$ , allot  $R_i(D_i)$  to  $N_j$  without violating the load capacity.
10     If  $R_i(D_i) > L_j$ , allot only  $L_j$  to  $N_j$  while the rest is moved to other nodes
        in random.
11     If there are any remaining requests, then queue them to next time stamp.
12     If all requests are served for  $D_i$ , then mark it as served.
13 } until (all requests of all data items are served)
14 // Queue is maintained at every timestamp with incoming requests at tail.
15 Based on the queue, assign the requests to nodes as follows:
16 {
17     At a time  $T_k(R_i)$  the requests are allotted to  $N_j$  based on  $L_j$ .
18     If there is a request  $R_i$  already in  $T_k(R_i)$ , then the new request  $R_{i+1}$  is
        allotted after the request which is currently served is completed.
19     Allot the requests to nodes to which they are assigned by (6) to (13)
        based on time stamps.
20     Based on the table built using (17) to (19), find the response time of
        every request  $R_i(D_i)$ , their waiting time and the average waiting time of
        each data item  $D_i$ .
21     Find the % of total profit gained by the nodes on serving the requests
        using:
22     Profit % of  $N_j = \Sigma(\text{Profit of } R_i \text{ assigned to } N_j) / \Sigma(\text{Profit of } R_i
        \text{ assigned to all nodes})$ 
23     Find the % of profit gained by each node against each item using:
24     Profit of  $D_i = \Sigma(\text{Profit of request } R_i(D_i) \text{ in all nodes})$ 
25     Profit % of each node  $N_j$  for serving
         $D_i = \Sigma(\text{Profit of request } R_i(D_i) \text{ in } N_j) / \Sigma(\text{Profit of request } R_i(D_i) \text{ in all}$ 
        nodes )
26 }
27 end

```

## 4.2 Strategy 2: Request Assignment with Orderly Circular Nodes

In this strategy, the storage servers are arranged in a circular fashion and the requests are assigned to the storage servers based on their order of arrangement and the placement of requests is restricted by the load capacity constraint. This method uses requests consolidation process and the assignment starts with maximum profitable requests first. The algorithm is presented below:



```

1  Algorithm Orderly_Circular(L,D,P,R,T)
2  begin
3  Stay time of Requests  $T_k(R_i)$  are allotted in random..
4  Sort the requests based on decreasing order of profit.
5  repeat
6  {
7      Start with the unserved data item with the highest profit.
8      Allot the requests  $X_j(R_i)$  one by one in circular order as in round robin.
9      If  $X_j(R_i) \leq L_j$ , allot  $X_j(R_i)$  to  $N_j$  without violating the load capacity.
10     If  $X_j(R_i) > L_j$ , allot only  $L_j$  to  $N_j$  and queue the remaining requests.
11     If all requests are served for  $D_i$ , then mark it as served.
12 }
13 until (all requests of all data items are served)
14 // This algorithm uses steps (17) to (25) of algorithm Random_assignment( )
15 end

```

### 4.3 Strategy 3: Request Assignment with Maximum Load Capacity

In this strategy, the requests are assigned based on the load capacity of the storage server as the constraint. The algorithm is presented below:

```

1  Algorithm Max_load(L,D,P,R,T)
2  begin
3  Stay time of Requests  $T_k(R_i)$  are allotted in random.
4  Sort requests by decreasing order of profit
5  Sort nodes by decreasing order of load capacity.
6  // Allot requests to specific nodes based on load capacity of nodes.
7  repeat
8  {
9      Start with the unserved data item with the highest profit.
10     Allot the request  $X_j(R_i)$  to the node  $N_j$  which has the highest load
        capacity.
11     If  $X_j(R_i) \leq L_j$ , allot  $X_j(R_i)$  to  $N_j$  without violating the load capacity.
12     If  $X_j(R_i) > L_j$ , allot only  $L_j$  to  $N_j$  and try to allot the rest in the
        succeeding nodes of the sorted order.
13     If all nodes are allotted, then queue the remaining requests.
14     If all requests are served for  $D_i$ , then mark it as served.
15 }
16 until (all requests of all data items are served)
17 // This algorithm uses the steps (17) to (25) of Random_assignment( )
18 end

```

#### 4.4 Strategy 4: Request Assignment Maximum Request

In this strategy, the storage servers are assigned with the order of maximum requests restricted with load capacity of the storage server as the constraint. The algorithm is presented below:

```

1  Algorithm Max_Request(L,D,P,R,T)
2  begin
3      Stay time of Requests  $T_k(R_i)$  are allotted in random.
4      Sort data items based on decreasing order of requests.
5  repeat
6      {
7      Start with the unserved data item with the highest request.
8      Split the requests equally into  $X_j(R_i)$ .
9      If  $X_j(R_i) \leq L_j$ , allot  $X_j(R_i)$  to  $N_j$  without violating the load capacity.
10     If  $X_j(R_i) > L_j$ , allot only  $L_j$  to  $N_j$  and try to allot the rest in other
        available nodes.
11     If all nodes are allotted, then move the remaining requests to the queue.
12     If all requests are served for  $D_i$ , then mark it as served.
13     }
14  until (all requests of all data items are served)
15  // This algorithm uses the steps (17) to (25) of the Random_assignment( )
16  end

```

#### 4.5 Strategy 5: Request Assignment with Equal Profit Distribution

In this strategy, the storage servers are arranged in a circular fashion. The total requests are equally divided for each server's proportion but restricted with load constraint. Requests with maximum profits are served first. The algorithm is presented below:

```

1  Algorithm Equal_Profit(L,D,P,R,T)
2  begin
3      Stay time of Requests  $T_k(R_i)$  are allotted in random.
4      Sort data items based on decreasing order of profit.
5  repeat
6      {
7      Start with the unserved data item with the highest profit.
8      Split the requests equally into  $X_j(R_i)$ .
9      If  $X_j(R_i) \leq L_j$ , allot  $X_j(R_i)$  to  $N_j$  without violating the load capacity.
10     If  $X_j(R_i) > L_j$ , allot only  $L_j$  to  $N_j$  and queue the remaining requests.
11     If all requests are served for  $D_i$ , then mark it as served.
12     }
13  until (all requests of all data items are served)
14  // This algorithm uses the steps (17) to (25) of the Random_assignment( )
15  end

```

## 5 Experimental Model and Results

The strategies are simulated and the application model is developed in Java using Netbeans. The model is experimented qualitatively with the following test data set. There are three server nodes  $N_1$ ,  $N_2$  and  $N_3$ . The load capacity is 3, 5 and 4 respectively. There are three data items A, B and C with the profit of cost unit 1, 2 and 3 respectively. At time stamp  $T_1$  requests are consolidated and A got 24, B got 30 and C got 18. Stay time of the request is randomly generated with the range of 1 to 5. The application is executed with the sample data input for all the five strategies and the results are compared as presented in the table 1, table 2, figure 1 and figure 2.

**Table 1.** % of Profit for the servers in total requests served

Nodes	Strategies									
	Random		Equal profit		Circular		Max requests		Max load capacity	
	Profit	%	Profit	%	Profit	%	Profit	%	Profit	%
<b>N1</b>	31	22.46	37	26.81	23	16.67	36	26.09	38	16.67
<b>N2</b>	62	44.93	55	39.86	49	35.51	52	37.68	55	35.51
<b>N3</b>	45	32.61	44	31.88	47	34.06	50	36.23	51	34.06

**Table 2.** % of Profit in serving individual data item

Nodes	Item	Strategies									
		Random		Equal profit		Circular		Max requests		Max load capacity	
<b>N1</b>	<b>A</b>	6	25.00	6	25.00	6	25.00	7	29.17	5	20.83
<b>N2</b>		8	33.33	8	33.33	11	45.83	11	45.83	11	45.83
<b>N3</b>		10	41.67	8	33.33	7	29.17	6	25.00	8	33.33
<b>N1</b>	<b>B</b>	16	26.67	16	26.67	18	30.00	14	23.33	18	30.00
<b>N2</b>		30	50.00	26	35.00	20	33.33	26	43.33	20	33.33
<b>N3</b>		14	23.33	18	30.00	16	26.67	20	33.33	22	36.67
<b>N1</b>	<b>C</b>	9	16.67	15	27.78	9	16.67	15	27.78	9	16.67
<b>N2</b>		24	44.44	21	38.89	18	33.33	15	27.78	24	44.44
<b>N3</b>		21	38.89	18	33.33	24	44.44	24	44.44	21	38.89

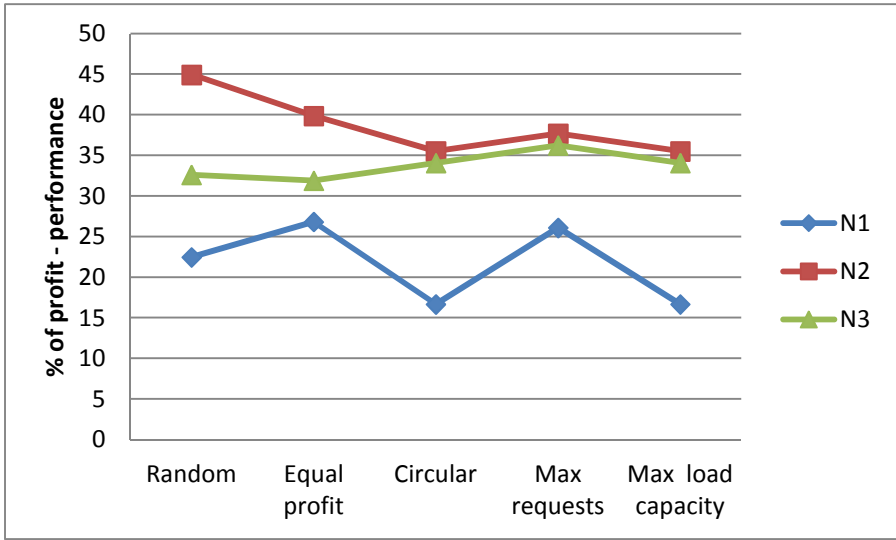


Fig. 1. Comparison of the profits using different strategies

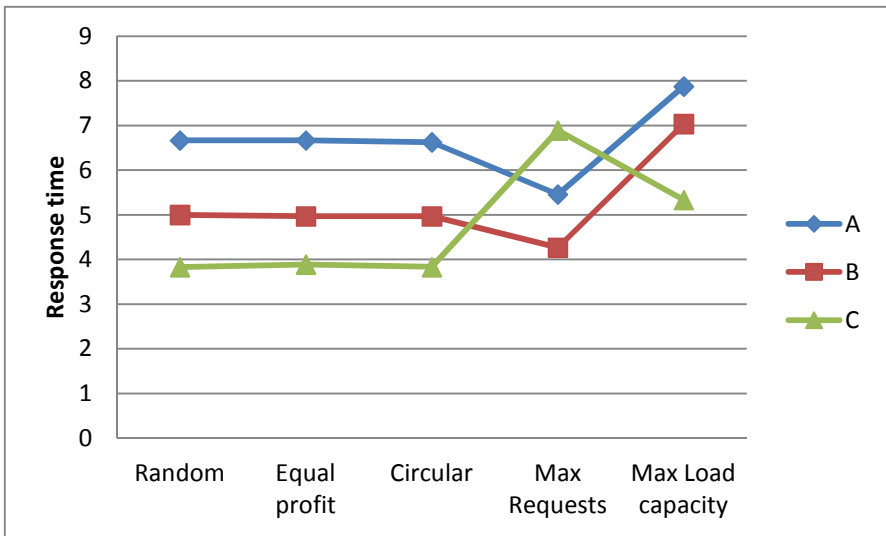


Fig. 2. Comparison of the response time of the requests

## 6 Conclusion and Future Works

Ultimate aim of any kind of service provider is to serve for more requests performing with lesser response time. The objective of the proposed strategies is to achieve this target in data management context. Even though these strategies are more suitable for persistent classified data but also generic in nature and that can be adaptable to any of the

storage centric systems particularly relevant for handling big data. Choosing and adapting the efficient strategy suitable for the business's architecture will highly improve the performance of the system with less capital investment on storage architecture.

If the strategies are enumerated with experimental results for big collection of data sets with inclusion of pertinent constraints and variants, it is quite logical to formulate and compare the efficiency of the strategies as well as suitably it can be fitted to the different system architectures like P2P, grid, storage clusters and cloud.

**Acknowledgement.** The authors would like to thank the enthusiasm and efforts put forth by Ms.S.Shanmugapriyaa for bringing out this work successfully.

## References

1. Black, L., Mandelbaum, J., Grover, I., Marvi, Y.: The Arrival of Cloud Thinking; How and Why Cloud Computing has come of age in large enterprises. White paper, Management Insight Technologies, USA (2010)
2. Sardari, M., Restrepo, R., Fekri, F., Soljanin, E.: Memory Allocation in Distributed Storage Networks. In: IEEE International Symposium on Information Theory Proceedings (ISIT), pp. 1958–1962 (2010)
3. Troppens, U., Erkens, R.: Storage Networks Explained; Basics and Application of Fibre Channel SAN, NAS, iSCSI and InfiniBand. John Wiley & Sons Inc., USA (2003)
4. Ding, J., Han, H.Y., Zhou, A.H.: A Data Placement Strategy for Data-Intensive Cloud. In: Advanced Materials Research, vol. 354-355, pp. 896–900 (2011)
5. Vaquero, L.M., Rodero-Merino, L., Buyya, R.: Dynamically scaling applications in the cloud. Newsletter ACM SIGCOMM Computer Communication Review Archive 41(1), 45–52 (2011)
6. Madathil, D.K., Thota, R.B., Paul, P., XieA, T.: Static Data Placement Strategy towards Perfect Load-Balancing for Distributed Storage Clusters. In: IEEE International Symposium on Parallel and Distributed Processing IPDPS 2008, pp. 1–8 (2008)
7. Bolor, K., Chirkova, R., Viniotis, Y., Salo, T.: Dynamic Request Allocation and Scheduling for Context aware Applications subject to a Percentile Response time SLA in a Distributed Cloud. In: 2nd IEEE International Conference on Cloud Computing Technology and Science, pp. 464–472 (2010)
8. Kashyap, S.R.: Algorithms for Data Placement, Reconfiguration and Monitoring in Storage Networks. Ph.D. dissertation report, University of Maryland (2007)
9. Kosar, T.: Data Placement in Widely Distributed Systems. Ph.D. dissertation report, University of Wisconsin-Madison, USA (2005)
10. Shachnai, H., Tamir, G., Tamir, T.: Minimal Cost Reconfiguration of Data Placement in Storage Area Network. In: Bampis, E., Jansen, K. (eds.) WAOA 2009. LNCS, vol. 5893, pp. 229–241. Springer, Heidelberg (2010)
11. Tsao, S.-L., Huang, Y.-M.: An Efficient Storage Server in Near Video-on-Demand Systems. IEEE Transactions on Consumer Electronics 44(1), 27–32 (1998)
12. Xu, Y., Wu, L., Guo, L., Chen, Z., Yang, L., Shi, Z.: An Intelligent Load Balancing Algorithm Towards Efficient Cloud Computing. In: AI for Data Center Management and Cloud Computing: AAAI Workshop (2011)
13. A Storage Architecture Guide. White paper by Auspex Systems (2000), <http://www.storagesearch.com/auspexart.html>

# A Generic Agent Based Cloud Computing Architecture for E-Learning

Samitha R. Babu<sup>1</sup>, Krutika G. Kulkarni<sup>1</sup>, and K. Chandra Sekaran<sup>2</sup>

<sup>1</sup> NITTE Meenakshi Institute of Technology, Computer Science and Engineering,  
Bangalore, Karnataka, India  
{samitha.r.babu,krutigk}@gmail.com

<sup>2</sup> National Institute of Technology Karnataka, Computer Science and Engineering,  
Mangalore, Karnataka, India  
kch@nitk.ac.in

**Abstract.** E-learning can be referred to as “education through electronic media”. It is one of the most emerging concepts in the field of technology. E-learning systems can be either synchronous or asynchronous. Agent based e-learning is helpful in managing the information overload, it can serve as academic expert and manages creation of programming environment for learners. There are many characteristics that an E-learning environment has to support; they are Interaction, Data Security, User Personalization, Adaptability, Intelligence, Interoperability, Accessibility and User Authentication. E-learning must also support a few other features like cost effectiveness, reusability, storage capacity, powerful computing and virtualization which can be provided by Cloud computing. Cloud computing is everywhere these days, pick any blogs, journals, papers. The Cloud computing Architecture is built using three models, Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS), which provides a variety of services. This proposed architecture integrates the features of Agents and Cloud Computing to produce an Agent based Cloud Computing Architecture to enhance e-learning.

## 1 Introduction

Electronic learning (E-Learning) has been widely adopted in universities and educational institutions in the last few decades. It offers a virtual online learning environment that facilitates the learning process. Cloud computing has also created a much hype in the IT industry in the past few years. It has the potential to play a vital role in the educational transformations. Software agents are also used widely to reduce system complexity. The following sub-sections give a detailed description about E-learning, Cloud computing and Software agents, which would help in the further understanding of our proposed architecture.

### 1.1 E-Learning

In recent years, E-learning has emerged to be the most effective and time saving way of learning. It is the most suited for distance learning and flexible learning.

It's providing education through the use of electronic media and information and communication technologies. E-learning can be either synchronous or asynchronous. Synchronous learning occurs in real time, with everyone interacting at the same time, while asynchronous learning is self-paced and allows participants to engage in exchange of ideas and information without depending on the involvement of others at the same time [1]. E-learning includes numerous types of media support that delivers text, audio, images, animation and streaming videos for interaction purpose. It also includes technology applications and processes such as audio or video tape, satellite TV and computer based learning. There are various technologies that facilitate e-learning; blogs, wikis, white boards, web casting and more. E-learning provides a better environment for learning. It also provides different learning courses based on the students' interest at any location and anytime away from classroom, which in turn maximizes the effectiveness of learning.

## 1.2 Cloud Computing

Cloud computing is a computing archetype where large number of systems are connected in a network to dynamically provide scalable infrastructure for applications, data and file storage. Cloud computing follows a practical approach of pay-per-use technique, where the consumers have to pay depending on what they use and how much they use. The basic idea of using cloud computing is a principle of reusing the IT capabilities. Cloud computing provides computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. It helps in improving the ability of handling and computing large scale data [1]. It is growing rapidly, with applications in almost every area, including education.

**Cloud Computing Service Models.** Cloud computing mainly provides three levels of services within the system, referred to as delivery models, they are, Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [1].

*Software as a Service.* This is a model in which an application is hosted and delivered as a service to customers who can access it via the internet. When the software is hosted off-site, the customer doesn't have to maintain it or support it. This delivers software applications like, email, social networking, ERP, etc., through cloud infrastructure.

*Platform as a Service.* This model is the most recent development and it refers to the availability of the application development platform services through cloud infrastructure. This helps the developers to build and deploy applications without having to invest on the infrastructure.

*Infrastructure as a Service.* Infrastructure as a Service is the delivery of computing services (hardware, storage, and networking) through cloud infrastructure. It is described as utility computing datacenters that make use of cluster technology to provide powerful and flexible computing resources.

### 1.3 Software Agents

Software agents are computational autonomous entities capable of acquiring information and producing and sending information in an environment to accomplish set of designated goals. These agents possess certain characteristics [2]:

**Autonomy:** The Agent acts without the direct intervention by humans or other agents; it has control over its own actions and internal state.

**Adaptivity:** The Agents react flexibly to the changes in its environment. They take goal directed initiatives.

**Sociability:** The Agent is capable of interacting in peer-to-peer manner with humans or other agents.

**Competency:** The capability to effectively manipulate the problem domain environment to accomplish the prerequisite tasks. Competency includes specialized communication proficiency.

**Amenability:** The ability to adapt behavior to optimize performance in an often non-stationary environment in responsive pursuit of the goals of the client.

We have designed this architecture using certain powerful characteristics of agents, where different agents are assigned different functions which are integrated into a single system. The agents deliver services to the client by acquiring the services from the cloud. These agents can be configured through an agent configuration interface to perform its job on behalf of the clients for a certain set of features of an E-learning system. It can retrieve information such as that about students' learning progress and study their behavior; aggregate the information, etc. One of the advantages of the agent-based approach is that many times a complex processing function can be broken into several smaller, simpler ones. Since each individual agent can be crafted to be an expert in solving a specific problem or performing a particular task, you can build systems that exhibit complex behaviors by using a collection of relatively simple agents. Incorporating Agents with the cloud make the cloud services smarter and an efficient application.

In this paper, section 2 presents the Related works which discusses about the previous architectures and their disadvantages, section 3, we propose our architecture and explain its functioning, section 4 we analyze our architecture and describe it's advantages, section 5 presents the conclusions and future enhancement.

## 2 Related Works

In this section we present a few earlier proposed architectures depicting various cases of using cloud computing but not agents and those using agents but not cloud computing.

- (1) Zhang Guoli and Liu Wanjum [3] developed an architecture based on cloud computing platform for E-learning. It is composed of three layers: Infrastructure layer of E-learning Platform, Layer of platform integration and Application layer. The Infrastructure layer provides system software, information management systems, scheduling combination and utilization of teaching resources. The layer



of platform integration integrates variety of resources belonging to different organization and various platforms in Infrastructure layer. The Application layer has the application of integrating the teaching resources in the cloud computing model including interactive courses and sharing teaching resources. This architecture mentioned the advantages of using cloud computing in E-learning and concluded that cloud computing platforms can be widely used in distance education and online learning.

- (2) K. Sakthiyavathi and K. Palanivel [4], proposed a generic Layered architecture which supported all the agents in one single system. The architecture is provided with a user interface layer, which provides adaptive interface for online learners, a middle layer which consists of all the agents needed by the system and a lower

**Table 1.** Drawbacks of different architectures

Architecture	Draw backs
A Novel Approach for Adopting Cloud based E-Learning system [6]	Fails to provide dynamism in E-Learning.
A Sharable E-Learning Platform based on Cloud Computing [7].	Since this architecture is modeled on the basis of cloud, it only concentrates in providing sharing, reusability and interoperability and fails to incorporate any other feature necessary for an E-Learning System.
Framework of an E-learning environment in continuing education institutions [8].	The services here are not hosted on the cloud so it fails to provide reusability, easy access, interoperability, scalability, versatile compatibility and serving automatic updates.
Research on E-learning system based on SOA [9].	This architecture does not provide a client interface and provides connection only through network service interface.
A New framework Semantic Web Technologies based E-learning [10].	It is platform dependent.
Towards an Effective Integrated E-learning System: Implementation, Quality assurance and Competency Models [11].	Does not provide reusability, sharing of information, adaptable framework.
The Applied Research of Cloud Computing Platform Architecture In the E-Learning Area [3].	Does not provide interoperability with the external content and social network.
A Generic Architecture for Agent based E-Learning System [4]	Does not provide Authentication facilities. Since it is not hosted in cloud, it does not provide any of the cloud features.

layer which contained a user repository. This paper provided a Generic Agent based architecture for E-learning that provides features like Intelligence, Distributed, Adaptive, Interactive, Extensible and Collaborative in a single system architecture using the web services.

- (3) Sh. Umar Khalid, Amna Basharat, Arshad Shabid and Syed Hassan [5] developed a conceptual architecture which focused on the core features required for an emerging E-Learning system; 1) Domain Specific Learning Services 2) Student Capability Analysis 3) Adaptive Lecture Authoring Tool 4) Intelligent Assessment Engine 5) User friendly E-Learning Portal. This framework is designed to make learning process class independent and to abate the distance between the teacher and the student. The highlight of this system is to judge the students' capability and adaptive management of learning process.
- (4) In Table 1. We have compared different architectures which were proposed earlier and listed out the drawbacks with those architectures.

On the basis of the study about different architectures given above, we infer that they fail to provide Dynamism in the system, Accessibility, Authentication, Interoperability, Data security and an Adaptive user interface. We also infer that hosting and delivering services through a cloud is in itself a major benefit.

### 3 Our Approach

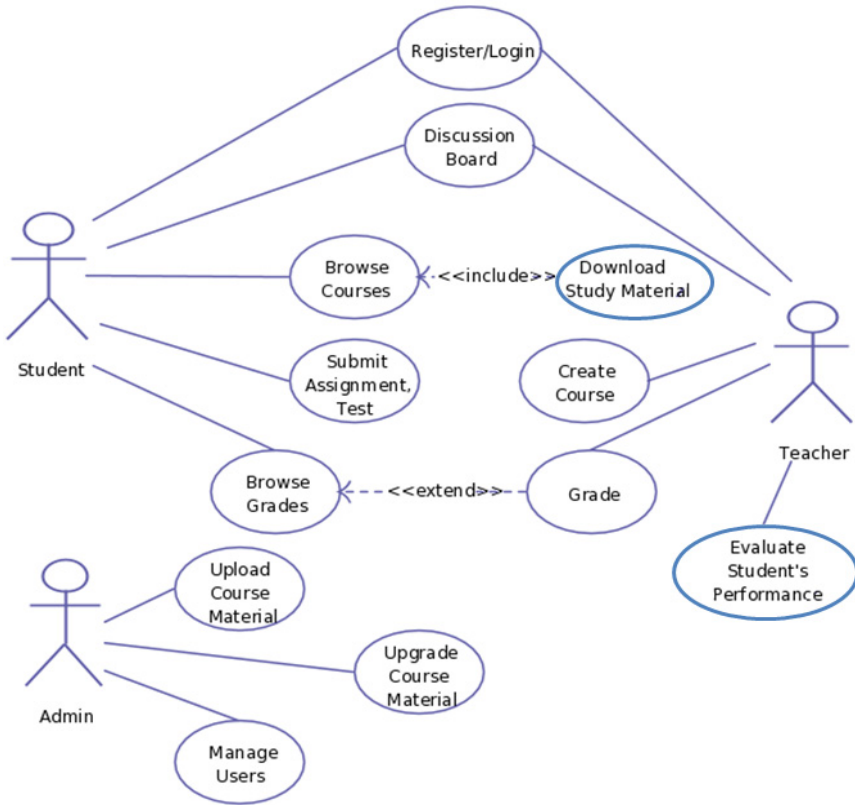
The main actors in the education process are students, teachers and the course coordinator. Most of the time, these actors are geographically and temporarily remote from each other, and they use the course website as the online environment for course delivery, assignment management and communication. Therefore, the course web server contains a lot of data about students' learning process and their study behavior [12].

The figure 1 represents a use case diagram for E-learning system. The student can login, browse through the course material, which facilitates download access, submit assignments and write tests, view his grades, and can have interactive sessions with the teachers through the discussion board. The teacher creates the course, evaluates the students based on their performance, finalize grades to the students. The administrator manages the system by uploading and upgrading the course materials, and also the user profiles.

As seen above, Agent based E-learning can manage the information overload, serve as academic experts, and create programming environment for the learners [4], which provides certain specific characteristics of E-learning system such as Interaction, Personalization, Adaptability, Intelligence, Interoperability, Accessibility and Security.

The applications that run on cloud have many advantages. Those in accordance with educational institutions will be that it is cost effective, that is, it reduces capital costs as it is pay-per-use. It minimizes licensing new software, has measured services, can also globalize the workforce in cheap. Any application in cloud will have unlimited storage, ubiquitous network, quick deployment, easy to access information,

efficient monitoring, scalability and speed, versatile compatibility, multi tenacity, on demand self-service, access to automatic updates, convenient, and also innovative [13]. This architecture is efficient, system-wise also, because the processes are streamlined, optimal resource utilization, high collaboration efficiency, flexible work practices, broad network access, rapid elasticity, and provides backup and recovery. The user can use the application anytime on any device, Personal Computers (PCs), mobile devices like smart phones and Personal Digital Assistants (PDA), having just the internet access as these applications are location independent. This also makes the application mobile.



**Fig. 1.** Use Case diagram for E-learning.

Having mentioned the advantages of generic agent based architecture for E-learning system and cloud computing applications, we combine these to get “A Generic Agent-Based Cloud Computing architecture for E-learning”. In this approach, the software agents are integrated with the cloud computing architecture. In this architecture, each functionality required by the E-learning system is provided by an agent. These agents interact directly with the cloud and provide the various services to the clients via internet.

In this architecture we are trying to overcome the drawbacks of previously mentioned architectures, namely; Dynamism, Accessibility to wide range of data, Data Security, Authentication, Interoperability with the external content and social networking and provide an effective user interface. Hence this generic architecture supports all the features to make the E-learning system efficient.

### 3.1 A Generic Agent-Based Cloud Computing Architecture for E-Learning

We introduce, E-Learning Software as a Service, ELSaaS. This is a cloud service which is basically for the students to interact with the cloud services, via agents. This will provision the students with their learning process. This architecture is as shown in figure 2.

The various layers described in the figure are:

**User Interface Layer:** This is the upper layer which provides adaptive interface for various users. The clients like students and teachers can avail the benefits of the system. The online learners can avail the E-courses that are offered, listen or watch

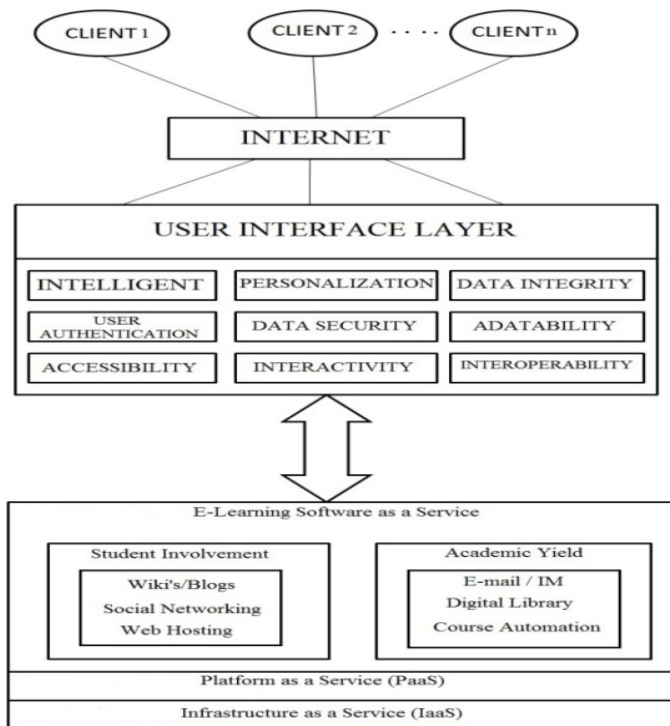


Fig. 2. A generic agent based cloud computing architecture for an E-learning System

lectures, write tests, etc. The study material, like notes, audios and videos of lectures, tests and certification exams, and the extra facilities are offered to the learners. The user interaction with their devices is facilitated by this layer. Any device with a browser and internet connection is enough to use this application.

**Agent Layer:** This is the middle layer, which contains software agents that support various functionalities. This is an important layer where the features of an E-learning system come into the picture. The agents here carry out those functionalities in a very efficient manner. It reduces the information overload in the network. The various agents here are as follows [4]:

*Intelligent Agent.* This agent uses Artificial Intelligence (AI) in the pursuit of providing dynamism in E-learning. It performs few specific tasks on behalf of the students and the teachers.

*Personalization Agent.* This agent provides a set of personalization functionalities to the user so that the learners have a better learning experience. This agent enables the dynamic insertion, customization or suggestion of content and adaptive instant interaction that is relevant to the individual user, based on the user's preferences and requirements. This enables personalized learning plans, learning materials, etc.

*Data Integrity Agent.* This agent provides maintenance and assurance of accuracy and consistency of data over its entire life cycle. Data integrity is identically maintained during any operation. When functions operate on the data, this agent ensures integrity.

*User Authentication Agent.* This agent provides user authentication, that is, it confirms the identity of the user as to what they claim to be.

*Data Security Agent.* This agent applies security framework and data privacy standards in the E-learning system. Security is a major issue for applications in the cloud and hence this agent ensures maximum security along with other agents like data integrity agent and user authentication agent.

*Adaptability Agent.* This agent provides adaptability functionalities in the E-learning system. This is an important factor for the efficiency and economic success. Adaptability increases the capability of monitoring the user activities, interpreting these on domain-specific model basis, infers user requirements and preferences, representing them in certain models, and finally, acting upon the users with the available knowledge, to dynamically facilitate the learning process.

*Accessibility Agent.* This agent makes the geographically dispersed content of the E-learning system easily accessible for its authenticated users. The main intention is to make the service available to as many people as possible.

*Interactive Agent.* This agent makes the system interactive to the users to provide a better learning environment. The users are allowed to modify certain objects in the learning environment according to their preference.

*Interoperability Agent.* This Agent helps to interoperate with the external content and the social network. This is brought up by using an open structure, and with external

content and social service, it can interoperate or inter-communicate with enterprise applications, like Groupware and consumer applications, like Gmail, Twitter, YouTube at data level by mapping mechanism [14]. The main motivation for this is by the increased importance of reusing and combining various learning elements in different ways. This also helps adding on to the economic benefits of the education system in the developed and developing nations.

**Cloud Services Layer.** This layer provides the cloud services for the agents, to provide the various functionalities to the users. The main service used here is the ELSaaS, as mentioned above, which enables the students to interact with the cloud services in educational institutes. The agents assist it by providing functionalities like interoperability and interactivity. This is definitely for the entrusted users as security is of major importance.

Digital Library has a wide variety of applications and content, multimedia and text files, on different topics available to the users, which can be downloaded. This has Online Storage to maintain and manage individual user data. This layer also helps the users to collaborate and generate contents like wikis, blogs, etc. Personalization tools are also available here which will be provided by the particular agent to control user profiles, access technical support, etc. Agents are used for authorization of applications and content which enables better security.

Next layer is the PaaS, which provides directory services, developer tools and database access. Then is the IaaS, which provides storage, virtual machine facilities, web services and service hosting functionalities.

This architecture provides Dynamism, Accessibility, Data security, Authentication, Interoperability, User interface, Adaptability, Interactivity, and overcomes all the drawbacks faced by the previous architectures, through the software agents using cloud services.

## 4 Analysis

Agent based E-learning can manage the information overload, serve as academic experts, and create programming environment for the learners [4]. As seen before there are a few specific characteristics in an efficient E-learning system such as Interaction, Personalization, Adaptability, Intelligence, Interoperability, Accessibility, User Authentication, Data Integrity and Data Security. Our architecture provides all these features in a single system. As this architecture hosts and delivers applications through cloud, it incorporates all the advantages of cloud computing. The major advantages like cost effectiveness, unlimited storage, reusability, ease of access, ubiquitous network, scalability and speed, versatile compatibility and convenience. E-learning has speeded up the knowledge transfer without restrictions on time and space in recent years. In order to achieve the goal of knowledge sharing and reusing interactively, many E-learning architectures have been proposed. But the main drawbacks of these architectures are that they are not efficient in many ways, like dynamism, interoperability, speed, personalization, intelligence, installation costs, and security as a whole.

Hence, we have come up with this architecture combining all the major features required by an efficient E-learning system. This architecture combines the agents providing the functionalities, in cloud, which accesses the cloud services like the ELSaaS, PaaS and IaaS and delivers them to the online learners. This architecture is better in every perspective that is, it is beneficial to the educational institute, the clients and cloud service providers.

## 5 Conclusion and Future Work

Cloud Computing has become one of the hottest buzzwords in the IT area over the past few years. It plays a vital role in education transformations which reaches a wide variety of users. This advantage can be used to educate people in developed and developing countries and eventually eradicate illiteracy in the world. This generic agent based cloud architecture for an E-learning system provides the most basic and primitive features that must be present in an E-learning system for a better learning experience. This architecture uses software agents to provide all the important functionalities to the online learners using the cloud services. This definitely contains the advantages of using software agents and cloud computing as well.

Our proposed architecture is just a conceptual idea for an E-learning system. This must be implemented in a real time E-learning application to validate our architecture, which will be the future work. The only concern is regarding the security issues that arise in cloud applications which have to be taken care of in the future.

## References

1. E-Learning, Cloud Computing and Software Agents, <http://www.wikipedia.org>
2. David Wallace Croft, Intelligent Software Agents: Definitions and Applications, <http://alumnus.caltech.edu/~croft/research/agent/definition/>
3. Guoli, Z., Wanjun, L.: The Applied Research of Cloud Computing Platform Architecture In the E-learning Area. In: The 2nd International Conference on Computer and Automation Engineering (ICCAE), February 26-28, vol. 3, pp. 356–359 (2010)
4. Sakthiyavathi, K., Palanivel, K.: A Generic Architecture for Agent based E-Learning System. In: International Conference on Intelligent Agent and Multi Agent System, July 22-24, pp. 1–5 (2009)
5. Talia, D.: Cloud Computing and Software Agents: Towards Cloud Intelligent Services. published in CEUR Workshop on Objects and Agents (2011)
6. Anwar, M., Masud, H., Huang, X.: A Novel Approach for Adopting Cloud based E-learning System. In: 11th IEEE/ACIS International Conference on Computer and Information Science, May 30-June 1, pp. 37–42 (2012)
7. Wang, C.-C., Pai, W.-C., Yen, N.Y.: A Shareable E-Learning Platform based on Cloud Computing. In: 3rd International Conference on Computer Research and Development (ICCRD), March 11-13, vol. 2, pp. 1–5 (2011)
8. Cheung, K.S., Lam, J., Im, T., Szeto, R.: Framework of an E-learning environment in Continuing Education Institutions. In: International Conference on Electronic Computer Technology, February 20-22, pp. 43–46 (2009)

9. Niu, H., Ma, X.: Research on E-learning system based on SOA. In: 2nd International Conference on Multimedia and Information Technology, April 24-25, vol. 1, pp. 148–150 (2010)
10. Shrivastava, G., Sharma, K., Bawankan, A.: A New framework Semantic Web Technologies based E-learning. In: 11th International Conference on Environment and Electrical Engineering (EEEIC), May 18-25, pp. 1017–1021 (2012)
11. Al-Sharan, S., Al-Hunaiyyan, A.: Towards an Effective Integrated E-learning System: Implementation, Quality assurance and Competency Models. In: 7th International Conference on Digital Information Management (ICDIM), August 22-24, pp. 274–279 (2012)
12. Choy, S.-O., Ng, S.-C., Tsang, Y.-C.: Software agents to assist in Distance Learning Environments. *EDUCAUSE(Quarterly)* (January 1, 2005)
13. Zhang, H., Yang, X.L.: Cloud computing architecture based on SOA. In: 5th International Symposium on Computational Intelligence and Design, October 28-29, vol. 1, pp. 369–373 (2012)
14. Yang, Z.: Study on Interoperable Cloud framework for E-education. In: International Conference on E-Business and E- Government (ICEE), May 6-8, pp. 1–4 (2011)
15. Talia, D.: Clouds Meets Agents: Towards Intelligent Cloud Services. *IEEE Internet Computing* 16(2) (March-April 2012)
16. Oppong, E., Khaddaj, S.: SOA and Cloud service Provisioning Framework. In: 11th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES), October 19-22, pp. 200–204 (2012)
17. Talia, D.: Cloud Computing and Software Agents: Towards Cloud Intelligent Services. published in CEUR Workshop on Objects and Agents (2011), [http://ceur-ws.org/Vol-741/INV02\\_Talia.pdf](http://ceur-ws.org/Vol-741/INV02_Talia.pdf)
18. Yang, X.L., Zhang, H.: Cloud Computing and SOA Convergence and Research. In: 5th International Symposium on Computational Intelligence Design, 28-29 October, vol. 1, pp. 330–335 (2012)
19. Mariya, S.: Cloud Computing- An Advanced E-Learning Platform of School Education? In: 14th International Conference on Interactive Collaborative Learning (ICL), Slovakia, September 21-23, pp. 569–570 (2011)



# Cache Based Cloud Architecture for Optimization of Resource Allocation and Data Distribution

Salim Raza Qureshi

Department of Computer Science,  
Model Institute of Engineering & Technology, Jammu  
salim.cse@mietjammu.in

**Abstract.** Cloud computing comes into focus when you think about what IT always needs: a way to increase capacity or add capabilities on the fly without investing in new infrastructure, training new personnel, or licensing new software. In the current architecture we don't have the facility to access offline data and resources also due to the bandwidth issue and geographical distances; the performance of the cloud infrastructure at the user end is low. We propose a new architecture which would allow caching of the highly utilized data and resources on a separate server Resource & Data Caching/Proxy Server (RDCPS).

**Keywords:** Cloud Computing, Caching, Indexing, Virtual Machine.

## 1 Introduction

The advancement of Cloud technologies in the last few years has opened up new possibilities to Internet applications developers. Previously, deployment and hosting of an application was one of the first and main concerns when designing an application for the Internet. But with the advent of the Cloud, now it is possible to solve this problem more economically and more flexibly using the powerful infrastructure services provided by a Cloud service provider on an as-required basis. Cloud computing is the delivery of computing as a service rather than a product. ) over a network (typically the Internet)[1]. Cloud computing architecture consists of 4 key components or layers: SaaS (Software as a Service), PaaS (Platform as a Service), IaaS (Infrastructure as a Service) and DC (Data Centers) as explained in Figure 1.

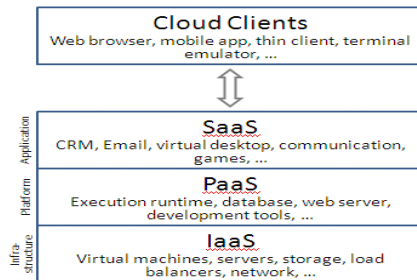
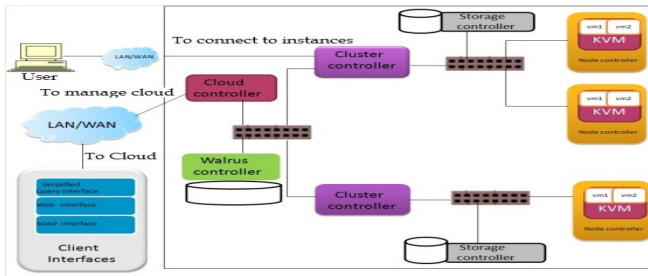


Fig. 1. [8]Cloud computing layers

- **SaaS (Software as a Service)** : provide software as a service on demand over the network. Maintenance and support are simplified .
- **PaaS (Platform as a Service)** : deliver platform as services. i.e. Google App Engine, Amazon S3
- **IaaS (Infrastructure as a Service)** : provide infrastructure, such as servers, software, data center space, network as a services. i.e. Amazon EC2
- **DC (Data Centers)** : Centralized data centers containing servers with special hardware and software design

## 2 Existing Architecture

We have taken the architecture of Eucalyptus i.e.is software available under GPL that helps in creating and managing a private or even a publicly accessible cloud. It provides an EC2 compatible cloud computing platform and S3 compatible cloud storage platform.



**Fig. 2.** Architecture design of private cloud developed using Eucalyptus (Source: <http://mdshaonimran.wordpress.com/2011/11/26/eucalyptus-and-its-components/>)

### 2.1 Node Controller (NC) [16]

A UEC node is a VT enabled server capable of running KVM as the hypervisor. UEC automatically installs KVM when the user chooses to install the UEC node. The VMs running on the hypervisor and controlled by UEC are called instances. Node Controller interacts with the OS and the hypervisor running on the node on one side and the CC on the other side.

### 2.2 Cluster Controller (CC) [16]

CC manages one or more Node Controllers and deploys/manages instances on them. CC also manages the networking for the instances running on the Nodes under certain types of networking modes of Eucalyptus. CC communicates with CLC on one side and NCs on the other side.

### 2.3 Walrus Storage Controller (WS3) [16]

WS3 provides a persistent simple storage service using REST and SOAP APIs compatible with S3 APIs.

### 2.4 Storage Controller (SC) [16]

SC provides persistent block storage for use by the instances. This is similar to the Elastic Block Storage (EBS) service from AWS.

### 2.5 Cloud Controller (CLC) [16]

The Cloud Controller (CLC) is the front end to the entire cloud infrastructure. CLC provides an EC2/S3 compliant web services interface to the client tools on one side and interacts with the rest of the components of the Eucalyptus infrastructure on the other side. CLC also provides a web interface to the users for managing certain aspects of the UEC infrastructure.

## 3 Problems with the Existing Architecture

In the current architecture following problems have been identified.

1. **Migration Problem:** If any of the nodes which are being used by the users go down, migration of VMs need to be done and that would result in down-time, no temporary server is available which could handle the application request for that period of time.
2. **Up-Scaling / Down-Scaling Problem:** Up-scaling and down-scaling request are most often sent by the users. Every time during scale-up or scale-down of these resources the whole VM needs to be redeployed, this would also result in down-time.
3. **Load Balancing Problem:** Instead of physical architecture, virtual architecture is used in cloud, i.e., users are assigned virtual machines which can handle request to an extent.

## 4 Proposed Architecture (RDCPS)

We propose a new architecture which would allow caching of the highly utilized data and resources on a separate server Resource & Data Caching/Proxy Server (RDCPS). This server would be stacked between the cluster controller and node controllers as well as the storage controllers. RDCPS would enable better performance and offline usage of the most commonly used data and resources.

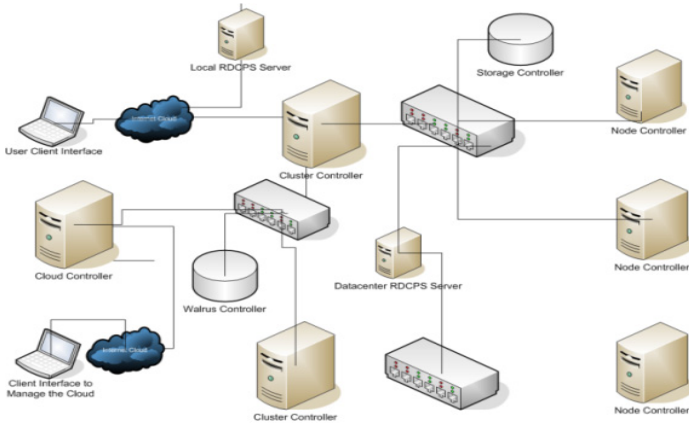


Fig. 3. RDCPS Installation Scenario in Eucalyptus Cloud Environment

## 5 Internal Architecture of RDCPS

RDCPS has a four layered architecture that caters to the problems listed in the paper.

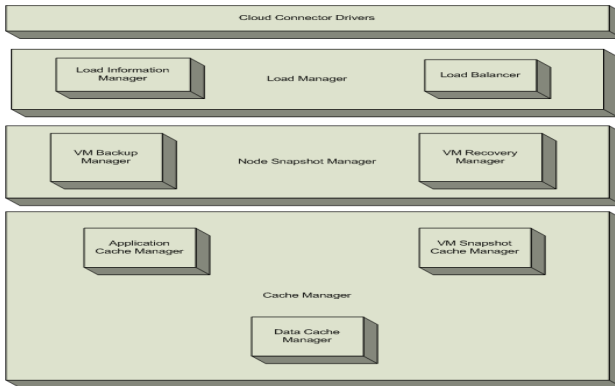


Fig. 4. Internal Architecture of RDCPS

### 5.1 Layer 1: Cache Manager

Cache Manager controls the high performance memory and disk cache which stores application snapshots, complete VM snapshots and data snapshots. Cache manager has three sub-components namely Data Cache Manager (DCM), VM Snapshot Cache manager (VMSCM) and Application Cache Manager (ACM). DCM interacts with various storage managers and caches the data. It would also help in minimizing the down-time while migration of the server or up/down scaling of VMs. ACM would cache the whole application and would also cater to the load balancing and down-time problems.

## 5.2 Layer 2: Node Snapshot Manager

This controls the backup and recovery of the VMs on various nodes. It comprises of two modules namely VM Backup Manager and VM Restore Manager.

## 5.3 Layer 3: Load Manager

Load Manager collects the information from each VM and balances the load if needed using the Node Snapshot stored at RDCPS. Load Managers consist of two components namely Load Information Manager (LIM) and Load balancer (LB). LIM fetches the current load and performance parameters from each VM and processes and shares the information with LB, which in-turn balances the load

## 5.4 Layer 4: Cloud Connector Drivers

This layer would consist of drivers needed to connect to different cloud platforms and retrieve information.

# 6 Class Diagram of RDCPS

RDCPS implementation on Cloud Analyst (A cloud-sim based simulator) has been depicted the Figure 5. Simulation results in the later section are based on the following class architecture.

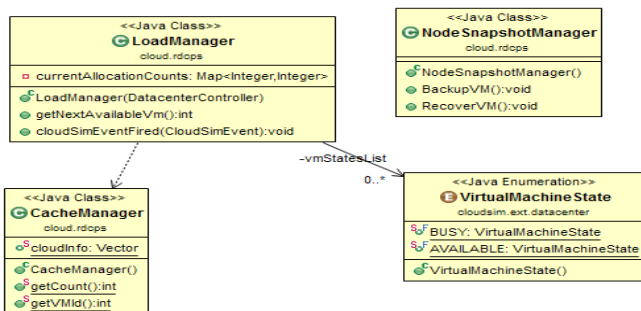


Fig. 5. Class Diagram of RDCPS

# 7 Program Code

Load Balancing /Caching Code for RDCPS

```

public int getNextAvailableVm()
{int vmId = -1;
for (int availableVmId : vmStatesList.keySet()) {

```

```

if (!currentAllocationCounts.containsKey(availableVmId)) {
    CacheManager.cloudInfo.add(availableVmId);}}
vmId=CacheManager.getVMId();
if (vmId<0)
{System.out.println("AllocatingActive using VMLoadbalancer");
if (currentAllocationCounts.size()< vmStatesList.size())
    {for (int availableVmId : vmStatesList.keySet())
    {if (!currentAllocationCounts.containsKey(availableVmId)) {
        vmId = availableVmId;
                                break; }}}
else {
    int currCount;
    int minCount = Integer.MAX_VALUE;
    for (int thisVmId : currentAllocationCounts.keySet()) {
        currCount = currentAllocationCounts.get(thisVmId);
        if (currCount < minCount){minCount = currCount;
            vmId = thisVmId;}}}}
else
    {System.out.println("Getting VM Snapshot from RDCPS :: VMID :
"+vmId);}allocatedVm(vmId);
    return vmId;}

```

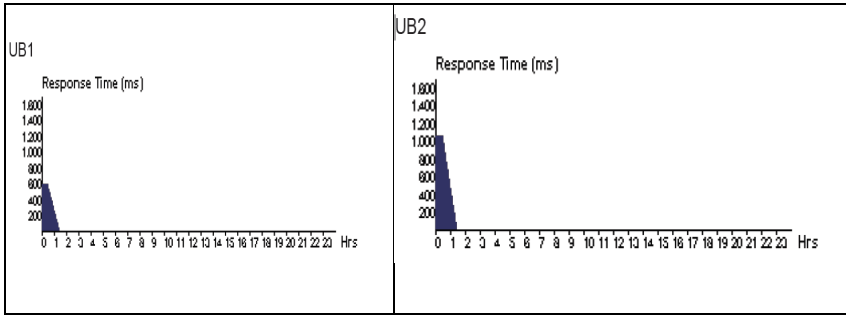
## 8 Simulation Results

**Scenario 1** – Simple Web Application Hosted on a Single Data Center with Round Robin VM loads balancing policy. Let us assume initially the application is deployed in a single location, in Region 3 (Asia). Assuming the application is deployed in 50 virtual machines (with 1024Mb of memory in each VM running on physical processors capable of speeds of 100MIPS)

**Table 1.** Overall Response Time

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	511.62	45.12	15210.71
Data Center processing time:	115.94	0.06	14730.70

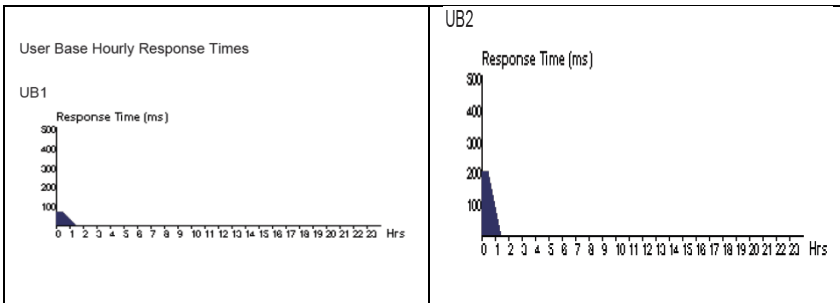
Please note that these numbers are based on all the parameters mentioned above and therefore may not be realistic. The response times experienced by each user base are depicted graphically as follows:



**Scenario 2:** Simple Web Application Hosted on a Single Data Center with RDCPS Servers hosted

**Table 2.** Overall response time

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	197.29	42.54	5803.91
Data Center processing time:	39.67	0.06	5751.39



## 9 Future Work

Results shown in the paper are simulated results, we would further work on implementing the RDCPS in the actual cloud environment and work on following aspects:

- Configuration of Local and Datacenter RDCPS, which would include the hardware and software platforms and No. of required RDCPS
- Determine the level and type of security to be imparted at local and datacenter instances.

## References

1. CloudAnalyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environments. Project Report Bhatiya Wickremasinghe
2. Foster, I., Zhao, Y., Raicu, I.: Cloud Computing and Grid Computing 360-Degree Compared. Dept. of Comp. Sci., Univ. of Chicago, Chicago, IL, USA
3. The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology by Peter Mell & Timothy Grance
4. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud Computing and Emerging IT Platforms: Future Generation Computer Systems, vol. 25(6), pp. 599–616. Elsevier Science, Amsterdam (2009) ISSN: 0167-739X
5. Buyya, R., Yeo, C.S., Venugopal, S.: Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, HPCC 2008, Dalian, China, September 25-27. IEEE CS Press, Los Alamitos (2008)
6. Introduction to Cloud Computing Architecture. White-paper, 1st edn. Sun Microsystems, Inc. (June 2009), <http://www.sun.com>
7. Buyya, R., Ranjan, R., Calheiros, R.N.: Proceedings of the 7th High Performance Computing and Simulation Conference, HPCS 2009, June 21-24. IEEE Press, New York (2009) ISBN: 978-1-4244-4907-1
8. U. o. E. Institute for Computing Systems Architecture. simjava (May 15, 2009), <http://www.dcs.ed.ac.uk/home/hase/simjava/>
9. Sun CTO: Cloud computing is like the mainframe (March 11, 2009), <http://Itknowledgeexchange.techtarget.com>
10. It's probable that you've misunderstood 'Cloud Computing' until now. TechPluto. 1496100&coll=&dl=ACM&CFID=21518680&CFTOKEN=1880080
11. Danielson, K.: Distinguishing Cloud Computing from UtilityComputing. Ebizq.net (March 26, 2008), [http://www.ebizq.net/blogs/saasweek/2008/03/distinguishing\\_cloud\\_computing/](http://www.ebizq.net/blogs/saasweek/2008/03/distinguishing_cloud_computing/)
12. Defining "Cloud Services" and "Cloud Computing". IDC (September 23, 2008), <http://blogs.idc.com/ie/?p=190>
13. Cloud Simulation Frameworks by Barry Lumpkin, Tuan Nguyen, Nguyen
14. Rodrigo, N., Calheiros, R.R.: CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services
15. Buyya, R.: GridSim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing
16. Ubuntu Enterprise Cloud Architecture By Simon Wardley. Etienne Goyer & Nick Barcet (August 2009)



# Implementing a Publish-Subscribe Distributed Notification System on Hadoop

Jyotiska Nath Khasnabish, Ananda Prakash Verma, and Shrisha Rao

International Institute of Information Technology,  
Bangalore 560 100, India  
{jyotiskanath.khasnabish,anandaprakash.verma}@iiitb.org,  
shrao@ieee.org

**Abstract.** Apache Hadoop is an open source framework for processing massive amount of data in a distributed environment. Hadoop services use a polling mechanism for event notifications. In this paper, we propose a distributed notification system for Hadoop based on the Publish-Subscribe model. Such a notification system can be used for message-passing among Hadoop services. It can also be used to chain multiple MapReduce jobs based on events occurring in a Hadoop cluster. This results in reduced cluster load and network bandwidth requirement. We have used two popular Publish-Subscribe-based messaging systems—Apache ActiveMQ and Apache Kafka—for implementation. Lastly, we have executed performance tests on both these messaging systems to monitor time taken for message delivery and reception.

## 1 Introduction

In recent years, Hadoop [2] has become a *de facto* standard for processing massive amounts of data in a distributed environment. With growing numbers of Hadoop-based services each year, the necessity to implement an event-based notification system has grown significantly. In the Hadoop Summit 2011 [13], Yahoo disclosed that their primary workflow manager ‘Oozie’ manages over 600,000 processed jobs per month internally on their cluster, with the total number of users being more than 300. According to their prediction, the number of jobs will grow to a larger number in coming years. Different Hadoop services like MapReduce [10] computations produce large number of jobs every hour. Often these services are run together on a Hadoop cluster with several other Hadoop services to perform complex data-intensive operations.

We have designed and implemented a notification system on Hadoop using the Publish-Subscribe [11] model, which provides high performance and scalable solution for passing messages between different services. In this system, one node or one service can play one of the following two roles, ‘Publisher’ or ‘Subscriber.’ The benefit of using the Publish-Subscribe model is that the Publishers are connected to the Subscribers through one or more than one message brokers rather

than directly; a Publisher need not know a Subscriber, and vice versa. The intermediate message broker performs the filtering procedure on the messages based on the ‘Topic,’ so that the Subscribers only get a relevant subset of messages instead of all messages sent by all the Publishers.

In Section 2 we describe in detail the Publish Subscribe model, how a specific Topic-based notification message is delivered from a Publisher and reaches the Subscribers who are interested in that specific Topic. In Section 3 we explain the entire architecture of our notification system and the different roles and the functions in it. We also discuss the JMS (Java Message Service) architecture and how JMS can be used to implement our notification system.

In Section 4 we briefly mention three possible use cases of our proposed notification system. First, we discuss how the system can be used to send notification between different Hadoop services. Second, we discuss how our system can be used to build an event-based job control framework for chaining multiple MapReduce jobs together and respecting dependencies among multiple jobs. Lastly, if one job waits for particular data which is to be produced as the output of another job, then the notification system can be used to notify about the data availability, instead of polling the NameNode.

In Sections 5 and 6, we have analyzed the performance and scalability of the notification system using both ActiveMQ and Kafka, both of which are open-source versions of Publish Subscribe pattern-based messaging systems. Based on the results, we have concluded how the notification system can be scaled up with increasing sizes of input message sets, and provide better delivery times. Lastly, we discuss the possible areas where the proposed notification system can be deployed in future.

## 2 Publish Subscribe Model

Publish-Subscribe [11] is a message passing model where senders, also known as ‘Publisher’ can send messages to receivers known as ‘Subscriber’. The publish-subscribe model differs from a traditional client server system in that the Subscribers do not need to be directly connected to Publishers to receive messages. Publishers publish their messages when events occur, and messages are typically classified by a taxonomy of predefined Topics. The intermediate message broker takes care of the guaranteed message delivery. Since the message broker performs Topic-based filtering on all incoming messages from the Publishers, the Subscribers do not get all the messages published by all the Publishers present. Instead, a Subscriber gets only the messages on Topics to which it is subscribed.

We use a Topic based publish-subscribe model where a Hadoop service or a Node can act as a Publisher and publish some message on a specific Topic. Nodes or services which act as Subscribers, subscribed to that specific Topic, receive the message. The advantage of this system is that Subscriber nodes do not have to be on the same physical rack, and services need not to originate from same node to receive a message. In fact, the Publishers may not even be aware of the

existence of a Subscriber. A Publisher just publishes its message and continue its own operation.

For example, let us say that one MapReduce job is dependent on the output of a Pig job. So, it waits until the data has arrived from the currently executing job. The Pig job, upon the successful completion, publishes a message on a sample Topic like ‘Job #413 Complete.’ The MapReduce job which is subscribed to that Topic receives the notification as soon as the message is published, and therefore can start its computation immediately, rather than checking for the status of the previous job again and again.

In this system, each process  $p_i$  can execute the following operations:  $publish_i(m)$  and  $subscribe_i(T)$ . A Publisher publishes a message  $m$  on a Topic  $T$ , and Subscribers which are listening on that Topic receive it from the message broker.

We have implemented our notification system twice, using both Apache ActiveMQ and Apache Kafka, and measured their performances with different input sets. ActiveMQ and Kafka both provide APIs for Topic-based messaging system using the Publish-Subscribe model and both of them are highly powerful, scalable and can be distributed over large networks.

## 2.1 Apache ActiveMQ

Apache ActiveMQ [1] is an open-source messaging system which uses JMS (Java Messaging Service) to send and receive messages. ActiveMQ provides high scalability, performance and security for large scale messaging system.

ActiveMQ architecture is divided into three components: Publisher, Broker and Subscriber. Following the Publish-Subscribe model, the Publisher publishes a message on a specific Topic to the ActiveMQ message broker. The broker may choose to store the messages which is known as Persistent ActiveMQ, or may not choose to store the messages which is known as Non-Persistent ActiveMQ. For persistence, ActiveMQ uses KahaDB to store the messages.

By default, ActiveMQ uses TCP (Transmission Control Protocol) for guaranteed and safe message delivery. Each time a connection needs to be established, a Publisher uses 3-way handshaking protocol which is the norm in TCP. This is the reason why ActiveMQ is highly scalable, has zero message loss and guarantees message delivery.

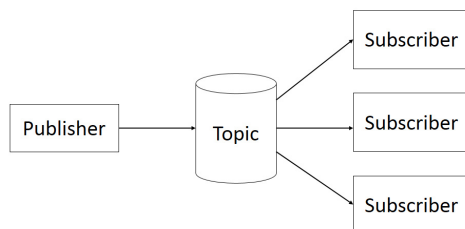


Fig. 1. ActiveMQ Messaging System Architecture

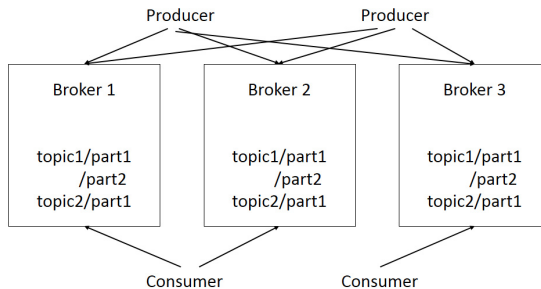
A Subscriber also follow the same method to create a connection using 3-way handshaking. Then it starts listening on specific Topics. In a persistent system, all the messages on that Topic are delivered to the Subscriber. In a non-persistent system, only those messages that are published after the Subscriber has started are delivered. After receiving a given message, the Subscriber may choose to close the connection or to keep listening for further messages.

## 2.2 Apache Kafka

Apache Kafka [14] is a distributed publish-subscribe messaging system designed by LinkedIn, the social media company. Kafka supports persistent messaging with  $\mathcal{O}(1)$  disk structures that provide constant-time performance (even with many TB of stored messages), high throughput, explicit support for partitioning messages over Kafka servers, and distributing consumption over a cluster of consumer machines while maintaining per-partition ordering semantics, support for parallel data load into Hadoop [3].

In Kafka, messages of a specific type are identified by Topics. A set of servers, known as ‘Brokers’ store the messages which are published by the Producers on a Topic. A Consumer can choose and subscribe to one or more one than Topics from the available brokers and consume the subscribed messages provided by the brokers.

In Kafka, each of the brokers gets a partition of the Topic after it is divided into several parts. This results in balancing the load so that more than one producers and consumers can send and retrieve messages simultaneously. Kafka uses a Zookeeper [5] instance to start its server. The default port for Zookeeper is 2181 and Kafka server is 9092. After the server has started, Kafka can create a new Topic which is then added to the list of Topics, or it can choose from the existing Topics. Kafka automatically persists messages with its default configuration which means that when Subscriber comes live, it is able to receive messages on a specific Topic from the beginning.



**Fig. 2.** Kafka Messaging System Architecture (from [14])

### 3 Architecture

Our notification system is based on JMS [7] publish-subscribe model and an intermediate message broker.

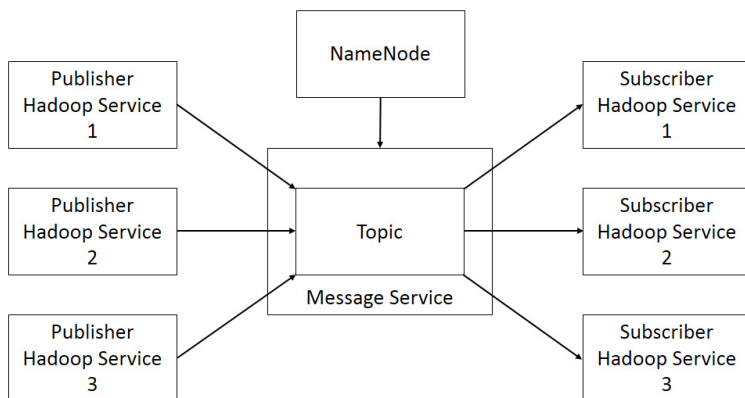
The Observer [12] pattern in a publish-subscribe channel helps to inform the Subscribers when there is a change in the system. In this case, the system administrator creates a channel for message passing. The Publishers create a Topic on which the Subscribers receive messages published by the Publishers. As a result, when a Publisher sends a message the channel makes sure that a copy of the message is sent asynchronously to all the Subscribers listening on that Topic. For details, see [12].

JMS [7] or Java Message Service is a *Java Message Oriented Middleware* API which is used to send and receive messages among multiple clients. JMS has two different implementation models: *Point to Point* and *Publish and Subscribe*. In our system, we have used ‘Publish and Subscribe’ implementation model. JMS is comprised of the following elements: Provider, Client, Producer/Publisher, Consumer/Subscriber, Message, Queue, Topic.

The model is implemented in the following steps in Hadoop:

- The NameNode creates a Publish-Subscribe Channel. (This is represented as a JMS Topic.)
- The Hadoop services acting as the Publisher creates a Topic to send messages on the channel.
- Each of the Hadoop services acting as the Subscriber subscribes to a Topic to receive messages on the channel.

For example, let us go back to the scenario mentioned in Section 2. Whenever the Pig job has finished its execution, it loads data into HDFS and the data



**Fig. 3.** Publish Subscribe Model of Notification System

is divided into blocks of fixed size. Then, the NameNode allocates the blocks to the DataNodes where the data gets written. As soon as the NameNode gets some acknowledgement from the DataNodes about the successful completion of the writing of data, it publishes a notification message of data availability to the Topic ‘Job #413 Complete’ using JMS (Java Message Service). The next MapReduce job depending on the output of the previous job and subscribed itself to the Topic receives the message of data availability. After receiving the notification message about the availability of data, it triggers its own workflow and begins the computation process. In this way, polling on the NameNode is removed, thus resulting in reduced network bandwidth usage and improvement in cluster performance.

## 4 Use Cases

Our notification system consists of two modules, ‘Publish’ and ‘Subscribe’. Each time a Hadoop service wants to publish a message based on a specific Topic on a channel, it just calls the ‘Publish’ module with necessary parameters such as a predefined Topic name, message and the broker URL. The ‘Subscribe’ module also performs the same work except it just needs a Topic name and the broker URL as parameter in order to receive the messages published by the Publisher.

We have identified the following three possible use cases where we can deploy our proposed notification system.

### 4.1 Passing Messages between Hadoop Services

As discussed before, we can use our notification system if we want to pass notification messages between different running Hadoop services. The messages can be status flags, or progress reports about a certain running job (busy, waiting, or complete). For example, the TaskTrackers keep updating their status to JobTrackers at certain intervals, which creates network overhead when there is no status change. Our notification system allows the TaskTrackers to only notify JobTrackers when there is a status change.

If a service wants to publish some message on a specific Topic, it may do so by calling the ‘Publish’ module with appropriate parameters (Topic, message, broker URL). In the same way, if a Hadoop service is waiting for certain notification message from another service then it will subscribe to a Topic and whenever the message arrives on the channel it is delivered to the Subscriber by the intermediate message broker.

### 4.2 Notification for Data Availability

In a large Hadoop environment, many of the jobs depend on the output of other currently executing jobs. If an organization produces large number of jobs every hour, many of which are dependent on other jobs this means that a job cannot start its execution until the jobs on which it is dependent finish executing and

put their output data into HDFS. In this scenario, we can use the notification system to inform currently waiting messages about the data availability.

For example, if one job *A* is waiting for one batch of log data regarding user activities in a website over a period of month, when the data becomes available, it is analyzed and a report is created. The job *A* is completely unaware of the availability of data or exactly when the data should be available. So it has no choice but to keep polling the NameNode. Now if thousands of jobs, like 10,000 or 15,000 jobs, keep polling the NameNode for such information continually, it not only cause unnecessary waste of bandwidth, but also increases the cluster load significantly.

Instead, if the jobs were to subscribe to a specific Topic, say ‘Data Available,’ then they do not have to poll the NameNode repeatedly. As soon as the data are available, the NameNode, as a ‘Publisher,’ publishes a message on the Topic ‘Data Available’ and the subscribed jobs are able to continue their execution as soon as they hear back. This approach saves both network bandwidth and reduces the load on the Hadoop cluster.

### 4.3 Event Based Job Chaining

If multiple MapReduce or any other Hadoop jobs need to be chained in order to accomplish a complex task, this can be achieved using our notification system also. Many complex problems need to be solved by writing several MapReduce steps which run in series to accomplish a goal. It is also very common in a large organization where thousands of jobs are created every hour, that many of the jobs are interdependent with one another. This means that there should be an efficient workflow manager within the cluster to handle the jobs. Existing workflow managers like Oozie [4] handles the workflow using a DAG (Directed Acyclic Graph) where the jobs and their dependencies are represented using edges and nodes. The chain of jobs can be depicted as following:

Map1 - Reduce1 - Map2 - Reduce2 - Map3 ...

There are existing workflow managers like LinkedIn’s Azkaban [8], Spotify’s Luigi [9] or Yahoo’s Oozie [4] which are capable of chaining jobs. But in certain cases it is better to employ a workflow manager using the notification system we have implemented to remove the overhead of chaining jobs one by one and trigger workflows automatically. In this system, to make a chain of jobs, the ‘Publish’ and ‘Subscribe’ modules can be used so that when a job has finished executing, it can trigger the next set of jobs automatically without having to set the job dependencies manually.

## 5 Performance Analysis

We have analyzed the performance with different sets of messages and observed how much time it takes to publish messages and also to receive messages. We also have observed whether any message is getting lost or not. Since Hadoop is

designed to run on commodity hardware, we have used day-to-day use computers and executed our tests within virtual machines (VMs) with limited processing power.

We used three VMs, one for ‘Publisher,’ one for ‘Subscriber,’ and one for ‘Message Broker.’ All three were running Hadoop v1.1.1 and ActiveMQ 5.8.0.

For our graph we have taken message sets of 100, 200, 500, 1000, 2000 and 5000 and noted down the time taken to send those messages. We ran our tests three times and used the average values to increase accuracy.

When the Publisher or Subscriber starts, it first creates a connection by the TCP 3-way handshaking protocol with the message broker using its URL and port number, which is by default 61616. After the session and Topic have been created, the Publisher prepares a message, sends the message to the broker, and closes the connection. The Subscriber keeps listening on some Topic; if a message is Published against that Topic, the broker sends the message to the Subscriber and upon receiving the desired message, the Subscriber may close the connection or keep listening for further messages on that or other Topics.

In Kafka, if there are multiple brokers running, then a Topic gets divided into multiple partitions to manage load. But in our tests, we have used a single broker running on Machine #1. All the messages published from the Publishers go to the broker first, then are sent to the Subscribers based on their Topics.

We have monitored the load on our cluster and overall network bandwidth consumption when our notification system was not in use and when we used our notification system. Based on the results received, we have shown in Figures 5 and 6 how the notification system can be used to bring down the cluster load and reduce network bandwidth by replacing the default polling mechanism.

**Table 1.** System Configuration used for Testing

	Machine #1	Machine #2	Machine #3
Processing Speed	2.3 GHz	2.3 GHz	2.3 GHz
Primary Memory (RAM)	2 GB	2 GB	2 GB
Disk Space	8 GB	8 GB	8 GB
Operating System	Ubuntu 12.04	Ubuntu 12.04	Ubuntu 12.04
Hadoop Version	1.1.1	1.1.1	1.1.1
ActiveMQ Version	5.8.0	5.8.0	5.8.0
Kafka Version	0.8	0.8	0.8

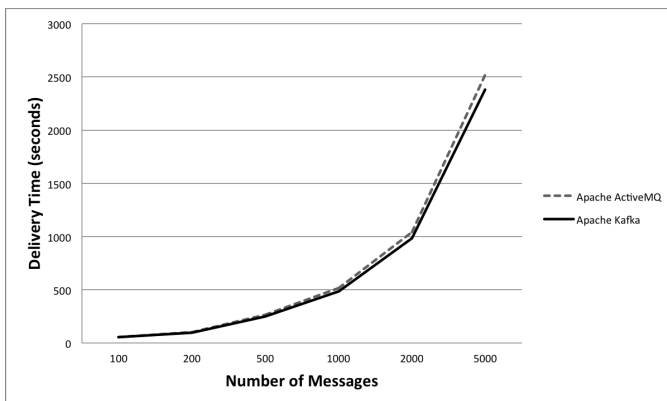
**Table 2.** Test Results using Apache ActiveMQ

Number of Messages	Iteration 1 (sec)	Iteration 2 (sec)	Iteration 3 (sec)	Average Time (sec)
100	51	57	55	54.33
200	104	100	104	102.67
500	265	270	261	265.33
1000	509	519	520	516
2000	1035	1048	1043	1042
5000	2522	2497	2528	2515.67



**Table 3.** Test Results using Apache Kafka

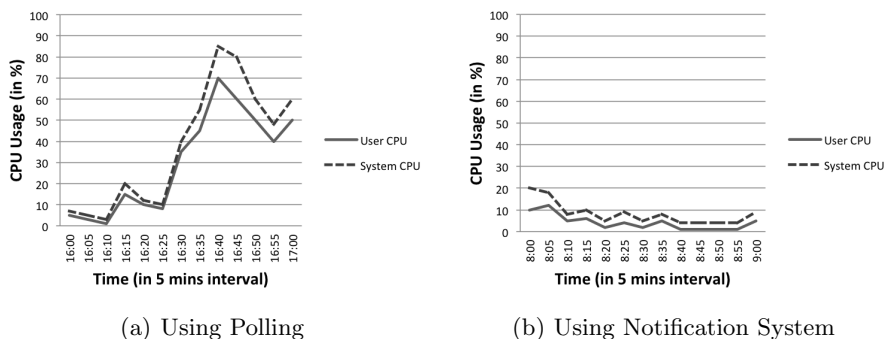
Number of Messages	Iteration 1 (sec)	Iteration 2 (sec)	Iteration 3 (sec)	Average Time (sec)
100	55	53	52	53.33
200	94	93	96	94.33
500	253	246	247	248.67
1000	478	489	493	486.67
2000	986	981	978	981.67
5000	2402	2377	2362	2380.33



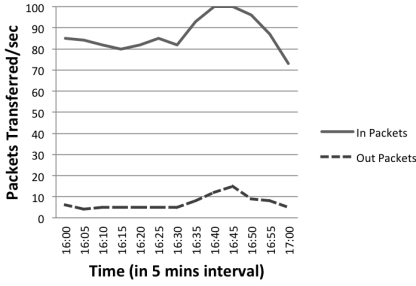
**Fig. 4.** Analysis of delivery times of messages using ActiveMQ and Kafka

Figure 5 shows how the load on the cluster was significantly reduced when we used our notification system instead of polling for a certain hour.

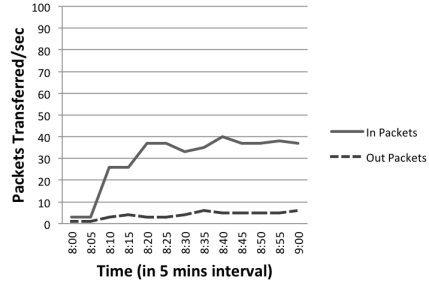
Figure 6 shows the network bandwidth usage before and after using the notification system. We have used the Ganglia Monitoring System [6] for monitoring cluster load and network bandwidth usage, and to retrieve the data to plot the



**Fig. 5.** Monitoring Cluster Load before and after using Notification System



(a) Using Polling



(b) Using Notification System

Fig. 6. Monitoring Network Bandwidth before and after using Notification System

graphs used to visualize the results found. In all our tests, we have experienced zero message loss. This is per the guarantee provided by both Kafka and ActiveMQ because of the underlying protocol (TCP) used by them. Apart from reliable transmission, TCP provides error detection, flow control, and congestion control, which helps ensure guaranteed message delivery. Also, ActiveMQ and Kafka both have almost similar message delivery times according to the results we found. This means that either of them can be used to implement our notification system without decrease in performance.

## 6 Conclusion

We have presented a distributed notification system for Hadoop based on the Publish-Subscribe model. Event-based notification systems like this not only reduce the load on Hadoop components such as the NameNode, but also increase the productivity of developers using Hadoop. As we have shown in our use cases, this notification system can be used to lower network bandwidth usage by reducing the number of redundant network packets; it can also chain multiple MapReduce jobs together to accomplish a complex task. High scalability and reliability are more reasons to deploy our notification system in different use cases we have mentioned, like message-passing between Hadoop services or chaining jobs.

Because of the extensibility of our proposed system, it can also be used with any Hadoop services in the future. Also the scalable feature of such system ensures that large Hadoop clusters (more than 100 nodes) benefit from it to a great extent. In such environments, efficient use of ‘Publish’ and ‘Subscribe’ modules can ensure that job control frameworks are able to handle heavily interdependent jobs and computations resulting in optimal use of hardware and resources. Also, this system can reduce network bandwidth and resource usage if integrated with the core Hadoop framework. Over the years, as the complexity of job control frameworks increases, our notification system will help Hadoop users address these difficulties.

## References

1. Apache ActiveMQ, <http://activemq.apache.org/>
2. Apache Hadoop Project, <http://hadoop.apache.org>
3. Apache kafka, <https://github.com/apache/kafka>
4. Apache Oozie, <https://oozie.apache.org/>
5. Apache zookeeper, <http://zookeeper.apache.org/>
6. Ganglia monitoring system, <http://ganglia.sourceforge.net/>
7. Java message service, [http://en.wikipedia.org/wiki/Java\\_Message\\_Service](http://en.wikipedia.org/wiki/Java_Message_Service)
8. Linkedin azkaban, <http://azkaban.github.io/azkaban2/>
9. Spotify luigi, <https://github.com/spotify/luigi>
10. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. In: Proceedings of the 6th Conference on Symposium on Operating Systems Design & Implementation, OSDI 2004, pp. 137–149 (2004)
11. Eugster, P.T., Felber, P.A., Guerraoui, R., Kermarrec, A.-M.: The many faces of publish/subscribe. *ACM Computing Surveys* 35(2), 114–131 (2003)
12. Hohpe, G., Woolf, B.: *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley (2012)
13. Islam, M.K.: Oozie: Scheduling workflows on the grid. In: *Hadoop Summit* (2011)
14. Kreps, J., Narkhede, N., Rao, J.: Kafka: a distributed messaging system for log processing. In: *NetDB: Networking Meets Database, NetDB 2011* (2011)

# Securing Public Data Storage in Cloud Environment

D. Boopathy and M. Sundaresan

Department of Information Technology, Bharathiar University,  
Coimbatore, Tamilnadu, India

{ndboopathy, bu.sundaresan}@gmail.com

**Abstract.** Cloud computing technology is a new concept of providing dramatically scalable and virtualized resources. It implies a service oriented architecture type (SOA), reduced information technology overhead for the end-user, great flexibility model, reduced total cost of ownership, on-demand service providing structure and many other things. One of the main concerns of customers is Cloud security and the threat of the unknown. The lack of physical access to servers constitutes a completely new and disruptive challenge for investigators. The Users are store, transfer or exchange their data using public cloud. This paper represents the encryption method for public cloud and also the cloud service provider's verification mechanism using the third party auditors.

**Keywords:** Secured Data Storage, Public Cloud Service Provider, Cloud Service Provider, Cloud Encryption, Cloud Decryption, Third Party Auditor.

## 1 Introduction

Cloud computing is a natural evolution of the widespread adoption of virtualization service, service-oriented architecture (SOA), autonomic, and also utility computing. Cloud computing is the broader concept of infrastructure convergence [5]. This results in reduced cost of the services. Quality of services also gets better as the organization can spend the saved amount and time on improving it [1]. In cloud environment, resources are shared among all of the servers, users and individuals [3]. As a disarray invention with foreseen implication, cloud computing is mending way it uses business with IT [4]. Security and Privacy issues are of more concern to cloud service providers who are actually hosting the services [14]. Cloud computing models are of two types: Deployment model and Service model. Deployment model is further classified into four type's namely private cloud, community cloud, public cloud and hybrid cloud. Cloud computing service providers provide their services in a number of fundamental models [13]. Cloud computing is a term which is often used with synonyms like grid computing, cluster computing, autonomic computing [12]. For the organization, the cloud service provider offers data centers to move their data globally [11]. It is up to the clients to decide the vendors and also depending on how willing they are to implement secure policies and be subject to 3rd party verifications [10]. Another factor of concern is that the cloud is still under development process and there are no set standards for the data storage and application communication [6].

In Private cloud, the cloud infrastructure is operated solely for an organization. This cloud model may be managed by the organization or a third party and may exist on premise or off premise. In Community cloud, the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns. In public cloud, the cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services. In hybrid cloud, the cloud infrastructure is a composition of two or more clouds.

### **1.1 Software as a Service (SaaS)**

The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The cloud applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including like network, servers, storage, operating systems or individual application capabilities, and with the possible exception of limited user-specific cloud based application configuration settings. It is also referred as Resource Code; provide (managed and scalable) resources as services to the user- in other words, the service providers basically provide enhanced virtualization capabilities. Accordingly, different types of resources may be provided via a service interface [6].

### **1.2 Platform as a Service (PaaS)**

The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the service provider. The consumer does not able to manage or control the underlying cloud infrastructure including like network, servers, operating systems, or storage, but has control over the deployed cloud applications and possibly application hosting environment configurations.

### **1.3 Infrastructure as a Service (IaaS)**

The capability provided to the consumer is to provision processing, networks, storage and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can also include the operating systems and applications. The consumer/client does not manage or control the underlying cloud infrastructure but has control over operating systems; deployed applications, storage and possibly limited control of select networking components (e.g., host firewalls).

## **2 Cloud Computing Types**

There are four different types of cloud are in use, they are:

### **2.1 Private Cloud**

The private cloud infrastructure is operated solely for an organization. It may be managed or maintained by the service acquiring organization or a third party and may exist on organization premise or off premise.

## **2.2 Public Cloud**

The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

## **2.3 Community Cloud**

The community cloud infrastructure is shared by the several organizations and supports a specific community purpose that has shared concerns.

## **2.4 Hybrid Cloud**

The hybrid cloud infrastructure is a composition of two or more clouds (i.e. private, community, or public) that remain unique resource entities but are bound together by standardized or proprietary technology that enables data and application portability [1].

The Security goals of cloud data storage include three important points namely: Data Availability, Data Confidentiality and Integrity.

# **3 Problem Statement**

Information security is a critical issue in cloud computing environments [9]. The public cloud is widely accessed and utilized by the many users for their own purpose, small organizations to manage their data and so on. The public cloud services providers are mostly make the trust among the users by some attraction advertisement and avail their service in very low cost. The data transferred, stored or exchanged in public cloud is mostly unsafe due to the untrust environment. The cloud Computing provides an undemanding and non ineffectual Solution for Daily Computing [5]. The aspect of security and confidentiality must intervene to protect the data from each of the enterprises [10]. Share resources, software and information are provided to computers and other devices on demand [8].

While the benefits of storage networks have been widely acknowledged, consolidation of enterprise data on networked storage poses significant security risks [2]. Hackers adept at exploiting network-layer vulnerabilities can now explore deeper strata of corporate information [2]. Its unauthorized disclosure could seriously and adversely impact the organization and its employees [16]. The Cloud Security Alliance's initial report contains a different sort of taxonomy based on different security domains and processes that need to be followed in general cloud deployment [15].

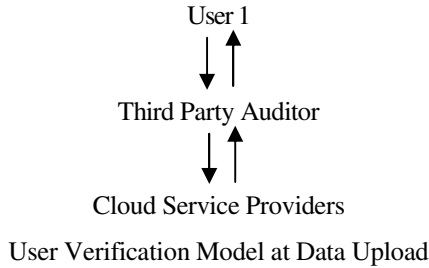
To bring the users data into safer side the secured data storage model is proposed.

# **4 Secured Data Storage**

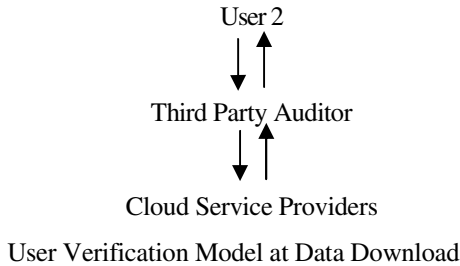
The secured data storage is used the third party auditor as an intermediate person to connect and helps to transfer the secure data transfer between the data destination

place to data accessing place. In this model both the data accessing user and data storing user are uses the secured process by using the encryption algorithms. The RSA algorithm is used to encryption style, but the process is made in moderate style. The data hijack and data stolen is not possible in this secured data storage model, due to the RSA algorithm moderate style.

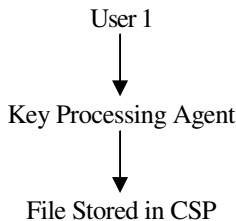
The user1 verifies the public cloud service provider using the third party auditor. When the third party auditor will issue the status of the selected service provider, user1 will ready to upload the files into that public cloud service provider.

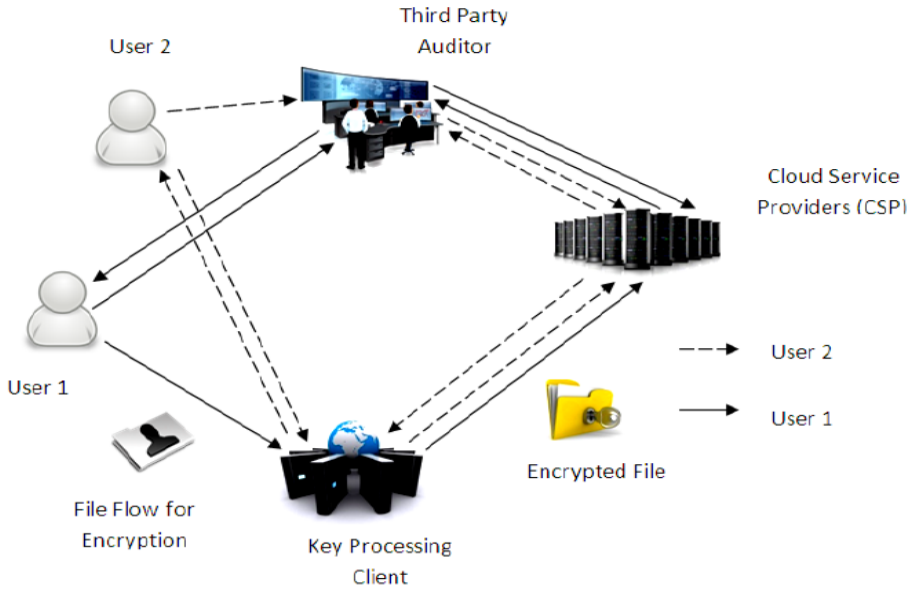


When user wants to download the file from the public cloud service provider again the verification process will be taken place to verify that the provider still in the live status or not.



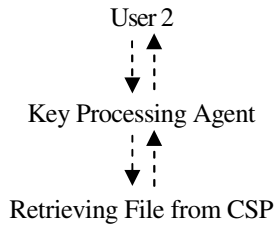
The User 1 encrypt the file using the private key and store it to the CSP with public key access, then the User 2 required the file which was stored by the User 1 in the CSP. The User 2 retrieving the file from the CSP using his/her private key to get the decrypted file. The access made through the public key and the encryption took place through the User 2 private key.





**Fig. 1.** Secured Data Storage

File stored in the public cloud service provider in the encrypted format using the RSA algorithm.



The user2 retrieve the file from the public cloud service provider and decrypt the file using his/her private key.

**4.1 The Mechanism Working Style**

Step 1: The RSA algorithm is used in this method. The User 1 send the file to the key processing client using his private key, then the file encryption taken place, public key was fixed and moved to CSP.

Step 2: When the user 2 required the encrypted file which was stored in the CSP by the User 1. So the User 2 used his/her private key to access the key processing client. The key processing client recognizes the User 2 request by his/her private key. Then the file was decrypted by using the public key then transfer to the User 2.



## 4.2 The Working Flow Code

User 2 = U2, User 1 = U1, Third Party Auditor = TPA, Public Cloud Service Provider = PCSP, Key Processing Client = KPC.

### User1 storing the file in public cloud service provider:

1. U1 send request to TPA.
2. TPA analysis PCSP whether the service is available or not.
3. If the service is available TPA replies to U1 as positive.
4. U1 send the file to KPC using private key and encrypt the file.
5. After encryption KPC stored the encrypted file in the PCSP with the public key access.

### User 2 access the file from the public cloud service provider:

1. U2 send request to TPA.
2. TPA analysis PCSP where the service is available If service is available TPA replies to U2 as positive.
3. U2 send the file to KPC using private key and decrypt the file.
4. The KPC access the file from the PCSP using the public key and decrypt the file using the U2 private key and send it to the U2.

$$\text{For Accuracy Level of Services} = \frac{\text{No. of Cloud Service Providers}}{\text{No. of Cloud Services Rendering}} \times \frac{\text{No. of Satisfied Users}}{\text{No. of Cloud Service Providers}}$$

The result of the accuracy level of services is calculated based on the third party auditor report format. If any error or mistakes was taken place in the upload or download of the file, it is also taken into the accuracy level report. According to the information collected at the time of data accessing, uploading the data by main user and also data downloaded by end user. All are taken into the accuracy calculation.

## 5 Discussion

The Secured Cloud Storage mechanism is using the RSA algorithm for encryption and decryption. So the users can encrypt the file using their private key and stored the file in the public cloud service provider using public key. The public key is used to store and access the file from the public cloud service provider. The important thing is the public key is only used to access the file. The user required private key to decrypt the file.

The processing methods take lesser steps when compared to the existing model.

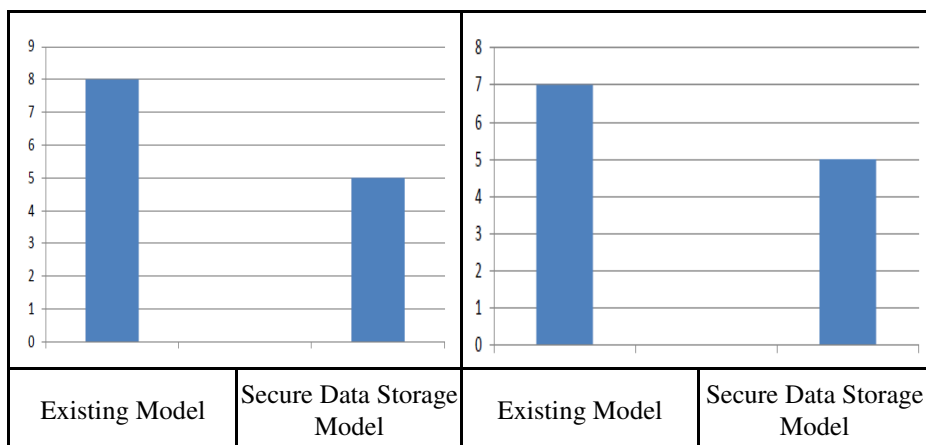


Fig. 2. No. of steps in Encryption the file

Fig. 3. No. of steps in Decryption the file

## 6 Conclusion and Future Enhancement

The Secured Cloud Storage takes fewer steps to encrypt and decrypt the files. The existing models are still not entered into the real world service environment. The many encryption models for cloud environment are still in developing stage. The Secured Cloud Storage model will provide the maximum level of Secured structure in the cloud environment. In future the Secured Cloud Storage, the remaining algorithms will take into development to provide the better security to this model. The remaining algorithms results will take into consideration and compared to provide the maximum level of security in the cloud environment.

## References

1. Tripathi, A., Yadav, P.: Enhancing Security of Cloud Computing using Elliptic Curve Cryptography. *International Journal of Computer Applications* (0975 - 8887) 57(1), 26–30 (2012)
2. El-Khameesy, N., Rahman, H.A.: A Proposed Model for Enhancing Data Storage Security in Cloud Computing Systems. *Journal of Emerging Trends in Computing and Information Sciences* 3(6), 970–974 (2012)
3. Nafi, K.W., Kar, T.S., Hoque, S.A., Hashem, M.M.A.: A Newer User Authentication, File encryption and Distributed Server Based Cloud Computing security architecture (IJACSA) *International Journal of Advanced Computer Science and Applications* 3(10), 181–186 (2012)
4. Govinda, K., Gurunathaprasad, V., Sathishkumar, H.: Third Party Auditing for Secure Datastorage in Cloud through Digital Signature using RSA. *International Journal of Advanced Scientific and Technical Research* 4(2), 525–530 (2012) ISSN 2249-9954

5. Kaur, M., Mahajan, M.: Implementing Various Encryption Algorithms to Enhance the Data Security of Cloud in Cloud Computing. *VSRD International Journal of Computer Science & Information Technology* 2(10), 831–835 (2012)
6. Kaur, S.: Cryptography and Encryption In Cloud Computing. *VSRD International Journal of Computer Science & Information Technology (VSRD-IJCSIT)* 2(3), 242–249 (2012); Kaur, A., Bhardwaj, M.: Hybrid Encryption for Cloud Database Security. *International Journal OF Engineering Science & Advanced Technology* 2(3), 737–741 (2012)
7. Sudha, M., Monica, M.: Enhanced Security Framework to Ensure Data Security in Cloud Computing Using Data Security. *Advances in Computer Science and its Applications* 1(1), 32–37 (2012)
8. Gampala, V., Inuganti, S., Muppidi, S.: Data Security in Cloud Computing with Elliptic Curve Cryptography. *International Journal of Soft Computing and Engineering (IJSCE)* 2(3), 138–141 (2012) ISSN: 2231-2307
9. Tebaa, M., El Hajji, S., El Ghazi, A.: Homomorphic Encryption Applied to the Cloud Computing Security. In: *Proceedings of the World Congress on Engineering (WCE 2012)*, London, U.K., July 4-6, vol. I, pp. 536–539 (2012)
10. Bhagat, A., Sahu, R.K.: Using Third Party Auditor for Cloud Data Security: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering* 3(3), 34–39 (2013)
11. Saravanan, N., Mahendiran, A., Venkata Subramanian, N., Sairam, N.: An Implementation of RSA Algorithm in Google Cloud using Cloud SQL. *Research Journal of Applied Sciences, Engineering and Technology* 4(19), 3574–3579 (2012)
12. Bhosale, P., Deshmukh, P., Dimbar, G., Deshpande, A.: Enhancing Data Security in Cloud Computing Using 3D Framework & Digital Signature with Encryption. *International Journal of Engineering Research & Technology (IJERT)* 1(8), 1–8 (2012)
13. Radhika, G., Satyanarayana, K.V.V., Tejaswi, A.: Efficient Framework for Deploying Information in Cloud Virtual Datacenters with Cryptography Algorithms. *International Journal of Computer Trends and Technology* 4(3), 375–380 (2013)
14. Marwaha, M., Bedi, R.: Applying Encryption Algorithm for Data Security and Privacy in Cloud Computing. *IJCSI International Journal of Computer Science Issues* 10(1(1), 367–370 (2013)
15. Nigoti, R., Jhuria, M., Singh, S.: A Survey of Cryptographic Algorithms for Cloud Computing. *International Journal of Emerging Technologies in Computational and Applied Sciences* 4(2), 141–146 (2013)
16. Mishra, A., Gupta, D.K., Sahoo, G.: BIT Mesra Ranchi, Jharkhand. The Secure Data Storage in Cloud Computing Using Hadamard Matrix. *International Journal of Engineering Science and Innovative Technology (IJESIT)* 2(2), 389–395 (2013)

# An Enhanced Strategy to Minimize the Energy Utilization in Cloud Environment to Accelerate the Performance

M. Vaidehi<sup>1</sup>, V. Suma<sup>2</sup>, and T.R. Gopalakrishnan Nair<sup>3</sup>

<sup>1,2,3</sup> Research and Industry Incubation Centre, Dayananda Sagar Institutions,  
Bangalore 560078, India

<sup>3</sup> Prince Mohammad University,  
ARAMCO International Endowed Chair,  
Technology and Information Management,  
Kingdom of Saudi Arabia

{dm.vaidehi, sumavdsce, trgnair}@gmail.com

**Abstract.** The need for power consumption prevailing in data centers and increased demand for cloud structures as service entity has raised several issues in power management or energy optimizations in large scale computing. The increased demand for virtualization of all resources demands further increased power utilization leading to more expensive operation cost. Therefore, an efficient strategy would resolve the aforementioned issue and also assure to provide the required Quality of Service (QoS) of the cloud computing system. This paper introduces an Enhanced Dynamic Voltage and Frequency Scaling (EDVFS) along with a Single Threshold (ST) value to minimize the energy utilization and Service Level Agreements (SLA). This paper provides a case study comprising of job arrival pattern which is recorded from a monitoring system in order to analyze the energy utilization especially at peak hours when there is a cloud burst. The investigation results have led to the introduction of a mathematical approach, where in we have arrived at an Enhanced Dynamic Voltage and Frequency Scaling for the cloud computing system (EDVFS).

**Keywords:** Cloud Computing, Cross layer, Cloud burst, Energy utilization, SLA Violations, VM Migration.

## 1 Introduction

The advancement in data centre engineering, system wise approaches and improved global networking have created the opportunity for realizing computing power or IT infrastructure support in a demand driven way. The possibilities of such systems capable of doing business in IT markets, led to the emergence of the cloud computing models.

The cloud computing is a pay-go-model providing services like IaaS (Infrastructure as Service), AaaS (Application as Service) and PaaS (Platform as Service). Virtualization is a mechanism of optimizing and integrating servers as one external

interface for end user. The challenge in virtualization is to maximize the resource efficiency and conserve energy.

Cloud system employing virtualization [17, 18, 19], provides services to the clients on a pay and use basis. IT organizations currently outsource the computational needs to the Cloud, rather than investing in procurement and maintenance of infrastructure, upgrading of software and hardware. The advantage of the Cloud system is that the customers are not concerned about the expenditure related to the resources. As energy is a costly resource, there is a need to optimize the operational cost of cloud datacenters by minimizing the energy utilization for various operations in the cloud. Virtualization being the key characteristics of Cloud, it provides a solution for energy inefficiency problem by creating multiple Virtual Machine (VM) instances on a physical system and thereby achieving efficient utilization of resources.

Minimization of energy utilization can be achieved by various other techniques such as load balancing, efficient scheduling techniques or by deactivating the idle VMs. In this paper, we have introduced an enhanced strategy to minimize the energy utilization and Service Level Violations. This is achieved by distributing the execution load among the VMs with efficient scheduling in an effective way. Further live migrations of VMs account for dynamic consolidation according to the requirement.

To achieve the aforementioned goal, the resources have to be allocated to the jobs in an efficient way so that the energy utilization is reduced. The paper is organized as follows. The section 2 provides the Research

Background. Section 3 specifies Research design. Section 4 discusses the Research Method, section 5 sections and 6 details the Simulation setup and Performance analysis and results respectively.

## **2 Research Background**

The technological advancement of cloud computing in the IT industry has created opportunities for many research avenues.

Anton Beloglazov et al. have implemented an energy efficient resource management system in Virtualized Cloud Data Centers for addressing the problem of reducing the operational cost in the virtualized Cloud data centers while providing the required Quality of Service (QoS) through Single-Threshold and double-threshold policy aiming at Minimization of Migrations. [1].

Lskrao Chimakurthi et al. have proposed a Power Efficient Resource Allocation for Clouds using Ant Colony Framework for allocating the cloud resources to the applications without violating the given Service Level Agreement (SLA) for performance such as throughput and response time. [2].

V.K. Mohan Raj et al. have suggested Power Aware Provisioning in Cloud Computing Environment by implementing a dynamic resource provisioning framework for virtualized server environment. This framework achieves power efficiency by using optimum power efficient allocation and workload forecasting scheme [3].

Shuo Liu et al. have proposed On-Line Real-Time Service Allocation and Scheduling for Distributed Data Centers by implementing an energy efficient, profit and penalty aware allocation and scheduling approach for distributed data centers in a multi-electricity-market environment. [4].

### 3 Research Design

An efficient architecture plays a vital role in minimizing the energy utilization in cloud datacenters. Further, the cloud service providers expect faster return on investment (ROI) with low cost of ownership while providing services to the clients. Henceforth, an efficient architecture for resolving these issues becomes a mandatory in the cloud computing environment

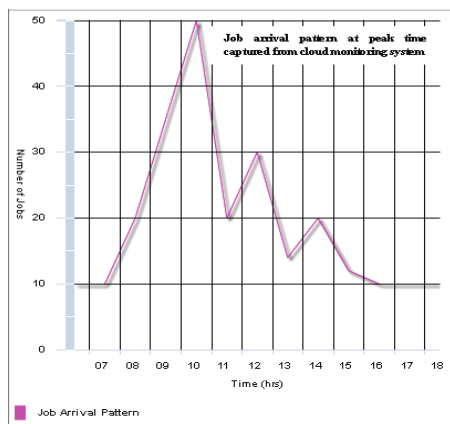
The secondary data provisioned by one of the leading cloud service providing industry has supported this research to analyze the efficiency of the service provider exclusively at peak hours with minimized energy utilization and SLA violation.

#### 3.1 Case Study

Figure 1 depicts the graph of job arrival pattern which is sampled and shown in Table I. It depicts the job arrival pattern captured from the cloud monitoring system. From the table, it is visible that the pattern of job arrival is not uniform throughout the day. It is observed from the graph that there is a spike in the demand for resources by the jobs for computation during the peak hour which consequently results in tremendous energy utilization. Henceforth, this needs to be resolved using an efficient strategy.

**Table 1.** Number Of Jobs Arrived At Peak Hours

Peak Time (hours)	Number of jobs arrived
7	10
8	20
9	35
10	50
11	20
12	30
13	14
14	20
15	12
16	10
17	10



**Fig. 1.** A Graph of Job Arrival Pattern At Peak Hours

## 4 Research Method: Enhanced DVFS Approach to Reduce Energy Utilization

### 4.1 DVFS (Dynamic Voltage and Frequency Scaling)

Dynamic voltage and frequency scaling has however proven to be a feasible solution to reduce processor power utilization. It is employed in microprocessors. DVFS is a dynamic technique in which both the clock frequency and power utilization vary during operations [9].

DVFS operates on the assumption that the processors need not always have to operate at full speed to process all the tasks. If the processor's workload does not require all available CPU performance, then the processor can be slowed down to the lowest available performance level to meet the required demand. Further, DVFS technique operates on the assumption that the performance varies with the Voltage (V) while power utilization varies as square of the Voltage ( $V^2$ ). A DVFS – enabled processor is operated on a set of supply voltage (V) and a set of processor frequencies (F).

$$V = \bigcup_{m=1}^M (v_m)$$

$$F = \bigcup_{m=1}^M (f_m)$$

M = Total number of processors in the computing system

m = The first processor in the computing system

And

$v_m$  is the operating voltage of the m-th processor.

$f_m$  is the operating frequency of the m-th processor.

However the energy utilization by processor is composed of dynamic energy utilization and static energy utilization.

Hence,

$$\text{Total Energy} = \text{Energy}_{\text{dynamic}} + \text{Energy}_{\text{static}}$$

The dynamic power utilization is given as:

$$\text{Power}_{\text{dynamic}} = A.C. V^2. F$$

Where,

A is the percentage of active gates. C is the total capacitance load.

V is the supply voltage.

$$\text{Power}_{\text{dynamic}} = A \cdot C \cdot V^2 \cdot F$$

F is the processor frequency.

Hence

$\text{Power}_{\text{dynamic}}$  is the dynamic power.

$\Delta t$  is the time period.

$$\text{Energy}_{\text{dynamic}} = \sum_{\Delta t} \text{Power}_{\text{dynamic}} \Delta t$$

#### 4.2 EDVFS (Enhanced Dynamic Voltage and Frequency Scaling)

EDVFS in the cloud computing environment is applicable to the virtual machines (VM). The EDVFS – enabled VM is operated on a set of operating voltages and utilized voltages and on set of operating frequencies and utilized frequencies.

$$V = \bigcup_{m=1}^M (v_m, v_{um})$$

$$F = \bigcup_{m=1}^M (f_m, f_{um})$$

Where,

$v_m$  is the operating voltage of the m-th VM

$f_m$  is the operating frequency of the m-th VM

$v_{um}$  is the utilized voltage of the m-th VM

$f_{um}$  is the utilized frequency of the m-th VM

Where operating voltage is the voltage used for creation of the VMs in the datacenter, for scheduling the jobs in queue, for allocating of the resources to the jobs and so on. Whereas utilized voltage is the voltage used only for computations. Similarly, operating frequency is the frequency required for creation of the VMs in the datacenter, for scheduling the jobs in queue, for allocating of the resources to the jobs and so on. Whereas utilized frequency is required only for computations.

The energy utilization by VM is composed of dynamic energy utilization and static energy utilization.

$$\text{Total Energy} = \sum \text{Energy}_{\text{dynamic}} + \text{Energy}_{\text{static}}$$

The dynamic power utilization is given as:  $\text{Power}_{\text{dynamic}} = A.C.V^2.F$

Where,

A = Percentage of active Virtual machines. C = Total Load.

V = Supply Voltage.

F = Virtual Machine Frequency.

Then

$$\text{Energy}_{\text{dynamic}} = \sum_{\Delta t} \text{Power}_{\text{dynamic}} \Delta t$$



Where,

Power dynamic is the Dynamic Power.

$\Delta t$  is the time period.

### 4.3 Single Threshold (ST) Value

As the nature of the cloud is VM creation which supports the feature of scalability and elasticity it is essential to have a threshold to avoid SLA violation.

It is therefore essential to have a threshold value for these features in order to minimize SLA violation. The significance of Single Threshold (ST) is to set the upper utilization threshold for service providers and placing of VMs while keeping the CPU's total utilization below this threshold.

At every time frame all VMs are reallocated using Modified Best Fit Decreasing algorithm with additional condition of keeping the upper utilization threshold not violated. The new placement is achieved by live migration of VMs [10].

Here we present a comparison study of the Enhanced Dynamic Voltage and Frequency Scaling with two different values Single Threshold (ST) i.e.  $ST=0.1$  and  $ST=0.8$ .

## 5 Simulation Setup

In our research we have proposed an Enhanced DVFS (EDVFS) having a Single Threshold=1.0, and EDVFS having a Single Threshold=0.8 to minimize the energy utilization and SLA violation. To prove the aforementioned, we implemented the same using CloudSim tool. To achieve the expected results five different configurations namely C1 to C5 has been considered comprising of Number of Hosts, Number of Virtual Machines and Number of Cloudlets. Table 2 presents the five different configurations. The Hosts shown in Table 2 comprise the following parameters.

*Host (MIPS Rating,  
RAM, Storage, Bandwidth)*

Table 3 presents the configuration of these parameters. Similarly, VMs shown in Table 2 comprise the following parameters.

*VM (MIPS Rating,  
Number of CPUs, RAM,  
Bandwidth)*

Table 4 present the configuration of these parameters. Similarly, Cloudlets shown in Table 2 comprise the following parameters.

*Cloudlets (Length,  
Number of CPUs, File  
Size, output Size)*

Table 5 presents the configuration of these parameters.

**Table 2.** Data Centre Configuration

Configuration No.	Number of Hosts	Number of VMs	Number of Cloudlets
C1	10	10	10
C2	10	20	20
C3	40	30	30
C4	30	10	40
C5	50	20	50

**Table 3.** Host Configuration

Configuration No.	MIPS Rating	RAM	Storage	Bandwidth
C1	1000, 2000, 3000	10000	1000000	100000
C2	1500, 2500, 3500	1000	100000	10000
C3	1500, 3000, 4500	100000	10000	100000
C4	500, 1500, 2000	10000	1000000	100000
C5	1000, 2000, 3000	100000	1000000	100000

**Table 4.** Virtual Machine Configurations

Configuration No.	MIPS Rating	Number of CPUs	RAM	Bandwidth
C1	250, 500, 750, 1000	1	128	2500
C2	500, 750, 1000, 1250	1	256	3000
C3	250, 750, 1250, 1750	1	1024	3000
C4	500, 1500, 2500, 3500	1	512	2500
C5	1000, 2000, 3000, 4000	2	1024	3000

**Table 5.** Cloudlet Configuration

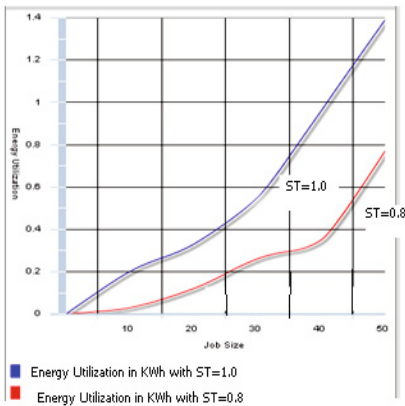
Configuration No.	Length	Number of CPUs	File Size	Output Size
C1	150000	1	150	150
C2	150000	1	300	300
C3	250000	2	300	300
C4	1500000	1	200	200
C5	2500000	2	300	300

## 6 Performance Analysis and Results

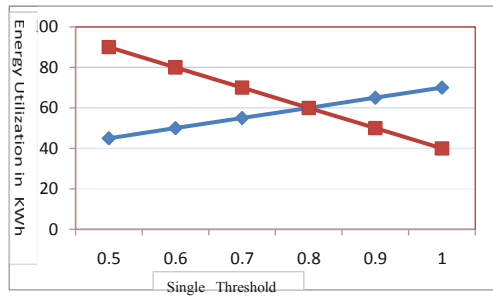
The Simulation of the proposed research method has given appreciable results in terms of energy utilization and

SLA violation. Table 6 presents a comparative analysis in terms of energy utilization and SLA violation.

It can be observed from the Figures 2 that the Energy utilization is less with a Threshold Utilization of 0.8



**Fig. 2.** Energy utilization with different Single Threshold



**Fig. 3.** Graph depicting the Service Level Agreement Violation and Energy Utilization at Single Threshold = 0.8

**Table 6.** Comparative Results of the two approaches

Configuration No.	Approach Parameters	ST EDVFS with Single Thresh Hold=1.0	ST EDVFS with Single Thresh Hold=0.8
C1	Energy Utilization (KWh)	0.18	0.08
	SLA violation (%)	97.02	17.14
	No. of VM Migrations	12	8
	Number of SLA Violation	2481	5
C2	Energy Utilization (KWh)	0.83	0.21
	SLA violation (%)	99.36	33
	No. of VM Migrations	21	5
	Number of SLA Violation	5398	3
C3	Energy Utilization (KWh)	1.43	0.49
	SLA violation (%)	99.77	33.33
	No. of VM Migrations	13	5
	Number of SLA Violation	5730	2

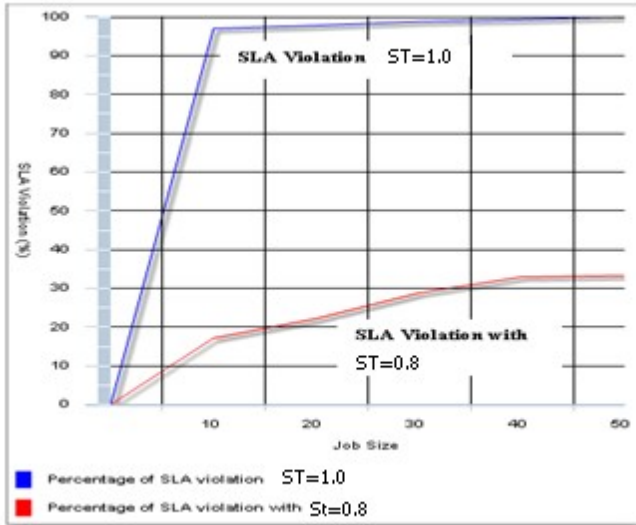


Fig. 4. SLA violation with Single Threshold Values (ST) =1.0 and ST=0.8

## 7 Conclusion

The advancement in Cloud technology has led to an increased demand for the Cloud Services. As Cloud operates on virtualization of resources, there is an increased demand for power utilization leading to high operation cost. To minimize the energy utilization by the large scale-virtualized datacenters, the implementation of EDVFS with Single Threshold values of 1.0 and 0.8 was implemented. The result analysis showed minimum energy utilization and SLA violation therefore enhancing the Cloud performance.

## References

1. Beloglazov, A., Buyya, R.: Energy Efficient Resource Management in Virtualized Cloud Data Centers. In: 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (2010)
2. Chimakurti, L., Madhu Kumar, S.D.: Power Efficient Resource Allocation for Clouds Using Ant Colony Framework. In: 2011 arXiv: 1102.2608v1 [cs.DC] (February 13, 2011)
3. Mohan Raj, V.K., Shriram, R.: Power Aware Provisioning in Cloud Computing Environment. In: International Conference on Computer, Communication and Electrical Technology, ICCET 2011, March 18-19 (2011)
4. Liu, S., Quan, G., Ren, S.: On-line Real Time Service Allocation and Scheduling for Distributed Data Centers. In: 2011 IEEE International Conference on Services Computing (2011)

5. Bai, Y., Liu, S., Sha, M., Lu, Y., Xu, C.: An Energy Optimization Protocol Based on Cross-Layer for Wireless Sensor Networks. 2008 Journal of Communications 3(6), 27–34 (2008)
6. Kim, K.H., Buyya, R., Kim, J.: Power Aware Scheduling of Bag – of – Tasks Application with Deadline Constraints on DVS – Enabled clusters. In: CCGRID, pp. 541–548 (2007)
7. Ge, R., Feng, X., Cameron, K.: Performance-Constrained Distributed DVS Scheduling for Scientific Applications on Power-Aware Clusters. In: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing. IEEE Computer Society, Washington, DC (2005)
8. Wang, L., von Laszewski, G., Dayal, J., Wang, F.: Towards Energy Aware Scheduling for Precedence Constrained Parallel Tasks in a Cluster with DVF. In: 2010 10th IEEE/CAN International Conference on Cluster, Cloud and Grid Computing (2010)
9. Wolf, W.: Modern VLSI design, 3rd edn. Pearson Education (2007)
10. Beloglazov, A., Abawajy, J., Buyya, R.: Energy-Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing. In: Future Generation Computer Systems (May 2011)
11. Mazzucco, M., Dyachuk, D., Deters, R.: Maximizing Cloud Provider’s Revenues via Energy Aware Allocation Policies. In: 2010 IEEE 3rd International Conference on Cloud Computing, pp. 131–138. IEEE (2010)
12. Rajkumar, B., Anton, B., Jemal, A.: Energy Efficient Management of Data center Resources for computing: Vision, Architectural Elements and Open Challenges. In: International Conference on Parallel and Distributed Processing Techniques and Applications (July 2010)
13. Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A.F., Buyya, R.: CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environment and evaluations of resource provisioning algorithms. In: Software: Practice and Experience, Wiley Press, NY (2010)
14. Kliazovich, D., Boury, P., Khan, S.U.: DENS: Data Center Energy Aware Network – Scheduling. In: 2010 IEEE/ACM International Conference on and Cyber, Physical and Social Computation, pp. 69–75 (2010)
15. Gajjar, S.H., Pradhan, S.N., Dasgupta, K.S.: Cross-Layer Architectural approaches for Wireless Sensor Networks. In: 2011 IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp. 557–562 (2011)
16. Suma, V., Deshpandey, B., Vaidehi, M., Gopalakrishnan Nair, T.R.: Cloud Computing for Microfinance. In: International Conference on Systemics, Cybernetics and Informatics (ICSCI 2012), Hyderabad, India Hyderabad, India, February 15-18 (2012)
17. Gopalakrishnan Nair, T.R., Vaidehi, M., Suma, V.: Improved Strategies for Enhanced Business Performance in Cloud-based IT Industries. In: 26th Indian Institute of Engineers Congress, IEI Congress 2011, Bangalore, December 16-18 (2011)
18. Gopalakrishnan Nair, T.R., Vaidehi, M., Rashmi, K.S., Suma, V.: An Enhanced Scheduling Strategy to Accelerate the Business Performance of the Cloud System. In: Satapathy, S.C., Avadhani, P.S., Abraham, A. (eds.) Proceedings of the InConINDIA 2012. AISC, vol. 132, pp. 461–468. Springer, Heidelberg (2012)
19. Gopalakrishnan Nair, T.R., Vaidehi, M.: Efficient Resource Arbitration and Allocation Strategies in Cloud Computing through Virtualization. In: 2011 International Conference on Cloud Computing and Intelligence Systems (CCIS). IEEE Xplore, China (2011)

# An Efficient Approach to Enhance Data Security in Cloud Using Recursive Blowfish Algorithm

Naziya Balkish<sup>1</sup>, A.M. Prasad<sup>2</sup>, and V. Suma<sup>3</sup>

<sup>1</sup> Computer Science and Engineering,  
Department of Information Science and Engineering, Dayananda Sagar College of Engineering,  
Bangalore, India

naziya.balkish@gmail.com

<sup>2</sup> Department of CSE, Dayananda Sagar College of Engineering, Bangalore, India  
prasaddsce@gmail.com

<sup>3</sup> Research and Industry Incubation Centre (RIIC), Department of Information Science and  
Engineering, Dayananda Sagar College of Engineering, Bangalore, India  
sumavdsce@gmail.com

**Abstract.** Cloud computing and Services have gained popularity both from industry and academia since both enable the rapid development of the distributed computing in areas of collaborative research and development, healthcare, grid-enabled applications, enterprise computing infrastructure, military applications and domestic Security. The existence of the Service providing system depends on the customer's requirements. Software Engineering is one of the disciplines which orients towards accomplishment of quality in technology through best principles and practices. Applying these software engineering quality attributes enables one to achieve customer satisfaction in cloud applications. Security which is one of the significant quality attributes is also deemed to be one of the biggest issues in cloud computing since it needs to be managed and control huge amount of data.

Cryptography is one of the popular methods of achieving data security, whose objective is to covert information from original form into unreadable form. Despite of advancement in technology and existence of various security models for realizing secured encryption process, there still prevail various security issues during data transfer. The objective of this paper is to enhance the Customers satisfaction by providing an improved level of security through reduced possibility for hacking through Blowfish algorithm by varying the number of rounds and decide the value of N, which is complex and confuses the hacker.

**Keywords:** Security, Cryptography, Software Quality, Blowfish, Customer satisfaction, Non-functional requirement.

## 1 Introduction

Quality and customer satisfaction are the responsibilities of any Service provider. This can be achieved through enhancing the functional and non-functional

requirements which are the prime areas of operation in software engineering. The functional requirements describe the behavior of the system as it depends on the system functionality. The non-functional requirements elaborate on the performance characteristics such as Accessibility, Privacy, Quality, Security, Efficiency, Extensibility and so on. The Cloud computing is one of the most significant and upcoming technology. This advancement in technology has enabled to handle huge amount of information with less investment on resources, time and cost.

One of the challenges in cloud environment is Security. Security in the cloud is achieved, in part, through third party controls and assurance much like in traditional outsourcing arrangements. Many cloud vendors implement their own proprietary standards and security technologies and implement different security models, which need to be evaluated on their own merits. The security models generally depend on the cryptographic technique applied. Cryptography converts the plaintext (original message) into cipher text (unreadable message) which is a process well known as Encryption for ensuring security issues (Stalling 1999). However, encryption should not be susceptible for attacks. Some of the encryption techniques that are popularly used include techniques such as bit level encryption like substitution box, permutation box, encoding and rotation, Caesar's cipher method, poly alphabetic substitution method.

Cryptanalysis technique is a study of cipher text. There are different cryptanalysis techniques accessible to crack the encryption algorithms. Most of the algorithms are either may be block cipher or stream cipher [6]. However, these types of algorithms can be easily attacked by performing different cryptanalysis techniques such as Brute force attack, man-in-the-middle attack, meet-in-the-middle attack and so on [8]. It is worth to note that few of the popular algorithms are broken by the eavesdroppers despite of technological betterment [6]. The majority of the existing systems maintain text encryption other than media types. As intruders and eavesdroppers had shown their excelling skills towards breaking the encryption algorithms almost in all essential and rational areas like banking, military, defense, network, a need for "virtually strong and infeasible to get attacked" algorithm becomes very important. Cryptography is well known and widely used method that operates information to crypt their extinction. It also protects information by transforming it into unreadable format (Stalling 1999). Since, decryption of text is through the knowledge of secret key, the cryptographic method ensures that it scrambles a message which cannot be understood.

One of the best way to enhance security is to strengthen a cryptographic method which is possible through increasing the strength of key. The objective of this work is to enhance Security in cloud by enhancing the Blowfish algorithm. This is done by varying the number of rounds and determining the value of N which is more complex [6]. This mode of implementing the algorithm creates ambiguity in predicting the key and thereby makes hacking process quite difficult. During the process of encryption by keeping the S-boxes and sub key as constant, every word is encrypted using different number of rounds (N). This further enhances the security of encryption process in cloud.



The organization of the paper is as follows. Section 1 provides introduction. Section 2 provides brief information about various Related work made during this investigation. Section 3 discusses the proposed system. Further, section 4 details the structure of the blowfish algorithm. Section 5 provides the working procedure of the blowfish algorithm. Section 6 specifies the advantage of this algorithm section 7 shows the results and snapshots of the algorithm while section 8 concludes the work and future enhancement.

## 2 Related Work

Research is progressing continuously towards achieving security in cloud. Many cryptographic algorithms prevail to provide security to the user such that message would remain safe at the time of communication. Some of the widely implemented and complex cryptographic algorithms include Advance Encryption Standard, Triple Data Encryption Standard, Rivert Cipher 4 and BLOWFISH which provides high security and liability towards network security. Nevertheless, hacking has become a common practice in society and thus challenges the security of cryptographic algorithms.

Authors made a comparative analysis of Encryption Algorithms for Data Communication. The experimental results show the comparison of three algorithm AES, DES and RSA and they authors state that RSA has very smaller output byte compared to AES and DES algorithm [7].

Authors discussed the popular symmetric key encryption algorithms such as DES, TRIPLE DES, AES, and Blowfish. They express that symmetric Key algorithms run faster than asymmetric Key algorithms such as RSA etc and the memory necessity of Symmetric algorithms is lesser than Asymmetric encryption algorithms. Further, the security aspect of Symmetric key encryption is advanced than Asymmetric key encryption [8].

Authors in provided a new approach to join two basic forms of ciphers, such as block cipher and stream cipher which was further enhanced on the plaintext. The authors have suggested using a method which uses one with the plaintext and other with key [11]

Author presented simulation results which depicts that Blowfish has a better performance than other common encryption algorithms used.

Authors in [4] proposed an approach to decrease the execution time of blowfish algorithm by using modified F function.

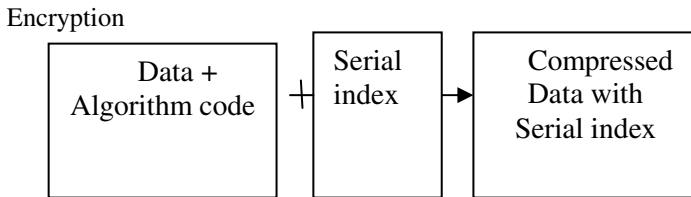
Authors in [5] have made an analysis of Energy Consumption of RC4 and AES Algorithms in Wireless LANs. Further, the authors had carried out a study for different secret key algorithms such as DES, 3DES, AES, and Blowfish. The results showed that Blowfish had comparatively a good performance than other algorithms. Thus, this research focused upon enhancing the Blowfish algorithm in order to reduce the probability of hacking.

**Existing Blowfish Algorithm**

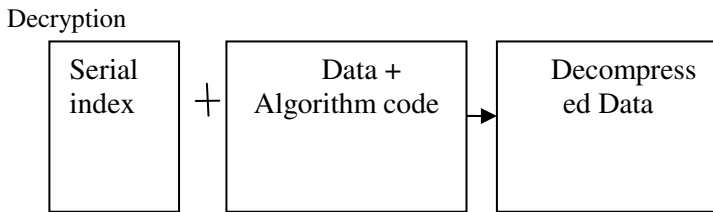
Blowfish is a symmetric block cipher that uses a modified Fiestel network structure which takes 16 rounds for encryption and decryption improvement. It takes a variable key length from 32bits to 448 bits. The strength of the Blowfish algorithm depends on its sub key generation and its encryption.

**3 Proposed System**

The proposed system comprises of the encryption and decryption models for the enhanced Blowfish algorithm. The figure 1 depicts the block diagram of encryption process for the proposed algorithm, here, the data to be protected is encrypted based on the enhanced algorithm, a Serial index is provided for the encrypted compressed data for further reference by the client or Cloud service provider. Similarly in decryption the data is decrypted and decompressed based on the proposed algorithm, figure 2 depicts the same.



**Fig. 1.** Data compression subsequently the service provider verifies user’s access right from the database



**Fig. 2.** Data decompression

**4 Enhanced Blowfish Algorithm**

Figure 3 and 4 depicts the encryption and decryption of data using enhanced approach. Figure 3 infers that initially user requests the security provider for the algorithm code and Serial index (SI). Security provider verifies the request and provides the same. Here algorithm code and SI index is used for encryption. The file

is compressed and is transmitted to cloud along with SI index and algorithm code, figure depicts the same.

### 5 Working Procedure

The encryption of the confidential data happens in  $N$  number of rounds. Here the client has the privilege to choose  $N$  which is the number of times the data has to be encrypted. The encryption is computed in three steps

- a. Generate random prime number  $P$ .
- b. Compute the primitive root  $Q$  of  $P$ .
- c. Multiply  $Q$  with  $P$  which is the number of rounds  $N$ .

While, random prime number is used in conniving the number of rounds, the file has to be encrypted such that the attacker will not accurately decrypt the message.

During the process of decryption the user requests the decrypted file by sending the SI index and algorithm code to the service provider. Here the service provider upon receiving the request verifies the authenticity of the request. Service providers decompress the file and decrypt it. The same is depicted in figure 3 and 4.

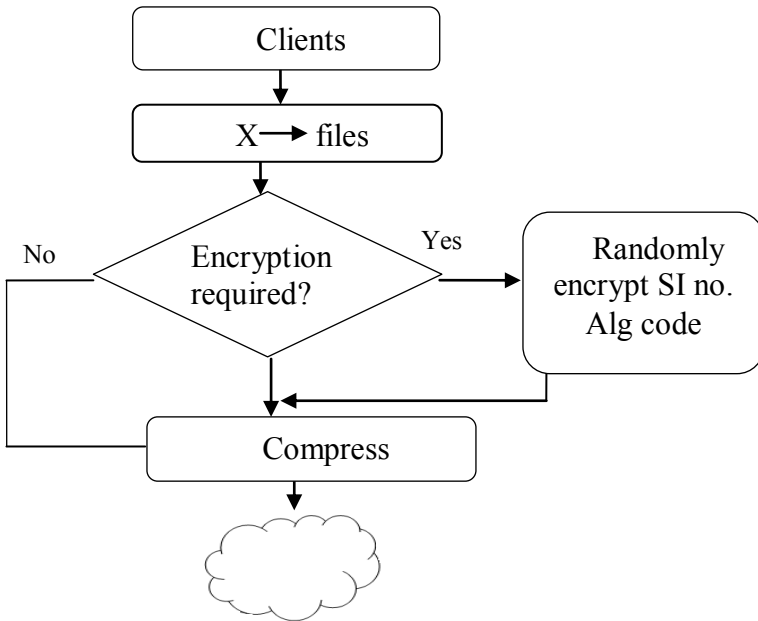


Fig. 3. Data encryption

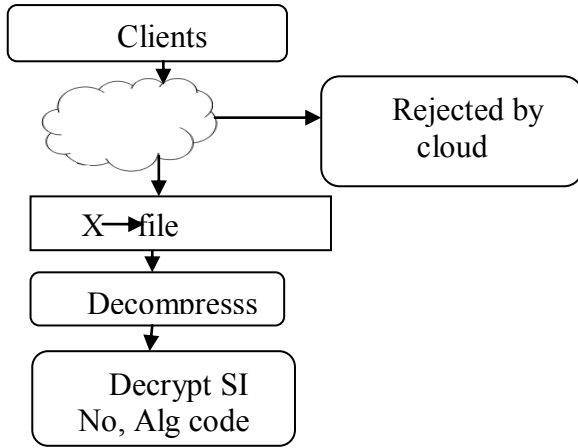


Fig. 4. Data decryption

## 6 Advantages of Enhanced Blowfish Algorithm

- The Enhanced Blowfish algorithm is more secure than normal encryption algorithm since the encryption is performed in N rounds.
- By varying the number of rounds of file encryption, even if hacker manages to decrypt, it results in obtaining a twisted file.
- The number of rounds of encryption and decryption is unpredictable.
- The size of the text remains same after the encryption thus avoiding distrust.

## 7 Results

Figure 5 and 6 depicts the snapshots of the simulation output for the enhanced blowfish algorithm. The encryption and decryption process is depicted in figure 5, where the value of N and file to be encrypted is specified. Figure 6 specifies the case of decryption failure.

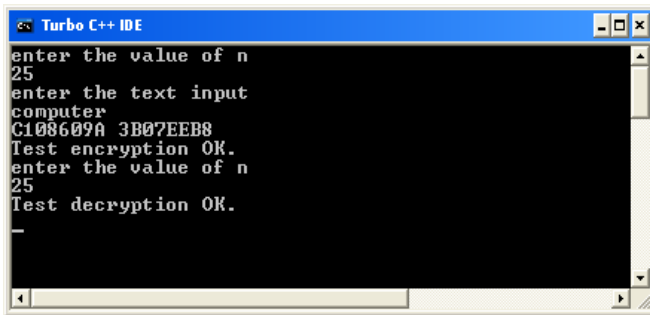
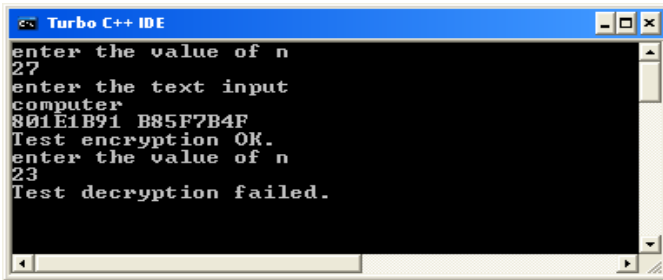


Fig. 5. Decryption with the correct secured key “n”



```

Turbo C++ IDE
enter the value of n
27
enter the text input
computer
801E1B91 B85F7B4F
Test encryption OK.
enter the value of n
23
Test decryption failed.

```

Fig. 6. Decryption with the incorrect secured key “n”

## 8 Conclusion

Most of the IT industries are inclining towards the Cloud technology. It provides on demand services to its clients. The Cloud models mainly depends on large and consolidate data centers in order to provide best services to its clients. This upcoming technology has few shortfalls like efficient scheduling, resource allocation, availability, reliability, security and disaster recovery. This paper focuses on providing enhanced Security to its clients since this model operates on shared pool of distributed resources. Security which is one of the non-functional requirements, this is achieved, in part, through third party controls and assurance much like in traditional outsourcing arrangements. But since there is no common cloud computing security standards there are additional challenges associated with this.

This paper proposes an Enhanced Blowfish algorithm. This recursive algorithm in this paper has major advantage more than the standard Blowfish algorithm. Through implementation of the proposed algorithm it can be observed that by varying the number of rounds and determining the values of N increases the complexity, hence by doing this the level of Security is enhanced. Therefore there is a reduced possibility of hacking. Here the S-boxes and sub keys are kept constant and every word is encrypted using different number of rounds. In the proposed method only one key is used, the strength of this algorithm can be increased by using different keys at each step. Incorporating the proposed strategy would enhance the Security thereby accelerating the business performance.

### Future Enhancement

As discussed the cloud computing system is a very promising model that can cope with the Security limitations generally occurring in a public cloud environment, while still being able to support many of the economic advantages of public cloud computing. In future the data Security can be enhanced just not in the data center but also considering the network Security, which is essential to transmit the protected data accurately to the desired client. This analysis can be done using advanced versions of simulators to get better results. MATLAB, Ns2, Ns3, OPNET, NetSim etc. can be used for better results of cryptographic applications. Enhanced encryption and decryption techniques would also improve the security service.

## References

- [1] Balkish, N., Prasad, A.M., Suma, V., Vaidehi, M.: A Survey on factors influencing security in cloud. In: 10th International Conference on Advanced Computer Science and Information Technology (ACSIT), Chennai, India,
- [2] Balkish, N., Prasad, A.M., Suma, V.: Comparative Analysis of Cryptographic Algorithms for Better Utilization. In: 2nd International Conference on Recent Trends in Engineering Technology (ICRTET 2013), Bangalore, India (April 30, 2013)
- [3] Al Tamimi, A.-K.: Performance Analysis of Data Encryption Algorithms
- [4] Vaithiyathan, V., Manikandan, G., Krishnan, G.: A novel approach to the performance and security enhancement using blowfish algorithm. *Inter. J. Adv. Res. Comp. Sci.* 1(4), 451–454 (2010)
- [5] Prasithsangaree, P., Krishnamurthy, P.: Analysis of Energy Consumption of RC4 and AES Algorithms in Wireless LANs. In: *The Proceedings of the IEEE GLOBECOM 2003*, pp. 1445–1449 (2003)
- [6] Manikandan, G., Sairam, N., Kamarasan, M.: A New Approach for Improving Data Security using Iterative Blowfish Algorithm. *Research Journal of Applied Sciences, Engineering And Technology* 4(6), 603–607 (2012) ISSN: 2040-7467
- [7] Seth, S.M., Mishra, R.: Comparative Analysis Of Encryption Algorithms For Data Communication. *IJCST* 2(2), 192–192 (2011)
- [8] Agrawal, M., Mishra, P.: A Comparative Survey on Symmetric Key Encryption Techniques. *International Journal on Computer Science and Engineering (IJCSSE)* 4(5), 877–882 (2012)
- [9] Manikandan, G., Manikandan, R., SundarGanesh, G.: A New Approach for generating strong key in RC4 algorithm. *J. Theory. Appl. Inf. Technol.* 24(2), 113–119 (2011a)
- [10] Manikandan, G., Krishnan, G., Sairam, N.: A unified block and stream cipher based file Encryption. *J. Global Res. Comp. Sci.* 2(7), 53–57 (2011b)
- [11] Manikandan, G., Manikandan, R., Rajendiran, P., Krishnan, G., SundarGanesh, G.: An Integrated block and stream cipher approach for key Enhancement. *J. Theory. Appl. Inf. Technol.* 28(2), 83–87 (2011c)
- [12] Minaam, D.S.A., Abdual-Kader, H.M., Hadhoud, M.M.: Evaluating the Effects of Symmetric Cryptography Algorithms on Power Consumption for Different Data Types. *International Journal of Network Security* 11(2), 78–87 (September)
- [13] Zunnurhain, K., Vrbsky, S.: Security Attacks and Solutions in Clouds. In: 2nd IEEE International Conference on Cloud Computing Technology and Science, Indianapolis, USA, November 30-December (2010)
- [14] Suma, V., Deshpandey, B., Vaidehi, M., Gopalakrishnan Nair, T.R.: Cloud Computing for Microfinance. In: *International Conference on Systemics, Cybernetics and Informatics (ICSCI 2012)*, Hyderabad, India Hyderabad, India, February 15-18 (2012)
- [15] Stallings, W.: *Cryptography and Network Security Principles and Practices*. Prentice Hall, New Delhi
- [16] Gopalakrishnan Nair, T.R., Suma, V.: Defect management using pair metrics, DI and IPM. *Crosstalk, The Journal of Defense Software Engineering* 24(6), 22–27 (2011)
- [17] Suma, V., Gopalakrishnan Nair, T.R.: Effective defect prevention approach in software process for achieving better quality levels. Paper presented at the Fifth International Conference on Software Engineering, Singapore (August 2008)

# Effective Disaster Management to Enhance the Cloud Stability

O. Mahitha<sup>1</sup> and V. Suma<sup>2</sup>

<sup>1</sup> Computer Science and Engineering,  
Department of Information Science and Engineering, Dayananda Sagar College of Engineering,  
Bangalore, India

Mahitha14@gmail.com

<sup>2</sup> Research and Industry Incubation Centre (RIIC), Department of Information Science and  
Engineering, Dayananda Sagar College of Engineering, Bangalore, India  
sumavdsce@gmail.com

**Abstract.** Cloud computing is the state-of-art technology in IT industry. One of the major contributions of the Cloud is its capacity to handle enormous amount of data for either processing or storing. The most significant concern in Cloud is that it is prone to Disaster. A short period of downtime would result in significant financial loss. Therefore, it is essential for the cloud service providers to function with efficient data management and data recovery approach at the data centers which are one of the customer requirements. The existence and continuity of any technology depends on customer satisfaction. The satisfaction for any service or products can be achieved by applying the basic principles of Software Engineering. Customer satisfaction depends on parameters like business needs and the resultant Service Quality provided. This paper focuses on a survey of a novel approach for Disaster management through Efficient Scheduling mechanism and Efficient Load balancing technique to control Disaster thereby enhancing the Quality.

**Keywords:** Cloud Computing, Disaster Management, Deadlock, Starvation, Aging, Load balance, Quality.

## 1 Introduction

Customer satisfaction is one of the prime intentions of any Service provider. Such satisfaction would bring long term business growth and better business returns. The depth of customer satisfaction depends upon the Quality of service provided this can be achieved through the basic principles of Software Engineering [10][11].

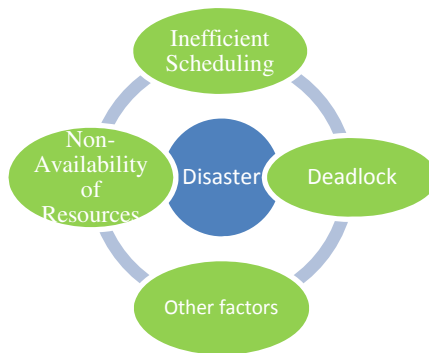
Quality means the client and the service provider agree upon a required level of quality to be achieved within a defined cost and time constrains. The requirements could be screened as functional requirement and non-functional requirement. Here, the functional requirements are associated with specific functions or Jobs the system must support. While the non-functional requirements are constrains like the availability, performance, scalability, reliability etc of the computing system[12][13].

The cloud maintains data storage to store enormous amount of data which is contained in various cloud services through different cloud models. Due to the growing demand for the cloud services a short period of downtime would result in significant financial loss and henceforth the business failure. It is highly essential to recover the data from disasters in cloud. This research focuses on minimizing the Disaster through efficient scheduling and Deadlock avoidance in the virtualized environment.

### *Disaster in Cloud*

Generally in the cloud environment, jobs of various sizes and different infrastructure requirement pop up concurrently. The major upcoming problem in cloud computing is effective disaster management. Disaster is an unexpected event that occurs in the cloud environment during its operational life time [1].

The major few possible rationales for the occurrences of disasters are presented in Fig 1 [1], it depicts Inefficient Scheduling, Deadlock, and Non - availability of resources and The other factors which influence Disaster are catastrophic failures (earthquakes, floods, fire, tsunami, etc).



**Fig. 1.** Factors Influencing Disaster in Cloud

The Cloud model performance is influenced by efficient scheduling and resource allocation in the virtualized environment. Identifying and implementing the appropriate scheduling techniques would prevent the occurrence of Disaster. In the computing model, jobs both high priority and low priority would arrive. Generally jobs of higher priority (Criticality and ROI) [3] will be provisioned with the requested resources. The lower priority jobs will be ignored. Ignoring lower priority jobs would lead to Starvation [4] and ultimately result in Disaster. Efficient Scheduling is one of the explication to enhance the system performance.. As the requested resources are provisioned on demand, it is necessary for the resources to be highly available, failure of which will lead to overloading. Henceforth to enhance the Cloud performance, efficient scheduling and Load balancing techniques are required.

Hence this paper presents a survey on the prime features which would enhance the stability of the Cloud through Disaster Management.



The organization of the paper is as follows Section 2 presents the Related work, section 3 presents the Research Work, Section finally in section 4 the Conclusion is presented.

## 2 Related Work

Vijaykumar Javaraiah proposed simple Linux box as a solution for the data backup and disaster recovery. Since Linux box does not have higher bandwidth, it is not possible to achieve complete server backup at a point of time. This approach also makes the process of migration from one cloud provider to another cloud provider [6].

However, Manish Pokharel et al. recommended a reliable Markov model approach for disaster recovery. This model has low cost to implement and further the model has lower energy loss. The authors were able to achieve high availability, high survivability and short downtime with less cost [7].

Sinung Suakanto et al. used remote monitoring system. The authors work presents how the remote sensing system could be run in cloud computing architecture. They further propose FTR\_HTTP method, which can be used to ensure the quality of the data transmitting from remote client into server [8].

Kruti Sharma and Kavita R Singh suggested remote server backup architecture to overcome back-up and to recover data. With this technique the authors have achieved data privacy, security, reliability, cost effectiveness, appropriate timing and easy migration from one server to another server [9].

## 3 Research Work

Cloud Computing is a pay as you go model. Disaster is one of the major reasons influencing the performance of the Cloud. An efficient resource model is tremendously required to avoid the Disaster in the cloud. To improve the Cloud Performance it is necessary to have an effective Job scheduling Strategy and an efficient load balancing strategy to minimize the Disaster. Hence, it is necessary that the computing model must support the utilization of available resources and execution of each task within a particular time in the cloud system to avoid Starvation and Deadlock of jobs. It is manifesting from the analysis that scheduling and load balancing strategies are the influencing factors in cloud to reduce job rejections and thereby increase the business performance of the system.

Here in the research design two interdependent modules are implemented for Disaster Management.

- *Efficient Scheduling*
- *Deadlock Avoidance*

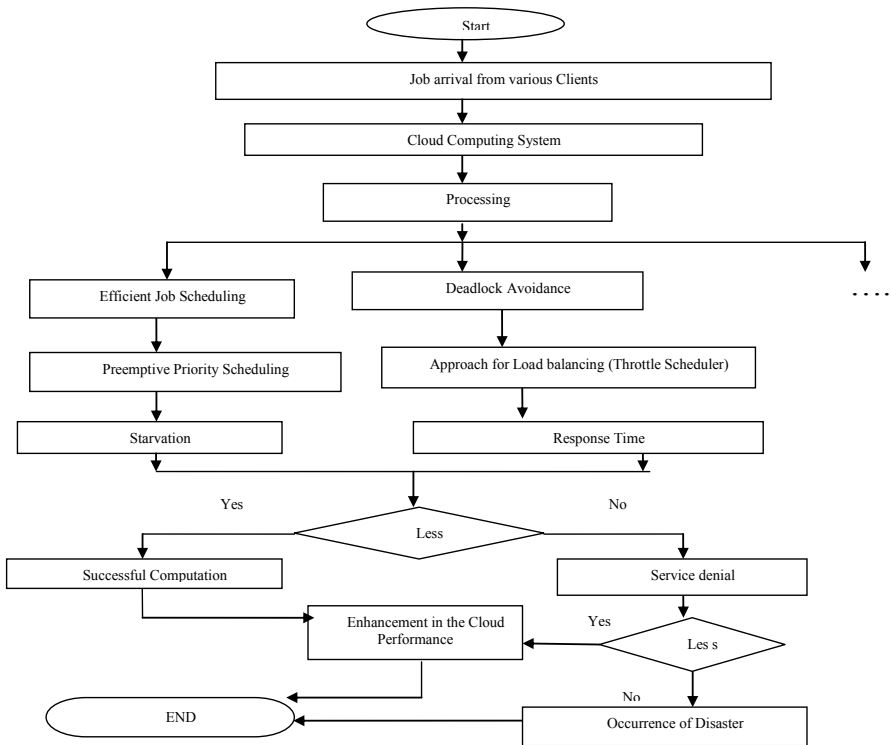
The identified parameters are non-functional which would enhance the system performance.

Fig 2 is the flow diagram of this analysis depicting the Job processing in cloud to control disasters. In the diagram different requests arrives from different clients to the computing system. The Cloud processes these requests and sends the results to its clients. The scheduling strategy applied here is Preemptive priority scheduling with aging and the throttled scheduling algorithm. The implementation of both results in minimizing starvation and balance the load with less response time. Subsequently, the reduction in starvation and response time results in controlling Disaster and accelerate the business performance of the Cloud.

Here the two suggested parameters are presented in two different modules.

*Module 1: Scheduling Strategy using Preemptive priority scheduling with aging*

*Module 2: Deadlock Avoidance through throttled scheduling algorithm*



**Fig. 2.** Job Processing in Cloud

*Module 1: Scheduling Strategy using Preemptive priority scheduling with aging*

Table 1, Table 2 and Table 3 depicts the Result Analysis of Preemptive Priority scheduling with Aging.

**Table 1.** Job Arrival Pattern in Cloud

Jobs	Job Arrival time “T” (min)	Burst time(min)	Job priority
J <sub>0</sub>	1	9	70
J <sub>3</sub>	3	10	1
J <sub>1</sub>	5	4	3
J <sub>4</sub>	9	11	6
J <sub>2</sub>	11	4	5

The Table 1 shows the arrival pattern of jobs with different Job Priority. Here for example, job J<sub>0</sub> having a priority 70 (Least priority) arrives first i.e. at time T=1, J<sub>1</sub> having a priority 3 i.e it arrives at time T=5, J<sub>2</sub> having a priority 5 arrives it arrives at time T=11, J<sub>3</sub> having a priority 1 (higher priority) arrives it arrives at time T=3, J<sub>4</sub> having a priority 6 arrives it arrives at time T=9.

**Table 2.** VM utilization based on Priority Preemptive Scheduling

VM utilization time (in Minutes) of preemptive
J <sub>0</sub> from 1 to 3 =2
J <sub>3</sub> from 3 to 13 =10
J <sub>1</sub> from 13 to 17 =4
J <sub>2</sub> from 17 to 21 =4
J <sub>4</sub> from 21 to 32 =11
J <sub>n</sub> (30 to ....) = ( Job J <sub>n</sub> having higher priority than J <sub>0</sub> will be scheduled)

Table 2 presents the comparison between the VM (Virtual Machine) utilization with respect to job priority. From the Table 1 at time T=3, job J<sub>3</sub> arrives with priority “1” (higher priority), it is observed in Table 2 that job J<sub>0</sub> is immediately preempted by job J<sub>3</sub>. Similarly the other jobs in the queue having higher priority than Job J<sub>0</sub> will be provisioned the requested resources thereby denying the resources for job J<sub>0</sub>.

Priority preemptive scheduling technique causes the job J<sub>0</sub> to enter starvation, there by leading to Disaster in the Cloud.

Table 3 presents the Virtual Machine utilization with respect to Priority Preemptive Scheduling with Aging.

**Table 3.** VM utilization based on Priority Preemptive Scheduling with Aging

VM utilization time (in Minutes) of preemptive with aging
$J_0$ from 1 to 3 =2
$J_3$ from 3 to 13 =10
$J_1$ from 13 to 17 =4
$J_2$ from 17 to 21 =4
$J_0$ from 21 to 28 =7 ( $J_0$ gets the VM $WT \gg WTT$ )
$J_4$ from 28 to 39 =11

According to Algorithm 2 the Wait Time Threshold (WTT) is assumed to be 20 minutes. Here Job  $J_0$  having the least priority is provisioned the requested resource once the WT of the  $J_0$  exceeds the WTT. The jobs with higher priority than  $J_0$  having less WT will be scheduled after  $J_0$ . Therefore from this approach the least priority jobs are also provisioned with the requested resources preventing them from starvation.

*Module 2: Deadlock Avoidance through throttled scheduling algorithm*

A Deadlock in Cloud is a situation when one or two processes are waiting for other process to complete the task.

*Throttle Scheduling Algorithm (TSA)*

The TSA maintains a record of the state of each virtual machine (busy/idle). If a request arrived concerning the allocation of virtual machine, the TSA sends the ID of ideal virtual machine to the data center controller and data center controller allocates the idle virtual machine. TSA schedules with the help of size of the Job and the available resources [5].

The proposed Load Balancing technique in Cloud is Throttled Scheduling Algorithm. In order to analyze the efficiency of TSA technique a simulation is carried using Cloud Analyst simulator [2]. Table 4, Table 5 and Table 6 depicts the simulation configuration.

Table 4 depicts the information of user base configuration, which provides the information related to requests or users per each hour (Req/User/Hour), data size for each request, peak hours start time and end time in the form of Greenwich Mean Time (GMT).

**Table 4.** User Base Configuration

Name	RUH	RS/R	PHS(GMT)	PH E(GMT)	APU	AOPU
UB1	10	100	1	3	1000	100
UB2	10	1000	3	9	1000	100
UB3	10	2000	2	6	1000	100
UB4	10	3000	3	9	1000	100
UB5	20	4000	2	9	1000	100

RUH=Req/User/Hour, RS/R=Resource Size/Req , PHS=Peak Hours Start, PHE=Peak Hours End, APU=Avg Peak Users, AOPU= Avg Off Peak Users.

**Table 5.** Datacenters Deployment configuration

Data centers	Arch	OS	VMM	#VMS	Memory Size	BW
DC1	X86	Linux	Xen	25	512	10000
DC2	X86	Linux	Xen	33	1024	100
DC3	X86	Linux	Xen	35	1024	1000
DC4	X86	Linux	Xen	20	512	100
DC5	X86	Linux	Xen	22	1024	10000

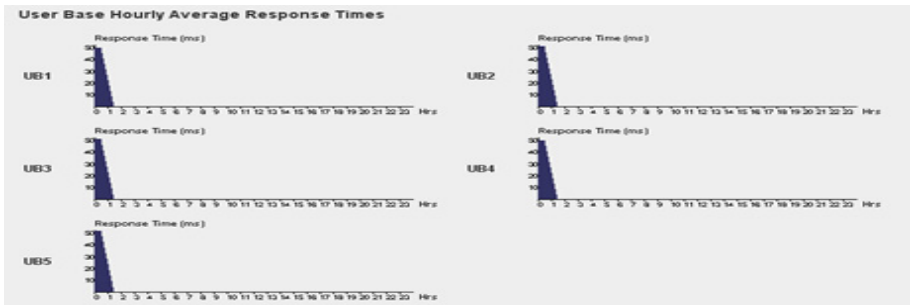
Arch= Architecture,OS= Operating System,VMM= Virtual Machine Manager,#VMS= Number of Virtual Machines,BW= Bandwidth.

Each Datacenter Compromises of Architecture, Operating System, Virtual Machine Manager, Number of Virtual Machines, Memory Size, and Bandwidth.

**Table 6.** Overall response time and processing time obtained using TSA load balancer

PM	Avg(ms)	Minimum(ms)	Maximum (ms)
ORT	51.42	41.25	61.28
DCPT	1.52	0.19	2.78

PM=Performance Metrics, ORT=Overall Response Time,, DCPT=Datacenter Processing Time.



**Fig. 3.** A sample graph of user base hourly average overall response time using the ThrottleLoad Balancer

From the Fig 3 it is observed that there is a less response time leading to less processing time and ultimately resulting in decreased Deadlock, by avoiding deadlock disaster is minimized.

## 4 Conclusion

The rapid development in the technology has lead to tremendous computing. The Cloud provides an assured Quality arena to its clients where in they can perform the computation with less infrastructure investment and maintenance. Though there is

advancement in technology, the Cloud is prone to Disaster. Disaster management is one of the performances metric which reflects on the Quality of Service of Cloud. Here this paper focuses on Customer satisfaction, which is achieved by applying the basic principles of Software Engineering.

Disaster management is a challenging task, it is essential for the Cloud Service providers to apply appropriate metrics to monitor the Cloud performance. Due to the growing demand for the cloud services a short period of downtime would result in significant financial loss. Here paper we have focused on Disaster management through two different interdependent parameters. Disaster Management is achieved through efficient Priority preemptive scheduling with Aging to avoid starvation of the low priority Jobs, and Efficient Throttled Load Balancing for Deadlock avoidance. This enhances the cloud performance and QoS of the computing model.

The techniques maintain the status of the resources periodically. It rejects the jobs exceeding the Job queue size. From the result analysis and simulation results it is observed that the Computing model has a better Throughput. By enhancing the Throughput disaster is under control in the computing model.

## References

- [1] Mahitha, O., Suma, V., Vaidehi, M.: Disaster Recovery in the Cloud Environment – An Analysis. In: International Conference on Advanced Computer Science and Information Technology (ICACSIT), Pune, India (2013)
- [2] Mahitha, O., Suma, V., Vaidehi, M.: A Comparative Study of Tools for Disaster Management in Cloud. In: 2nd International Conference on Recent Trends in Engineering Technology (ICRTET 2013), Bangalore, India (2013)
- [3] Mahitha, O., Suma, V.: Disaster Management in Cloud through Enhanced Scheduling Strategy. In: International Conference on Electrical Engineering and Computer Science (EECS), Bangalore, India (2013)
- [4] Mahitha, O., Suma, V.: Preemptive Priority Scheduling with Aging Technique for Effective Disaster Management in Cloud. Graduate Research in Engineering and Technology (GRET) An International Journal 1(2) (2013)
- [5] Mahitha, O., Suma, V.: Deadlock Avoidance through Efficient Load Balancing to Control Disaster in Cloud Environment. In: The Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), India (2013)
- [6] Brocade, V.J.: Backup for Cloud and Disaster Recovery for Consumers and SMBs. In: IEEE 5th International Conference on Advanced Networks and Telecommunication Systems (ANTS) (2011)
- [7] Pokharel, M., Lee, S., Park, J.S.: Disaster Recovery for System Architecture using Cloud Computing. In: 2010 10th Annual International Symposium on Applications and the Internet, pp. 304–307 (2010)
- [8] Suakanto, S., Supangkat, S.H., Suhardi, Saragih, R., Nugroho, T.A., Nugraha, I.G.B.B.: Environmental and Disaster Sensing Using Cloud Computing Infrastructure. In: International Conference on Cloud Computing and Social Networking (ICCCSN), pp. 1–6 (2012)
- [9] Sharma, K., Singh, K.R.: Online Data Back-up and Disaster Recovery Techniques in Cloud Computing: A Review. International Journal of Engineering and Innovative Technology (IJEIT) 2(5), 249–254 (2012)

- [10] Gopalakrishnan Nair, T.R., Suma, V.: Defect management using pair metrics, DI and IPM. *Crosstalk, The Journal of Defense Software Engineering* 24(6), 22–27 (2011)
- [11] Suma, V., Gopalakrishnan Nair, T.R.: Effective defect prevention approach in software process for achieving better quality levels. Paper presented at the Fifth International Conference on Software Engineering, Singapore (August 2008)
- [12] Suma, V., Gopalakrishnan Nair, T.R.: Enhanced approaches in defect detection and prevention strategies in small and medium scale industries. Paper presented at the Third International Conference on Software Engineering Advances (IEEE), Malta, Europe (October 2008b)
- [13] Suma, V., Gopalakrishnan Nair, T.R.: Defect management strategies in software development. In: *Book on Recent Advances in Technologies*, Vienna, Austria, pp. 379–404. Intecweb Publishers (2009) ISBN 978-953- 307-017-9

# An Efficient Job Classification Technique to Enhance Scheduling in Cloud to Accelerate the Performance

M. Vaidehi<sup>1,2</sup>, T.R. Gopalakrishnan Nair<sup>1,2</sup>, and V. Suma<sup>2,3</sup>

<sup>1</sup> ARAMCO International Endowed Chair,  
Technology and Information Management, Kingdom of Saudi Arabia

<sup>2</sup> Research and Industry Incubation Centre,  
Dayananda Sagar Institutions, Bangalore 560078, India

<sup>3</sup> Prince Mohammad University, Kingdom of Saudi Arabia  
{dm.vaidehi, trgnair, sumavdsce}@gmail.com

**Abstract.** As cloud computing is becoming more ubiquitous with increasing espousal of advanced technologies, more and more efficient techniques are required to enhance the system performance. The computing systems in the cloud comprise heterogeneity of components or resources. Challenge here is efficient scheduling and resource allocation to jobs requesting the computing devices in order to achieve customer satisfaction. Retention of customer satisfaction is one of the primary factors for an organization to exist. To achieve the aforementioned goal, it is required to implement the principles of software engineering in every task that is accomplished in the organization. Thus, with organizations marching towards cloud environment, the jobs are initially clustered or grouped and subsequently scheduled for the resource arbitration and allocation. This paper focuses on clustering or grouping of jobs. The unsupervised technique or the clustering of the jobs is done based on the logistic regression approach. As this approach is more robust and the parameters considered for classification are more independent, simulation evidence suggests the classification technique cluster the jobs more effectively and provide a consistent utilization of the available resources. This ensures that the non functional requirement of availability of jobs to their customers is achieved thereby enhancing the business performance.

**Keywords:** Cloud, Grouping, Logistic Regression, Utilization, Optimality, Priority, Customer Satisfaction.

## 1 Introduction

Customer satisfaction depends on many factors which are associated with the business needs, the kind of services (Software, Platform, and Infrastructure as Services) and the resultant quality of services. The customers look for added value to benefit the business operations within a defined timeframe at an affordable price hence the customers priority is an overall successful business [10]. The service provider has to deliver the services within an agreed cost plan to satisfy the customer requirements.



The Customer requirements can be achieved by verifying whether the developed product or Service satisfy both the Functional and Non-functional requirements. The functional requirements involves in ensuring that the functionality specified in the requirement specification is achieved. Generally the product or service is tested in different environments (homogeneous environment) to ensure Customer satisfaction.

Similarly the required product or service is tested to verify the Non-functional properties. The Non-functional requirements generally relate to the Quality goals or Quality of Service requirements of the Product or Service [10]. In Cloud which is a non homogeneous computing environment, the Non- functional parameters tested here are Accessibility, Availability, Fault Tolerance, Disaster recovery, Efficiency in terms of resource utilization, Performance etc.

In a non homogeneous computing environment a collection of various resources are integrated to provide the computing capability to the jobs in the system requesting for resources. The virtualized hardware or software resources are provisioned to the requesting jobs based on the availability and business returns. The conventional system-centric resource management architecture cannot process the resource assignment task and dynamically allocate the available resources in a cloud computing environment [1]. The simple cloud architecture queues the job in a central location, and at each time slot a central resource manager helps in the formation of the virtual machines and later the scheduler schedules the jobs as shown in figure 1.

Task scheduling, one of the most prominent combinatorial optimization problems, it is one of the prime factor which plays the key role to enhance the performance of flexible and reliable systems. The main objective is to schedule tasks to the adaptable resources in accordance with adaptable time, which involves finding out a proper sequence in which tasks can be executed under transaction logic constraints [2].

The criticality here is to allocate the resources to the jobs in the request queue for optimal utilization of the available computing system.

Here to achieve the optimality, grouping of the jobs is done and later the resources are allocated. The advantage of grouping or clustering of the jobs assists in identifying the identical jobs requesting for similar kind of resource, accordingly the resources can be provisioned. Upon utilization of the resource, the next job in the cluster is provided the same set (size) of resource. That is when the current job releases the resource, the resource is not immediately torn off or put to idle state, and instead it is provisioned to the next job in the cluster or group.

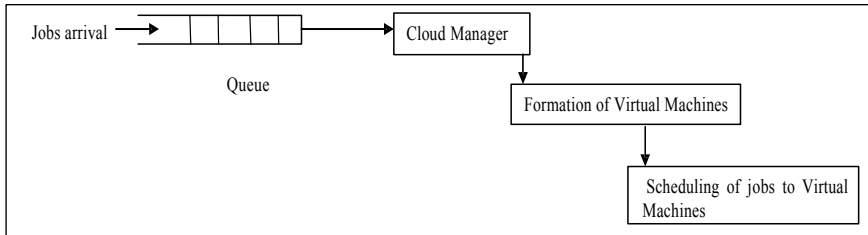
## 1.1 Logistic Regression

Logistic regression is used for predicting the result of an unambiguous dependent variable based on one or more predictor variables. The probabilities describing the possible outcome of single trial are modeled, as function variables using a logistic function.

A logistic function or logistic curve is a sigmoid function. A generalized logistic curve is the “S” shaped model exhibiting the behavior of growth of some population “P”.

The initial stage of growth is approximately exponential, then as saturation begins, the growth slows, and at maturity growth stops [3]

This paper deals with the application of the logistic regression approach for grouping or clustering of the jobs prior to scheduling.



**Fig. 1.** Conceptual Diagram of Cloud Architecture

The organization of the paper is as follows, Section 1 provides a brief Introduction to heterogeneous systems, scheduling and resource allocation. Section 2 provides the Literature Survey. This is followed by section 3 which defines the Problem, section 4 presents the Design and then in section 5 the Simulations and Result Analysis are presented. Finally section 6 presents the Conclusion.

## 2 Literature Survey

The advancement and the advantages in Cloud computing has unwrapped several research avenues to improve the performance of the computing system in order to enhance the business performance.

D. Armstrong and K. Djemame discuss that with the emerging interest on Cloud Computing and the new paradigm that it has introduced into Distributed Computing, QoS and pay-per-use models, the economy is transforming from a product oriented economy to a service oriented economy. This trend is boosting the economic exploitation of Distributed Computing environments like Grid Computing and High Performance Computing and Cloud Computing [10]. The authors also discuss that without the management of resources the computing model would be unable to function. Resource management encompasses the dynamic allocation of tasks to computational resource and requires the use of a scheduler (or broker) to guarantee performance. QoS is enabled by the efficient scheduling of tasks, this guarantees that resource requirements of an application are strictly supported but resources are not over provisioned and used in the most efficient manner possible [10].

Monika et.al have proposed a scheduling algorithm. Here the incoming tasks are grouped on the basis of task requirement like minimum execution time and minimum cost. The research work only says that the jobs are grouped or clustered, but does not emphasize on the grouping or the clustering technique [3].

The research work of Sandeep Tayal presents the implementation of fuzzy GA (genetic algorithm) optimization which makes a scheduling decision by evaluating the

entire group of task in job queue, the algorithm is based on the accuracy of the predicted execution time of each task [4].

Mousumi Paul et.al have used credit based scheduling decision to evaluate the entire group of task in the task queue to find the minimal completion of all tasks. Here cost matrix has been generalized as the fair tendency of task to be assigned in a resource [5].

Selvarani S et.al have proposed an algorithm based on cost with user task grouping. Their work focuses on scheduling. The proposed algorithm employs a cost-based scheduling algorithm for making efficient mapping of tasks to available resources in cloud. Their algorithm measures both resource cost and computation performance. Their work focuses on the computation and communication ratio by grouping the user tasks according to a particular cloud resource processing.

Senthil K. et.al have proposed an algorithm using modified linear programming problem transportation based on task scheduling and resource allocation for decentralized dynamic cloud computing. Their algorithm utilizes task historical values like success rate, failure rate of each task in each cluster and also the previous execution time and total cost .The proposed algorithm shows an enhancement in the reliability of the cloud environment [7].

Zhong et.al have proposed an optimized scheduling algorithm to optimize the cloud scheduling. They have used Improved Genetic algorithm (IGA) for automated scheduling policy. Their approach shows an increase in the utilization rate of the resources in the cloud [8].

### 3 Problem Definition

As the cloud computing is highly dynamic, the tasks encountered will not be alike. In such situations, techniques like ANN (artificial neural network) or algorithms like ant colony, honey bee algorithm is heuristic and they need lot of time to get trained and react on the situation. As the cloud is dynamic same clients and similar kind of jobs submitted may be frequent.

An efficient job classification technique plays a vital role in efficient utilization of resources in the cloud datacenters. Further, the cloud service providers should achieve faster return on investment (ROI) with low cost ownership while providing services to the clients. Henceforth, an efficient job classification for resolving these issues becomes a mandatory in the cloud computing environment.

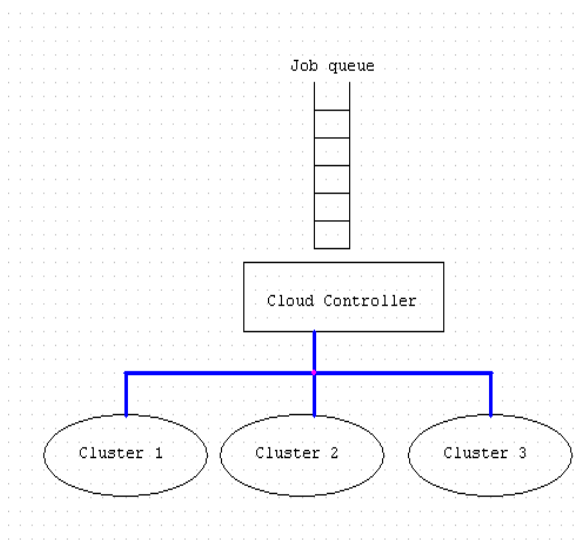
Data collection for this research consists of the secondary data for analyzing the efficiency of cloud service providers for successfully providing services to the clients exclusively at peak hours with utilization of the available resources and minimized SLA violations.

Table 1 depicts the job arrival pattern captured from the cloud monitoring system. From this it is observed that there is a high demand for resources by the jobs for computation during the peak hour which consequently results in tremendous resource utilization. Henceforth, this issue is resolved using logistic regression clustering technique.

**Table 1.** Number of jobs arrived at peak hours

Peak Time (hours)	Number of jobs arrived
7	10
8	20
9	35
10	50
11	20
12	30
13	14
14	20
15	12
16	10
17	10
18	10

## 4 Design

**Fig. 2.** Jobs clustering done by Cloud Controller

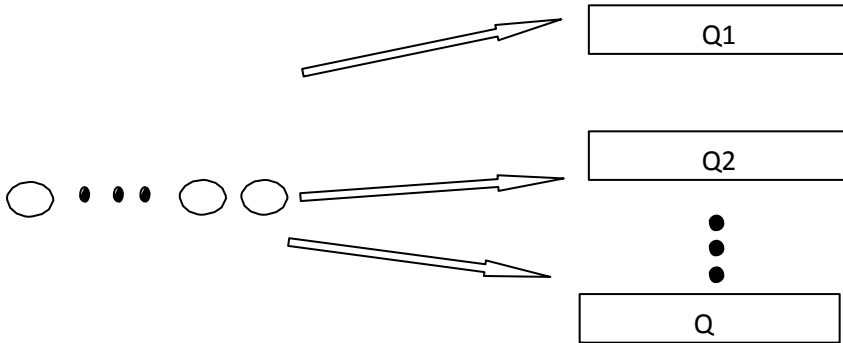


Fig. 3. Analogous to Multi Queuing

### 4.1 Logistic Regression Algorithm in Cloud Computing

This algorithm runs at the Cloud controller level as shown in Figure 2. The proposed technique is analogous to multi queuing approach as shown in Figure 3, here the queue “Q1” will have the highest priority jobs based on the resource size requested and criticality.

Mathematical Model: This Model presents the arrival pattern of the Jobs which are further classified according to the model below

$$Q_N = \frac{J_n^t}{J_n^t + 1} = \frac{1}{1 + J_n^{-t}} \tag{1}$$

$J_n^t$  is the arrival time of Job  $J_n$

As the jobs  $J_n$  arrive they are classified and

$Q_N$  represents the Queue with Nth priority, ie if a job arrives at time,  $t$  then, the priority of the job will be verified and accordingly sorted or grouped and transferred to the queue as per the Classification Algorithm 1.

#### Classification Algorithm 1

The algorithm facilitates the transformation of the arithmetic expression in Equation 1 to implementation model

Here only two parameters are considered for classification

1. The Job Size ( $J_S$ )
2. Job Priority ( $J_P$ )
3. Threshold Size of the job ( $T_S$ )

The Job priority is generally set based on the Criticality and Business returns

Job Arrival and Classification or Clustering  
If

```

    { Job Size>>
      Js >> Ts
If
      Jp >> High
    }
Then
  Q1=1++
Else
  If
    {
      Js >> Ts
If
      Jp >> High-1
    }
Then
  Q2=1++
  .
  .
  .
Else
  QN= 1++

```

## 5 Simulations and Result Analysis

For the job arrival pattern captured as shown in Table 1, it is observed that by implementing the proposed algorithm, the jobs are grouped according to their Data size as shown in Table 2.

The Table 2 comprises of the User Base, geographical region from where the users are sending the requests, number of requests sent to the Data Centre, it also specifies the the peak hours of operations where in more requests flood in. The graph in Figure 4 and Figure 5 depicts the utilization of the resources after grouping or classification of the jobs according to the proposed algorithm.

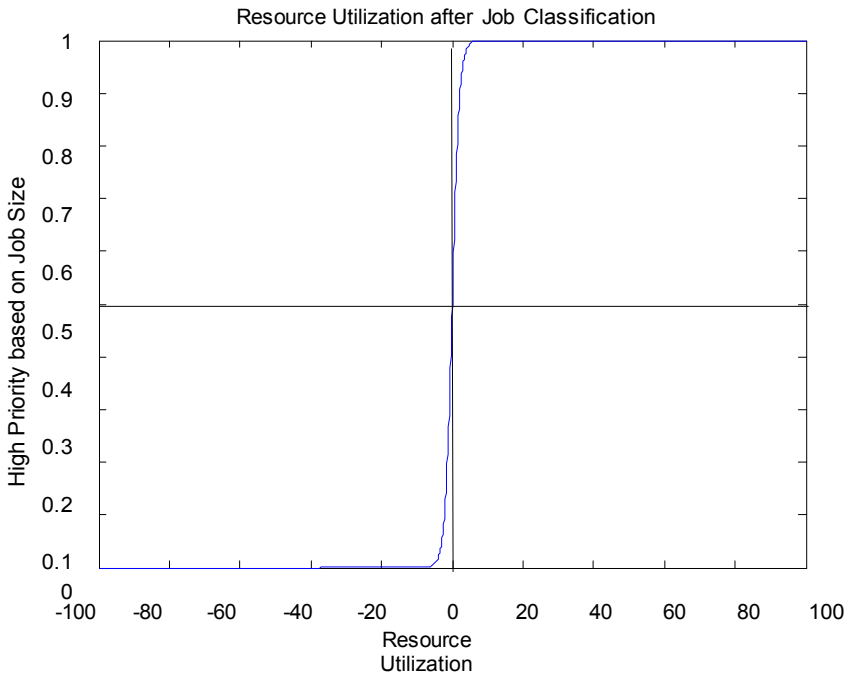
It can be observed from Figure 4 that after classification of the jobs (based on the Logistic Regression algorithm) according to the job size and setting the priority high there is optimal resource utilization.

Here the jobs are classified and a priority is assigned according to the Data request size (Resource Size). It is observed from the graph that the jobs are classified and assigned priority ranging from 0, 0.1.....1, the job with priority “1” implies that it is provisioned a larger resource based on its request, there is maximum utilization. This Classification technique prevents the under utilization and over utilization of the resources [9].

From Figure 5 it can be observed that the jobs of same size are classified based on the Logistic regression technique, there is 99 % to 100 % utilization of the resource.

**Table 2.** Classification of Jobs Based on the Algorithm

Name	Region	Requests/ User/hour	Data Size /Request	Peak Hrs Start	Peak Hrs End	Average Peak Users	Average -Off Peak Users
UB1	2	10	100	7	12	100	10
UB2	2	20	100	7	12	100	10
UB3	2	10	100	8	12	100	10
UB4	2	40	90	8	12	100	10
UB5	2	10	90	7	12	100	10
UB6	2	40	90	10	12	100	10
UB7	2	20	80	9	12	100	10
UB8	2	10	80	7	12	100	10
UB9	2	10	70	8	12	100	10
UB10	2	20	70	7	12	100	10



**Fig. 4.** Resource Utilization

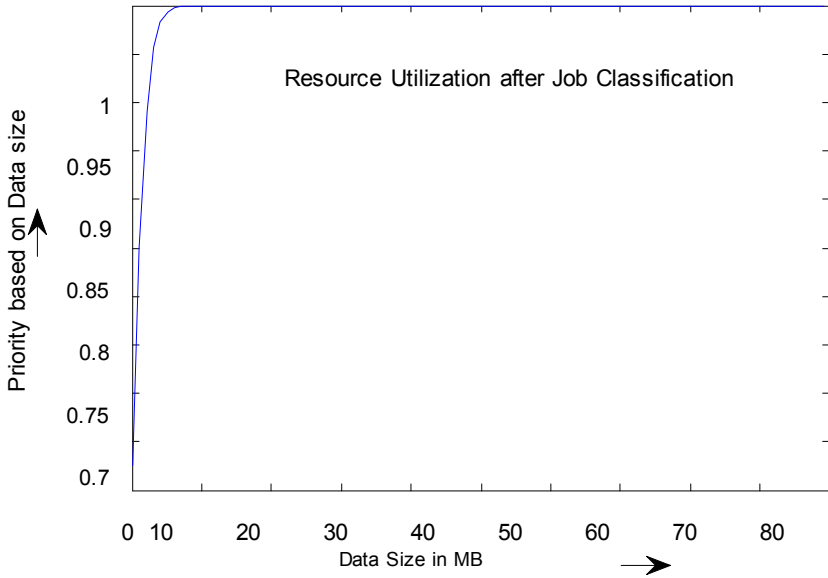


Fig. 5. Resource Utilization

### Samples of Simulation

Define Internet Characteristics
Main Configuration
Data Center Configuration
Advanced

Simulation Duration:  min

User bases:

Name	Region	Requests per User per Hr	Data Size per Request (bytes)	Peak Hours Start (GMT)	Peak Hours End (GMT)	Avg Peak Users	Avg Off-Peak Users
UB1		2	60	100	3	12	1000
UB2		2	60	100	3	12	1000
UB3		2	60	100	3	12	1000
UB4		2	60	90	3	12	1000
UB5		2	60	90	3	9	1000

Application Deployment Configuration:

Service Broker Policy: Closest Data Center

Data Center	# VMs	Image Size	Memory	BW
DC1	5	10000	512	100



### Configure Simulation

Main Configuration | Data Center Configuration | Advanced

Simulation Duration:   ▾

User bases:

Name	Region	Requests per User per Hr	Data Size per Request (bytes)	Peak Hours Start (GMT)	Peak Hours End (GMT)	Avg Peak Users	Avg Off-Peak Users
UB6	2	60	80	3	9	1000	100
UB7	2	60	80	3	9	1000	100
UB8	2	60	70	3	9	1000	100
UB9	2	60	70	3	9	1000	100
UB10	2	60	80	3	9	1000	100

Application Deployment Configuration:

Service Broker Policy:  ▾

Data Center	# VMs	Image Size	Memory	RW
DC1	5	10000	512	1000

## 6 Conclusion

The transformation of IT industries from the Capex model to an Opex model is a huge revolution in the current computing era which has posed several challenges. The Cloud model conceptually developed from the Data centers which supported resource sharing, concurrent users through advanced architecture and highly virtualized environment requires more attention towards enhancing the system performance through effective resource utilization for computational easiness.

Focusing on effective resource sharing and utilization would enable various features such as Availability, Speed, Cost optimization etc. The Logistic Regression technique has given better results with respect to classification of jobs based on their Data size or Resource request size. Due to classification, the reformation of the resources based on the request size is prevented and henceforth there is a better utilization and increase in speed of operation.

## References

1. Yeo, C.S., Buyya, R.: Pricing for Utility-driven Resource Management and Allocation in Cluster. *International Journal of High Performance Computing Applications* 21(4), 405–418 (2007)
2. Polo, J., Castillo, C., Carrera, D., Becerra, Y., Whalley, I., Steinder, M., Torres, J., Ayguadé, E.: Resource-Aware Adaptive Scheduling for MapReduce Clusters. In: Kon, F., Kermarrec, A.-M. (eds.) *Middleware 2011*. LNCS, vol. 7049, pp. 187–207. Springer, Heidelberg (2011)

3. Choudhary, M., Peddoju, S.K.: A Dynamic Optimization Algorithm for Task Scheduling in Cloud Environment. *International Journal of Engineering Research and Applications (IJERA)* 2(3), 2564–2568 (2012) ISSN: 2248 -9622, <http://www.ijera.com>
4. Tayal, S.: Tasks Scheduling optimization for the Cloud Computing systems (IJAEEST) *International Journal of Advanced Engineering Sciences and Technologies* 5(2), 111–115
5. Paul, M., Sanyal, G.: Task-Scheduling in Cloud Computing using Credit Based Assignment Problem. In: *International Journal on Computer Science and Engineering (IJCSE)* (2011)
6. Selvarani, S., Sudha Sadhasivam, G.: A Novel SLA based Task Scheduling in Grid Environment. *International Journal of Applied Information Systems, IJAIS Journal* (2012)
7. Senthil Kumar, S.K., Balasubramanie, P.: Dynamic Scheduling for Cloud Reliability using Transportation Problem. *Journal of Computer Science* 8(10), 1615–1626 (2012) ISSN 1549-3636
8. Zhong, H., Tao, K., Zhang, X.: An Approach to Optimized Resource Scheduling Algorithm for Open- Source Cloud Systems. In: *2010 Fifth Annual ChinaGrid Conference*, July 16-18, pp. 124–129 (2010), doi:10.1109/ChinaGrid.2010.37
9. Gopalakrishnan Nair, T.R., Vaidehi, M.: Efficient Resource Arbitration and Allocation Strategies in Cloud Computing through Virtualization. In: *Cloud Computing and Intelligence Systems (CCIS)*, vol. 262, pp. 258–262 (2008, 2011)
10. Suma, V., Nair, T.R.G.K.: Effective Defect Prevention Approach in Software Process for Achieving Better Quality Levels. *World Academy of Science, Engineering and Technology (WASET)* 42, 258–262 (2008)
11. Armstrong, D., Djemame, K.: Towards Quality of Service in the Cloud. In: *Proc. of the 25th UK Performance Engineering Workshop*, Leeds, UK (2009)

# A Study on Cloud Computing Testing Tools

M.S. Narasimha Murthy<sup>1</sup> and V. Suma<sup>2</sup>

<sup>1</sup> Department of Computer Science & Engineering,  
Acharya Institute of Technology, Bangalore  
narasimhamurthys@acharya.ac.in

<sup>2</sup> RIIC, Dayananda Sagar Institutes, Bangalore  
sumav\_dsce@gmail.com

**Abstract.** Today's IT industries are growing rapidly and demands for new technology for handling on-line requirements of their customers in a very efficient and cost effective way. With the invent of popular cloud computing technology, IT enterprises have started moving their service capabilities to cloud model to satisfy the needs of the customers by providing variety of services under single window. Further, it is imperative to understand the level of services extended by cloud computing model for diverse applications. Therefore it is required to adopt proper testing of cloud services and also strategies for testing applications running on different cloud computing models. This paper deals with the need of testing application in cloud, comparative study of different tools available to test the application in the cloud for various abilities, essential quality attributes necessarily tested for some sample applications and finally an inference is drawn from the study that all testing tools are not supporting testing of every quality attributes and also each and every quality attribute is not necessarily required to be tested against quality of application.

**Keywords:** Cloud Computing, Cloud Testing, Testability, Web Applications Mobile Applications, Multimedia Applications, Functional Testing, Non-Functional Testing, Performance Testing.

## 1 Introduction

Cloud computing is a shared computing platform available over the network used to run a variety of business or personal applications. The concept of cloud computing is new and it has its roots in providing varieties of services to big data centers and utility computing. With the help of cloud computing services, models users are running different applications on the web. This saves cost, stretched reach, and improved quality by running applications in the cloud such that it leads to the business success [1].

Software testing is an important phase in the software design life cycle and many software firms are spending 40% of their resources on testing process and it is highly expensive process. Software testing is the process of executing a program with the goal of finding errors [2].

Conventional software testing requires high investment on hardware, software and its maintenance in order to satisfy the users' requirements for the applications which are running from different geographic locations. Sometimes applications works with random and varied number of users and deployment environment depending on client requirements, in such cases cloud testing is more effective.

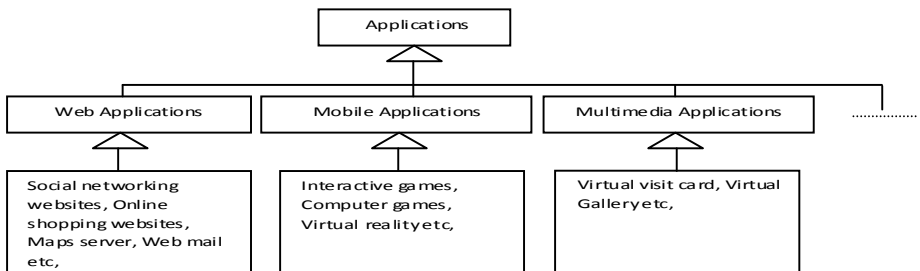
Cloud Testing is a process of testing the applications, which have been migrated or are to be migrated to the cloud so as to guarantee their performance, security and reliability matches or exceeds expectations in view of the changing delivery methods (Testing Cloud).

Cloud testing may also be considered as, to leverage the cloud-based hardware infrastructure and computing resources to perform traditional testing like performance, load, stress, security and compatibility testing for regular, on-premise applications (Testing using Cloud).[3]

Testing an application in the cloud is a process of testing a web application that uses cloud computing environment which look for real world user traffic for load testing and stress testing the web sites, with this an application running in the cloud is benefited by unlimited computing and storage resources. Hence forth testing process is available as a service and is called TaaS (Testing as a Service) [4]

Cloud computing is very much same as traditional web hosting with respect to services extended at SaaS and PaaS service layers, but the key difference lies at the service provided at IaaS layer, that is provisioning of required amount of computing resources such as RAM, CPU, Cache memory etc. In traditional web hosting there is a need to monitor, install, upgrade and configure programs, add sites deal with potential hacks, and troubleshoot systems. Therefore, there is a need for the business to have an employee.

Fig.1 provides the details of various applications which are considerable to be moved on to the cloud environment.



**Fig. 1.** Detail of categories and sub categories of applications

Fig.1 infers that any application will generally fall in one of the major categories such as web applications, mobile applications, and multimedia applications and so on.

Fig.2 (a) & (b) depicts the way in which the processing of application takes place when it is hosted into the conventional web services and when the applications are hosted into cloud computing environment.

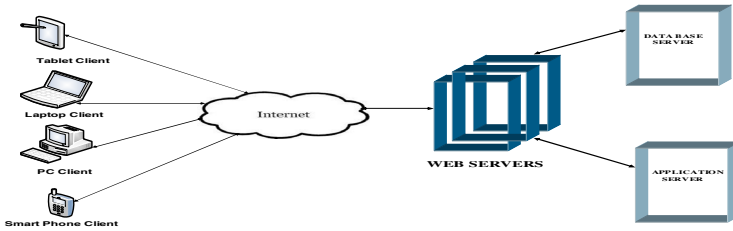


Fig. 2(a). Requests are handled by Web processing



Fig. 2(b). Requests are handled by Cloud computing

Fig.2 (a) & (b) shows that in conventional web services provide responses to the user’s request by processing the request in the server. Further, it is required to maintain the application and only the requests are serviced. However, a request, which is posted by a user in cloud based environment, will be processed by service provider capabilities. Hence, it is not required to maintain the application. Additionally, the cloud service providers ensure enhanced processing and storage abilities.

## 2 Literature Survey

Research is taking place continuously in cloud computing and testing the applications in the cloud ever since the invention of cloud in industries.

Author in [5] describes the popularity of testing an application in the cloud, they bring out the key trends in cloud adoptions, market potential and key drivers for testing in cloud , impact on testing, challenges in testing. The author further deals with the testing methodologies like functional, and non functional testing along with testing methodology process.

Author in [1] brings out the importance of cloud computing and its key characteristics for building the applications in the cloud and its necessity along with the cloud computing deployment and ownership model. Further author mentions the importance of measuring the cloud and its services in order to meet the expectations of cloud consumers for various parameters like data centre offline, servers are overloaded and network is slow.

Author in [6] indicates the drawback of traditional on premises testing scenarios for small-medium business as they have to set up infrastructure for testing the applications along with the additional overhead in maintaining them. He further feels

the importance of cloud computing for testing the application in the form of services, such that industries can focus only on their core business instead of setting up an infrastructure in advance and maintain it later. He also put forth some requirements of dedicated planning and rigorous testing of applications in cloud to satisfy the business need of the customers to greater extent. He also brings out the key benefits of testing an application in the cloud, type of testing to be performed in cloud, key considerations which are very important before moving testing to cloud and associated challenges in testing in cloud.

Author in [7] describes the hype and reality of cloud computing technology in modern enterprise infrastructure, such that more organizations are realizing the benefits of remote hosting of IT services over local IT management. He indicates the requirement of widespread testing and management capabilities in order to ensure the requisite levels of service availability. However he says testing and management of cloud services differs from traditional IT services. Further author elaborates on the challenges in testing (i.e., testing service availability, service assurance and service efficiency), testing end-to-end service availability and virtual servers.

Article [8] describes when to test cloud applications, the need of testing applications in cloud to validate and verify specific cloud functionality such as redundancy, performance and scalability. The article further focuses on the importance of Apica Load test for load balancing and front end cache system management for various projects.

Author in [9] explains the top cloud testing tools for different aspects of testing including test management, performance testing etc. The testing tools like SOASTA, CloudTest Parasoft SOAtest, Cloud Testing's, GCLoads' Proxy PractiTest, Blaze Meter, LoadStorm, Keynote, TestMaker Persistent CLAP have been elaborated along with the benefits of using cloud based testing tools.

The authors in [10] provide the introduction to cloud computing, characteristics of cloud computing and categories of research issues in cloud testing. In addition to this authors focus on the need of testing in the cloud and three facts associated with it. Further they have done a study by interviewing personals from eleven organisations to identify new theories and concepts in order to make cloud testing a valid choice. Finally authors speak in the result section of the paper the different issues associated with cloud testing like, application issue, management issues and legal & financial issues.

This paper is organised as follows, Section-3 describes Research Methodology, Section-4 explains Research Work and Section-5 Conclusion.

### **3 Research Methodology**

Cloud has taken over all the processing of any application from conventional practise scheme. So, it now becomes very critical to test every application against the expected quality of service and not just upon the satisfaction of service.

The main intention of this research is to investigate the successful nature of processing of applications in cloud based environment. Hence, this research aimed to

analyze various tools that are capable to test various qualities attributes of applications. In order to achieve this objective, applications are initially classified into various major categories and subsequently they are analyzed against critical non functional behaviour comprising of essential testability, test factors and test success criteria. The analysis has led to bringing in awareness for the dire need to enhance the existing cloud testing tools.

## 4 Research Work

Since, the emergence of cloud in technology, various industries have marched themselves towards processing their applications in cloud based environment. Hence, it is essential not to just satisfy processing of applications and also needs the processing efficiency.

Thus, the functional behaviour of the applications is tested first, when it is moved on to the cloud environment and subsequently, it is important to test various non functional behaviours of the application for its level of performance. The diverse application that are running in current IT enterprises are like Web applications, Mobile applications and Multimedia applications needs to be better tested for its non functional characteristics.

This research is therefore aimed at identifying the applications and its suitability in cloud in addition to analyze the types of testing needed for those applications. The research design that is followed for this part of investigation is as below.

**Step 1.** Applications are identified for various type of testing since all applications are not suitable for cloud computing environment.

**Step 2.** Tools which supports various types of testing for different applications is recognised.

**Step 3.** Further, the applications and the associated tools are analysed against various quality attributes such as, scalability, availability, interoperability etc.

**Step 4.** Analysing if the tools satisfies the test success criteria, test factors and testability

Test success criteria includes, functionality, likeability, security, usability, maintainability, interoperability, configurability.

Testability comprise of, controllability, observability, operability, stability, simplicity, scalability, stability, accessibility, context sensitivity, suitability, understandability, navigability etc,

Test factors consists of, coupling, performance, portability, correctness, audit trail, reliability, compliance, ease of use, ease of operation etc,

Table.1 illustrates a sample of applications which are hosted in cloud based environment and the respective tools that are available for testing those applications. Table.1 further indicates the types of testing that are carried out by the tool in the respective applications.

Table.1 infers that some of the applications are suitable for cloud environment and thus they need accurate testing against various quality attributes. As an instance, it is

always appreciable for web based applications to be tested for functionalities and also to be tested against the desired level of performance. Additionally, web applications also need to be tested against certain quality attributes such as load test, cross browser test, quality management test, message protocol testing and so on. However, the table

**Table 1.** Comparative Study of different Cloud Testing Tool

Application	Tools	Testing
Web Applications & Services	SOASTA CloudTest	Functional and performance testing
	Parasoft SOAtest	Application testing, Message/protocol testing
	Cloud Testing's	Functional and Cross Browser testing
	GCLoads Proxy	Web/Application servers, Load testing
	PractiTest	Quality Management testing
	LoadStorm	Performance testing
	Persistent CLAP	Load-test
Mobile Application & Services	SOASTA CloudTest	Functional and Performance testing
	Keynote	Performance testing
	SandStorm CE	Performance, Reliability, Scalability
Multimedia Applications & Services	TestMaker	Performance testing
	WebLOAD	Load testing
	Citrix UI	Load testing

**Table 2.** Quality attributes tested for Sample applications

Testability's	Quality Attribute	Web Applications	Mobile Applications	Multimedia Applications
Test success criteria	Functionality	ME	ME	ME
	Interoperability	ME	ME	E
	Usability	E	ME	ME
	Configurability	E	E	E
	Security	E	ME	O
Testability	Scalability	E	E	O
	Stability	E	ME	E
	Suitability	E	ME	E
	Accessibility	ME	ME	E
Test factors	Performance	E	ME	E
	Portability	ME	ME	E
	Reliability	E	E	O
	Maintainability	E	ME	O

E-Essential; ME-Most Essential; O-Optional



further indicates that not all cloud based testing tools are appreciable enough to test all the testability. Further it is true that only some of the test success criteria, test factors and testability's have been taken care by most of the cloud testing tools.

Therefore, this investigation has further focused towards analysing the critical testability factors such as, aaccessibility, functionality, security, interoperability, usability, configurability, scalability, stability, suitability, performance, portability, reliability, maintainability that are mandatory for testing any application to yield optimal customer satisfaction. Table.2 thus, depicts the mandatory testability factors that need to be accurately tested and satisfied in the sampled applications as provided in Table.1.

Table.2 infers that, all quality attributes are not necessarily to be tested against checking the quality of applications. Further it reveals that some of the attributes are most essential and some are optional. Hence in order to claim that a particular testing is successful, it should have tested certain quality abilities which are as depicted in Table 2. Therefore, an inference is drawn after the thorough investigation of Table.2 is that,

**If PQ test has to be successful, XT tool should have tested various attributes like, scalability, availability, security etc., in X application.**

## 5 Conclusion

Testing applications in cloud is an important task once an application is moved into the cloud computing environment. Further it is essential to identify the proper testing tool and necessary quality attributes to test the application for its quality. The study indicated that no single cloud testing tool is available to test all the quality attributes for an application and further, it is not necessary to test all testability factors in order to know the quality of an application.

## References

1. Saylor, M.: New Technologist EXFO Service Assurance, Testing the cloud. White Paper-23rd 2013, EXFO-Assessing Next Generation Networks, New Technologist EXFO (2013), <http://www.lightwaveonline.com/content/lw/en/whitepapers/2013/01/testing-the-cloud.whitepaperpdf.render.pdf>
2. Jovanovic, I.: Software Testing Methods and Techniques. Transactions on Internet Research, 30–41 (January 2009), <http://www.internetjournals.net/journals/tir/2009/January/paper06.php>
3. Gantayat, N.: Testing in the Cloud and its challenges. Asian Journal of Research in Social Science & Humanities 2(4), 226–239 (2012) ISSN 2249 7315, <http://www.aijsh.org>
4. Jun, W., Meng, F.: Software Testing Based on Cloud Computing. In: International Conference on Internet Computing and Information Services, ICICIS, Hong Kong, pp. 176–178 (2011), doi:10.1109/ICICIS.2011.51

5. Kothandaraman, H.: Testing Applications n Cloud. In: Computer Sciences Corporation 2011(CSC) (2011), [http://assets1.csc.com/.../LEFBriefing\\_TestingApplicationsCloud\\_021011.pdf](http://assets1.csc.com/.../LEFBriefing_TestingApplicationsCloud_021011.pdf)
6. An article by Sahoo, K.: Overview of Testing in Cloud, in codeproject.com (April 18, 2013), <http://www.codeproject.com/Articles/580167/Overview-of-Testing-in-Cloud>
7. Barry, D.J.: Cloud Computing: Testing the Cloud. TechNetMagazine (September 2011), <http://technet.microsoft.com/en-us/magazine/hh395480.aspx>
8. An article on Scalability Testing and Cloud Load Testing Services, by Apica Cloud Testing, <http://www.apicasystem.com/cloud-testing/cloud-testing.aspx>
9. Rao, R.: 10-Cloud Based Testing Tools. Tools Journal on Cloud Tools (January 26, 2012), <http://www.toolsjournal.com/testing-lists/item/404-10-cloud-based-testing-tools>
10. Riungu, L.M., Taipale, O., Smolander, K.: Research Issues for Software Test-ing in the Cloud. In: 2nd IEEE International Conference on Cloud Computing Technology and Science, Indianapolis, Indiana USA, November 30-December 03, pp. 557-564., doi:10.1109/CloudCom.2010.58.

# Co-operative Junction Control System

Sanjana Chandrashekar, V.S. Adarsh, P. Shashikanth Rangan, and G.B. Akshatha

Dept. of ISE, RV College of Engineering, Bangalore

{sanjanac1120, aadirao, shashikanth.rangan, akku.bhat}@gmail.com

**Abstract.** An efficient solution is essential for the effective management of the movement of vehicles in a junction. This paper offers an algorithm that leverages the power of wireless vehicle to vehicle communication and vehicle to infrastructure communication to create a system that manages vehicular movement at a junction. An evaluation of this algorithm is carried out using a distributed simulation framework that combined both, network communication and traffic parameter computation. The dynamic simulation platform allowed user level interaction with the display of statistics that depict information exchange between the moving vehicles. A clear mapping of the real world scenario, as observed on Indian roads to the simulation, was brought out and it is shown that a 42% reduction in waiting times were produced as a result of this method.

**Keywords:** VANET, Traffic Management, Wireless V2V communication, Simulation Framework, Indian roads.

## 1 Introduction

The future of current transportation systems lies in the use of technology. Intelligent Transportation Systems (ITS) are vehicle technologies for reducing congestion, increasing safety and convenience and reducing disastrous environmental impact. These systems work by embedding transportation utilities such as vehicles, traffic lights, roads, street sign boards with sensors and communication units. The primary aim is to make these elements intelligent enough and give them enough computational power to make correct critical decisions [1],[2]. ITS elements work by communicating wirelessly with each other and establishing one large communicative network.

Vehicular Ad-hoc Networks (VANETS) are an emerging technology that supports vehicle to vehicle and vehicle to infrastructure communications. VANETS have their own standard set of short range communication protocols and standards known as Dedicated Short Range Communications (DSRC) [3]. DSRC are communication channels in the 5.9GHz band with a bandwidth of 75MHz set aside specifically for use in automobile communication [4]. The name changed from DSRC to WAVE (Wireless Ability in Vehicular Environments), which is also known as 802.11p, in 2003 when the standardization moved to IEEE Forum [5]. A VANET following the above mentioned standards usually comprises of an On-Board Unit (OBU) and a

Road Side Unit (RSU). The OBU is a device placed on the vehicle which processes data collected by the sensors fitted on the vehicles. It facilitates the communication of the vehicle with the outside network i.e., enables the vehicle to communicate with other vehicles and the roadside infrastructure. The RSU, generally placed on the pavement entities such as the traffic signal and sign boards, behaves like a wireless LAN access point and enables the infrastructure to be able to be a part of the network [6]. The RSU also allocates channels to the OBU. VANETs also consist of a third type of element- Public Safety OBU (PSOBU). The PSOBU serves as a mobile RSU. These PSOBU's are put to use mainly in emergency services vehicles such as police cars and ambulances [7].

Testing a VANET technology by direct implementation can be very expensive; hence software simulators are used in imitating real world situations. The simulation of VANET applications requires two kinds of simulators: a Network simulator which simulates underlying network protocols and communication patterns and a Mobility simulator which simulates the vehicle movement patterns. The usage of the Network simulator and the Mobility simulator together results in a VANET simulator [8].

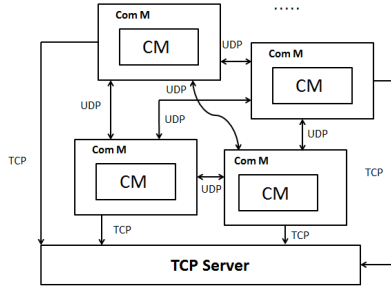
The current use of physical traffic light systems can be very expensive and not adaptive to the changing traffic conditions on the streets. We propose that the use of wireless communication systems between vehicles at a junction can help manage traffic at intersections by creating an in-vehicle traffic light instead of the physical traffic light infrastructure. This paper discusses a new algorithm that has been developed for a Cooperative Junction Control System (CJCS) in a VANET environment. This scheme keeps in mind the high levels of reliability, safety awareness, precision and synchronization needed in the area of managing traffic at junctions.

## 2 Design and Implementation

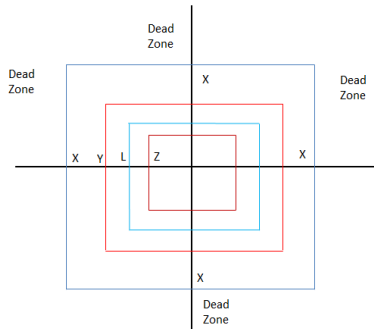
### 2.1 System Architecture

The system architecture mainly contains two components: the Computation Module (CM) and Communication Module (ComM). Each vehicle is equipped with both these modules which take care of the two paradigms of operation which are implementation of the algorithms and communication. The ComM facilitates wireless communication with similar modules in other vehicles present in the range. This wireless communication is necessary for the transfer of relevant data and thus helping in the successful execution of the algorithm. The CM performs necessary parametric computation in each vehicle on the values communicated between the vehicles. This computation is followed again by the communication of the computed values between the vehicles using the ComM. The system architecture of the CJCS is diagrammatically represented in Fig. 1.

Each street in a junction is fitted with an RSU that performs the role of assigning each vehicle that enters into the street an IP address. The RSU also stores Street ID, the dimensions of the street and the co-ordinates of the vital points on the street. When a vehicle enters a street, the vehicle's ComM unit starts broadcasting a message with its registration number. The RSU, on receiving this message, assigns the car a



**Fig. 1.** System Architecture of the CJCS



**Fig. 2.** Points on the road vital to the algorithm for decision making

reusable IP address within the range allocated to it and stores a mapping between the vehicle’s registration number and the IP address assigned to the vehicle.

The system works on the concept of state changes in the functions of the vehicle. The streets in the junction are divided into multiple imaginary regions that play a major role in triggering these state changes in the vehicle. Fig. 2 depicts this division of the streets.

As the vehicle approaches the street and moves in the communicating range of the RSU, the vehicle enters into “registration state” and obtains information about the street, like Street ID, coordinates of vital points on street and the dimensions of the street stored in the RSU. The vehicle wirelessly communicates its registration number as a message to the RSU through its ComM unit. The RSU, on receiving this message, assigns the car a reusable IP address within the range allocated to it and stores a mapping between the vehicle’s registration number and the IP address assigned to the vehicle. This activity occurs before the vehicle reaches region ‘X’ on the street.

The Lane Representative (LR) selection algorithm [explained in Section 2.2.1] starts execution once the vehicle crosses region ‘X’ on the lane. The position of region ‘X’ should be such that it should allow sufficient time for the CM and ComM to be able to complete their respective task before reaching the junction. By the time the vehicle reaches point ‘Y’ a Lane Representative (LR) is chosen which takes part in the election for the representative of the junction. The vehicle elected the LR

moves into the “LR state” while the other vehicles go into a “neutral state” and listen to information about the traffic signal broadcasted by the LR. At point ‘Y’ the LRs from each street communicate with each other to select the Junction Representative (JR). The process finishes at the point ‘Z’. The LR vehicle elected the JR moves into the “JR state” whereas the other non-JR LR’s enter the “LR-waiting state”. The Junction Representative (JR) elected then is in charge of issuing appropriate signals to the LRs who in turn notify the vehicles in their lane. The process is repeated continuously as and when new vehicles enter the range.

## 2.2 Algorithms

There are two algorithms that are implemented in the system. These algorithms have definite timeline and are active in a particular range on the street.

### 2.2.1 Lane Representative Selection Algorithm

This algorithm is active from region ‘X’ to region ‘Y’. The main objective of this algorithm is to elect a Lane Representative (LR) from every street at the junction which is in an optimal position to communicate with LR’s in other streets. The algorithm is based on the timestamp of the vehicle at the point when it reaches ‘X’. The vehicle with minimum timestamp indicates that it is closer to the junction and thereby is in an optimal position to communicate with the LR’s of the other streets and communicate with vehicles in its own street. Information about the number of vehicles taking part in the election is recorded within each vehicle after the communication and computation process, and will later be used by the LR in JR election. This is represented in Equation (1).

$$\text{Lane Representative} = \text{Vehicle}_{\text{Min Timestamp}} \quad (1)$$

Every vehicle broadcasts its timestamp to all other vehicles in its street when it reaches region ‘X’. The broadcast message consists of the vehicle’s IP address, street ID to which the vehicle belongs, its coordinates obtained from a GPS system, its dimensions which is used for calculation of traffic density by the LR and timestamp. The broadcast of the messages uses the UDP protocol. A periodic broadcast of messages is issued every 5 seconds so as to account for the possibility of packet loss due to UDP protocol. Messages from other streets and duplicate or redundant messages are discarded without processing the entire message to improve the performance.

The Lane Representative selection algorithm runs within each vehicle in the particular street present between regions ‘X’ and ‘Y’. This computation, performed by the CM of every vehicle, involves finding the IP address of the vehicle with the least timestamp. This vehicle with the least timestamp is ultimately the LR of the street. Each vehicle in the street will hence have knowledge of the LR’s IP address. To achieve a general consensus among the vehicles on the knowledge of the LR and to account for the vehicles with faulty CM’s, a signal is sent to the ComM in each vehicle to broadcast the LR’s IP address to all vehicles in the LR election state. The vehicle assigned with the LR’s IP address also receives this broadcast message and moves into the “LR state”.

### 2.2.2 Junction Representative Election Algorithm

This algorithm deals with the election of the vehicle which acts as the authority for the entire junction. This JR has the responsibility of directing and clearing the traffic. From point 'Y' to 'L' the LRs broadcast their vehicle's IP address, the density of traffic calculated by taking into account vehicles per unit area and the number of vehicles in their street for which this vehicle acts as the LR. This broadcast also happens periodically at an interval of 5 seconds to account for packet loss.

At point 'L' each LR involved in the JR election has information about other LR's and their street information. A hash table is used to store this information. Once each LR has all information about the other streets, each LR in its CM, calculates the JR probability value of a street using Equation (2).

$$JRPV=(RT+Den)*(TV/VIL) \quad (2)$$

Where: JRPV-Junction Representative Probability Value, RT-Red Timestamp, Den-Density, TV-Total Vehicles, VIL- number of Vehicles In Lane.

This equation is designed so that the street with the maximum number of vehicles, greatest density and one in which the vehicles have been waiting in a Red Signal (RS) state for a long time is the first to receive a Green Signal (GS). The vehicle with the least JRPV is elected as the JR. The street with the JR technically acts as the controlling street and has the lowest chance of receiving a GS.

The value of the Red Timestamp (RT) indicates the total amount of time that the vehicles in a street have been waiting in a Red Signal (RS). The election process takes into consideration the RT parameter so as to avoid starvation i.e., the same street receiving a RS every time the algorithm runs. Higher the RT, lower is the probability of the LR in that street becoming the JR and higher probability of that street receiving a GS. The LR vehicles that have not been elected to be JR, wait for the signal information from the JR. The JR calculates the green light period G(T) for each of the other lanes using the parameters it has gathered. This period is dynamic based on the number of cars and density of traffic in each lane. There is a base G(T) of 10 seconds and then it is further incremented based on the number of vehicles in the lane by the formula in Equation (3).

$$G(T)=10+(Num*2) \quad (3)$$

Where: 'Num' is number of vehicles in the lane.

The Green Signal (GS) message, consisting of the G(T) value and the street ID that receives the GS, is sent by the JR to the LR's of the other street. The street ID with the highest JRPV is the first street to receive a GS. This GS message is also sent as a periodic broadcast to by the JR to the LR's at an interval of 2 seconds. The LR of the street that receives the GS, broadcasts the GS message to all the vehicles that were involved in the LR selection process. The LR's of the other streets that have not received a GS, broadcast a Red Signal (RS) message to the vehicles in its street. Once a street receives a GS, the LR and other vehicles of the street switch to the moving state and move forward. The whole process of LR selection and JR selection repeats again as new vehicles enter the junction.

At times of emergency in a passenger vehicle, an ambulance or a police chase, the vehicles have to notify the localised database hosted in the RSU with the vehicle’s IP address and street ID, which in turn broadcasts it. The JR then has to recompute the values using a special case condition that can be defined, and issue a GS to that particular street until the vehicle has passed by. The vehicle sends a leaving beacon to the RSU to notify that the vehicle has passed and that the normal course of action can be ensued. The person in charge of the vehicle can then be questioned by the authorities to determine whether the reason for the emergency call was authentic.

If a vehicle breaks the RS then the vehicles registration number and IP address automatically gets registered with the local database in the RSU for further action. This can be seen as an alternative to the traditional detection using cameras to save resources and also place a better system in place.

The communication between the vehicles is based on the UDP protocol. The transfer of information between the vehicles and the localised database on the street hosted in the RSU is via the TCP transport layer protocol. The IEEE standard used is 802.11p.

Currently, the algorithm does not include allowances for right and left turn, U-turn signals and pedestrian signals. The algorithm will be enhanced with these as a part of our future work.

### 3 Results

Each node is a system comprising of the CM and ComM. These nodes connect to a central system that is running the Traffic Simulation Module (TSM) through a reliable TCP connection. Every node that is connected is detected by the TSM and is shown as a moving dot to view a bird’s eye view of the simulation. Fig. 3 shows the user interface of the node running the traffic simulator. All the vehicles that participate in the CJCS algorithm are shown on the map of the junction. The colours of the dots representing the vehicle change with the changes in the vehicle’s states. In the Fig. 3, it is seen that street 3 has received the green light and the vehicles in that street are moving. Street 1 and street 2 are seen to have a red signal. The vehicles depicted in black are still in the broadcasting state and haven’t started implementing the algorithm.

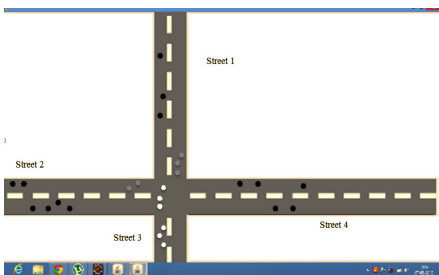


Fig. 3. Traffic Simulator Module

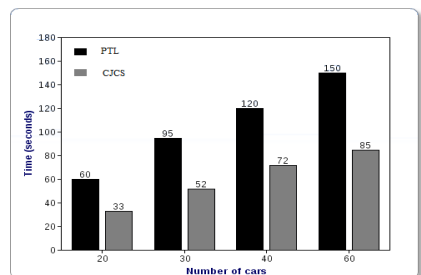


Fig. 4. Comparison of traffic clearance time between PTL and CJCS



An analysis of the results of the simulation showed promising results. A video camera was placed on a building close to the junction to capture the vehicular movement at the Kathriguppe traffic junction in Bangalore, India. The junction was chosen to conduct the study as it had 4 separate streets forming the junction at right angles to each other. The junction is a true representative of a busy Indian road junction with approximately 150 vehicles participating in the traffic signal at the junction. The building was chosen to be at an optimal height so as not to impair the quality of the video and yet capture a view of the whole junction and vehicle movement patterns.

The vehicular movement patterns captured in this video recording were replicated in close approximation in the simulation setup that was built. The vehicles in the simulation were assumed to move in straight line path. It was also assumed that every vehicle at the junction was fitted with an OBU and the conditions of the roads were ideal. On each node, 10 virtual machines were hosted with different IP addresses each running an instance of the same program, thus behaving as 10 different vehicles. A group of 15 different systems were connected to a network and the simulation setup was built. A timer was triggered on starting the simulation of the CJCS on this setup. All the different vehicular densities patterns that were observed were replicated in the simulation. The times taken to clear the traffic in the simulation and the video recording data were compared for each of these 4 simulations carried out and the graph was plotted. The graph in Fig. 4 shows a comparison between the time taken by physical traffic lights and CJCS to clear the same amount of traffic. The physical traffic lights take a longer time because the time allocated for a particular street is constant and due to this, traffic in other streets is not able to move even if the currently green light allocated street has no traffic. The CJCS on the other hand allocates green light dynamically and thus almost eliminates the wastage of time. On an average there is around 42% decrease in the amount of time taken to clear a specified amount of traffic by using CJCS.

Comparisons of the costs and expenses involved were also carried out. While the electrical Physical Traffic Light (PTL) setup can cost anywhere between 4-6lakh and borne by the governing agency, the price of the OBU on every vehicle that will participate in the CJCS will be included in the cost of the vehicle itself. This can lead to substantial reductions in cost for the governing agencies.

## 4 Conclusion

In this paper, we briefly reviewed the role of ITS and VANETS in helping improve the field of traffic management. We introduced the CJCS algorithm developed as a solution for managing traffic at road junctions in a VANET environment. The CJCS is a novel and efficient method of traffic management and can drastically improve the traffic scenario.

In this work, a simulation framework was built to evaluate the algorithm and analyse the working of the algorithm by simulating real world situations. A case study of the Indian roads was used to compare the results of the simulation against real world statistics. Varied simulation setups with large number of nodes, high chances of packet drop, single nodes, and malfunctioning nodes were tested to study the

complete working of the algorithm. The CJCS algorithm showed a significant reduction in waiting times of vehicles and proved to be 42% more efficient than its physical traffic light counterpart. Additionally, the cost of using this system was also found to be lower than the physical traffic light infrastructure. Hence we propose that this system can effectively substitute the current system and completely change the traffic control and management procedures in junctions.

## References

1. Zeadally, S., Hunt, R., Chen, Y.-S., Irwin, A., Hassan, A.: Vehicular ad hoc networks (VANETS): status, results, and challenges. In: Proceedings of VANET 2004, Philadelphia, PA USA (October 2012)
2. Kamini, R.K.: VANET Parameters and Applications: A Review. *Global Journal of Computer Science and Technology* 10(7), Ver. 1.0, 112–123 (2010) ISSN 0348-1756
3. Chandrasekaran, G.: VANETs: The Networking Platform for Future Vehicular Applications. In: *Mobile Networking for Vehicular Environments*, Los Angeles, pp. 109–114 (March 2009)
4. Rawashdeh, Z.Y., Mahmud, S.M.: Communications in Vehicular Networks. *International Journal of Research and technology (IJERT)* 2(4), 34–45 (2009) ISSN 4589-8231
5. Campolo, C.: On vehicle-to-roadside communications in 802.11p/WAVE VANETs. In: *Proceedings of Wireless Communications and Networking Conference (WCNC)*, pp. 91–102. IEEE Press, Houston (2011)
6. Oprea, A.: MAC protocols for VANETS. *International journal on AdHoc Networks* 4(5), 82–97 (2010) ISSN 8794-0965
7. Cristea, V., Gradinescu, V., Gorgorin, C., Diaconescu, R., Iftode, L.: Simulation of VANET Applications. Paper presented at 1st IEEE International Symposium on Wireless Vehicular Communications, pp. 56–67. Phoenix (January 2011)
8. Khairnar, V.D., Pradhan, S.N.: Comparative Study of Simulation for Vehicular Ad-hoc Network. *International Journal of Computer Applications* 4(10), 70–81 (2010)

# Identifying Best Products Based on Multiple Criteria Using Decision Making System

Swetha Reddy Donala, M. Archana, and P.V.S. Srinivas

Department of Computer Science and Engineering,  
TKR College of Engineering & Technology, Hyderabad, A.P.-500 097, India  
{swethareddy.donala,mogullaarchana7,  
pvssrinivas23}@gmail.com

**Abstract.** The necessity of dominance and skyline analysis has been developed in multi-level solution creating applications. Earlier researches stare on the way to help users to get a set of efficient possible products from a stream of available products. In this paper, we explored different crisis of a problem, retrieving top-k preferable products, which are not explored in previous researches. Available a set of products in the existing system, we are in need to hunt a set of k efficient available products incase these new ones are not influenced by the products which are already present in the old market. We research couple of problems of getting top-k preferable products. In the first crisis instance, we are in situation to set the cost of these products in that way the complete profit is increased. Those products which are capable known as top-k profitable products. Approaching to another problem or crisis, we are in need to get k products in that way these k products are capable of attracting the huge number of users. Such products are referred as top-k products. Moreover, there are a multiple number of available subsets. In this paper, we prefer solutions to get the top-k profitable products which are of accurate in nature and also the top-k popular products efficiently. An extreme working research by utilizing both synthetic and real data sets is referred to examine the accuracy and efficiency of referred algorithm.

**Keywords:** P.V.S.Srinivas, M.Archana, Data Mining, Finding best products.

## 1 Introduction

The main in lots of multicriteria decision-making apps are dominance analysis. Taking skyline one of the examples. Take a reality situation if customer want to buy a new product like mobile, he filters the brand and get top two branded companies p and q. He checks the properties or known as features of both products.

Those features are like battery life, camera pixel,model, technology used etc. Then the comparison goes between top two products p and q. If any features of p one are less than the q, then p was dominated by q lets us take a three different brands: p1; p2,

p3 and p4. As per customer view that a lower price and a long battery life are more preferable.[10]

Brands	Long battery life	Price
P1	6 hours	3000
P2	4 hours	4000
P3	2 hours	5000
P4	3 hours	6000

Here P4 has a longer battery life than the p3 but it has high price than the p3. So p4 was dominated by p3. These both are dominated by p2 because it has longer battery life than the p3 and p4 and also it has lowest price than the both of them.[5][3] Based on complete table most preferable and also top one is p1 because it has longer battery life than the others and also it has lower price than the others. So p1 dominates all p2, p3 and p4.

### 1.1 Getting Most Accurate and Top-k Profitable Products

The first instance is called finding top-k profitable products. A naive way for this instance/problem is to enumerate all possible subsets of size k from Q, calculate the sum of the profits of each possible subset, and choose the subset with the greatest sum. However, this approach is not scalable because there are an exponential number of all possible subsets. This motivates us to propose efficient algorithms for problem TPP.[2][8][14].

Although how we set the price of a new package may affect how we set the price of another new package and there are an exponential number of possible subsets, interestingly, we propose a dynamic programming approach which finds an optimal solution when there are two attributes to be considered.[7][11] But, we show that this problem is NP-hard when there are more than two attributes to be considered. Thus, we propose two greedy algorithms for this problem. One greedy algorithm has a theoretical guarantee on the profit returned while the other greedy algorithm performs well empirically.[9]

Finding top-k profitable products is common in many real-life applications. Other applications include finding profitable laptops in a new laptop company, finding profitable delivery services in a new cargo delivery company and finding profitable e-advertisements in a webpage. [10] In some cases, data sets are dynamic and change from time to time.

In this paper, we also study how to find top-k profitable products when data sets change. For example, some new products are launched in the existing market while some products which were present in the existing market become unavailable.

Besides, the prices of existing products in the market may change due to various reasons, such as inflation and cost increase.[19]

## 1.2 Getting Top-k Popular Products

The second instance is called finding top-k popular products. In some cases, if we know how many customers are interested in some potential products, we can better find potential products. One well-known application which allows customers to provide their preferences is —Name Your Own Price developed by —Priceline.com. If customers indicate their preferences on some hotels, —Name Your Own Price service will return some potential hotels to customers. Similarly, a naive way is to enumerate all possible subsets of size  $k$  from  $Q$ , calculate the total number of customers interested in some packages in this subset, and choose the subset with the greatest number of customers. But, it is not scalable.[18] We show that this problem is NP-hard. But, interestingly, we propose a 0.63-approximate algorithm which runs in polynomial time.

## 2 System Evaluation

The skyline of a given data set  $D$  is denoted by  $SKY(D)$ . We have a set  $P$  of  $m$  tuples in the existing market, namely  $p_1; p_2; \dots; p_m$ . Each tuple  $p$  has  $l$  attributes, namely  $A_1; A_2; \dots; A_l$ . The domain of each attribute is  $\mathbb{R}$  where a smaller value is more preferable. The value of attribute  $A_j$  for tuple  $p$  is given and is denoted by  $p:A_j$  where  $j \in \{1, \dots, l\}$ . In particular, the last attribute  $A_l$  represents attribute price and all other attributes represent the attributes other than price. Besides, we have a set  $Q$  of  $n$  potential new tuples, namely  $q_1; q_2; \dots; q_n$ . Similarly, each tuple  $q$  has the same  $l$  attributes, namely  $A_1; A_2; \dots; A_l$ . The value of attribute  $A_j$  for tuple  $q$  is denoted by  $q:A_j$  where  $j \in \{1, \dots, l\}$ . However, the value of attribute  $A_l$  for tuple  $q$  is not given and the value of each of the other attributes is given. We assume that no two potential new tuples in  $Q$  are identical (i.e., no two tuples in  $Q$  have the same attribute values for  $A_1; A_2; \dots; A_{l-1}$ ). In addition to these  $l$  attributes, each tuple  $q$  is associated with one additional cost attribute  $C$ . The value of attribute  $C$  for  $q$  is denoted by  $q:C$ . We assume that for any two tuples in  $P, Q$ , they have at least one attribute value different among the first  $l-1$  attributes. This assumption allows us to avoid several complicated, yet uninteresting, —boundary cases. If this assumption does not hold, the proposed algorithms can be modified accordingly.[19][16]

### *Popular Products Dataset*

The product dataset will be formed based upon the criteria of existing products which are already present in the market by different vendors. Let us take an example of cars. For that research purpose on car products, the existing cars, features, cost etc. has to be gathered from different vendors

### *Frequent Feature Set*

Frequent feature set identification on the dataset of popular products. The high utility mining refers to the feature set extraction, by satisfying certain conditions. The

condition in the work is high profitability.

Considering the high profitability, identify the feature set that occurred more frequently by the other vendors.

**Finding the Smaller Frequency Sets**

The Frequency set provides various levels of sets.

**Ex:level1:** with single item

**level2:** with two combinations of items

**level3:** with three combinations of items Have to build the least sets to build a product.

**Setting Optimal Value**

Setting the optimal value to dominate existing market, by its price, with Dynamic programming.The outcome of Dynamic Programming is the set of size't' which provides greater profit.

**Price Correlation**

For the Price Correlation among the optimal feature.Set 't' and the existing feature and their cost.

Here NP-Hard with Greedy Approach is applied.

**Algorithm**

With Algorithm 1, intra-dominance checking steps are removed if the scenario has the at-most-one merging attribute characteristic. Thus, |TQ|<sup>2</sup> checks are avoided. Since we focus on processing T0 Q instead of TQ (where T0 Q = TQ), the total number of interdinance checking steps is reduced from |TE| × |TQ| to |TE| × |T0 Q|.

---

**Algorithm 1** Algorithm for Creating Competitive Products

---

**Input:** a set  $T_E$  of products in the existing market and a set  $T'_Q$  of all possible products from  $T'_1, T'_2, \dots, T'_k$

**Output:** the set  $O$  of competitive products

- 1:  $O \leftarrow \emptyset$
- 2: **for each**  $q \in T'_Q$  **do**
- 3:     // if  $q \in SKY(T_E \cup \{q\})$
- 4:     **if**  $q$  is not dominated by any tuple in  $T_E$  **then**
- 5:          $O \leftarrow O \cup \{q\}$
- 6: **return**  $O$

---

The total number of checks in this algorithm is utmost |TE|× |T0 Q|. Similarly, we can find T0 E = SKY (TE) so that the number can be reduced to |T0 E| × |T0 Q|.

In the following, when we write TE, we mean T0 E. Although Algorithm 1 helps us to derive an efficient algorithm, a naïve implementation still *materializes* all

possible products generated from  $T_0 1, T_0 2, \dots, T_0 k$  and obtains a set  $T_0 Q$ , which is computationally expensive. As we described before, if is the size of each table  $T_0 i$ , the total number of tuples in  $T_0 Q$  is  $k$ . In the following, propose techniques to avoid materializing  $T_0 Q$ .[5][8]

### 3 Related Work

Our paper adds to a growing literature on sentiment analysis. Similar to almost any other sentiment mining technique, the first step involves selecting the set of features to use. Our approach to feature selection is very close to the one presented by Hu and Liu [15]. Hu and Liu used a POS-tagger and association miner to find frequent item-sets, which are then treated as candidate features [16][7][13].For each instance of a candidate feature in a sentence, the nearby adjective (if there is such) was treated as the effective opinion. Note that effective opinions used by Hu and Liu are direct counterparts of evaluation words in our study. A major difference of our approach is that whereas Hu and Liu were interested in the effectiveness of feature extraction and high recall values, we are concerned about gathering a small set of major features and evaluation words[7][2]. In order to ensure that gathered features reflect hedonic characteristics of the products, we performed manual post-processing of frequent itemsets. Contrary to Hu and Liu, who performed additional research to identify infrequent features, we intentionally discarded such features to obtain robust estimates of model coefficients.[4][8][2].

## Experimental Result for Finding Top-k Profitable Products

### i) Result over Synthetic Data Sets

In this section, we conducted experiments over both small and large synthetic data sets to study the scalability of GR1 and GR2. We varied  $j, q, d, l, k, \_$ , and  $h$  in our experiments. The values of each parameter used over large synthetic data sets are given in Table 4 where the default values are in bold. For the sake of space, we show the results when we varied  $j, q$  only in Fig. 1. Other experimental results over small synthetic data sets and large synthetic data sets can be found in [19][2].

**Execution Time:** Fig. 1a shows the effect on the execution times of all algorithms. GR2 runs slower than GR1. As we discussed, the time complexity of GR2 is higher than that of GR1.

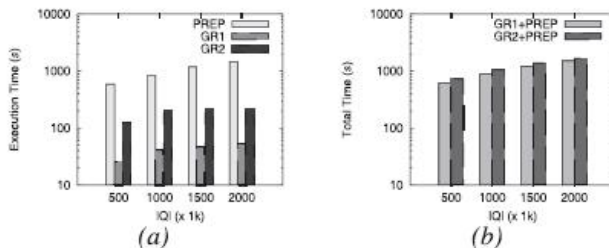


Fig. 1(a). Execution Time

**Total Time:** Fig. 1b shows the effect on the total time of each greedy algorithm which corresponds to the sum of the execution time of the greedy algorithm and the execution time of the preprocessing time of the greedy algorithm. We denote the total times of GR1 and GR2 by —GR1 PREP and —GR2 PREP, respectively. It is clear that when  $jQj$  increases, the total times of the algorithms increase. Memory cost. Fig. 1c shows the effect on the memory cost of the algorithms. Since the memory cost of both GR1 and GR2 is the memory occupied by the R-tree on data sets P Q, when  $jQj$  increases, the memory cost increases, as shown in Fig. 1c. Profit. Fig. 1d shows the effect on the profit returned by the algorithms. In most cases, GR1 and GR2 give similar profits.[17]

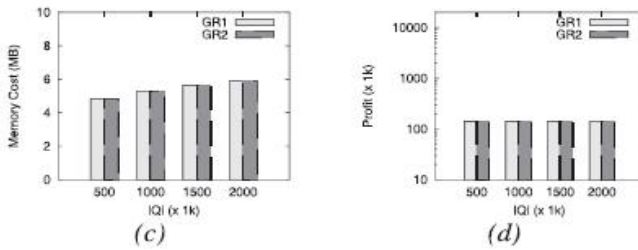


Fig. 1(b). Total Time

ii) **Result over Real Data Sets**

We also conducted experiments on real data sets. We varied four factors, namely  $h$ ;  $k$ ;  $d$ , and  $\_$ . For the sake of space, we only show the results with two factors  $h$  and  $k$ . The default setting configuration is:  $k \frac{1}{4} 150$ ,  $h \frac{1}{4} 20$ ,  $d \frac{1}{4} 0:6$ , and  $\_ \frac{1}{4} 50$ . The results for real data sets are similar to those for synthetic data sets. Note that there is a big difference in execution times between GR1 and GR2 in the real data sets but this big difference cannot be found in the synthetic data sets. [3][7]As we described the time complexity of GR2 is quadratic with respect to  $k$  but the time complexity of GR1 is not. Thus, when  $k$  is larger, then the difference in execution times between GR1 and GR2 is larger. Compared with the synthetic data sets where  $k$  is set to 10, 20, 50, or 100, in the real data sets, since  $k$  is set to a larger value (e.g., 100, 150, 200, and 250), the difference in execution times between GR1 and GR2 is larger. [5][9][4]

**4 Conclusion**

In this paper, we find and tackled the problem of getting top- $k$  preferable products, which are not gone through previously. We researched couple of instances which are of preferable products. Those are referred as profitable products and another one is popular products. We explores methods to tackle top- $k$  profitable products and also top- $k$  popular products perfectly. Research of an extensive performance by using both synthetic and real data sets results to check its accuracy and efficiency. For further



researches, we will go through other aspects of the crisis of finding top-k preferable products by maintaining the utility function to other relative objective functions. This complete description explains single promising utility function is the function which gives back the sum of the unit profits of the choosed products multiplied by the multiple number of users preferred in these products.

## References

- [1] Borzsonyi, S., Kossmann, D., Stocker, K.: The Skyline Operator. In: Proc. Int'l Conf. Data Eng. (ICDE) (2001)
- [2] Bently, J.L., Kung, H.T., Schkolnick, M., Thompson, C.D.: On the Average Number of Maxima in a Set of Vectors and Applications. *J. ACM* 25(4), 536–543 (1978)
- [3] Bently, J.L., Clarkson, K.L., Levine, D.B.: Fast Linear Expected-Time Algorithms for Computing Maxima and Convex Hulls. In: Proc. First Ann. ACM-SIAM Symp. Discrete Algorithms (SODA) (1990)
- [4] Barndorff-Nielsen, O., Sobel, M.: On the Distribution of the Number of Admissible Points in a Vector Random Sample. *Theory of Probability and Its Application* 11 (1966)
- [5] Hockhbaum, D.S.: Approximating Covering and Packing Problems: Set Cover, Vertex Cover, Independent Set, and Related Problems. In: *Approximation Algorithms for NP-Hard Problems*. PWS Publishing Company (1997)
- [6] Jiang, B., Pei, J., Lin, X., Cheung, D.W.-L., Han, J.: Mining Preferences from Superior and Inferior Examples. In: Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining (2008)
- [7] Kang, J.M., Mokbel, M.F., Shekhar, S., Xia, T., Zhang, D.: Continuous Evaluation of Monochromatic and Bichromatic Reverse Nearest Neighbors. In: Proc. Int'l Conf. Data Eng. (ICDE) (2007)
- [8] Korn, F., Muthukrishnan, S.: Influence Sets Based on Reverse Nearest Neighbor Queries. In: Proc. ACM SIGMOD Int'l Conf. Management of Data (2000)
- [9] Kossmann, D., Ramsak, F., Rost, S.: Shooting Stars in the Sky: An Online Algorithm for Skyline Queries. In: Proc. 28th Int'l Conf. Very Large Data Bases (VLDB) (2002)
- [10] Li, B., Ghose, A., Ipeirotis, P.G.: Towards a Theory Model for Product Search. In: Proc. 20th Int'l Conf. World Wide Web (WWW 2011), pp. 327–336 (2011)
- [11] Lin, X., Yuan, Y., Zhang, Q., Zhang, Y.: Selecting Stars: The k Most Representative Skyline Operator. In: Proc. Int'l Conf. Data Eng. (ICDE) (2007)
- [12] Papadias, D., Tao, Y., Fu, G., Seeger, B.: Progressive Skyline Computation in Database Systems. *ACM Trans. Database Systems* 30(1), 41–82 (2005)
- [13] Sacharidis, D., Papadopoulos, S., Papadias, D.: Topologically-Sorted Skylines for Partially-Ordered Domains. In: Proc. Int'l Conf. Data Eng. (ICDE) (2009)
- [14] Stanoi, I., Riedewald, M., Agrawal, D., Abbadi, A.E.: Discovery of Influence Sets in Frequently Updated Databases. In: Proc. Int'l Conf. Very Large Data Bases (VLDB) (2001)
- [15] Tan, K.-L., Eng, P., Ooi, B.: Efficient Progressive Skyline Computation. In: Proc. Int'l Conf. Very Large Data Bases (VLDB) (2001)
- [16] Tao, Y., Ding, L., Lin, X., Pei, J.: Distance-Based Representative Skyline. In: ICDE 2009: Proc. IEEE Int'l Conf. Data Eng., pp. 892–903 (2009)

- [17] Wan, Q., Wong, R.C.-W., Ilyas, I.F., Ožsu, M.T., Peng, Y.: Creating Competitive Products. In: Proc. VLDB Endowment, vol. 2, pp. 898–909 (2009)
- [18] Wan, Q., Wong, R.C.-W., Peng, Y.: Creating Top-K Profitable Products. technical report (2010), <http://www.cse.ust.hk/~raywong/paper/createTopKProfitableProduct-technical.pdf>
- [19] Wan, Q., Wong, R.C.-W., Peng, Y.: Finding Top-K Profitable Products. In: Proc. Int'l Conf. Data Eng. (ICDE) (2011)

# Design and Performance Analysis of File Replication Strategy on Distributed File System Using GridSim

Nirmal Singh and Sarbjeet Singh

Computer Science and Engineering,  
UIET, Panjab University, Chandigarh, India  
nsinghin@gmail.com, sarbjeet@pu.ac.in

**Abstract.** Distributed Computing Systems like Peer-to-Peer, Web and Cloud are becoming popular day by day and are being used in a wide variety of data intensive applications. Data replication is an important concept which increases the availability and reliability of data in these systems. This paper presents the design and performance analysis of simulated implementation of data replication strategy on a distributed file system using GridSim toolkit. The parameters taken for the performance analysis are aggregate bandwidth, successful execution rate and system byte effective rate. The results indicate that the distributed file system making use of data replication strategy performs better, with respect to the parameters mentioned above, than the distributed file system which is not making use of data replication strategy. The integration of data replication scheme with distributed file system can greatly improve the availability and reliability of data.

**Keywords:** Distributed File System, Replication Strategy, Web, Cloud, Peer-to-Peer, GridSim.

## 1 Introduction

Distributed Computing systems like Grid and Cloud enable storage of large amount of data and provide high speed access to that data. Data is made available to users according to their requirements and demand. Moreover, data is made available to users at different geographical locations. Reliability and availability of data are the major concerns for the user who accesses data and play crucial role in making these systems popular and widely acceptable. These parameters can be guaranteed by replicating the data at several sites over the system but this introduces other problems like maintaining different replicas and ensuring consistency among them. Zero replication limits availability and reliability of data whereas large number of unnecessary replicas makes the process of updating the data complex and error prone. So number and location of replicas must be chosen very carefully. A survey of important data replication strategies being used in wide area distributed systems is given in [1].

This paper presents the performance analysis of Dynamic Data Replication Algorithm (D2RS) [2] on Hadoop [3] like distributed file system. The simulated

implementation has been done using GridSim [4] toolkit. The paper is organized as follows: Section 2 describes Hadoop Distributed File System, Section 3 briefs Dyanmic Data Replication Strategy (D2RS), Section 4 describes the integration of D2RS with distributed file system, Section 5 presents implementation details and finally Section 6 concludes the work with future scope.

## 2 Hadoop Distributed File System [HDFS]

Hadoop is an open-source system consisting of a distributed file system and its programming model. The underneath file system is called Hadoop Distributed File System (HDFS). It is similar to the Google File System (GFS) and is designed to run on commodity hardware. It provides high throughput access to data and is suitable for applications that make use of large data sets. The two major components of this architecture are the NameNode and the DataNode. In HDFS, file system metadata and application data are stored separately. The dedicated server which stores metadata is called NameNode and the servers which store application data are called DataNodes [3], [5].

In HDFS, a data file is stored as a sequence of blocks. Thus for storage in HDFS, a file is divided into blocks. The information of file to block mapping is stored in NameNode whereas blocks itself are stored on various DataNodes. The NameNode maintains the HDFS namespace. The operations like opening a file, renaming a file, closing a file, deletion of file etc. are handled by NameNode. The NameNode also maintains a log of various operations performed on metadata. The DataNodes store blocks of file in local file system. The read and write requests of files are served by DataNodes. The blocks of a file are replicated over different DataNodes to provide fault tolerance. The number of replicas of a block and the size of the block are configurable in HDFS.

## 3 Dynamic Data Replication Strategy

The D2RS (Dynamic Data Replication Strategy) proposed in [2] describes the way to dynamically replicate data files in distributed file system. In D2RS, a dynamic data replication strategy is described which works by taking into consideration the following points:

- Which data file should be replicated?
- When data file should be replicated?
- How many replicas should be created?
- Where the replicas should be placed?

Data replication is an effective approach to ensure reliability and availability. Various data replication algorithms, strategies and schemes have been discussed in [6-16]. [2] define several terms like availability, file availability, popularity degree and replica factor etc. which are used in the algorithm and with the definitions it also provides the formulas to derive numerical value of that parameter. These parameters

determine when and where to replicate a data file and also how many copies of the data file further needs to be made.

#### 4 Integration of D2RS with Hadoop Distributed File System

The workflow of HDFS can be divided into two stages: data saving/creation stage and data retrieval/access stage. In data creation stage, a request from client is sent to NameNode to save the data in file system. NameNode replies back to client with location of DataNode where data can be saved. In data retrieval stage, a client sends an access request to NameNode with the filename. When NameNode receives the request, it searches its namespace and retrieves the DataNode that is closest to this client and responds back to request. Data retrieval in HDFS with D2RS is different and has more probability of data availability. In this case, most of the operations remain same but difference comes when all the current replicas for a particular data file are leased out to other clients and more requests come for the same data file. If number of requests exceeds a predefined threshold then further replication is triggered. Fig. 1 shows the complete workflow in this case.

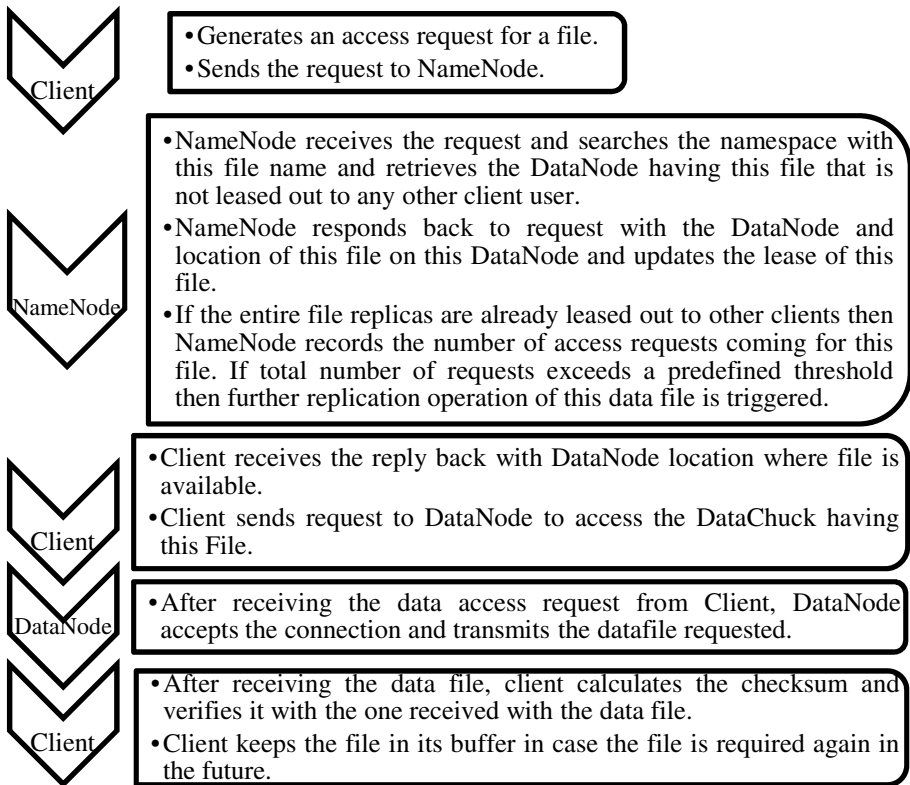


Fig. 1. Workflow of HDFS with D2RS

Following section describes the implementation details and the results of performance analysis.

## 5 Implementation Details and Performance Analysis

The implementation has been done using GridSim toolkit. The pseudocodes of existing HDFS like distributed file systems and HDFS with D2RS are described below:

### 5.1 Pseudo-code of Existing HDFS Like Distributed File System

- a. *A new file access request is generated by a file user client, it is sent to the NameNode of the file system.*
- b. *NameNode receives the request and searches its namespace with this file name and retrieves the DataNode having this file and is not leased out to any other client user.*
- c. *NameNode responds back to request with the DataNode and location of this file on this DataNode and updates the lease of this file in its namespace.*
- d. *Client receives the reply back with DataNode location where file is available.*
- e. *Client sends request to DataNode to access the data chunk having this file.*
- f. *After receiving the data access request from client, DataNode accepts the connection and transmits the datafile requested.*
- g. *After receiving the data file, client calculates the checksum and verifies it with the one received with the data file.*

### 5.2 Pseudo-code of HDFS with D2RS

- a. *A new file access request is generated by client. This request is sent over to NameNode of the file system.*
- b. *NameNode receives the request and searches its namespace with this file name and retrieves the DataNode having this file which is not leased out to any other client user.*
- c. *NameNode responds back to the request with the DataNode and location of this file on the DataNode and updates the lease of this file in its namespace.*
- d. *If the all file replicas are already leased out to other clients then, NameNode records the number of access requests coming for this file. If total number of requests exceeds a predefined threshold then the further replication of this data file is triggered and request is addressed with this new replication.*
- e. *Then client receives the reply back from NameNode with the location of DataNode where file is available.*
- f. *Further Client sends request to DataNode to access the data chunk having this file.*

- g. After receiving the data access request from client, DataNode accepts the connection and transmits the datafile requested.
- h. After receiving the data file, client calculates the checksum and verifies it with the one received with the data file.

The simulation strategy adopted is to simulate the Hadoop like distributed file system on GridSim simulation toolkit and then simulating another version of file system with D2RS. Test runs are performed on both the simulations and results are recorded. Performance is measured on the basis of results received in test runs. Flow chart of Fig. 2 describes the general approach adopted for simulation.

Performance analysis has been done on the basis of aggregate bandwidth distribution, successful execution rate and system byte effective rate, as mentioned in the following sections:

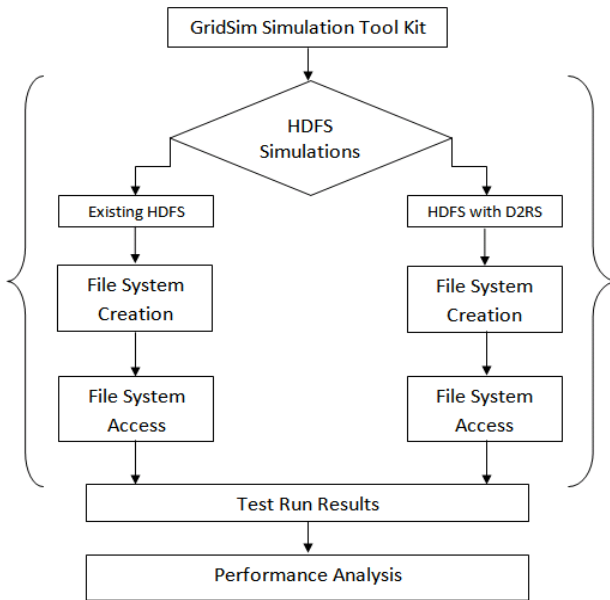


Fig. 2. Simulation strategy

### 5.3 Aggregate Bandwidth Distribution

Aggregate bandwidth [16] is the sum of the bandwidths of DataNodes in the File System that are consumed at all terminals in a file system network. Fig. 3 shows the experimental results for Aggregate Bandwidth Distribution obtained by running the HDFS and HDFS with D2RS on the GridSim. In Fig. 3, solid portion depicts the bandwidth distribution of the HDFS with D2RS where as the lined portion shows the bandwidth distribution for normal HDFS. We can clearly see that the solid portion is more evenly distributed over the range whereas the lined portion shows very uneven

distribution. This simulation depicts that HDFS with D2RS utilizes the resources more efficiently and distributes the load evenly over the DataNodes. Furthermore the number of requests handled by the HDFS with D2RS is 326 which is double than the request that were handled by existing HDFS. In this simulation the replication factor used is 2 and number of DataNodes and files are 50.

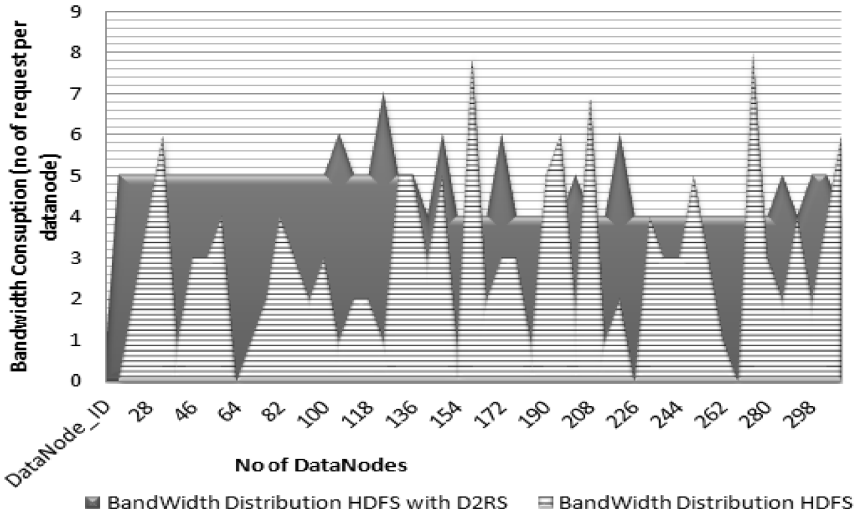


Fig. 3. Aggregate Bandwidth Distribution Comparison

Below are the detailed specifications that were used in the simulation:

Table 1. Experimental Specifications

No. of Data Nodes	50
No. of Files	50
Total requests	368
Number of Requests Handled by HDFS	151
Number of Requests Handled by HDFS with D2RS	326
Replication Threshold	2

#### 5.4 Successful Execution Rate

Successful execution rate can be defined as the file access request that is answered by the NameNode correctly. When D2RS algorithm is not used, the successful execution starts from 1 which is 100% completion rate but it drops as the number of requests increases. In the case of D2RS, the successful execution rate initially drops but eventually rises back as the file replication gets triggered when number of unsuccessful requests crosses a threshold. This can be clearly observed from Fig. 4.



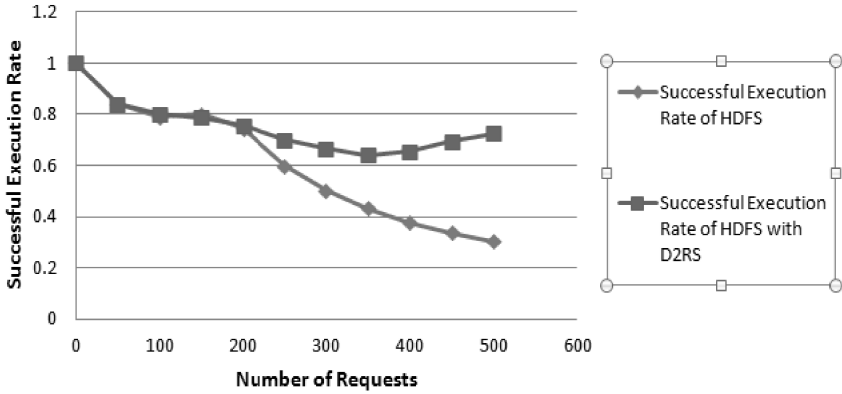


Fig. 4. Successful Execution Rate Comparison

### 5.5 System Byte Effective Rate

System Byte Effective Rate (SBER) is the rate of the number of bytes potentially available and the total number of bytes requested by all the tasks in the system. As shown in Fig. 5, as the number of replication factor increases, the SBER tends to reach near 1. The line showing the SBER graph for HDFS with D2RS algorithm reaches 1 earlier than the line showing SBER graph for existing HDFS. This means that even with lower replication factor, HDFS with D2RS algorithm can sustain more efficiently and hence reduces the resource over-utilization.

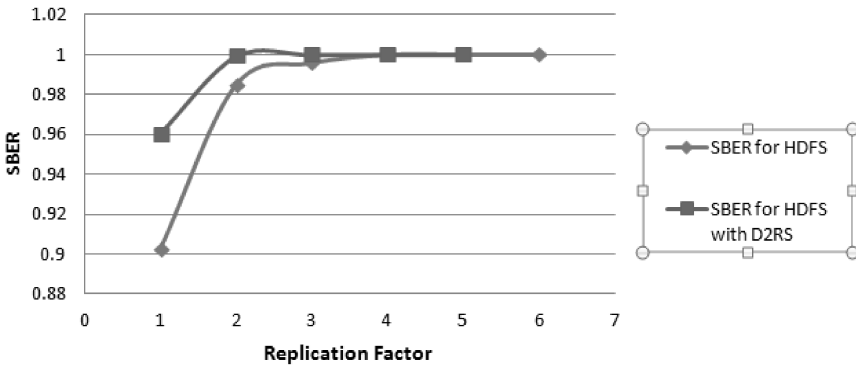


Fig. 5. System Byte Effective Rate Comparison

## 6 Conclusion and Future Scope

This paper presents the performance analysis of simulated implementation of data replication strategy on Hadoop like distributed file system using GridSim toolkit. Performance from various perspectives have been considered e.g. from aggregate bandwidth distribution point of view, successful execution rate point of view and system byte effective rate point of view. From the results obtained, it is clear that D2RS improves existing Hadoop like distributed file system and increases the reliability and availability of data.

## References

1. Goel, S., Buyya, R.: Data Replication Strategies in Wide Area Distributed Systems, <http://www.buyya.com/papers/DataReplicationInDSCchapter2006.pdf>
2. Sun, D., Chang, G., Gao, S., Jin, L., Wang, X.: Modeling a Dynamic Data Replication Strategy to Increase System Availability in Cloud Computing Environments. *Journal of Computer Science and Technology* 27(2), 256–272 (2012)
3. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The Hadoop Distributed File System. In: *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, May 03-07, pp. 1–10 (2010)
4. Buyya, R., Murshed, M.: GridSim: A toolkit for the modeling and simulation of distributed resource management and scheduling for Grid Computing. *Concurrency and Computation: Practice and Experience* 14(13-15) (2002)
5. Borthakur, D.: HDFS Architecture, The Apache Software Foundation, [http://hadoop.apache.org/docs/r0.20.0/hdfs\\_design.pdf](http://hadoop.apache.org/docs/r0.20.0/hdfs_design.pdf)
6. Rahman, R.M., Barker, K., Alhaji, R.: Replica placement design with static optimality and dynamic maintainability. In: *Proceedings of the 6th IEEE International Symposium on Cluster Computing and the Grid*, pp. 434–437 (2006)
7. Dogan, A.: A study on performance of dynamic file replication algorithms for real-time file access in data grids. *Future Generation Computer Systems* 25(8), 829–839 (2009)
8. Lei, M., Vrbsky, S.V., Hong, X.: An on-line replication strategy to increase availability in data grids. *Future Generation Computer Systems* 24(2), 85–98 (2008)
9. Litke, A., Skoutas, D., Tserpes, K., Varvarigou, T.: Efficient task replication and management for adaptive fault tolerance in mobile grid environments. *Future Generation Computer Systems* 23(2), 163–178 (2007)
10. Dobber, M., Van der Mei, R., Koole, G.: Dynamic load balancing and job replication in a global-scale grid environment: A comparison. *IEEE Transactions on Parallel and Distributed Systems* 20(2), 207–218 (2009)
11. Yuan, D., Yang, Y., Liu, X., Chen, J.: A data placement strategy in scientific cloud workflows. *Future Generation Computer Systems* 26(8), 1200–1214 (2010)
12. Rood, B., Lewis, M.J.: Grid resource availability prediction-based scheduling and task replication. *Journal of Grid Computing* 7(4), 479–500 (2009)
13. Latip, R., Othman, M., Abdullah, A., Ibrahim, H., Sulaiman, M.N.: Quorum-based data replication in grid environment. *International Journal of Computational Intelligence Systems* 2(4), 386–397 (2009)

14. Wei, Q., Veeravalli, B., Gong, B., Zeng, L., Feng, D.: CDRM: A cost-effective dynamic replication management scheme for cloud storage cluster. In: Proceedings of the 2010 IEEE International Conference on Cluster Computing, Heraklion, pp. 188–196 (2010)
15. Bonvin, N., Papaioannou, T.G., Aberer, K.: A self-organized, fault-tolerant and scalable replication scheme for cloud storage. In: Proceedings of the 1st ACM Symposium on Cloud Computing, Indianapolis, pp. 205–216 (2010)
16. Peter, B., Ishai, M., Mosharaf, C., Pradeepkumar, M., David, A.M., Ion, S.: Surviving failures in bandwidth-constrained datacenters. In: Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, August 13-17 (2012)

# Extended Goal Programming Approach with Interval Data Uncertainty for Resource Allocation in Farm Planning: A Case Study

Bijay Baran Pal<sup>1,\*</sup> and Mousumi Kumar<sup>2</sup>

<sup>1</sup> Department of Mathematics, University of Kalyani,  
Kalyani-741235, West Bengal, India  
bbpal18@hotmail.com

<sup>2</sup> Department of Mathematics, Alipurduar College,  
Alipurduar Court-736122, West Bengal, India  
mousumi886@gmail.com

**Abstract.** This paper presents an extended version of goal programming (GP) approach for modeling and solving farm planning problems having objectives with interval parameter sets by utilizing farming resources in the planning horizon. In the model formulation of the problem, the defined goals with interval parameters are converted into conventional goals by using interval arithmetic technique in interval programming and introducing under- and over-deviational variables to each of them. In the decision process, extended GP (EGP) approach, i.e. convex combination of both the modelling aspects, *minsum* GP and *minmax* GP are addressed in the achievement function for minimizing the possible regret towards goal achievement from the optimistic point of view in the inexact decision making environment. The potential use of the approach is demonstrated via a case example.

**Keywords:** Farm Planning, Goal Programming, Interval arithmetic, Interval Programming.

## 1 Introduction

The farm planning problems are actually management science/operations research problems with multiplicity of objectives of optimizing them for production of crops towards meeting the need of food in society. The general mathematical programming (MP) model for allocation of arable land and thereby optimal production of crops was first presented by Heady [1]. A comprehensive bibliography on the *state-of-the-art* of modeling aspects of agriculture planning problems was prepared by Nix [2]. The conventional GP [3] approach based on the satisficing philosophy [4], as one of the most promising tool for multiobjective decision analysis, has been introduced to the field of crop planning problems in [5, 6].

The methodological aspects of fuzzy programming [7] as well as fuzzy goal programming (FGP) [8], as an extension of the conventional GP method, have been

---

\* Corresponding author.

studied extensively and implemented to different real life problems [9, 10] including farm planning problem [11].

Now, in the real-world decision situations, it is to be observed that the setting of fuzzy goal values in an imprecise environment may not be always possible in practice. For accommodation of such an imprecision, model parameters can be represented by taking parameter sets in the form of intervals instead of crisp descriptions of them. Here, interval programming (IVP) approach [12, 13] has appeared as the robust tool for solving decision problems in an inexact environment. Again, to solve multiobjective decision making (MODM) problems, the conventional GP approaches to IVP problems known as interval goal programming (IVGP) have been studied by Inuguchi and Kume [14]. The methodological aspects of IVGP studied in the past have been surveyed by Olivera and Antunes [15] in 2007. The potential use of IVGP to different real-life MODM problems has also been demonstrated [16] in the recent past. However, deep study on methodologies of IVP as well as IVGP and implementation to practical problems is at an early stage.

In the present context of modeling the agricultural land allocation problem, the various model goals are defined with coefficient intervals and target intervals to make a satisfactory decision in the decision making situation. In the model formulation of the problem, first the goals with interval parameters are converted into the standard goals in conventional GP formulation by using interval arithmetic technique [17]. Then, EGP approach [18], which is the convex combination of both *minsum* [3] and *minmax* [19] modeling aspects of GP are addressed to construct the achievement function for minimizing the possible regret towards achieving the goal values within the specified intervals from the optimistic point of view of DM.

The proposed approach is illustrated through the case example of Bardhaman District of West Bengal (W.B.), India. A comparison of the model solution with other approaches [3, 19] is also made there to demonstrate the potential use of the approach.

Now, the general IVGP formulation of the problem is presented in section 2.

## 2 General IVGP Problem Formulation

The general structure of an IVGP problem can be presented as follows [14].

Find  $\mathbf{X}$  so as to:

$$\text{satisfy } Z_i : [a_i^L, a_i^U] \mathbf{X} + [\alpha_i^L, \alpha_i^U] = [t_i^L, t_i^U], \quad i \in I_1 \tag{1}$$

$$Z_i : \frac{[a_i^L, a_i^U] \mathbf{X} + [\alpha_i^L, \alpha_i^U]}{[b_i^L, b_i^U] \mathbf{X} + [\beta_i^L, \beta_i^U]} = [t_i^L, t_i^U], \quad i \in I_2 \tag{2}$$

$$\text{subject to } \mathbf{X} \in S = \{ \mathbf{X} : \mathbf{X} \in R^n \mid C\mathbf{X} \begin{pmatrix} \leq \\ \geq \end{pmatrix} h, \mathbf{X} \geq 0, h \in R^m \}, \tag{3}$$

where  $\mathbf{X}$  is the vector of decision variables and  $[a_i^L, a_i^U], i \in I_1 \cup I_2, [b_i^L, b_i^U], i \in I_2$  are vectors of coefficient intervals,  $[\alpha_i^L, \alpha_i^U], i \in I_1 \cup I_2, [\beta_i^L, \beta_i^U], i \in I_2$  are the vectors of constant intervals and  $[t_i^L, t_i^U], i \in I_1 \cup I_2,$  represent target intervals

associated with the  $i$ -th interval goal and where  $I_1 \cup I_2 = \{1, 2, \dots, I\}$  and the superscripts  $L$  and  $U$  stand for lower- and upper-ends, respectively, of the define intervals.

### 2.1 Construction of Planned Interval Goals

Following the rules of interval arithmetic operation, the planned interval goals associated with the expression in (1) can be presented as:

$$\left[ \sum_{j=1}^n a_{ij}^L x_j + \alpha_i^L, \sum_{j=1}^n a_{ij}^U x_j + \alpha_i^U \right] = [t_i^L, t_i^U], \quad i \in I_1 \tag{4}$$

Similarly, the fractional interval goal expression in (2) takes the form:

$$\frac{\left[ \sum_{j=1}^n a_{ij}^L x_j + \alpha_i^L, \sum_{j=1}^n a_{ij}^U x_j + \alpha_i^U \right]}{\left[ \sum_{j=1}^n b_{ij}^L x_j + \beta_i^L, \sum_{j=1}^n b_{ij}^U x_j + \beta_i^U \right]} = [t_i^L, t_i^U], \quad i \in I_2 \tag{5}$$

Then, the linear form of expression in (13) can be obtained by using direct multiplication technique and following the operation rules in interval arithmetic as:

$$\left[ \sum_{j=1}^n (a_{ij}^L - t_i^L b_{ij}^U) x_j + (\alpha_i^L - t_i^L \beta_i^U), \sum_{j=1}^n (a_{ij}^U - t_i^U b_{ij}^L) x_j + (\alpha_i^U - t_i^U \beta_i^L) \right] = [0, 0], \quad i \in I_2 \tag{6}$$

Now, in the process of formulating IVGP model, the defined goals in (4) and (6) are to be converted into the flexible goals by using the interval arithmetic technique and introducing under- and over-deviational variables to each of them.

### 2.2 Description of Flexible Goals

The linear interval goals in (3) can be presented in standard form of goals as:

$$\begin{aligned} \sum_{j=1}^n a_{ij}^U x_j + \alpha_i^U + d_i^{L-} - d_i^{L+} &= t_i^L \\ \sum_{j=1}^n a_{ij}^L x_j + \alpha_i^L + d_i^{U-} - d_i^{U+} &= t_i^U, \quad i \in I_1 \end{aligned} \tag{7}$$

Similarly, the linear interval goals in (6) can be presented as:

$$\begin{aligned} \sum_{j=1}^n (a_{ij}^U - t_i^U b_{ij}^L) x_j + (\alpha_i^U - t_i^U \beta_i^L) + d_i^{L-} - d_i^{L+} &= 0 \\ \sum_{j=1}^n (a_{ij}^L - t_i^L b_{ij}^U) x_j + (\alpha_i^L - t_i^L \beta_i^U) + d_i^{U-} - d_i^{U+} &= 0, \quad i \in I_2 \end{aligned} \tag{8}$$

where  $(d_i^{L+}, d_i^{U+}), (d_i^{L-}, d_i^{U-}) \geq 0$  with  $d_i^{L-} \cdot d_i^{L+} = 0$  and  $d_i^{U-} \cdot d_i^{U+} = 0$  represent the sets of under- and over-deviational variables, respectively, associated with the respective goals.

The GP model formulation of the problem is presented in section 3.

### 3 GP Model Formulation

In the present GP formulation both the aspects of GP, *minsum* GP [3] for minimizing the sum of the weighted unwanted variables as well as *minmax* GP [19] for minimizing the maximum of the deviations, are simultaneously taken into account as a convex combination of them to reach a satisfactory solution within the specified target intervals of the defined goals.

The GP model of the problem can be presented as:

$$\text{Minimize } Z = \lambda \sum_{i=1}^I w_i (d_i^{L-} + d_i^{U+}) + (1 - \lambda)V$$

and satisfy the goal expressions in (7) and (8)

subject to

$$d_i^{L-} + d_i^{U+} - V \leq 0, \quad i = 1, 2, \dots, I$$

and system constraints in (3),

$$\text{where } V = \max_{i=1}^I (d_i^{L-} + d_i^{U+}) \tag{9}$$

where  $w_i (> 0), i = 1, 2, \dots, I$  with  $\sum_{i=1}^I w_i = 1$ , denote the numerical weights of importance of achieving the goals within the respective target intervals, and  $0 < \lambda < 1$ .

Now, the description of variables, parameters associated with the farm planning problem is discussed in the following section.

### 4 Farm Planning Problem Formulation

The decision variables corresponding to various sources for cultivating crops together with the model parameters, different types of resource utilization and farm output levels are introduced first to formulate the model of the problem.

- Decision variables

- $x_{cs}$  = Allocation of land for cultivating the crop  $c$  during the season  $s$ ,  
 $c = 1, 2, \dots, C; s = 1, 2, \dots, S$ .
- $CW_s$  = Supply of canal-water during season  $s, s = 1, 2, \dots, S$ .
- $GW_s$  = Abstraction of groundwater during season  $s, s = 1, 2, \dots, S$ .

- Productive resource utilization parameters
- Interval resource parameters

$[MD^L, MD^U]$  = Estimated interval of total man-days (in days) required for all the seasons during the plan period.

$[P_c^L, P_c^U]$  = Interval for annual production level (in qtls) of crop  $c$ .

$[F_f^L, F_f^U]$  = Estimated interval for total amount of the fertilizer  $f$  ( $f = 1, 2, \dots, F$ ) (in quintals (qtls)) required during the plan period.

$[CW^L, CW^U]$  = Estimated interval of total amount of canal-water (in million cubic metres (MCM)) supplied during the plan period.

$[GW^L, GW^U]$  = Estimated interval associated with total amount of groundwater (in MCM) abstracted during plan period.

$[RS^L, RS^U]$  = Estimated interval of total amount of cash (in Rupees (Rs.)) required in the plan period for the purchase of productive resources.

$[MP^L, MP^U]$  = Estimated interval of total market value (in Rs.) of all the crops yields during the plan period.

$[R_{ij}^L, R_{ij}^U]$  = Target interval to maintain the ratio of annual production of  $i$ -th and  $j$ -th crop ( $i, j = 1, 2, \dots, C; i \neq j$ ).

- Crisp resource parameters

$LA_s$  = Total farm land (in hectares (ha)) available for cultivating the crops in season  $s$ .

$RW_s$  = Expected amount of rainwater (in MCM) precipitated during a cropping season  $s$ .

$W_{cs}$  = Water consumption (in cubic metre (CM) / ha) per ha of land for cultivating the crop  $c$  during the season  $s$ .

- Interval coefficients

$[MD_{cs}^L, MD_{cs}^U]$  = Interval of manpower (in days) required per ha of land for cultivating crop  $c$  during season  $s$ .

$[P_{cs}^L, P_{cs}^U]$  = Interval of estimated production of crop  $c$  per ha of land during season  $s$ .

$[F_{fcs}^L, F_{fcs}^U]$  = Interval for the amount of fertilizer  $f$  ( $f=1, 2, \dots, F$ ) utilized per ha of land for cultivating the crop  $c$  during the season  $s$ .

$[A_{cs}^L, A_{cs}^U]$  = Specified interval of the estimated cost for purchasing seeds and different farm assisting materials per ha of land for the crop  $c$  cultivated during the season  $s$ .

$[MP_{cs}^L, MP_{cs}^U]$  = Specified interval of market price (Rs/qlt) of the crop  $c$  during season  $s$  at the time of harvest.

The interval-valued model goals of the problem are described in section 4.1.



### 4.1 Description of Planned Interval Goals

(i) *Manpower requirement goal*

An estimated total manpower within a specified interval must be employed for smooth cultivation activities throughout the plan period.

The planned interval manpower goal takes the form

$$\sum_{s=1}^S \sum_{c=1}^C [MD_{cs}^L \cdot x_{cs}, MD_{cs}^U \cdot x_{cs}] = [MD^L, MD^U] \tag{10}$$

(ii) *Fertilizer requirement goal*

To achieve the optimal level of crop production different types of fertilizers need be used in different seasons in the plan period.

The planned interval goals for fertilizer requirement take the form

$$\sum_{s=1}^S \sum_{c=1}^C [F_{fcs}^L \cdot x_{cs}, F_{fcs}^U \cdot x_{cs}] = [F_f^L, F_f^U] \quad f = 1, 2, \dots, F \tag{11}$$

(iii) *Cash expenditure goal*

An estimated amount of money (in Rs.) is involved for the purpose of purchasing seeds, fertilizers and various productive resources.

The planned interval goals take the form

$$\sum_{s=1}^S \sum_{c=1}^C [A_{cs}^L \cdot x_{cs}, A_{cs}^U \cdot x_{cs}] = [RS^L, RS^U] \tag{12}$$

(iv) *Production achievement goals*

To meet the demand of agricultural crops in society, a minimum achievement level of production of each type of the crops is needed.

The goal expression can be stated as

$$\sum_{s=1}^S [P_{cs}^L \cdot x_{cs}, P_{cs}^U \cdot x_{cs}] = [P_c^L, P_c^U], \quad c = 1, 2, \dots, C, \tag{13}$$

(v) *Production-ratio goals*

To meet the demand of the main food products in society, certain ratios of total production of major crops should be maintained. Therefore, without loss of generality, certain ratios of total productions of major crops are taken into account with target intervals.

The production-ratio goals in interval form appear as

$$\left[ \sum_{s=1}^S (P_{is}^L x_{is} - R_{ij}^U P_{js}^U x_{js}), \sum_{s=1}^S (P_{is}^U x_{is} - R_{ij}^L P_{js}^L x_{js}) \right] = [0, 0], \quad i, j = 1, 2, \dots, C; i \neq j. \tag{14}$$

(vi) *Annual Profit achievement goal*

The level of earning a farm depends on the market price of individual production of crops, which in turn differs from period to period, and also both the quantity and the quality of the crops. Hence, profit achievement levels of crops are to be considered with target intervals.

The associated goal appears as:

$$\sum_{s=1}^S \sum_{c=1}^C [(P_{cs}^L MP_{cs}^L - A_{cs}^U) \cdot x_{cs}, (P_{cs}^U MP_{cs}^U - A_{cs}^L) \cdot x_{cs}] = [MP^L, MP^U] \tag{15}$$

(vii) *Canal-water goal*

Canal-water supplied from River-barrage is imprecise in nature owing to capacity limitation of barrage, and after a certain level release of water is not permissible for preserving the eco-system of earth.

The interval goal appears as

$$\sum_{s=1}^S CW_s = [CW^L, CW^U] \tag{16}$$

(viii) *Groundwater goal*

Ground water is very scarce and after a certain limit, it cannot be abstracted to prevent from harmful mineral contamination.

The interval goal appears as

$$\sum_{s=1}^S GW_s = [GW^L, GW^U] \tag{17}$$

**4.2 Description of Constraints**

(i) *Land utilization constraints*

Utilization of cultivable land generally differs from season to season due to duration of yielding the crops as well as soil conditions.

The land utilization constraints appear as:

$$\sum_{c=1}^C x_{cs} \leq LA_s \quad s = 1, 2, \dots, S \tag{18}$$

(ii) *Water supply affinity constraints*

Since, both the canal-water and groundwater as reserved water are very limited, utilization of individual total capacity of each of them during major cropping seasons are considered with great care and constrained to some extent to other seasons depending on local climatic conditions.

The affinity constraints associated with reserved water supply can be presented as

$$CW_{s_r} \leq p_r \sum_{s=1}^S CW_s, \quad GW_{s_r} \leq q_r \sum_{s=1}^S GW_s, \quad s = 1, 2, \dots, S; \quad r \in \{1, 2, \dots, S\} \tag{19}$$

where  $p_r, q_r$  designate the certain percentage of supply of total canal-water and groundwater, respectively, during  $r$ -th season.

(iii) *Total water supply constraints*

Since, all the three water sources are scarce; irrigation water demands in all the seasons are always constrained in the planning horizon.

The water supply constraints appear as

$$\sum_{c=1}^C W_{cs} \cdot x_{cs} - (CW_s + GW_s) \leq RW_s, \quad s = 1, 2, \dots, S \tag{20}$$

The modeling aspect of the proposed problem is demonstrated via the case example in section 5.

## 5 An Illustrative Case Example

The land-use planning problem for production of principal crops of the Bardhaman district of W.B., India is considered to demonstrate the application potential of the proposed approach. The district Bardhaman is predominantly an agricultural district, which is properly known as the granary of W.B. The principal commodity for trade in the district is Rice. Main rivers of the district are Ajay, Bhagirathi (or Hooghly) and Damodar.

The data associated with the model of the problem were collected from different agricultural planning units: District Statistical Hand Book [20], Economic Review [21], Basak [22] and Department of Agri-Irrigation [23] of Bardhaman District.

The decision variables and various types of data involved with the problem are summarized in the Tables 1–3.

**Table 1.** Summary of decision variables and crisp resource coefficients

Season (s)	Pre-Kharif (1)				Kharif (2)		Rabi (3)		
Crop (c)	Jute (1)	Aus (2)	Aman (3)	Boro (4)	Wheat (5)	Mustard (6)	Potato (7)	Pulses (8)	
Variable ( $x_{cs}$ )	$x_{11}$	$x_{21}$	$x_{32}$	$x_{43}$	$x_{53}$	$x_{63}$	$x_{73}$	$x_{83}$	
$W_{cs}$ (in CM)	520	864	1270	1787	382	255	457	286	
Rainwater (in MCM)	975.77		5072.52			540.86			
Total Arable land (in '000 ha)	458.20		458.20			458.20			

**Table 2.** Data description of target interval goals

Goal	Target Interval
1. Manpower (in '000 man-days):	[39379, 43848]
2. Fertilizer requirement (in metric ton):	
(a) Nitrogen	[66.20, 85.00]
(b) Phosphorus	[54.30, 65.80]
(c) Potassium	[38.00, 45.00]
3. Budget allocation (in Rs Lac)	[128800, 154550]
4. Production (in '000 metric ton):	
(a) Jute	[28.30, 50.80]
(b) Rice	[1468.80, 1965]
(c) Wheat	[7.14, 12.80]
(d) Mustard	[30.20, 42.45]
(e) Potato	[922.30, 1335.80]
(f) Rabi pulse	[2.20, 4.50]
5. Profit (in Rs. Lac)	[159850.50, 177615.80]
6. Production ratio goals (Rice : Potato)	[1.5, 2.3]
7. Water supply (in MCM )	
(a) Canal water	[1645.50, 1840.00]
(b) Groundwater	[1440.45, 1620.10]

\* Rs = Rupees in Indian currency.

**Table 3.** Data description of interval resource goals

Crop	MD <sub>cs</sub> <sup>L</sup> , MD <sub>cs</sub> <sup>U</sup>	[F <sub>csf</sub> <sup>L</sup> , F <sub>csf</sub> <sup>U</sup> ]			[P <sub>cs</sub> <sup>L</sup> , P <sub>cs</sub> <sup>U</sup> ]	[A <sub>cs</sub> <sup>L</sup> , A <sub>cs</sub> <sup>U</sup> ]	[MP <sub>cs</sub> <sup>L</sup> , MP <sub>cs</sub> <sup>U</sup> ]
		N	P	K			
Jute	[86, 94]	[45, 56]	[19, 22]	[18, 20]	[2731, 2755]	[18655, 18845]	[870, 1050]
Aus Paddy	[57, 64]	[58, 63]	[26, 33]	[22, 28]	[4324, 4401]	[21050, 21350]	[925, 1150]
Aman Paddy	[56, 63]	[75, 85]	[30, 34]	[30, 34]	[4323, 4365]	[22250, 22750]	[720, 800]
Boro Paddy	[55, 65]	[145, 155]	[78, 82]	[78, 82]	[5115, 5148]	[36430, 36970]	[950, 1105]
Wheat	[36, 43]	[105, 115]	[52, 58]	[47, 53]	[2380, 2432]	[24965, 25035]	[1350, 1475]
Mustard	[27, 34]	[75, 85]	[35, 45]	[25, 35]	[1025, 1078]	[18520, 18786]	[2965, 3065]
Potato	[66, 74]	[260, 290]	[165, 190]	[165, 190]	[27490, 27537]	[59245, 61390]	[500, 600]
Rabi pulse	[26, 34]	[45, 55]	[22, 28]	[12, 18]	[813, 835]	[14170, 14595]	[2920, 3160]

\*\* N = Nitrogen, P = Phosphorus, K = Potassium.

Now, following the expressions in (16) - (19) and using the data tables the model goals can be constructed in the following section.

### 5.1 Construction of Model Goals

#### (i) Manpower requirement goal

$$94x_{11} + 64x_{21} + 63x_{32} + 65x_{43} + 43x_{53} + 34x_{63} + 74x_{73} + 34x_{83} + d_1^{L-} - d_1^{L+} = 39379000,$$

$$86x_{11} + 57x_{21} + 56x_{32} + 55x_{43} + 36x_{53} + 27x_{63} + 66x_{73} + 26x_{83} + d_1^{U-} - d_1^{U+} = 43848000.$$

(21)

#### (ii) Fertilizer requirement goal

$$56x_{11} + 63x_{21} + 85x_{32} + 155x_{43} + 115x_{53} + 85x_{63} + 290x_{73} + 55x_{83} + d_2^{L-} - d_2^{L+} = 66200000,$$

$$45x_{11} + 58x_{21} + 75x_{32} + 145x_{43} + 105x_{53} + 75x_{63} + 260x_{73} + 45x_{83} + d_2^{U-} - d_2^{U+} = 85000000.$$

(Nitrogen)

$$22x_{11} + 33x_{21} + 34x_{32} + 82x_{43} + 58x_{53} + 45x_{63} + 190x_{73} + 28x_{83} + d_3^{L-} - d_3^{L+} = 54300000,$$

$$19x_{11} + 26x_{21} + 30x_{32} + 78x_{43} + 52x_{53} + 35x_{63} + 165x_{73} + 22x_{83} + d_3^{U-} - d_3^{U+} = 65800000.$$

(Phosphorus)

$$20x_{11} + 28x_{21} + 34x_{32} + 82x_{43} + 53x_{53} + 35x_{63} + 190x_{73} + 18x_{83} + d_4^{L-} - d_4^{L+} = 38000000,$$

$$18x_{11} + 22x_{21} + 30x_{32} + 78x_{43} + 47x_{53} + 25x_{63} + 165x_{73} + 12x_{83} + d_4^{U-} - d_4^{U+} = 45000000.$$

(Potassium)

(22)

#### (iii) Cash expenditure goal

$$18.85x_{11} + 21.35x_{21} + 22.75x_{32} + 36.97x_{43} + 25.04x_{53} + 18.77x_{63} + 61.39x_{73} + 14.60x_{83}$$

$$+ d_5^{L-} - d_5^{L+} = 12880000,$$

$$18.65x_{11} + 21.05x_{21} + 22.25x_{32} + 36.43x_{43} + 24.97x_{53} + 18.52x_{63} + 59.25x_{73} + 14.17x_{83}$$

$$+ d_5^{U-} - d_5^{U+} = 15455000.$$

(23)

(iv) *Production achievement goals*

$$\begin{aligned}
 27.55x_{11} + d_6^{L-} - d_6^{L+} &= 28300, & 27.31x_{11} + d_6^{U-} - d_6^{U+} &= 50800, \\
 44.01x_{21} + 43.65x_{32} + 51.48x_{43} + d_7^{L-} - d_7^{L+} &= 1468800, \\
 43.24x_{21} + 43.23x_{32} + 51.15x_{43} + d_7^{U-} - d_7^{U+} &= 1965000, \\
 24.32x_{53} + d_8^{L-} - d_8^{L+} &= 7140, & 23.80x_{53} + d_8^{U-} - d_8^{U+} &= 12800, \\
 10.78x_{63} + d_9^{L-} - d_9^{L+} &= 30200, & 10.25x_{63} + d_9^{U-} - d_9^{U+} &= 42450, \\
 275.37x_{73} + d_{10}^{L-} - d_{10}^{L+} &= 922300, & 274.90x_{73} + d_{10}^{U-} - d_{10}^{U+} &= 1335800, \\
 8.35x_{83} + d_{11}^{L-} - d_{11}^{L+} &= 2200, & 8.13x_{83} + d_{11}^{U-} - d_{11}^{U+} &= 4500.
 \end{aligned}
 \tag{24}$$

(v) *Production-ratio goals*

$$\begin{aligned}
 72.67x_{21} + 134.56x_{32} + 256.34x_{43} - 307x_{73} + d_{12}^{L-} - d_{12}^{L+} &= 0, \\
 69.46x_{21} + 117.89x_{32} + 248.70x_{43} - 413.85x_{73} + d_{12}^{U-} - d_{12}^{U+} &= 0.
 \end{aligned}
 \tag{25}$$

(vi) *Annual Profit achievement goal*

$$\begin{aligned}
 10272.50x_{11} + 29651.50x_{21} + 21400x_{32} + 20455.40x_{43} + 10970x_{53} + 14520.70x_{63} + 73832x_{73} + \\
 12216x_{83} + d_{13}^{L-} - d_{13}^{L+} &= 1598505000, \\
 4914.7x_{11} + 18647x_{21} + 17021.60x_{32} + 11525.50x_{43} + 7095x_{53} + 11513x_{63} + 48205x_{73} + \\
 9569x_{83} + d_{13}^{U-} - d_{13}^{U+} &= 1776158000.
 \end{aligned}
 \tag{26}$$

(vii) *Canal -water goal*

$$CW_1 + CW_2 + CW_3 + d_{14}^{L-} - d_{14}^{L+} = 1645.50, \quad CW_1 + CW_2 + CW_3 + d_{14}^{U-} - d_{14}^{U+} = 1840. \tag{27}$$

(viii) *Groundwater goal*

$$GW_1 + GW_2 + GW_3 + d_{15}^{L-} - d_{15}^{L+} = 1440.45, \quad GW_1 + GW_2 + GW_3 + d_{15}^{U-} - d_{15}^{U+} = 1620.10. \tag{28}$$

## 5.2 Construction of System Constraints

(i) *Land utilization constraints*

$$x_{11} + x_{21} \leq 458200, \quad x_{32} \leq 458200, \quad x_{43} + x_{53} + x_{63} + x_{73} + x_{83} \leq 458200. \tag{29}$$

(ii) *Water supply affinity constraints*

In W.B., the major crops are cultivated during rainy (Kharif) and winter (Rabi) seasons. Consequently, reserved water supply constraints appear as:

$$CW_1 \leq 0.05 \sum_{s=1}^3 CW_s, \quad GW_1 \leq 0.06 \sum_{s=1}^3 GW_s \tag{30}$$

(iii) *Total water supply constraints*

Total water supply constraints in different seasons are presented as:

$$\begin{aligned}
 5.08x_{11} + 8.636x_{21} - (CW_1 + GW_1) &\leq 975.767 && \text{(Pre-Kharif)} \\
 12.7x_{32} - (CW_2 + GW_2) &\leq 5072.424 && \text{(Kharif)} \\
 17.78x_{43} + 3.81x_{53} + 2.54x_{63} + 4.572x_{73} + 2.54x_{83} - (CW_3 + GW_3) &\leq 540.857 && \text{(Rabi)}
 \end{aligned}
 \tag{31}$$

Now, the executable GP model of the problem can be presented as

$$\text{Minimize } Z = \lambda \sum_{i=1}^{15} w_i (d_i^{L-} + d_i^{U+}) + (1 - \lambda)V$$

and satisfy the goal expressions in (21) - (28),  
 subject to  $d_i^{L-} + d_i^{U+} - V \leq 0, \quad i = 1, 2, \dots, 15$   
 and system constraints in (29) - (31),  
 where  $V = \max_{i=1}^{15} (d_i^{L-} + d_i^{U+})$  (32)

Following the procedure and giving equal weights to all the goals and considering  $\lambda = 0.5$ , the problem is solved by using the *Software Lingo* (Ver. 12.0).

The resultant land allocation (in '000 ha) decision is obtained as:

$$(Jute, Rice, Wheat, Mustard, Potato, Pulses) = (29.04, 632.30, 29.36, 41.36, 110.71, 5.227).$$

The crop production (in '000 metric ton) achievement is as follows:

$$(Jute, Rice, Wheat, Mustard, Potato, Pulses) = (65.59, 1979.40, 71.40, 44.59, 2048.76, 43.65).$$

The achieved profit = Rs 177452.38 Lac.

**Note 1:** If the problem is solved by using *minsum* GP [3] approach for consideration of  $\lambda = 1$  in (32) solution of the problem is found as:

$$(Jute, Rice, Wheat, Mustard, Potato, Pulses) = (19.80, 596.67, 24.66, 38.32, 106.37, 6.98).$$

The crop production (in '000 metric ton) achievement is as follows:

$$(Jute, Rice, Wheat, Mustard, Potato, Pulses) = (50.80, 1807.15, 62.68, 40.98, 1886.98, 48.02).$$

The achieved profit = Rs 162545.44 Lac.

**Note 2:** If the problem is solved by using *minmax* GP [19] approach for consideration of  $\lambda = 0$  in (32) solution of the problem is found as:

$$(Jute, Rice, Wheat, Mustard, Potato, Pulses) = (19.80, 593.37, 24.66, 41.68, 96.08, 6.98).$$

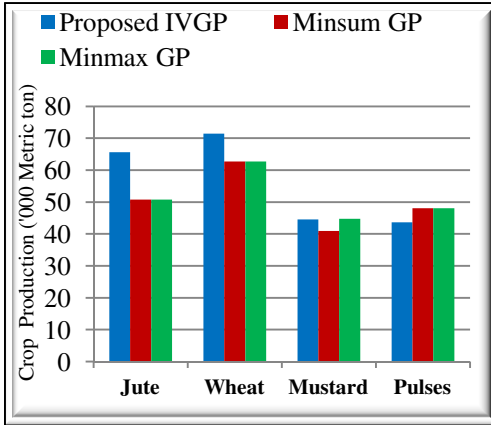
The crop production (in '000 metric ton) achievement is as follows:

$$(Jute, Rice, Wheat, Mustard, Potato, Pulses) = (50.80, 1796.58, 62.68, 44.79, 1754.55, 48.02).$$

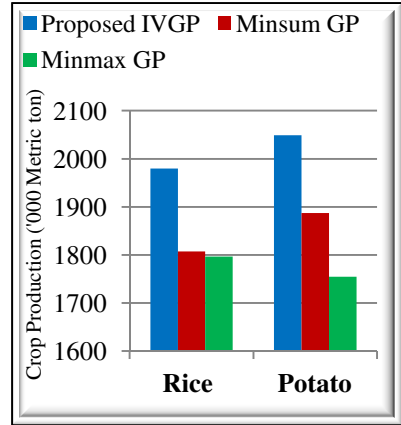
The achieved profit = Rs 160399.50 Lac.

The result reflects that the proposed farm planning model is better than individual *minsum* GP and *minmax* GP approach from the view point of achieving a more satisfactory decision in the context of crop production.

The graphical representation of the model solution and the solutions obtained by using other approaches regarding crop production decision are displayed in Figure 1 and Figure 2.



**Fig. 1.** Production achievement of four crops under different approaches



**Fig. 2.** Production achievement of two major crops under different approaches

The graphical comparisons reflect that the proposed approach is superior over the other ones in the same decision environment with a view to achieve a more acceptable crop production decision in the farm planning context.

## 6 Conclusions

The approach presented in this paper for production of different crops by utilizing various resources properly provides a new look into the way of analyzing the various farm management activities in an uncertain decision making environment. Under the flexible nature of the proposed approach, other different parameters (crisp or inexact) as well as other different regional based environmental constraints can be incorporated within the framework of the model, which may be problem in future study to improve crop production system as well to make a hunger free society in the current uncertain decision environment.

**Acknowledgement.** The authors would like to thank Prof. J. K. Mandal, the Special Session Chair of CSI-2013, and the anonymous reviewers for their useful comments and suggestions that has helped to improve the quality of presentation of the paper.

## References

1. Heady, E.O.: Simplified Presentation and Logical Aspects of Linear Programming Technique. *J. Farm Econ.* 36, 1035–1048 (1954)
2. Nix, J.: Farm Management - the State of the Art (or Science). *J. Agri. Eco.* 30, 277–292 (1979)
3. Ignizio, J.P.: *Goal Programming and Extensions*, Lexington, D.C. Health, Massachusetts (1976)

4. Simon, H.A.: *Administrative Behavior*. Free Press, New York (1957)
5. Wheeler, B.M., Russell, J.R.M.: *Goal Programming and Agricultural Planning*. *Oper. Res. Quar.* 28, 21–32 (1977)
6. Pal, B.B., Basu, I.: Selection of appropriate priority structure for optimal land allocation in agricultural planning through goal programming. *Indian Journal of Agricultural Economics* 51, 342–354 (1996)
7. Zimmermann, H.-J.: *Fuzzy Sets, Decision Making and Expert Systems*. Kluwer-Nijhoff Publishing, Dordrecht (1987)
8. Pal, B.B., Moitra, B.N., Maulik, U.: A Goal Programming Procedure for Fuzzy Multiobjective Linear Fractional Programming Problem. *Fuzzy Sets and Sys.* 139, 395–405 (2003)
9. Biswas, A., Pal, B.B.: Application of Fuzzy Goal Programming Technique to Land use Planning in Agricultural System. *Omega* 33, 391–398 (2005)
10. Pal, B.B., Kumar, M., Sen, S.: A Linear Fuzzy Goal Programming Approach for Solving Patrol Manpower Deployment Planning Problems -A Case Study, pp. 244–249. *IEEE Xplore Digital Library* (2009)
11. Pal, B.B., Kumar, M.: A Linear Fuzzy Goal Programming Method for Solving Optimal Power Generation and Dispatch Problem. *Int. J. Adv. Comp. Res.* 3, 56–64 (2013)
12. Bitran, G.R.: Linear Multiobjective Problems with Interval Coefficients. *Manag. Sci.* 26, 694–706 (1980)
13. Jiang, C., Han, X., Liu, G.R., Liu, G.P.: A Nonlinear Interval Number Programming Method for Uncertain Optimization Problems. *Euro. J. Oper. Res.* 188, 1–13 (2008)
14. Inuiguchi, M., Kume, Y.: Goal Programming Problems with Interval Coefficients and Target Intervals. *Euro. J. Oper. Res.* 52, 345–361 (1991)
15. Oliviera, C., Antunes, C.H.: Multiple Objective Linear Programming Models with Interval Coefficients - An Illustrated Overview. *Euro. J. Oper. Res.* 118, 1434–1463 (2007)
16. Pal, B.B., Kumar, M., Sen, S.: A Priority - Based Goal Programming Method for Solving Academic Personnel Planning Problems with Interval - Valued Resource Goals in University Management System. *Int. J. Appl. Manag. Sci.* 4, 284–312 (2012)
17. Moore, R.E.: *Interval Analysis*. Prentice-Hall, New Jersey (1966)
18. Romero, C.: A General Structure of Achievement Function for a Goal Programming Model. *Euro. J. Oper. Res.* 153, 675–686 (2004)
19. Romero, C.: *Handbook of Critical Issues in Goal Programming*. Pergamon Press, Oxford (1991)
20. *District Statistical Hand Book, Burdwan*. Department of Bureau of Applied Economics and Statistics. Govt. of West Bengal, India (2011)
21. *Economic Review*. Department of Bureau of Applied Economics and Statistics. Govt. of West Bengal, India
22. Basak, R.K.: *Soil testing and fertilizer recommendation*. Kalyani Publishers, New Delhi (2000)
23. <http://bardhaman.nic.in/agri/agriculture.htm>



# Design and Performance Analysis of Distributed Implementation of Apriori Algorithm in Grid Environment

Priyanka Arora<sup>1</sup> and Sarbjeet Singh<sup>2</sup>

<sup>1</sup> Department of Information Technology,  
Guru Nanak Dev Engineering College, Ludhiana, Punjab, India  
arorapriyanka29@gmail.com

<sup>2</sup> Computer Science and Engineering,  
UIET, Panjab University, Chandigarh, India  
sarbjeet@pu.ac.in

**Abstract.** This paper presents the design and performance analysis of distributed implementation of Apriori algorithm in grid environment. Apriori algorithm is very important algorithm in data mining discipline that enables organizations to mine large amount of historical data that they gather over period of time and discover hidden patterns in that data. Data mining techniques enable organizations to analyze market trends and user behavior. If the data set to be mined is very large then varying the basic algorithm for execution in a distributed environment makes sense because distributed technologies generally offer performance benefits. Grids have gained wide popularity in executing a task in distributed fashion and offer performance benefits. So in this paper we have made an attempt to implement distributed version of basic Apriori algorithm in a grid environment. The Grid environment has been constructed using Globus® Toolkit. Experimental results show that distributed version offers performance benefits over basic version of Apriori algorithm and hence is a good implementation choice if the data to be mined is really large and distributed.

**Keywords:** Data Mining, Grid Environment, Apriori algorithm, Globus® Toolkit.

## 1 Introduction

A distributed system can be defined as a collection of independent computers that appear to its user as a single system [1]. Distributed approach solves a problem by dividing the problem into smaller parts and then executing the independent parts on different nodes of a distributed system. The distributed systems can be classified as Distributed Computing Systems, Distributed Information Systems and Distributed Pervasive Systems [1]. The examples of distributed computing systems are cluster computing systems, grid computing systems and the examples of distributed

information systems are transaction processing systems [1]. Distributed pervasive systems are categorized by being small, mobile and generally wireless. All distributed systems and computing technologies, in one or other way, work towards the solution of a large problem by dividing the large problem into smaller parts.

Grids are generally heterogeneous networks. Grid computing environment consists of combination of heterogeneous resources from multiple administrative domains which collectively work to achieve a common goal. Grid nodes may have different hardware, software and network connections [2-6]. Grid uses middleware to divide and distribute independent parts of a program among available nodes. It involves computation in a distributed fashion and may involve other computing systems also. Grids provide performance benefits to compute and storage intensive jobs compared to their execution on single system. Mining large amount of complex data is storage and compute intensive job. So using grid computing environment for the efficient execution of this job makes sense, and must be tried, as storage and compute requirements can be easily met by grid computing environment. Mining of association rules between sets of items in large databases deals with identifying related data items in that set. It is a basic requirement in many applications and is described in [7].

In this paper, we have made an attempt to implement Apriori data mining algorithm in a distributed fashion in a grid environment. Apriori is a classic algorithm for association rule mining. Association rule mining involves finding interesting relationships among items in large data set. In the proposed approach, the data has been horizontally divided among grid nodes. A test-bed has been created for the evaluation and performance analysis. Experimental results show that executing distributed version of Apriori algorithm offers significant performance benefits over non-distributed approach.

The rest of the paper is organized as follows: Section 2 presents the related work in the area of association rule mining in distributed environments. Section 3 presents the proposed distributed version of Apriori algorithm. Section 4 describes implementation details along with experimental results and Section 5 concludes the paper with future scope.

## **2 Related Work**

The problem of mining association rules was proposed in [7]. The mining of data using association rules has its application in several different fields. Mining frequent item-sets has been demonstrated to be beneficial in several fields e.g. [8-9]. As the size of the database grows, the mining of frequent item-sets from such a large data becomes a challenging task. Various algorithms such as EM, ECLAT, Apriori, FP-Growth etc. exist for mining association rules [10]. Apriori is the highly cited one and used algorithm for mining association rules. It is widely used algorithm because of its features of simplicity and loosely coupled nature. We feel it is better for mining on large databases in distributed environment and therefore has been chosen for mining association rules in grid environment. The Apriori discovers candidate item-sets used in iterations, compute the support, and prune the candidate item-sets to mine the frequent item-sets that are above minimum support threshold.

[11] describes a parallel and distributed version of the Apriori algorithm using Globus® middleware by making use of Message Passing Interface extended with Grid Services (MPICHG2). The case study described in [11] presents the design and implementation of local and global mining of frequent item-sets. The parallel mining algorithm along with grid technology reduces the computation time and increases the speed of the application [11]. In [12], the R package *arules* is presented which provides basic framework for creating and manipulating input data sets and analyzing the resulting item-sets and rules. The package includes interfaces to two fast mining algorithms which can be used to mine frequent item-sets and association rules [12].

Grid implementation of Apriori algorithm based on virtual organizations is presented in [13]. In this, the approach used for distributed association mining is based on OGSA. The design of the Apriori Grid Service in the open grid service architecture is also presented. [14] proposes an architecture for data mining grid named DMGA and discusses the implementation of its architecture, named *wekaG*. The architecture presented is based on the main phases of data mining process. It is based on generic data grid and specific data mining grid services. The tool described provides the functionality of Weka (a well known data mining tool) in a grid environment [14]. [15] discusses the development of parallel and distributed Apriori algorithm in grid environment. Apriori algorithm along with FP-growth (frequent pattern growth) is implemented on grid network. Grid nodes find the local support count and prune all infrequent item-sets.

In the proposed distributed version of Apriori algorithm, a transactional matrix has been used which represents the entire database. It is prepared by scanning the database once. This avoids the costly step of scanning the database repeatedly by every node. In the later steps, until the frequent item-sets are generated, only this transactional matrix needs to be scanned. Next Section describes the details of the distributed Apriori algorithm.

### **3 Distributed Approach to Apriori Algorithm**

As described in [10], Apriori algorithm finds all sets of items (item-sets) that have support no less than a minimum support. The support for an item-set is generally the ratio of the number of transactions that contain the item-set to the total number of transactions [10]. Item-sets satisfying the minimum support constraint are called the frequent item-sets. The algorithm takes multiple passes over the data. During first pass, support of individual items is counted. During this pass, frequent items are determined. In each subsequent passes, a seed set of item-sets found to be frequent in the previous pass is used for generating new frequent item-sets, which are called candidate item-sets [10]. Their actual support is counted during the pass over the data. At the end, those satisfying minimum support constraints are collected, i.e., frequent item-sets are determined, and these become the seed for the next pass. The process is repeated till no new frequent item-sets are found [10].

The proposed work implements the task of mining association rules in a distributed fashion. Following is the pseudo-code for the proposed distributed approach of Apriori Algorithm:

1. *Generate valid user credentials. (These are required in a grid environment for making secure communications)*
2. *Request the server to start generating association rules in a distributed fashion.*
3. *Prepare the transactional matrix. (Transactional matrix is a representation of the entire database and is prepared by scanning the database once)*
4. *Distribute the transactional matrix's data equally among available grid nodes using ReliableFileTransfer service. The data is partitioned horizontally.*
5. *On each grid node, do the following:*
  - a. *Calculate the local candidates*
  - b. *Generate frequent local item-sets by pruning the non-frequent item-sets*
  - c. *Mine frequent local association rules*
  - d. *Send mined rules back to server*
6. *Receive the data from all the grid nodes and generate the global association rules*
7. *Send the mined global association rules to the client*

The implementation of the above steps in grid environment is shown graphically in Fig. 1 below:

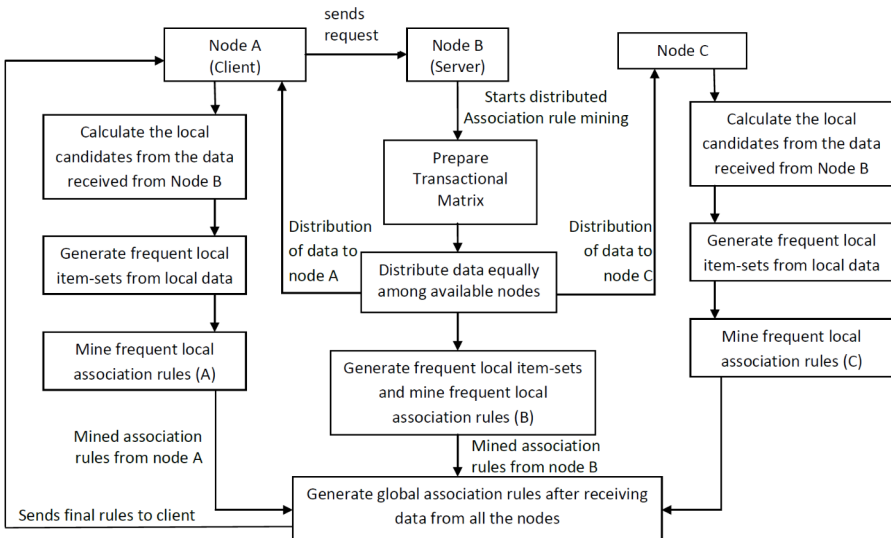


Fig. 1. Working of proposed Distributed Apriori Algorithm

The various steps involved in the working of proposed distributed approach are described below:

- i. Requesting the server:** The client requests the server to generate valid user credentials.
- ii. Generation of credentials:** The server accepts the request (if valid) and generates the credentials for the client.
- iii. Request for mining:** After receiving valid credentials, the client requests the Distributed Apriori Web Service, deployed on the server, for mining association rules.
- iv. Processing of mining request:** The server accepts the request and Distributed Apriori web service is invoked. The service scans the entire database and prepares a transactional matrix representing the dataset.
- v. Distribution of data:** The data in this matrix is distributed equally among all the available grid nodes for mining association rules using ReliableFileTransfer Service of Globus® Toolkit [16].
- vi. Mining on local data at each node:** Each of the grid node mine its local candidates and generates local frequent item-sets and then the local association rules.
- vii. Sending the generated local rules back to the server:** Each of the grid nodes sends the generated local association rules to the server.
- viii. Generation of global rules:** After receiving all the association rules from all the grid nodes, the server generates the global association rules.
- ix. Replies to the client with rules:** After the generation of global rules, these are sent to the client.

The approach first requires the generation of valid user credentials for mining association rules. The credentials are generated by the server which acts as Certificate Authority and can also act as grid node for mining rules. A distributed Apriori web service has been created and is deployed on the server as well as every grid node that participates in the job execution. As soon as the Server accepts the request, distributed Apriori web service gets invoked. The service starts scanning the database and prepares a transactional matrix. The data in this matrix is then distributed equally among all the available grid nodes in the environment using the ReliableFileTransfer Service of Globus® Toolkit [16]. Then the mining of local data on each node takes place and local frequent item-sets and association rules are generated. After processing, each grid node sends back its local association rules to the server. After receiving data from all the nodes, the server prepares the global transaction rules which are then sent to the client from which the request originated.

## 4 Implementation Details and Experimental Results

The implementation work has been divided into four phases which are:

- a) Setting up of the Grid Environment
- b) Setting up of Certificate Authority & Security Credentials

- c) Implementation of support services in GT4 for Apriori Algorithm
- d) Implementing the distributed Apriori Algorithm

The Grid environment was setup by installing Globus® Toolkit 4.0 [16] on Fedora Core 4 machines. Security Credentials were setup using the Grid Security Infrastructure (GSI). Globus supports many inbuilt services such as Reliable File Transfer, gridFTP, postgresQL service etc. The proposed implementation makes use of these services. Apriori Service is built and deployed in the container representing the distributed version of the Apriori Algorithm. This service is invoked whenever a client queries to mine association rules.

The distributed Apriori service has been implemented as a web service in Globus® container. The dataset used in this work is the ‘World Census Data’ taken from UCI Machine Repository [17]. The dataset contains 48998 instances having 13 columns such as: age, workclass, education, edu\_num, martial\_status, occupation, relationship, race, sex, gain, loss, country. The missing values were filled and then the dataset was converted into the postgresQL database. The distributed service is written using ECLIPSE IDE. The service interface is defined using WSDL and a DistributedApriori.wsdl file is created. The service is written and a DistributedApriori.java file for the same is created. The deployment parameters are defined using the WSDD language and JNDI deployment files named deploy-server.wsdd and deploy-jndi-config.xml. After the creation of all the files, a GAR file is generated. A GAR file is a collection of all the files created into a single file. This is done using the ANT tool and the ‘globus-build-service.sh’ script using the build.xml file.

```

----- All association rules -----
rule 0: workclass(State gov) ==> education(Bachelors) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 1: education(Bachelors) ==> workclass(Self-emp-not-inc) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 2: workclass(Private) ==> education(HS-grad) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 3: education(11th) ==> workclass(Private) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 4: workclass(Private) ==> edu_num(13) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 5: edu_num(14) ==> workclass(Private) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 6: workclass(Private) ==> edu_num(5) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 7: edu_num(9) ==> workclass(Self-emp-not-inc) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 8: martial_status(Never-married) ==> workclass(Private) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 9: workclass(Private) ==> martial_status(Married-civ-spouse) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 10: martial_status(Married-civ-spouse) ==> workclass(Private) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 11: workclass(State-gov) ==> martial_status(Married-civ-spouse) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 12: workclass(Private) ==> occupation(Adm-clerical) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 13: occupation(Sales) ==> workclass(Private) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 14: workclass(Private) ==> occupation(Craft-repair) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 15: occupation(Transport-moving) ==> workclass(Private) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 16: workclass(Self-emp-not-inc) ==> relationship(Own-child) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 17: relationship(Unmarried) ==> workclass(Private) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 18: workclass(Private) ==> relationship(Husband) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 19: relationship(Unmarried) ==> workclass(Self-emp-not-inc) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 20: edu_num(16) ==> education(Doctorate) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 21: education(HS-grad) ==> edu_num(9) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 22: edu_num(5) ==> education(9th) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 23: education(11th) ==> edu_num(7) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 24: martial_status(Divorced) ==> education(HS-grad) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 25: education(HS-grad) ==> martial_status(Separated) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 26: martial_status(Never-married) ==> education(HS-grad) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 27: education(Bachelors) ==> martial_status(Married-civ-spouse) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 28: occupation(Machine-op-inspct) ==> education(HS-grad) support : 0.3333333333333333 (1/3) confidence : 1.0
rule 29: education(Masters) ==> occupation(Evac-manual) support : 0.3333333333333333 (1/3) confidence : 1.0

```

Fig. 2. A snapshot of mined association rules

After following all the steps discussed in the previous paragraph, the evaluation has been done using a test census data taken from UCI Machine Learning Repository [17] for different sizes of the database. The rules returned by the server for various databases are as follows:

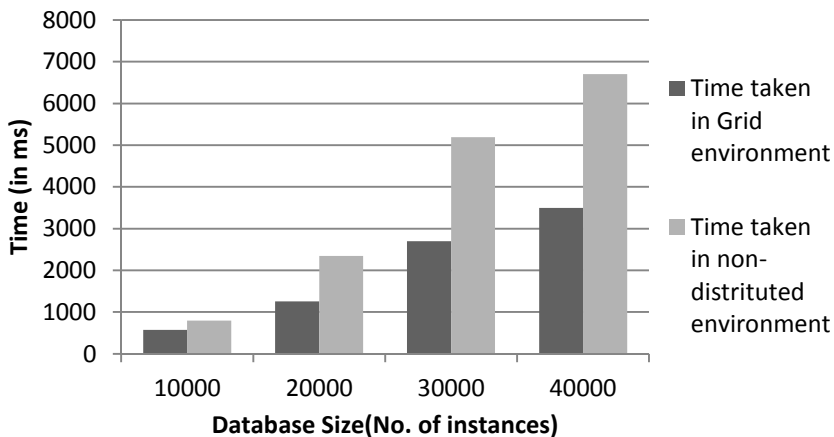
For database containing 10000 instances, 2543 rules were generated.

For database containing 20000 instances, 6971 rules were generated.

For database containing 30000 instances, 18175 rules were generated.

For database containing 40000 instances, 45755 rules were generated.

A snapshot of mined association rules is given in Fig. 2. The time taken to mine association rules is as follows:



**Fig. 3.** Time taken to mine association rules by distributed Apriori algorithm in Grid environment and non-distributed environment

This time includes the server time (i.e. the time spent by the server to execute mining algorithm), service connection time (i.e. the time spent by the client in obtaining service reference), client time (i.e. the time spent by the client in obtaining results) and RFT time (i.e. the time spent in transferring the data among grid nodes).

From the experimental results it is clear that the system takes linear increase in time with the increase in database size. This performance is much better than the performance of the algorithm on a single system as the execution of the same job took approximately double time on the single system.

## 5 Conclusion and Future Scope

The work presents the design and implementation of a distributed version of popular Apriori algorithm on grid environment. The results have been taken on a virtual test-bed and show a linear increase in time with the increase in database size. Though the distributed approach give performance benefits over single system approach, still, some issues that require further investigation are pending. The comparison of

performance between horizontal and vertical partitioning is required which requires a totally different approach. The comparison of performance using the grid and other distributed computing technologies can also be done. Work can also be started in the direction of incorporating security and privacy related issues as described in [18]-[20] to make the current work more secure with respect to privacy and security requirements. The effect of increasing the grid resources and network bandwidth can also be analyzed on the performance of the distributed approach.

## References

1. Tanenbaum, A.S., Steen, M.V.: *Distributed Systems: Principle and Paradigms*. Pearson Education, India (2010)
2. Foster, I., Kesselman, C., Tuecke, S.: *The Anatomy of Grid: Enabling Scalable Virtual Organizations*. *International Journal of Supercomputer Applications* 15(3), 200–222 (2001)
3. Foster, I., Kesselman, C.: *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco (1999)
4. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: *The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration*. Technical report, Global Grid Forum (2002)
5. Jacob, B., Brown, M., Fukui, K., Trivedi, N.: *Introduction to Grid Computing*. IBM Redbooks (2005)
6. Joseph, J., Fellenstein, C.: *Grid Computing*. Pearson Education, India (2004)
7. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: *ACM SIGMOD Conference*, Washington DC, USA (1993)
8. Duru, N.: An Application of Apriori Algorithm on a Diabetic Database. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) *KES 2005*. LNCS (LNAI), vol. 3681, pp. 398–404. Springer, Heidelberg (2005)
9. Hu, L., Zhuo, G., Quin, Y.: Application of Apriori Algorithm to the Data Mining of the Wildfire. In: *Sixth International Conference on Fuzzy Systems and Knowledge Discovery* (2009)
10. Wu, X., Kumar, V.: *The top ten algorithms in Data Mining*. CRC Press, Taylor & Francis Group (2009)
11. Sakthi, U., Hemalatha, R., Bhuvaneshwaran, R.S.: Parallel and Distributed Mining of Association Rule on Knowledge Grid. *World Academy of Science, Engineering and Technology* 42, 316–320 (2008)
12. Hahsler, M., Gruen, B., et al.: Arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. *Journal of Statistical Software* 14(15), 1–25 (2005)
13. Alfori, C., Craus, M.: Grid Implementation of Apriori Algorithm. *Advances in Engineering Software* 38(5), 295–300 (2006)
14. Perez, M.S., Sanchez, A., et al.: Design and Implementation of a Data Mining Grid-aware Architecture. *Future Generation Computer Systems* 23, 42–47 (2007)
15. Rawat, S.S., Rajamani, L.: Performance of Distributed Apriori Algorithms on a Computational Grid. In: *Asia-Pacific Services Computing Conference*, pp. 163–167. IEEE (2009)
16. Globus® Toolkit, <http://www.globus.org/toolkit>



17. UCI Machine Repository,  
<http://archive.ics.uci.edu/ml/datasets/Census+Income>
18. Singh, S., Bawa, S.: A Privacy Policy Framework for Grid and Web Services. *Information Technology Journal* 6, 809–817 (2007)
19. Singh, S., Bawa, S.: A Privacy, Trust and Policy based Authorization Framework for Services in Distributed Environments. *International Journal of Computer Science* 2, 85–92 (2007)
20. Singh, G., Singh, S.: A Comparative Study of Privacy Mechanisms and a Novel Privacy Mechanism [Short Paper]. In: Qing, S., Mitchell, C.J., Wang, G. (eds.) *ICICS 2009*. LNCS, vol. 5927, pp. 346–358. Springer, Heidelberg (2009)

# Criticality Analyzer and Tester – An Effective Approach for Critical Components Identification and Verification

Jeya Mala Dharmalingam, Balamurugan, and Sabari Nathan

Thiagarajar College of Engineering, Madurai, Tamil Nadu, India  
djmcse@tce.edu,  
{balamsg4u, sabarinathan4you}@gmail.com

**Abstract.** Nowadays, software industries spend considerable amount of time and cost to satisfy users' needs in developing quality software. To ensure the proper functioning of software after delivery, the software developers need to identify and test the critical components in the software rigorously which will otherwise make serious impact on users' requirements. As this is an important and time-consuming process, this paper proposes a novel approach for Criticality Analysis based on Sensitivity and Severity metrics and Testing using the efficient test cases generated using Bee Colony Optimization (BCO) based intelligent search algorithm. The proposed approach is also compared with Genetic Algorithm (GA) based approach and proved its efficiency. Further, an automated tool is developed for the identification and testing of any given software under test (SUT).

**Keywords:** Software Testing, Critical Components, Sensitivity Metrics, Severity Metrics, Bee Colony Optimization (BCO), Genetic Algorithm (GA), Software under Test (SUT).

## 1 Introduction

In software industries, customers are the most important stakeholders. The crucial and most important factor of such organizations is customer satisfaction. This requires a paramount importance over the quality of software. As any of the industrial strength software has more number of critical components than any other software, there is a need for rigorous effort in identifying such components and testing them is a crucial factor in quality software development. Generally, critical components are the ones that hold the key in the overall quality of the software, and the failure of such components lead to huge loss in terms of cost and life.

In the proposed approach, a novel methodology namely "Criticality Analysis" is applied to identify the Critical Components of the software. It comprises of two parts namely Sensitivity analysis and Severity analysis. Here, Sensitivity analysis is the process of extracting the coupling and cohesion metrics of the components in the software. It identifies how a component will impact its dependent components. To do this Sensitivity analysis, the proposed approach uses the existing metrics such as Fan-In, Fan-Out [1], and Information Flow [2]. In addition to these metrics, some novel

metrics such as Weightage of Methods of a Component, Weakness of Methods of a Component and Ratio of Pure Inherited Methods are introduced in this paper to perform Sensitivity analysis.

Severity analysis is the process of analyzing the impact of failure of each of the components in the software. It identifies what kind of impact each component might have on the software over their failure. The impacts are categorized as per the approach proposed by Garousi [3].

Based on these two analysis, the ‘Criticality Analysis’ yields the critical components present in the software. Then these identified components are rigorously tested using Bee Colony Optimization algorithm (BCO). As BCO is an optimization approach which was derived from the intelligent search behavior of honey bees to solve complex problems, the effective test cases are generated using it to cover all the branches of such components in less time without compromising quality.

To ensure the efficiency of the proposed approach, the results are compared with the existing GA based approach. Also, the proposed methodology is implemented as a tool to automate the entire process and to produce various test reports both graphically and textually.

## 2 Critical Components Identification

### 2.1 Sensitivity Analysis

Sensitivity analysis is the process of identifying the probability value that might sensitive to the dependent component. This process extracts the following 6 source code based metrics for each component.

- **Fan-In** – Number of Classes calling a given class
- **Fan-Out** – Number of Classes being called by a given class
- **Information flow** – It represents the flow of data in collective procedures in the processes of a concrete system. It can be calculated as follows:

$$\text{Information Flow} = (\text{Fan-In} * \text{Fan-Out})^2 . \quad (1)$$

- **Weightage of Methods in a Class** - This is used to show the complexity of a given Component by counting the number of independent paths in each of the methods of the given Component, whereas the Cyclomatic complexity [6] refers to the number of independent paths in a component.

$$\text{Weightage of Methods} = \sum CC_i . \quad (2)$$

Where,

$CC_i$  – Cyclomatic complexity of a method in a component,

$\forall i=1$  to  $m$

$m$  – Total number of methods in a component

- **Weakness of Methods of a Component (WM)** – The Weakness of methods of a component is the sum of Weakness of an each method. This can be represented as following:

$$WM(C_i) = \frac{\sum W_{mi}}{m} . \tag{3}$$

Where,

$W_{m_i}$  - Weakness of a method in a component,

$\forall i=1$  to  $m$

$m$  - Total number of methods in component

- **Ratio of Pure Inherited Methods (RPIM)** - It is the ratio of the number of pure inherited methods from a component by the dependent components.

$$DRIM(C_i) = \frac{p}{m} . \tag{4}$$

Where,

$p$  - Number of pure inherited methods

$m$  - Total number of methods in an inherited component.

## 2.2 Sensitivity Analysis

The Severity analysis is that the tactic of estimating the implications of failure and prioritizing the components per the severity level of implications. The components with higher severity value could cause the functionalities of the system. To mitigate those failures, the high severity components will be tested fastidiously.

$$SV(C_i) = \sum_{k=1}^p SV(M_k(C_i)) . \tag{5}$$

Where,

$SV(M_k(C_i))$  – Severity Value of method  $k$  in the component  $C_i$ .

$p$  – Number of methods in component  $C_i$

$SV(M_k(C_i))$  is assigned with the values as **0.95, 0.75, 0.50, 0.25** based on the four characteristic described by Garousi [3].

### 2.3 Calculating Critical Value

Critical value decides the critical level of each component. The results of Sensitivity analysis and severity analysis are considered for the calculation of critical value. In addition, the time taken to execute each component and execution count of each component also considered for calculating the critical value.

$$CI(C_i) = P(CV(C_i)) * P(SV(C_i)) * P(E(C_i)) * \text{Time-taken}(C_i) . \tag{6}$$

### 2.4 Critical Test Path Generation

This module analyzes the critical level of each component and an intelligent search process based on BCO is initiated on the graph to find out the interconnection among the critical components and with other components to generate the critical test paths to cover them. Then the selected components are assigned with a flag value to indicate their priority for testing.

For Critical Component selection in a path is calculated as follows:

$$Re(C_{i,j}) = \left| \frac{CI(C_i)}{\sum(CI(C_i))} \right| . \tag{7}$$

$$Re(P_j) = \sum Re(C_{i,j}) \forall j=1 \text{ to } q . \tag{8}$$

Paths with highest Risk Exposure are selected first and test cases are generated based on message passing for Unit testing and pair-wise testing.

## 3 Critical Component Verification

### 3.1 Unit Testing

The unit testing is performed by means of Branch coverage. The components are tested until all the branches are covered or it reaches the maximum number of test case generations. To achieve unit testing of components, we use an optimization technique, Bee Colony Optimization to produce efficient test cases to cover the branches of the component. To trace which branches are covered, we instrument an additional code to existing source code that doesn't affect the original functionalities of the system. Sample set of test cases are shown in Table 1.

**Table 1.** Sample test cases generated for Unit Testing

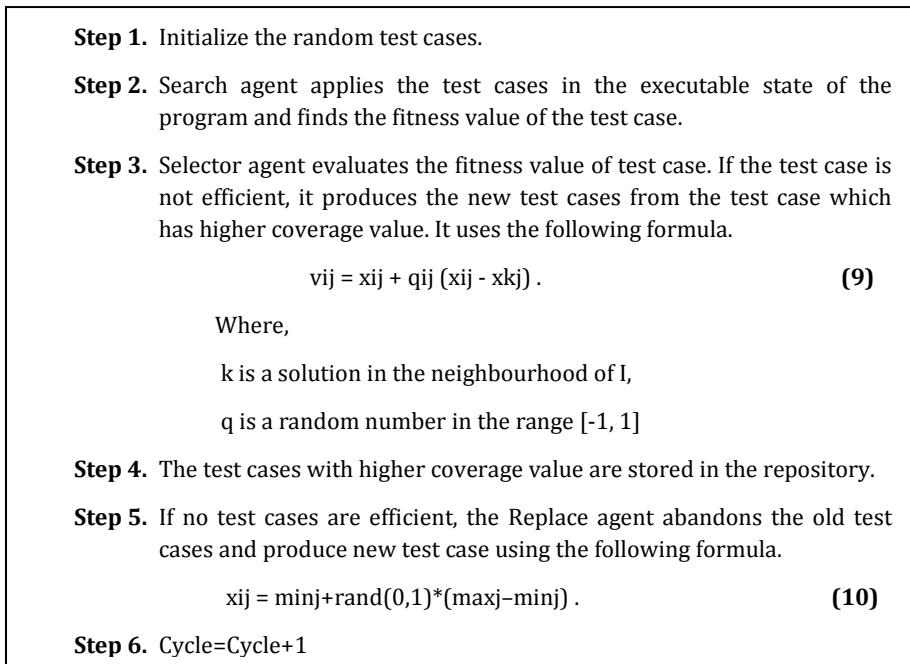
Parameter Type	Parameter Value
int	1290994490
class java.lang.String	fd
int	2036439598
class java.lang.String	null

### 3.2 Bee Colony Optimization Algorithm

Bee Colony Optimization algorithm is most recently defined by Dervis Karaboga, which was implied from the intelligent search behavior of honey bees. There are three groups of bees: employed bees, scouts bees and onlookers. We incorporate these groups of bees in our approach as the following agents [5].

- Search Agent – Employed Bee
- Selector Agent – Onlooker Bee
- Replace Agent – Scout Bee

These three agents work parallel to produce the optimized test cases by which the testing process can be reduced. The Branch coverage value of each component is considered for fitness value of Bee Colony Optimization algorithm.



**Fig. 1.** Algorithm of Bee Colony Optimization

### 3.3 Pair-Wise Testing

Pair-wise testing, also known as Integration testing, checks whether the functionalities of a component doesn't affect the integrated component. The pair-wise testing uses the connected components list extracted in initial phase, and the optimized test cases produced at the end of Unit testing. Sample test cases for pair-wise testing are shown in Table 2.

**Table 2.** Example of Pair-wise Testing

Source Class	Method that invoke the Integrated Class	Destination Class	Method that invoke the Integrated Class
Employee	Employeeedetails	Receipient	ReceipientDetails
Employee	Employeeedetails	Admin	Validate
Employee	Employeeedetails	Stock	Sno

## 4 Observation of the Proposed Approach

Critical value determines whether a component is critical to the system or not. Critical value is calculated as in equation 7 for each component in the SUT. The automated tool “JCTester” calculates the critical value of the component in the system “Library Management System”. The critical values of “Library Management System” are listed in the Table 3.

**Table 3.** Critical Value of Components in Library Management System

Class Name	Sensitivity Analysis	Severity Analysis	Execution Analysis	Time Taken (ms)	Critical Value
Issuebook	0.1141	0.1862	0.1428	210	0.6403
Bookinfo	0.1490	0.0490	0.2142	280	0.4595
Returnbook	0.1146	0.1862	0.0714	265	0.4040
Studentinfo	0.1517	0.1470	0.0714	183	0.3003
Fineinfo	0.0964	0.0490	0.1428	228	0.1496
Orderbook	0.1153	0.1862	0.0714	90	0.1372
Bookshop	0.0633	0.0490	0.0714	311	0.0680
Paymentinfo	0.0651	0.0490	0.0714	224	0.0490
Renewbook	0.0646	0.0490	0.0714	216	0.0472
Member	0.0654	0.0490	0.0714	212	0.0463

By looking on the Table 3, we can say that the components in Table 4 are most critical to the system, “Library Management System”.

**Table 4.** Critical Components List

S. No.	Component Name	Critical Value
1	Issuebook	0.6403
2	Bookinfo	0.4595
3	Returnbook	0.4040
4	StudentInfo	0.3003

The optimization technique, Bee Colony Optimization algorithm is applied for verifying the identified critical components. It reduces the time taken to complete the testing process than any other existing tools and almost cover the entire component within less number of test cases. The details are given in the Table 5.

**Table 5.** Verification Details using BCO

Sl. No.	Category	Value
1	Time taken to complete the testing	17.1 sec
2	Branches Covered by the BCO algorithm	95 %*
3	Total Number of Test cases Generated	322

\* Component may contain infeasible branches

## 5 Conclusion

Critical components are most significant to the system. It should be carefully designed and implemented by the development team before it released into the market. The proposed work has been implemented to automate the process of identifying and verifying the critical components. This approach optimizes the testing process by incorporating the intelligent search based algorithm, Bee Colony Optimization. In future, this work can be extended on web based applications also.

**Acknowledgement.** This work is a part of UGC Major Research project “Critical Components Identification and Verification using Bee Colony Optimization based Optimization approach”.

## References

1. Niclas, O., Helander, M., Wohlin: Quality improvement by identification of fault-prone modules using software design metrics. In: Proceedings of the International Conference on Software Quality, pp. 1–13 (1996)
2. Hendry, S., Dennis, K.: Software Structure Metrics Based on Information Flow. IEEE Transactions on Software Engineering 7(5), 510–518 (1981)
3. Garousi, V., Briand, L.C., Labiche, Y.: Analysis and visualization of behavioral dependencies among distributed objects based on UML models. In: Wang, J., Whittle, J., Harel, D., Reggio, G. (eds.) MoDELS 2006. LNCS, vol. 4199, pp. 365–379. Springer, Heidelberg (2006)
4. Balamurugan, S., Jeyamala, D., Jalila, A., Sabari Nathan, K.: Fault-prone Components Identification for Real time Complex systems Based on Criticality Analysis. International Journal of Computer Science and Informatics (IJCSI) 3(2), 17–23 (2013) ISSN: 2231-5292
5. Mala, D.J., Mohan, V.: BCO Tester—Bee Colony Optimization Based Software Test Suite Optimization Approach. International Journal of Software Engineering 2, 15–43 (2009)
6. Aditya, P.: Mathur.: Foundations of Software Testing. Pearson Education, India



## Appendix

The screenshots of the tool “JCTester” are given below:

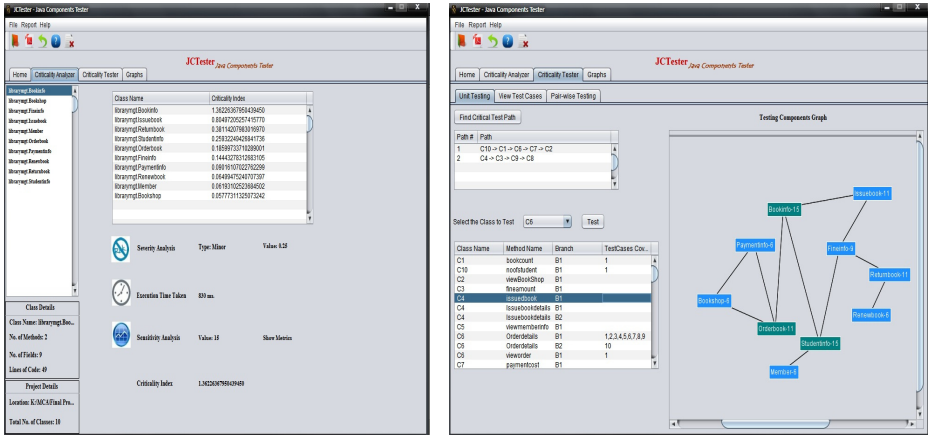


Fig. A1. Critical Components List identified and verified by the automated tool "JCTester" for the case study "Library Management System"

# Real Time Collision Detection and Fleet Management System

Anusha Pai\*, Vishal Vernekar, Gaurav Kudchadkar, Shubharaj Arsekar,  
Keval Tanna, Ross Rebello, and Madhav Desai

Department of Computer Engineering, Padre Conceicao College of Engineering,  
Verna - Goa, India 403722  
anusha.pai@gmail.com

**Abstract.** In most of the accidents occurring in remote areas information about their occurrence does not reach the emergency services on time. This can lead to fatalities or severe mental trauma to the accident victim till they are attended. In this paper development of real time collision detection and fleet management system is explained. The system has been developed adhering to the Software Engineering framework of systematic analysis, design, implementation, testing and modification. On the hardware front, an accelerometer has been used as a crash or rollover detector of the vehicle during and after a crash. With signals from an accelerometer, a severe collision is recognized and the vibration sensor will send a signal to microcontroller which in turn will activate GPS-GSM module. GPS module will send the coordinates that it receives from the satellite on a real time basis of the vehicle via GSM module to the website, where the operator can view the locations of the accident and send help appropriately. The entire system is simulated to understand its effectiveness in handling collision detection.

**Keywords:** Software Engineering, Real Time system, Accident Detection, Global Positioning System, Impact Sensor, Fleet Management.

## 1 Introduction

Software Engineering is the process in which customer requirements are systematically translated into a software product. This has to be achieved in an efficient manner with optimum use of resources and satisfying cost criteria. The attributes of good software are dependability, efficiency, acceptability and maintainability. The software engineering methodology has been used in this paper with an application to a real time system. A real time system is a system which has to respond to an input signal within a specified amount of time and send appropriate responses. Real time systems are implemented on microprocessors which are critically constrained in memory space and execution time. Efficiency is the primary

---

\* Corresponding author.

quality attribute in these systems. The effectiveness of decision making process in real time system can be improved by integrating this system with software engineering approach.

An increase in the number of vehicles on the road has considerably increased the number of accidents. As a result of which the casualties have increased many fold. Presently, whenever there is an accident the fellow vehicle users call up the emergency services. The time gap between the occurrence of accident and the reaching of fellow vehicle user on the road is a matter of chance. Hence there is a need for real time information dissemination to the emergency services.

## 2 Related Work

Lots of research on application of software engineering practices is done by various authors. Authors in [1] have compared various software life cycle process models in terms of their common, distinct and unique features. The use of architecture throughout the software development life cycle has been discussed by the authors in [2]. Experiences of applying new approach called Decision Based Software Development (DBSD) have been discussed by authors in [3] and concluded that DBSD provided life cycle support for complex software systems. Authors in [4] have analyzed data from various software firms and concluded that defect prevention involves proactive, reactive and retrospective moves to reduce defects considerably. Authors in [5] applied different defect detection and prevention methods to identify majority of defects. The need for efficient real time systems has focused research to integrate software engineering practices with real time system development. Application of software engineering practices into real time system development in the area of hydrological data acquisition and ecological monitoring has been discussed by authors in [6].

Software engineering approach is widely used for developing real time system for collision detection. The authors in [7] have proposed the use of component based software engineering approach in an application involving distributed embedded real time systems. Authors in [8] have developed a tool to capture the distance between objects and vehicles in order to help the driver of a vehicle of an impending collision. Authors in [9] applied method of detecting and reporting accident using RF transceiver module to communicate to emergency service provider. As reported in this literature this system has a range upto 100 meters. Use of image compression analysis and RFID system in vehicle accident detection is discussed by authors of [10]. Authors in [11] have developed an application of accident detection and reporting using GPS, GPRS and GSM technology. However in this work, vehicles with speed lesser than a specified threshold are assumed to have met with accident which is not true in real world. Authors in [12] developed an auto surveillance system which uses traffic video to detect abnormal instances of traffic. Use of biomedical smart sensors and microcontroller based mobile technology is discussed by the authors in [13] wherein the information about the status of the victim from the site of the accident is communicated to the emergency care center through SMS.

In this paper, the development of automated system for real time detection of accidents and fleet management is discussed. The system sends the information through the website to first aid center in a few seconds covering geographical coordinates.

### **3 Case Study on Automated Collision Detection System and Fleet Management**

The real time collision detection system provides solution in terms of vehicle security, emergency services and fleet management system. The objectives of our study are:

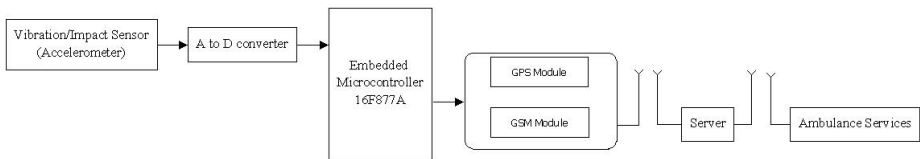
1. In case of an emergency situation like an accident, the vehicle should send messages to the registered numbers and the emergency services would hence be notified.
2. In the event of collision occurring in an area without range, the driver should be provided with a manual switch inside the vehicle, which would enable the emergency services to reach the vehicle and help the person in need.
3. Provide the exact location of the vehicle and help the emergency services to reach the location faster
4. Provide the essential parameters like vehicle's latitudinal & longitudinal position obtained from the GPS module.
5. Website should provide complete details about the vehicles essential parameters which are required during vehicle tracking.
6. Provide the above information to the fleet management to track the status of their vehicles

The present accident handling system involves passerby observing the accident and informing the medical team. There are three drawbacks to this system. One is the delay between the time of accident and the arrival of a passerby which is a matter of chance. Second is the availability of communication device or the availability of signal to the communication device of the passerby. Third is the mental state of passerby on observing the collision. Due to these reasons, this system has become highly unreliable. To overcome this, we have proposed a system which is devoid of any human interference in taking decisions about communicating the collision. In the proposed system whenever a collision occurs, the vibrations are captured and sent via the GPS-GSM module to the website. The operator on the computer then calls up the emergency service to immediately proceed to the site of the collision.

#### **3.1 System Development**

A vibration sensor (accelerometer) is used as an accident detector. When the vehicle meets with an accident, the vibration sensors generate the signal; this signal is compared with the threshold values. If the value generated exceeds the predefined threshold it is recognized as a collision. The Microcontroller then activates the GPS-GSM module, in which the GPS hardware obtains the coordinates via satellite

and then sends the values to the GSM hardware. Subsequently, the module sends the received coordinates to the server via predefined hardware protocols, which are then available to the emergency services on the website. Fig. 1 shows the system architecture in the form of a block diagram. Fig. 2 shows the proposed approach to real time collision detection and fleet management. When the module is started it will first begin with its system initialization. Then it will continuously check if an accident has occurred or not. If an accident has occurred, it will check whether the impact is higher than the threshold set. If the value is less, there will be an emergency switch which can be used to send the message, else it again checks for collision. If the impact is higher than threshold it will sense that an accident has occurred. The GPS module will then be activated to receive the current coordinates and send them to the GSM module. The GSM module will now form a message along with the information about owner and the vehicle number and send it to the server. Once it is received by the server the operator can see the details of the accident and alert the rescue team to go to that location.



**Fig. 1.** Block diagram showing components of proposed system

The system consists of modules like Impact sensors, Microcontroller, GPS and GSM.

### 3.1.1 Impact Sensors

An Impact Sensor is an electromechanical device that will measure acceleration forces caused by moving or vibrating the accelerometer. In this work, a triple (3) - axis accelerometer called ADXL 335 is used, because it measures the g-forces along all the three the axis i.e. X, Y and Z. ADXL 335 requires a constant input voltage of 5V which is provided by the automotive battery but is then stabilized using a voltage stabilizer and the required voltage is supplied. The input to the sensor is a +5V supply provided by a voltage regulator circuit that takes an input from the +12V supply of the vehicle. The output of the sensor is an analog voltage on all 3-axis.

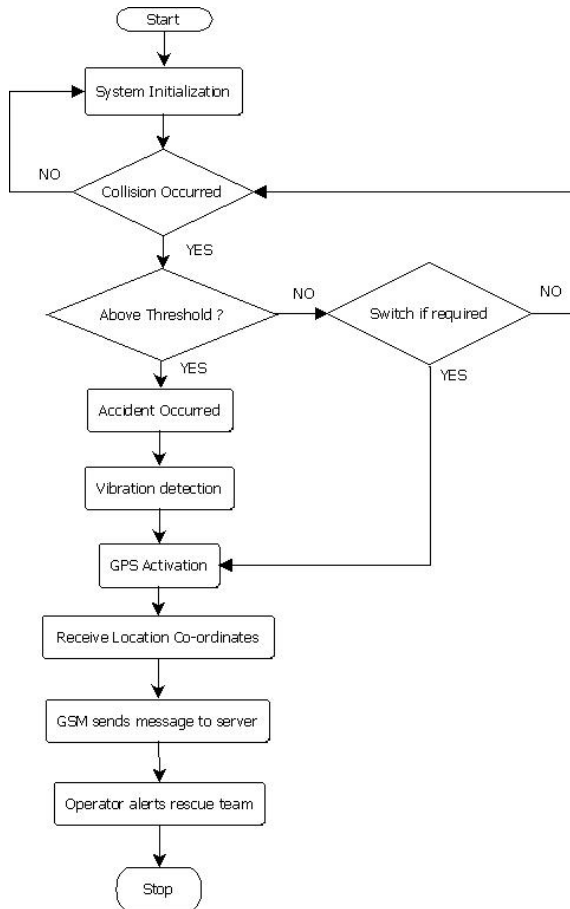
### 3.1.2 PIC Microcontroller16F877A

The input to the microcontroller is the analog output of the impact sensors. Three inputs can be taken anywhere at the input pins of the chip from X, Y, Z axis respectively. The output provided by the microcontroller is a digital output which enables the panic switch on the GPS-GSM module.

### 3.1.3 GPS Module

The output of the GPS module will be the coordinates which it receives from the satellites on real time basis. The GPS module used in the work is SiRF STAR III GPS Module. It will track up to 20 satellites at a time while providing fast time-to-first-fix.

Its far reaching capability meets the sensitivity & accuracy requirements of car navigation as well as other location-based applications.



**Fig. 2.** Flow chart of proposed system

### 3.1.4 GSM Module

The inputs to the GSM module are the coordinates provided by the GPS module after which it sends the same via packets to the website. The GSM module used in the work is SIM 340. SIM340 delivers GSM/GPRS850/900/1800/1900MHz performance for voice, SMS, Data, and Fax in a small form factor and with low power consumption.

## 3.2 Software Implementation

Making use of predefined protocols the hardware module is integrated with the software. The output of the software is a website where the coordinates sent by the hardware module are plotted along with other vital parameters of the vehicle like

vehicle number, velocity, date/time etc. The geographic location is obtained using Google Maps. By using the Google Maps API, Google Maps are embedded into an external website, on to which site specific data can be overlaid. A service for retrieving static map images and web services for performing geocoding, generating driving directions, and obtaining elevation profiles is provided.

The website developed by the authors provides two modes of operation: Emergency Services and Fleet Management. In Emergency Services mode, only one central operator can view the vehicles which are in need of emergency situation. A special feature is incorporated on the website which filters only the vehicles that have signaled an emergency situation to the website via the hardware module through which the centralized operator can carry out necessary action. Only the vehicles having the State = 9 appear on this mode of the website. The hardware module can be in any one of the following states. START- 1, MOVE- 2, STOP- 3, OVER SPEED- 4, ENGINE ON & VEHICLE STOP – 5, MAIN POWER CUT- 6, MAIN POWER ON- 7, DEVICE RESTART- 8, PANIC- 9, LOW BATT- 10, GPS not found -11. Fleet Management mode is predominantly used for private users to track their vehicles. It enables the registered user to track their vehicle in real time. Users can view essential vehicle parameters such as fuel quantity, velocity etc. along with the location. Every user registered here creates his/her personal profile through which only their vehicles can be tracked and real-time information such as Vehicle Number, Latitude and Longitude, Velocity/Speed, State and Amount of Fuel Left.

## 4 Results

The code that was required for the programming of the PIC Microcontroller was written on the VSM studio. In this code threshold value is specified so that the PIC Microcontroller gives an output of 5V on the output line. The code is compiled to get a .asn file and a .dsn file. When the .asn file is run on the VSM studio the corresponding .dsn file runs on the PROTEUS software. The PROTEUS is used to simulate the working of the microcontroller. The PCB layout was designed using the Express PCB. This is used to connect the microcontroller to the GPS-GSM module. When the PIC microcontroller gives 5V output, the GPS-GSM module is activated and it sends an emergency message to the server. On the server side, a website was developed to be used as an user interface with the operator where the later will see the locations of the accidents on the map. The basic design of the webpage is made up of HTML, PHP and JavaScript. The website consists of a registration page where a new user will enter details like name, address, vehicle number, and contact number. The validation on the registration page is done by using JavaScript validation on the backend. Once the user has registered he can log on to the main page where he can track his vehicle and also see the current location and condition of his vehicle. This page consists of a map obtained by integrating the Google map API. The Google map API uses javascript functions as an interface for various functions on the map. For the emergency services, the operator will only see vehicles that have met with an accident along with the timestamp of when the vehicle met with an accident. At the backend MySQL is used to store the data, wherein there are three tables, one for storing the

user details, second to store data that a GSM sends and a third that contains the details of the accident. Fig. 3 shows the tracking of the vehicles using Google Maps and Fig. 4 depicts the essential vehicle parameters on the website.



Fig. 3. Location of Collision

Name	Vehicle	Lat	Long	DateTime	Velocity	State	Fuel	IO
100	PCCE-GOA-CP4	15.326602	73.933441	2013-07-13 09:20:03	0	3	0	0,0,0,0
100	PCCE-GOA-CP4	15.326602	73.933456	2013-07-13 09:18:00	0	3	0	0,0,0,0
100	PCCE-GOA-CP4	15.326549	73.933532	2013-07-13 09:16:05	0	3	0	0,0,0,0
100	PCCE-GOA-CP4	15.326541	73.93354	2013-07-13 09:14:03	0	3	0	0,0,0,0
100	PCCE-GOA-CP4	15.326525	73.933555	2013-07-13 09:12:00	0	3	0	0,0,0,0
100	PCCE-GOA-CP4	15.326561	73.933532	2013-07-13 09:10:32	0	3	0	0,0,0,0
100	PCCE-GOA-CP4	15.326601	73.933433	2013-07-13 09:05:04	0	3	0	0,0,0,0
100	PCCE-GOA-CP4	15.32656	73.933486	2013-07-13 09:03:01	0	3	0	0,0,0,0
	PCCE-			2013-				

Fig. 4. Location Parameters on the website

## 5 Conclusion

Integration of software engineering practices in the development of real time system for collision detection and fleet management has been discussed in this paper. The impact from the vibration sensor was captured by the microcontroller which in turn activated the GPS-GSM module which then sent message to the website wherein the operator could inform the emergency service. This approach provides means to an efficient and reliable solution for decreasing the mortality rate in the event of collision. This system will not only be used in saving lives but also can be used in private sectors e.g. Fleet Management in companies in the form of Geofencing, Theft Protection etc. The limitation of this work is the range of the GSM module. If there is no range and the vehicle meets with an accident the GSM module will not be able to send the message to the server. In this case the server will show the last received coordinates. Future prospects of this paper would be in the form of an Event Data Recorder (Black Box) which would help in speeding up the judicial procedures in case of collisions for the reconstruction of events.

## References

1. Rodríguez, L.C., Mora, M., Alvarez, F.J.: A Descriptive Comparative Study of the Evolution of Process Models of Software Development Life Cycles (PM-SDLCs). In: Proc. of Mexican International Conference on Computer Science 2009, pp. 298–303 (2009)



2. Unphon, H.: Making use of architecture throughout the software life cycle - How the build hierarchy can facilitate product line development. In: Proc. of ICSE Workshop on Sharing and Reusing Architectural Knowledge 2009, pp. 41-48 (2009)
3. Wild, C., Maly, K., Zhang, C., Roberts, C.C., Rosca, D., Taylor, T.: Software engineering life cycle support-decision based systems development. In: Proceedings of IEEE Region Ninth Annual International Conference TENCON 1994, vol. 2, pp. 781-784 (1994)
4. Suma, V., Nair, T.R.G.: Enhanced Approaches in Defect Detection and Prevention Strategies In Small And Medium Scale Industries. In: Proc. of Software Engineering Advances 2008, pp. 389-393 (2008)
5. Suma, V., Nair, T.R.G.: Effective Defect Prevention Approach in Software Process for Achieving Better Quality Levels. In: Proc. of World Academy of Science, Engineering & Technology 2008, pp. 258-262 (2008)
6. Vieira, R.G., Romano, B.L., Braga e Silva, G., de Campos, H.F., da Cunha, A.M.: Using Best Practices of Software Engineering into a Real Time System Development. In: Proceedings of Sixth International Conference on Information Technology: New Generations 2009, p. 1681 (2009)
7. Hammer, D.K., Chaudron, M.R.V.: Component-based software engineering for resource-constrained systems: what are the needs? In: Proc. of Sixth International Workshop on Object-Oriented Real-Time Dependable Systems 2001, pp. 91-94 (2001)
8. Bhonsle, S., Trivedi, M., Gupta, A.: Database-centered architecture for traffic incident detection, management, and analysis. In: Proc. of IEEE Intelligent Transportation Systems, pp. 149-154 (2000)
9. Kannan, R., Nair, M.R.N., Prakhya, S.M.: Wireless Vehicular Accident Detection and Reporting System. In: IEEE Proceedings of International Conference on Mechanical and Electrical Technology (ICMET 2010), pp. 636-640 (2010)
10. Kantawong, S., Phanprasit, T.: Intelligent traffic cone based on vehicle accident detection and identification using image compression analysis and RFID system. In: IEEE Proceedings of International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 1065-1069 (2010)
11. Amin, M.S., Jalil, J., Reaz, M.B.I.: Accident Detection and Reporting System using GPS, GPRS and GSM Technology. In: IEEE Proceedings of International Conference on Informatics, Electronics & Vision (ICIEV 2012), pp. 640-643 (2012)
12. Bo, L., Qimei, C., Fan, G.: Freeway Auto-surveillance From Traffic Video. In: IEEE Proceedings of: ITS Telecommunications, pp. 167-170 (2006)
13. Prabakar, S., Porkumaran, K., Samson Isaac, J., GunaSundari, J.: An enhanced accident detection and victim status indicating system: Prototype. In: Proceedings of India Conference (INDICON), pp. 351-356 (2012)

# An Effective Method for the Identification of Potential Failure Modes of a System by Integrating FTA and FMEA

Samitha Khaiyum<sup>1</sup> and Y.S. Kumaraswamy<sup>2</sup>

<sup>1</sup> R & D Centre, MCA (VTU) Department, DSCE, Bangalore-78

samitha.athif@gmail.com

<sup>2</sup> MCA (VTU), DSCE, Bangalore-78

yskldswamy2@yahoo.com

**Abstract.** Development and maintenance of software projects which are failure free is one of the core challenges in software industries. There exist various system reliability engineering approaches to identify and recover from failures. Failure Mode Effect Analysis (FMEA) and Fault Tree Analysis (FTA) have provided their clear merits for reliability analysis. They are used to derive failures, prioritize these failures based on user perception to perform sensitivity analysis of different components of the system architecture. This paper presents an extension to reliability analysis by integrating FMEA and FTA to analyze and diagnose the different causes of failure. It further provides a case study comprising of the gas leak detection system using FMEA (Failure Modes and Effects Analysis) and FTA (Fault Tree Analysis) tools to investigate the failures. The study further provides a comparative analysis of above stated tools and integrates them to enhance the efficiency of the system reliability.

## 1 Introduction

Embedded systems have given rise to high industrial competition and advances in hardware and software technology which has an increased demand for much improved functionality. These embedded systems are integrated with the network environments that affect these systems in ways that might not have been foreseen during their development [12]. Since, these systems are complicated in both design and implementation, reliability is a major concern. Software failures happen due to memory corruption, incomplete execution, timing errors (early execution / late execution) etc. The development of high assurance systems must therefore make sure that it has no faults and even if they are out of reach and are highly unlikely to occur. There are many formal methods being used in the software industry that compare between the stated properties and the operational description of a given system [2].

The FTA and FMEA analysis aids in identifying the effects of these failures and look for the protection provided in the code for some of the critical failures. These tools identify failure modes of system components and have a great impact on the system reliability. Importance has to be given to analyze failures earlier in the life cycle, at the software architecture design level. [11].

The FTA is a Deductive top-down approach and FMEA is an Inductive bottom-up approach. Both these analyses provide input in testing, verification and validation which address the effect of failures at the system functionality-level. The outcome of the analysis provides inputs for component testing, integration tests and system testing in order to achieve reliability. However, the graphical representation of FTA defines relationships between faults and provides assistance in identifying undesirable system states that can lead to system failure. The root of the fault tree (Top node) represents the system failure and the leaf nodes represent faults [2].

The objective of this work is to develop a comparative analysis of these two tools. The work further intends to develop a failure domain model by analyzing and categorizing the different domains of failure. FMEA describes different failure scenarios in form of worksheets, which help to prioritize these failures. The severity of these failures is further revealed through the aid of a case study carried out in a leading product based industry.

Since FMEA requires lot of effort and time as well as FTA is a graphical tool to casually connect relations between states and faults, this research aims to integrate the FMEA and FTA to analyze the tree in the top down fashion to diagnose potential faults that may lead to failure. These identified faults can then be fed to FMEA to work on risk management techniques to avoid these faults.

## 2 Literature Survey

Authors in [10] have stated that reliability and safety analysis are important for any company, regardless of industry. They further state that safety analyses help to decrease the risk of a product will may cause any undesirable consequences during operation.

Authors in [7] have discussed differences between FMEA and FTA namely the boundaries of analysis, direction of analysis and presentation of analysis process results. They also throw light on the fact that FMEA and FTA are rarely combined but if done so can minimize the shortcomings of both the methodologies.

Author of [8] has discussed the advantages of using FMEA and FTA for safe control software design. He individually discusses the two methodologies to derive safety requirements and come out with a tool for safety related review of user and software requirement documents.

Authors of paper [9] have proposed Bi Directional Analysis (BDA), a core assessment technique by which safety critical software can be certified. This is a combination of forward and backward analysis. The forward analysis in BDA has its roots in software FMEA while the backward analysis has its roots in software FTA.

## 3 Integrated FTA / FMEA Safety Model

FMEA is carried out on components, functions, operating modes/conditions to analyze the causes, effects and its protective modes. FTA contributes precipitates, cascades the conditions /events / relationships of the root causes for the identified failures and thereby provides a preventive /mitigating protection modes [2].

The FMEA is represented conventionally in tabular format while FTA is in a graphical representation form [4][5]. It is worth to recall that input to FMEA is always specific in terms of failure modes and output of FMEA is generic depicting the system effects. However, the input to FTA is generic in terms of system faults while the output is specific to root cause. Thus, FTA is carried out in parallel but FMEA is carried out sequentially.

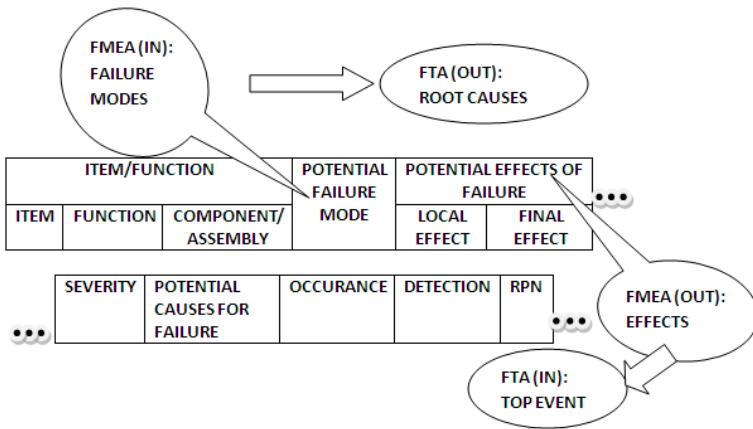


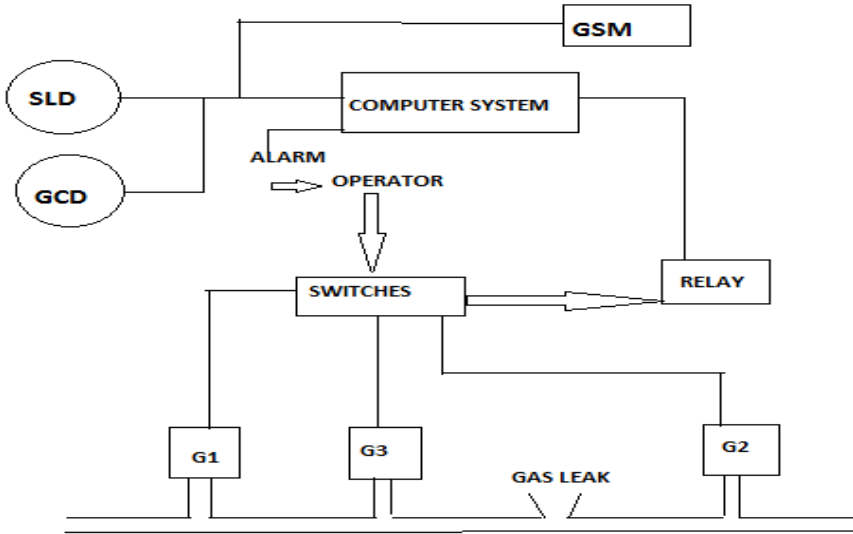
Fig. 1. Integrated model of FMEA and FTA

Thus, integrated FTA and FMEA model as in figure 1 provides input for analysis of interdependencies both causal and temporal. Further, the integration of the two approaches provides justification for prioritization of verification, validation and testing. The benefits of this process can be penned as being Systematic, well Structured and ordered in nature. It is Qualitative, Quantitative as well as Comparative. It acts as an indicator which is both documented and analytically justified in terms of input and output. It can be considered Complementary, Supplementary and Synergistic. This model is broader than specific analysis.

This can be extended to any Purpose and Scalable to accommodate many.

#### 4 Application of Integration of FMEA and FTA: A Case Study

The integration of FMEA and FTA upon safety critical applications is investigated through a case study comprising of gas leakage detection through automated mode. Figure 2 illustrates the gas leak detection system. It provides information of various components such as Control Gate G1, Control Gate G2, Blow down Gate G3, Sonic Leak Detector SLD, Gas concentration Detector GCD and Global system for mobiles GSM.



**Fig. 2.** Gas Leak Detection System

Figure 2 indicates the design to detect any undesired presence of gas which is achieved in two ways. Initially, it is achieved by the isolation of the sections so that the size of the leak is limited to the inventory between the two control gates, which is controlled by closing the control gates G1 and G2. Secondly, releasing the pressure in the section by flaring the gas, this is achieved by opening the Blow down gate G3.

Figure 2 further indicates that the gas leak detection is realized by two sensors namely Sonic Leak Detector (SLD) and Gas Concentration Detector (GCD) which triggers the gas concentration. The computer system is incorporated in the gas leak detection system such that upon the detection of leakage, the computer system drops off the relay and thus shut down the power supply to all three control gates. Further, it is programmed to automatically call the stored emergency numbers using the Global System for Mobile (GSM).

In addition to the automated preventions, measures are further considered to alert the Alarm to inform the Operator in case of gas leak who manually activates the switches to cut off power to the control gates.

## 5 Application of FTA and FMEA on Automated Gas Leakage System

Many critical accidents in the past years show increased risk in the modern systems which are complex due to advancement in technology. Safety has become one of the critical issues during any system development demanding improved and better safety developing strategies.

**Table 1.** FMEA worksheet for automated gas leakage system

Process Function	Potential Failure Mode	Potential Effects of Failure	Potential Causes of Failure	Current Controls	Occ.	Sev.	Det.	RPN
Gas Leak Detection system failure	Control gate G1 fails to close	Does not isolate the size of leak	Control gate G1 operation failure	G3 to open	1	7	2	14
	Control gate G2 fails to close	Does not isolate the size of leak	Control gate G2 operation failure	G3 to open	1	7	2	14
	Blow down gate G3 fails to open	No Flaring of gas	Blow Down gate G3 operation failure	Relay to trip	1	7	2	14
	Switches are stuck to close state	Manual contact does not break	Switches Contacts free zed	Dial emergency calls	3	8	1	24
	Relay is stuck to close state	Power to gates does not break	Relay Contacts don't drop out	Dial emergency calls	2	8	1	16
	Modem Failure	Call forwarding fails	Technical / Physical failure	Alarm as alternate measure	2	5	4	40
	Line Busy	Emergency number is not dialled	Connection failure	Alarm as alternate measure	2	3	4	24
	No Answer	No response to emergency calling	No availability for help	Alarm as alternate measure	2	3	4	24
	Operator unavailable	Does not manually break contact to power	Operator not present to cut off power	Dial emergency calls	2	3	6	36
	Alarm fails to sound	Operator is not warned	Technical failure of alarm	Dial emergency calls	1	3	3	9
	Computer fails to process Trip condition	Relay contacts are not activated	Computer does not activate relay	Dial emergency calls	1	7	4	28
	Gas Concentration Detector fails to register leak	Gas leak not detected	Concentration Detector does not detect change in gas concentration	Signal to computer for trip signal	1	6	4	24
	Sonic Leak Detector fails to register leak	Gas leak not detected	Concentration Detector does not detect change in gas concentration	Signal to computer for trip signal	1	6	4	24

This case study involves the application of fault trees to graphically represent faults or failures that contribute to some hazard or accident in automated gas leakage system. Figure 3 depicts the logical structure depicted as an upside-down tree with the hazard (called top-event) at its root to represent possible occurrences of failures in the automated gas leakage system.

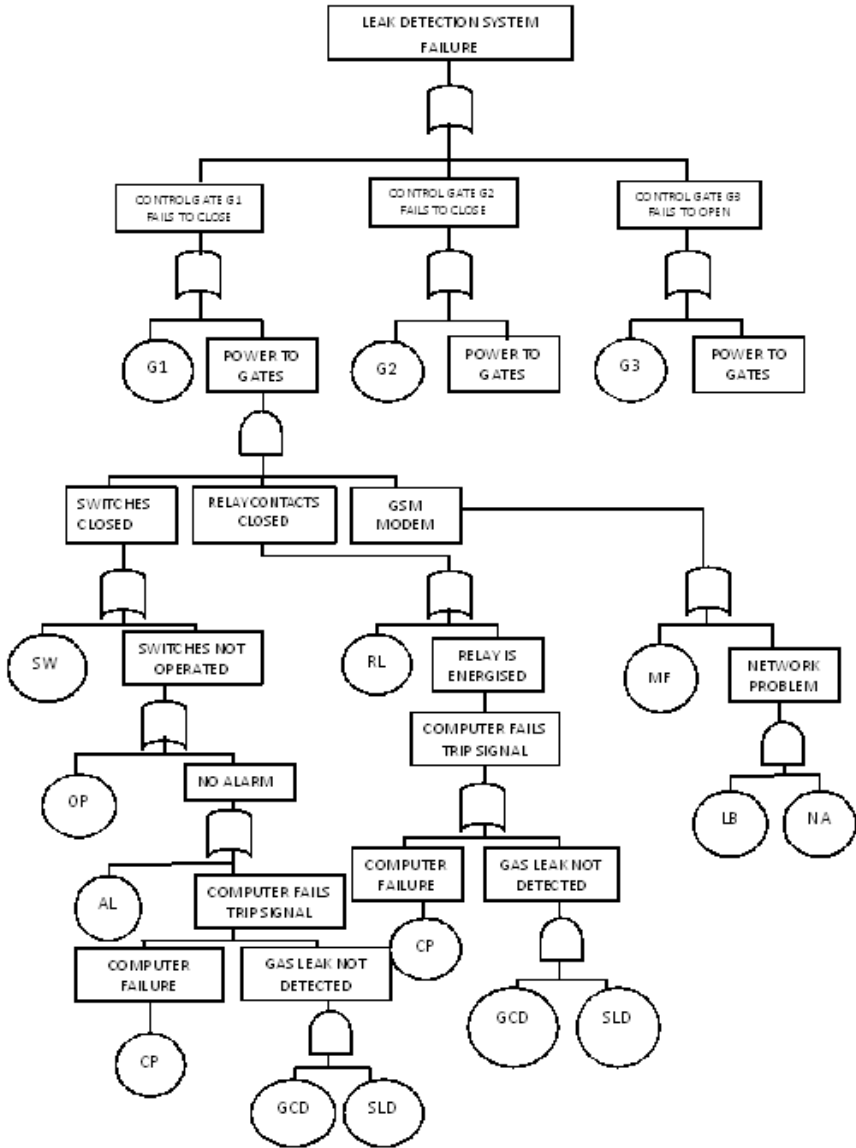


Fig. 3. Graphical representation of FTA on automated gas leakage system

Possible root causes for identified failures in automated gas leakage system as inferred from leaves in figure 3 are Control gate G1 fails to close (G1), Control gate G2 fails to close (G2), Blow down gate G3 fails to open (G3), Switches are stuck to close state (SW), Relay is stuck to close state (RL), Modem Failure (MF), Line Busy (LB), No Answer (NA), Operator Unavailable (OP), Alarm fails to sound

(AL), Computer fails to process Trip condition (CP), Gas Concentration Detector fails to register leak (GCD) and Sonic Leak Detector fails to register leak (SLD).

FMEA enables a team to identify FMEA tool has the potential to reduce cost and human error. It provides a standardized systematic approach to failure analysis while gathering model information. Without any tool, the user may repeatedly choose general failure modes which are rarely adequate for FMEA worksheet development, subsequent systems and safety analysis. It further helps to drive out the complete and most descriptive choices of applicable common failure modes. This encourages full consideration of potential failure modes and therefore more thorough and accurate analysis. [3]

Thus identified root causes are now given as input to FMEA where it analyzes what might go wrong and develop appropriate mitigation plans based on the probability, severity and ease of detection of the various 'failures'. Metrics for Occurrence (Occ.), Severity (Sev.) and Detection (Det.) are assigned numeric ranging from 1 to 10 in compliance with the standard values [1]. The product of Occ, Sev and Det yields a Risk Priority Number (RPN). The mitigation plans are prioritized based on the RPN of each failure [1]. Table 1 depicts the FMEA worksheet which is obtained upon integration of FTA output as an input to FMEA. It provides information about all the potential failure modes that were considered as input to FMEA, its corresponding effects, causes, and existing controls in addition to occurrence, severity, detection and RPN values. It infers that lower the value for occurrence, severity and detection, risk is less. Correspondingly, higher the values of occurrence or severity or detection which ultimately provides a product of these values as RPN demands high priority risk mitigation plans.

Complex systems are difficult to analyze using only FMEA as it becomes difficult to identify the potential failure modes.[6] However, when integrated with FTA, it simplifies the method of potential failure mode identification. The integrated approach of FTA and FMEA on the sampled case study ensures reduction in human effort with increased efficiency in identification and further corrective measures to be incorporated instantaneously.

## 6 Conclusion

Retention of reliability and safety in critical applications is one of the core needs of the day. There exist several tools, methods and policies to achieve the same in all industries. However, with advent of technology, it is imperative to reduce human efforts and achieve improved efficiency over the above said issues. This research therefore aims at failure detection and correction measures through integrated approach of Fault Tree Analysis (FTA) and Failure Mode Effect Analysis (FMEA) upon an automated gas leakage system as a case study.

The integrated FTA / FMEA approach ensures to be effective risk identification and reduction which can be applied on simple to complex applications. Further, this approach results in generation of risk priority values based on which mitigation plans can be formulated. The approach thus reduces human effort and yields efficiency in achieving higher reliability during system development process.



## References

1. de Queiroz Souza, R., Álvares, A.J.: FMEA and FTA analysis for application of the reliability centred maintenance methodology: case study on hydraulic turbines. In: ABCM Symposium Series in Mechatronics, vol. 3, pp. 803–812
2. Desovski, D., Cukic, B.: A Component Based Approach to Verification and Validation of formal software models. In: Architecting Dependable Systems IV, ch. 10, pp. 89–114
3. Flores, M., Malin, J.T.: Failure Modes and Effects Analysis Assistant Tool Feasibility Study, [http://ntrs.nasa.gov/archive/nasa/casi.ntrs../20130013157\\_2013012931.pdf](http://ntrs.nasa.gov/archive/nasa/casi.ntrs../20130013157_2013012931.pdf)
4. Huang, H.Z., Tong, X., Zuo, M.: Posbist fault tree analysis of coherent systems. *Reliab. Eng. Syst. Safety* 84(2), 141–149 (2004)
5. Pickard, K., Muller, P., Bertsche, B.: Multiple failure mode and effects analysis: an approach to risk assessment of multiple failures with FMEA. In: Annual Reliability and Maintainability Symposium, pp. 457–462. Institute of Electrical and Electronics Engineers Inc., Piscataway (2005)
6. FMEA: A Guide for Continuous Improvement for the Semiconductor Equipment Industry by Sematech, <http://www.sematech.org/docubase/document/0963beng.pdf>
7. Bluvband, Z., Beit-Dagan, P.R., Grabov, P.: Bouncing failure analysis (BFA): the unified FTA-FMEA methodology. In: Proceedings of the Annual Reliability and Maintainability Symposium, pp. 463–467 (2005) ISSN : 0149-144X
8. Maier, T.: FMEA and FTA to support safe design of embedded software in safety-critical systems. In: *Safety and Reliability of Software Based Systems*. Springer (1997)
9. Lutz, R.R., Woodhouse, R.M.: Bi-directional analysis for certification of safety-critical software (1999), <http://trs-new.jpl.nasa.gov>
10. Huffman, D.L., Bowman, K., Akers, J.: What we can learn about reliability and safety analyses from different industries. In: 2013 Proceedings-Annual Reliability and Maintainability Symposium (RAMS), pp. 1–6 (2013) ISSN:0149-144X, Print ISBN:978-1-4673-4709-9
11. Lutz, R.R., Shaw, H.-Y.: Applying adaptive safety analysis techniques [for embedded software]. In: Proceedings of the 10th International Symposium on Software Reliability Engineering (1999) ISSN: 1071-9458 Print ISBN: 0-7695-0443-4
12. Schmidt, D.C.: Middleware for real-time and embedded systems. *Magazine Communications of the ACM - Adaptive Middleware CACM Homepage Archive* 45(6), 43–48 (2002)

# Impact of Resources on Success of Software Project

N.R. Shashi Kumar<sup>1</sup>, T.R. Gopalakrishnan Nair<sup>2</sup>, and V. Suma<sup>1</sup>

<sup>1</sup> Advanced Software Engineering Research Group,  
RIIC Dayananda Sagar Institutions Bangalore, India  
{nrshash, sumavdsce}@gmail.com

<sup>2</sup> Advanced Software Engineering Research Group,  
RIIC, DSIAramco Endowed Chair - Technology, PMU, KSA, Bangalore, India  
RIIC, Dayanada Sagar Institute, Bangalore, India  
trgnair@ieee.org2

**Abstract.** Success level of the project depends on various factors such as cost, time, and availability of resources in terms of technology and personnel etc. This paper elucidates an empirical study of several projects developed over a period of time in a product and service based CMMI Level 5 Software Company. The investigation results present impact analysis of critical resources such as cost, time, number of developers towards the successful completion of the project as allocated by the project manager during the developmental process. This analysis further throws light on the need for improvements in the project management process in terms of right choice of resource allocation.

**Keywords:** Software Process, Project Manager, Software Engineering, Project Management, Software Quality, Defect Management.

## 1 Introduction

Software has become one of the core requisites of any application domain which demands huge investment in terms of cost, time, and resources by software organizations in order to develop high quality software. Demand for quality increases whenever application domain is mission dependent such as software that controls aircrafts, steers ships, runs banking system, medical diagnostics, and radar guided missiles so on [Watts s Humphrey, 2009]. Software engineering principles aim to develop customer satisfied high quality products. Nevertheless the advancement in technology, engineering a completely customer satisfied high quality software is yet a challenge. However, development of projects is dynamic in nature in order to meet the organization's long term business goals [Walker D, 2009]. Therefore, investment on software projects are made to attain long term cost savings to the organization on the expenditure [Blanchard, B. S., et al].

Santhosh Dubey [2011] feels that development of software projects is complex due to technical factors and anomalies that are environment-contingent where human resource influences other resources. However, organizations in the modern world are compelled to cope up with rapid progress of IT in addition to manage increasing complexities in larger projects which demands enhanced life cycles deliverables

[Davenport, T. H. et al, 1998]. Further, it is illustrious fact that product emerges out through process which is driven by people. Software products are never an exception for the same.

The main objective of this paper is to analyze the impact of critical resources such as cost, time and number of developers towards the successful completion of the project as allocated during the developmental process. The organization of the paper is as follows. Section 2 of the paper briefs about the background work for this investigation. Section 3 presents the research Methodology, Section 4 provides the empirical analysis of several projects through a case study and Section 5 summarizes the paper.

## 2 Related Works

Research in software development process is progressive ever since the evolution of software. A survey report indicates nearly 50 percent of projects are deemed to be failure projects due to over budget and lagging schedule [Debbie Tesch et al, 2007, Fairiey, R. E et al, 2003]. However, Debbie Tesch et al suggests that failures of projects in IT systems can be categorized for several reasons such as cost, time, performance and other such quality associated attributes.

Effective software development is achievable through well-defined process and efficient developing team. N. Ehsan et al, (2010) state that the skill of project manager influences the success or failure of projects. According to Yingxu Wang et al (2000), corporate management in addition to project managers influences the software development process by developing projects using professional management and software quality assurance methodologies. Shashi Kumar N.R et al (2013), states that significant project influencing parameters such as time, cost, developers and defect count will have an effect on the software project management process.

This investigation therefore aims to study the impact of project influencing parameters such as cost, time, and number of developers towards the realization of successful projects.

## 3 Research Methodology for Empirical Study

This article elucidates an empirical study of several projects developed over a period of time in a product and service based CMMI Level 5 Software Company. The investigation results present impact analysis of critical resources such as cost, time, number of developers towards the successful completion of the project as allocated by the project manager during the developmental process. This analysis further throws light on the need for improvements in the project management process in terms of right choice of resource allocation.

The research methodology followed for this investigation is explained below

### Step 1

Investigation of effeicnecy of project mangment proecess in leadirng software industries.To accomplish this objective, a study was conducted over industries holding the certifications.

### Step 2

Further there existed huge number of projects in the above said industries and since analyzing every project was a big challenge due to the time constraints involved for this investigation.

### Step 3

Several projects were investigated to analyze the role and efficiency of project management process towards the realization of effective software development. In order to carry out the investigation, the Secondary data was collected from quality assurance department and from the document management center. Further, the mode of data collection includes log reports, face to face communication and interviews.

### Step 4

Having obtained the data from the above 3 steps, this work continued towards analyzing the information.

### Step 5

The next steps of investigations lead towards analyzing the efficiency of project management process.

Observation results indicate the need for enhancing the project management process in terms of accurate estimation and subsequent allocation of project influencing parameters and henceforth realization of development of a successful project.

## 4 Case Study

This research includes several investigations of a software industry. However, this paper focuses on a case study of a sampled software industry which is CMMI Level 5 and ISO certified service/product based industry. The company operates their business on the areas like Business Intelligence, data warehouse, Enterprise Resource Planning, Business Process Outsourcing, Banking, Finance, Airlines and Energy Utilities.

Table 1 depicts a sampled data of twenty five projects developed since 2009 to 2012. These projects are developed in object supporting technology using languages such as .net and Java. Table 1 provides the resource information as estimated and actual resource information as occurred during developmental process in addition to the variation observed between the values. The resource information projected in the table includes cost, time, and number of developers assigned, number of defects in addition to the success level of the complete project.

Table 1 infers that estimated resource information and actual resources utilized are not same during the project development. Variation in the resources is not desirable since it affects the success of the project.

From the above observations, it infers that success level of the projects which are investigated has varying project success levels. Since estimation, control and prediction of the efficiency of any process and people effort requires the measurement of success level of the projects, this research led towards the introduction of new metric namely Success Level Index (SLI). SLI is a quality measurement metric aimed towards measuring the success level of the projects.

**Table 1.** Variation in Different parameters and Success Level of a project

PROJ ECT	NO. OF DEVELOPERS			DEFECTS			TIME			COST(\$)			Success Level of a Project
	EST.	ACT.	%	EST.	ACT.	%	EST.	ACT.	%	EST.	ACT.	%	
P1	1	1	0	36	30	15.6	132	128	2.9	3296	3200	2.9	94.64
P2	1	2	100	48	45	6.9	218	174	20	5438	4350	20	63.28
P3	1	1	0	64	67	-4	278	232	16.7	6960	5800	16.7	92.66
P4	1	1	0	67	65	2.5	288	240	16.7	7200	6000	16.7	91.04
P5	1	1	0	69	72	-4.5	298	248	16.7	7440	6200	16.7	92.8
P6	1	2	100	100	95	5	432	360	16.7	10800	9000	16.7	65.42
P7	1	1	0	144	145	-0.4	582	520	10.7	14560	13000	10.7	94.74
P8	1	2	100	162	159	2	701	584	16.7	17520	14600	16.7	66.17
P9	1	2	100	200	201	-0.5	907	720	20.6	22680	18000	20.6	64.81
P10	1	2	100	200	198	1	792	720	9.1	19800	18000	9.1	70.2
P11	1	2	100	211	215	-1.8	836	760	9.1	20900	19000	9.1	70.92
P12	1	2	100	211	215	-1.8	912	760	16.7	22800	19000	16.7	67.13
P13	1	2	100	356	349	1.8	1472	1280	13	36800	32000	13	68.02
P14	1	2	100	367	353	3.7	1478	1320	10.7	36960	33000	10.7	68.71
P15	2	3	50	456	459	-0.8	2017	1640	18.7	50430	41000	18.7	78.34
P16	2	2	0	478	481	-0.7	2133	1720	19.4	53320	43000	19.4	90.49
P17	2	3	50	536	535	0.1	2314	1928	16.7	57840	48200	16.7	79.14
P18	2	2	0	589	595	-1	2332	2120	9.1	58300	53000	9.1	95.71
P19	3	5	66.7	1011	1100	-8.8	4404	3640	17.4	110110	91000	17.4	76.85
P20	3	5	66.7	1111	1200	-8	4200	4000	4.8	105000	100000	4.8	82.95
P21	3	3	0	1133	1120	1.2	4570	4080	10.7	114240	102000	10.7	94.35
P22	5	7	40	1700	1665	2.1	7344	6120	16.7	183600	153000	16.7	81.15
P23	5	7	40	1789	1652	7.7	7599	6440	15.3	189980	161000	15.3	80.46
P24	6	8	33.3	1944	1532	21.2	7700	7000	9.1	192500	175000	9.1	81.82
P25	6	7	16.7	1989	1851	6.9	8592	7160	16.7	214800	179000	16.7	85.77

P1...P25 – Projects; Est. - Estimated .....

Success Level Index (SLI) = (Achieved Success Level in the Project)/ (Expected Success Level in the Project) (1)

Figure 1 through Figure 4 indicates the impact analysis of the project influencing resources such as number of developers, defect profile, time and cost on SLI.

Figure 1 infers that high variation in terms of estimated number of developers and subsequent actual number of developers assigned by the project manager for the successful completion of the project is observed to have higher impact on SLI. It is apparent from the figure that projects having no variation in estimated and actual number of developers has resulted in ideal SLI. As an instance projects such as P1, P3, P5, P4, P16, P21 etc proves the same. Similarly, it is also observed that whenever variation in estimation and actual requirement of number of developers is very high, does not lead towards complete customer satisfied products. This indicates that efficiency of project management process in right estimation and allocation of number of developers has higher impact on the success level of the project.

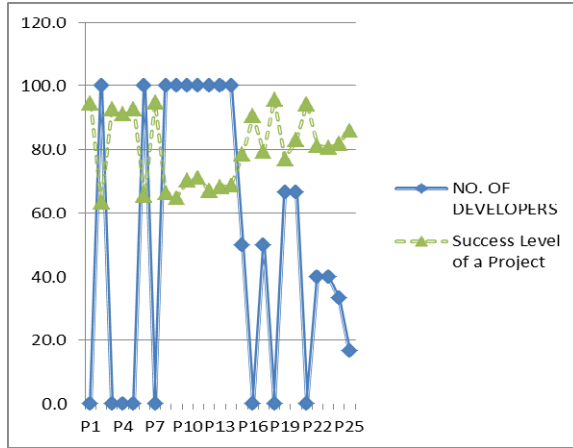


Fig. 1. Effect of Variation in No. of Developers and SL

However, Figure 2 infers that defect count also has an impact which is dependent on the developers who are allocated by the project managers. The observations further indicate that even with no variation in estimated and actual number of developers has resulted in high success level. The analysis infers that it is not just the number of developers who influence defect count but it is the right number of developers with right skill set who play a vital role in injection of defects, which influences the success level of projects.

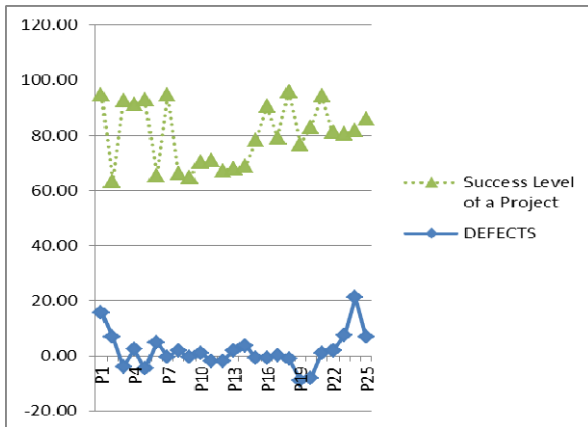


Fig. 2. Effect of Variation in No. of Defects and SL

Figure 3 and Figure 4 infers that time and cost has high impact on the SLI of the project. This is because time and cost which are mutually dependent on each other is further influenced by the efficiency of the developers. However, it is worth to recall

that choice of developers towards development of the project is decided by the project managers. Hence, apt choice of developers influences the time, cost and defect count in the project which also has an impact on the SLI.

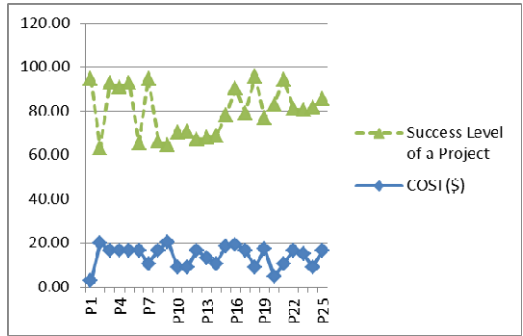


Fig. 3. Effect of Time on SL

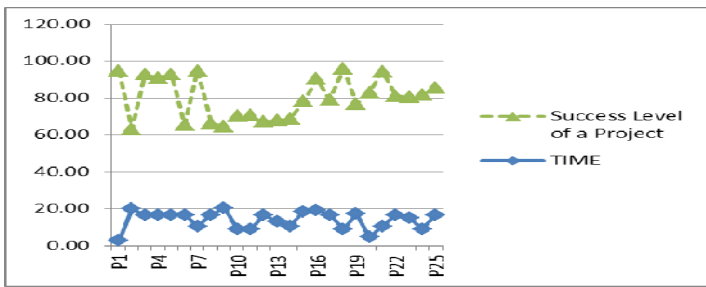


Fig. 4. Effect of Cost on SL

## 5 Conclusion

Project management process plays a very significant role in selection, optimizing and allocation of resources to the project. It also determines the Success Level of a Project and a company. This awareness among PMs and a company enable to achieve quality and maturity in software process.

There is a scope for continuous improvement in the Project Management process in development of quality software for the success of a project and a company. However there is a need for measuring the impact of variation in resources which are selected and allocated on the success of a software product and a company.

This work is an empirical study of several projects in a service/product based global software company to study the impact of variation in major resources such as Time, cost, No of developers and defects on the Success of a Project. The observational inferences drawn in the paper indicate the need to evaluate the impact of aforementioned parameters analytically on the success level.

**Acknowledgments.** The authors would like to sincerely acknowledge all the industry people for their immense help in carrying out this work. The authors would like to thank all the referees and the editor for their constructive and helpful comments.

## References

1. Munns, A.K., Bjeirmi, B.F.: The role of project management in success of project. *International Journal of Project Management* 14(2), 81–87 (1996)
2. Bessant, J., Caffyn, S.: High-involvement innovation through continuous improvement. *International Journal of Technology Management* 14(1), 7–28 (1997)
3. Blanchard, B.S., Fabrycky, W.J.: *Systems engineering and analysis*, 4th edn., p. 31. Prentice Hall, New Jersey (2006)
4. Choo, A.S., Linderman, K.W., Schroeder, R.G.: Method and context perspectives on learning and knowledge creation in quality management. *Journal of Operations Management* 25(4), 918–931 (2007)
5. Chan, C.W.-N., Cheng, C.-H., Gunasekaran, A., Wong, K.-F.: A framework for applying real options analysis to information technology investments. *Int. J. of Industrial and Systems Engineering* 2012 10(2), 217–237 (2012)
6. Davenport, T.H., Prusak, L.: *Working Knowledge*. Harvard Business School Press, Boston Massachusetts (1998), Budgen, D.: *Software Design*. Pearson Publication (2008)
7. Tesch, D., Kloppenborg, T.J., Erolick, M.N.: IT Project Risk factors: The Project Management Professional Perspectives. *Journal of Computer Information Systems* (Summer 2007)
8. Petersen, D.: *Redefining Project Management, PMP* (2009)
9. Ehsan, N., Waheed, K.Z., Asghar, U., Nawaz, M.T., Mirza, E., Sarwar, S.Z.: Effects of Project Manager's Competency on Project Success. In: *Proceedings of the 2010 IEEE ICMIT* (2010)
10. Fairiey, R.E., Willshire, M.J.: Why the Vasa sank: 10 problems and some antidotes for software projects. *IEEE Software* (March/April 2003)
11. Gray, J., Anantatmula, V.: Managing six Sigma projects through the integration of Six Sigma and project management processes. *J. of Six Sigma and Competitive Advantage* 5(2), 127–143 (2009)



# An Optimized Approach for Density Based Spatial Clustering Application with Noise

Rakshit Arya<sup>1</sup> and Geeta Sikka<sup>2</sup>

<sup>1</sup> Department of Electronics & Information Technology, National Informatics Center (NIC)  
New Delhi 110053, India  
arya.rakshit@nic.in

<sup>2</sup> Department of Computer Science, National Institute of Technology, Jalandhar (PB), India  
sikkag@nitj.ac.in

**Abstract.** The density based algorithms such as DBSCAN is considered as one of the most common and powerful algorithms in data clustering with the noise datasets. DBSCAN based algorithm's is able to find out clusters with the different shape and variable size. However it is failed to detect the correct clusters, if there is density variation within the clusters. This paper presents new way to solve the problem of detecting the clusters of varying density which most of the DBSCAN based algorithms can't deal with it correctly. Our proposed approach is depending on oscillation of clusters which is obtained by applying basic DBSCAN algorithm to conflation it in a new clusters, the proposed algorithm help to decide whether the different density regions belong to the same cluster or not. The experimental results showed that the proposed clustering algorithm gives satisfied results on different Data sets.

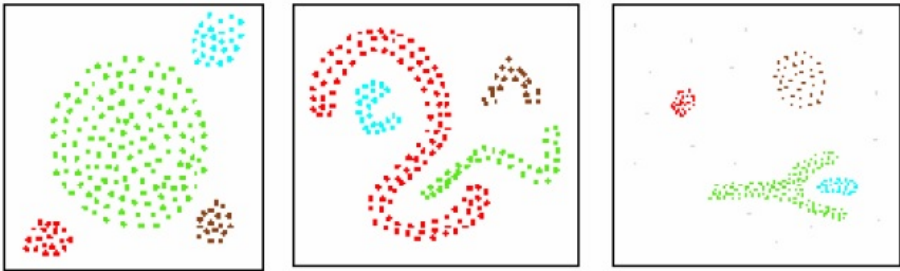
**Keywords:** Core point, DBSCAN, Eps, RegionQuery, MinPts.

## 1 Introduction

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data points that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Today clustering is mostly using in data classification, Segmentation, Data analysis and so on.

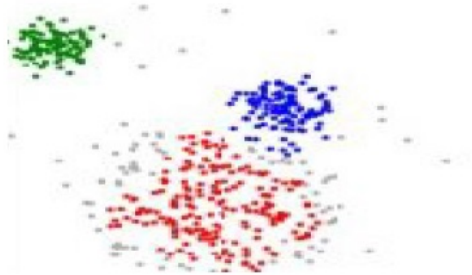
DBSCAN algorithm is exists within one of the most famous density-based clustering Algorithm (Ester et al., 1996). For each point within a cluster its eps-neighborhood for some given  $\text{eps} > 0$  has contained at least the minimum points. The "density" in the eps-neighborhood of points has to exceed some threshold to become the core point. The main advantages of the using DBSCAN are the, first; DBSCAN does not require a pre determination of the number of cluster on datasets, the second, DBSCAN can find out arbitrarily shaped clusters on the Datasets. It can even find a cluster completely surrounded by a different cluster. Due to the MinPts parameter, single-link effect is probably reduced, and the third, DBSCAN is not sensitive for noise and outliers.

The classification classes of DBSCAN points are core points, border points and outlier points. A point is a core point if it has more than specified number of points (MinPts) within Eps. Core point is exists within the interior of a cluster. A border point has fewer points than MinPts within Eps but it exists within the neighborhood of a core point. A noise point in DBSCAN is any point which is neither eligible for a core point nor a border point. Fig-1, reflects the DBSCAN clustering results on the artificial data sets with the property of uniform cluster's density.



**Fig. 1.** Examples of Uniform Density Clusters

Due the main disadvantage of the DBSCAN (Ester et al., 1996), that it cannot cluster data sets with varied density present in different cluster's, since the MinPts-eps combination cannot then be chosen effectively for all clusters. As shown in Figure-2, there are two clusters with different region density, after we applied DBSCAN, it detects four clusters.



**Fig. 2.** Clusters with Varied Density Regions

This paper proposes two mechanisms to solve the problem of varied density within clusters. These mechanisms depend on inter-dependency and distance closeness within the two sub-clusters which obtained from the basic DBSCAN algorithm. The rest of this paper is organized as follows: Section 2 introduces related work for this paper. Section 3 proposes and discusses the new algorithm. Section 4 presents experiment results and analysis. Finally section 5 the conclusion and the future work.

## 2 Related Works

Many papers proposed clustering algorithms which is depend on the density-based, Partitioned, and Hierarchical types methods. As a density-based clustering method, DBSCAN (Ester et al., 1996) is an algorithm which is based on typical density and also traverses the Eps neighbour's. DBSCAN tries to find out clusters according to user-defined criteria for the threshold density "k", neighbourhood radius (Eps) and a number of points on the threshold Eps-neighborhood (MinPts), and then find out the core object which satisfies the minimum MinPts criteria within a given radius Eps. However DBSCAN can find clusters of arbitrary shape and handle noise well, but due to unnecessary RegionQuery call for each data objects, it is slow in comparison to other methods. The main problem of DBSCAN is the selection criteria in determining the appropriate density threshold, and the main weakness in dealing with different density level of dataset.

To resolve these limitations of DBSCAN, many algorithms has been introduced such as DENCLUE and OPTICS etc. OPTICS (Ankerst et al., 1999) (Ordering Points to Identify the Clustering Structure) is improved version of DBSCAN, which practically takes the same execution time and same in the process, but represents clusters in the order of objects in the database. It is optically able to display information from the deepest point of view, which is based on the density of clusters and it work fine on varied density datasets. But OPTICS weakness is to find the information in clusters which exists in the outside of the sparse data sets even though it is good for dense datasets.

DD\_DBSCAN algorithm is another enhancement of DBSCAN, which finds the clusters which has a property of different shapes, sizes and differ in local density. But the algorithm is unable to handle the density variation within the cluster. DDSC (A Density Differentiated Spatial Clustering Technique) is proposed, which is an extension of the DBSCAN algorithm. It detects clusters having non-overlapped spatial regions with reasonable homogeneous density variations within them.

ROCK (Guha, 1999), a Robust hierarchical-clustering algorithm is top down hierarchical clustering based on the narration of links. The number of common neighbours in the dataset is called Number of links between two clusters. After ending the calculation of the initial number of links among the data objects, the algorithm considered each cluster as a single object and keeps track on integration of clusters based on the some goodness scale to integrate the sub-clusters.

CHAMELEON [3] finds the clusters in a data set by applying two-phase algorithm. In the first phase, it uses a traditional method to generate a k-nearest neighbor graph. In the second phase, it uses a top down hierarchical clustering algorithm to find the cluster by merging the sub-clusters.

Jain (1988) explores a density based approach to identify clusters in k-dimensional point sets. The data set is partitioned into a number of non-overlapping cells and histograms are constructed. Cells with relatively high frequency counts of points are the potential cluster centers and the boundaries between clusters fall in the "valleys" of the histogram. This method has the capability of identifying clusters of any shape.

However, the space and run-time requirements for storing and searching multidimensional histograms can be enormous. Even if the space and run-time requirements are optimized, the performance of such an approach crucially depends on the size of the cells.

### 3 Proposed Algorithm

Proposed algorithm is based on oscillating the data objects, obtained from the basic DBSCAN algorithm. The next subsection reflects how the oscillation works on data-Sets.

#### 3.1 Oscillation Process

**Lemma 1:** The total density function of data point reflects the difference between the data point with respect to core point. So  $E_d$  for the data point  $y_1$  is the difference between data point  $y_1$  and the core point of its cluster.

$$E_d = \text{distance}(Y_i, C_j) \quad (1)$$

Where  $1 \leq i \leq n$ ,  $1 \leq j \leq k$ ,  $n$  is the no of data points and  $k$  is the no of cores.

**Lemma 2:** In addition to the initial cluster received from the basic DBSCAN method, we can use the total density function  $E_d$  to calculate the distance function (Euclidean-function  $E$ ) of data points by taking difference of data points  $E_d$  and its respective core points  $E_d$ .

$$E_i = E_d(X_i) - E_d(C_i) \quad (2)$$

The main idea behind it is to oscillate the data points with respect to core point. A core point represents the specific cluster and measure the distance function  $E_i$  of each data point as per eq. (2). After calculating, if data-point's  $E_i$  is greater than  $E_i$  for some other cores-points then oscillate all data points in that cluster towards the core point which has maximum distance measure on that object point as per eq.(3).

$$X(i+1) = X(i) + \mu * ((X(c) - X(i)) * (e^{-1/2\alpha})) \quad (3)$$

Where,

$$\alpha = e^{-i/T}$$

$X(i)$  is tested data-point and  $X(c)$  is tested core point.

$\mu$  is learning rate normally selected by user.

$T$  is control in reduction of sigma. Normally takes the length of dataset.  $T$  is used in oscillation function is to control reduction in sigma i.e. as moment of the point towards new core is reduced as the time increases.

### 3.2 Algorithm

The oscillating algorithm works in two phases that is described below:

**Phase 1: Finding Core-Points:** The purpose of this step is to gather the basic understanding of initial data points and find the corresponding core object's of the clusters. In the first sub step, we apply DBSCAN to find initial clusters. In the second subsequent step we compute the value of density function  $E_d$  for each data points.

**Phase 2: Oscillation Process:** In this step we calculate the  $E_i$  for the data points according to eq. (2). In the Next step we continuously check, for each data point if this  $E_i$  with respect to core density function is greater than  $E_i$  for some other core-point density functions, then oscillate all the points in that cluster towards the core which has maximum distance function.

## 4 Simulation and Results

We simulated this algorithm onto some *Artificial* and *Real* datasets. These experiments have simulated on MATLAB R2012a environment. The finding and results are presented here in terms of error index, which is misclassified samples divided by total number of samples.

### 4.1 Artificial Datasets

In this section we will discuss about the results that have obtained from the application of optimized algorithm onto artificial datasets. These datasets was selected randomly. The first data set that is named ball dataset contains 1 cluster. When we applied traditional DBSCAN on ball dataset fig.3, we got 2 clusters this is due to 2 level density distributions of data points. But on the other hand our proposed algorithm gives the optimal result that is 1-cluster in fig.4. This is due to oscillation that merges the two level distributions of data points.

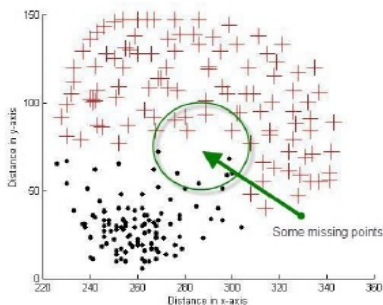


Fig. 3. Two Clusters by Basic DBSCAN

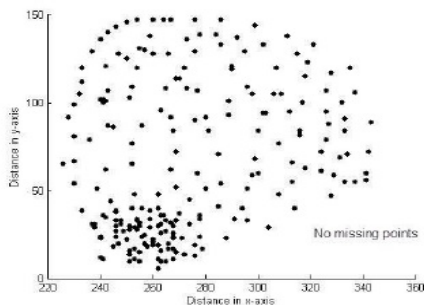


Fig. 4. One Cluster by Optimized DBSCAN

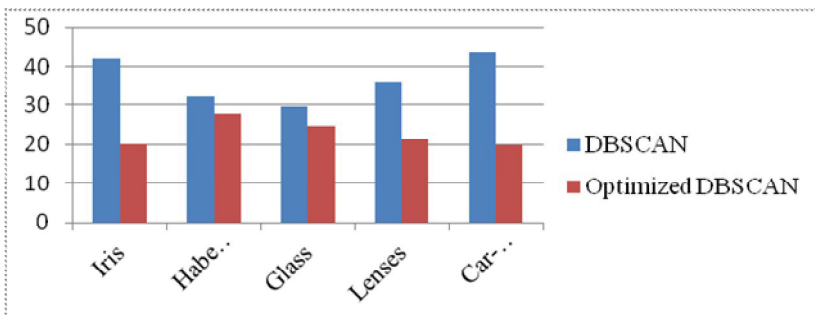
## 4.2 Real Datasets

We used the five real data sets which have taken from UCI [16] repository. Table-1 shows the characteristics of these data sets. Our first dataset *IRIS* contains 50 instances for the data of iris plant having three different classes. Each class is linearly separable from the others. Second dataset *Haberman* contains 306 instances with two classes and 3 attributes for patient age, operation time and no of positively affected node detected. Third dataset *Glass* contains 214 instances with 7 classes and 10 attributes of chemical composition. Fourth dataset *Lenses* contains 24 instances with 4 attributes per instance. Finally last dataset *Car-Evaluation* contains 1728 instances with 6 attributes per instance.

**Table 1.** Real Datasets Characteristics

Dataset Name	Dataset Characteristics	Number of Instances	Number of Attributes	True Clusters
Iris	Multivariate	150	4	3
Haberman	Multivariate	306	3	2
Glass	Multivariate	214	10	7
Lenses	Multivariate	24	4	3
Car evaluation	Multivariate	1728	6	4

We applied our proposed algorithm on Iris dataset with  $\mu = 0.00006$ , we got average error index of 20%. On the Haberman dataset with  $\mu = 0.00007$ , average error index of 27.78% is recorded. On the Glass dataset with learning rate  $\mu = 0.00007$ , average error index of 24.68% is resulted. When we applied optimized algorithm on Lenses dataset with  $\mu = 0.000068$ , an average error index of 21.35% is gathered. On the last dataset Car-evaluation with  $\mu = 0.000068$ , average error index of 19.68% is recorded. Fig.5 depicts the basic comparison with basic DBSCAN on the basis of average error index.



**Fig. 5.** Average-Error Index Comparison

Fig.6 reflects that proposed algorithm takes some amount of extra time to setup oscillation to adjust the varied density on multivariate datasets.

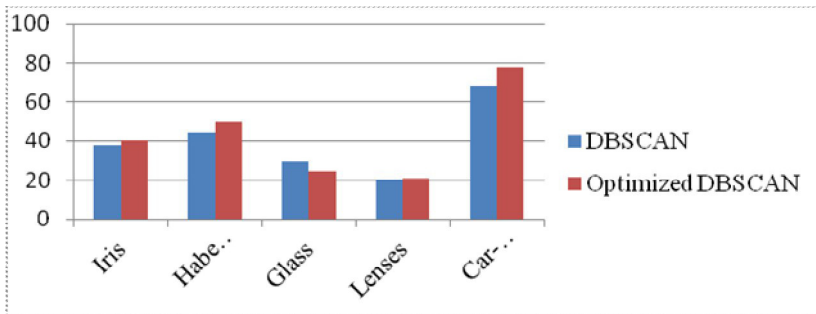


Fig. 6. Runtime Comparison with DBSCAN

## 5 Conclusion

In this paper we have introduced a new method to optimize the DBSCAN algorithm. Here we were oscillating the data points to fix the density distribution in dataset and then merge the sub clusters into new cluster. The optimization was done after obtaining the sub-cluster results from DBSCAN algorithm after that algorithm merged the clusters by using the concept of closeness and interconnectivity. In our experiments we have tested it on the different dataset to check its precision and to find whether it is working well or not. Experiment results prove that the proposed algorithm is more robust and effective on varied density datasets compared with fundamental DBSCAN.

The results show that our algorithm detected the intended cluster with more accuracy on Iris, Haberman, Glass, Lenses and Car-evaluation datasets.

## References

1. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers (2006)
2. Guojun, G., Ma, C., Wu, J.: Data Clustering: Theory, Algorithms, and Applications. SIAM, Philadelphia, ASA, Alexandria, VA. ASA-SIAM Series on Statistics and Applied Probability (2007)
3. Karypis, G., Han, E.H., Kumar, V.: CHAMELEON: A hierarchical clustering algorithm using dynamic modeling. *Computer* 32(8), 68–75 (1999)
4. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, KDD 1996, pp. 226–231. Portland, Oregon (1996)
5. Ankerst, M., Breunig, M., Kriegel, H.P., Sander, J.: OPTICS: Ordering points to identify the clustering structure. In: Proc. 1999 ACM-SIGMOD Int. Conf. Management of Data, SIGMOD 1996 (1999)

6. Hinneburg, A., Keim, D.A.: An efficient approach to clustering in large multimedia databases with noise. In: Proc. 1998 Int. Conf. Knowledge Discovery and Data mining, KDD 1998 (1998)
7. Berkhin, P.: Survey of Clustering Data Mining Techniques, Accrue Software, Technical Report, nnnn (2002)
8. Parimala, M., Lopez, D., Senthilkumar, N.C.: A Survey on Density Based Clustering Algorithms for Mining Large Spatial Databases. *International Journal of Advanced Science and Technology* 31 (June 2011)
9. Ng, R., Han, J.: Efficient and effective clustering method for spatial data mining, Santiago, Chile, pp. 144–155 (September 1994)
10. Guha, S., Rastogi, R., Shim, K.: CURE: An efficient clustering algorithm for large databases, Seattle, WA, pp. 73–84 (June 1998)
11. Guha, S., Rastogi, R., Shim, K.: ROCK: A robust clustering algorithm for categorical attributes, Sydney, Australia, pp. 512–521 (March 1999)
12. Roy, S., Bhattacharyya, D.K.: An approach to find embedded clusters using density based techniques. In: Chakraborty, G. (ed.) ICDCIT 2005. LNCS, vol. 3816, pp. 523–535. Springer, Heidelberg (2005)
13. Borah, B., Bhattacharyya, D.K.: DDSC: A Density Differentiated Spatial Clustering Technique. *Journal of Computers* 3(2) (February 2008)
14. Borach, B., Bhattacharyya, D.K.: A Clustering Technique using Density Difference. In: Proceedings of International Conference on Signal Processing, Communications and Networking, pp. 585–588 (2007)
15. Ram, A., Jalal, S., Jalal, A.S., Kumar, M.: DVBSCAN: A Density based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases. *International Journal of Computer Applications* (0975 – 8887) 3(6) (June 2010)
16. UCI Machine Learning Repository,  
<http://archive.ics.uci.edu/ml/datasets>



# A Kernel Space Solution for the Detection of Android Bootkit

Harsha Rao and S. Selvakumar\*

CDBR-SSE Project Laboratory,  
Department of Computer Science and Engineering,  
National Institute of Technology Tiruchirappalli,  
Tamil Nadu, India – 620 015  
harsha.cancer@gmail.com, ssk@nitt.edu

**Abstract.** A Rootkit is a malicious software that damages the operating system at user and kernel levels. A bootkit is a kernel rootkit which affects only the kernel space. The bootkit starts executing as soon as BIOS selects the appropriate boot device which is residing in Master Boot Record (MBR). Main feature of bootkit is that it cannot be easily detected since the existing antivirus solutions start detecting only after the boot process is completed. *Currently there is no solution for the detection of bootkit in android operating system.* The proposed detection module is deployed to get activated during the booting process itself. This detection technique is not only useful for detecting android bootkit but also useful for detecting any kind of illegal process which gets activated during the booting of android operating system. The proposed solution was tested by implementing a bootkit attack program and found to detect and kill the malicious process that got activated by the attack, during boot process.

**Keywords:** Boot process, Init process, kernel Rootkit, Android OS, Bootkit.

## 1 Introduction

The rootkit is a type of malicious software designed to get legal access to the operating system and to hide some illegal process or program from the general detection method [1]. Generally operating system works in two modes, viz., user mode and kernel mode, which leads to two types of rootkit, user mode rootkit and kernel mode rootkit respectively. Kernel mode rootkits are more harmful than the user mode rootkits since they may damage system files. Bootkit is the kernel mode type of rootkit. Bootkit derives its name from the concatenation of words “BOOT” booting process, and “KIT” which means software which implements some tools. Its main purpose is to get activated during booting process and get access to system even before the system has been initialized for other application. Further, the bootkit is more subtle and more difficult to detect.

---

\* Corresponding author.

Recently in March 2012, NQ mobile security centre found out bootkit (DKFBootKit) [2] on the android operating system. DKFBootKit repackages legal apps by enclosing its own malicious payloads in them. However, the victim apps it chooses to infect are utility apps which require the root privilege to work properly. The infected apps range from the ones managing apps installed on the phone, unlocking popular games, to others providing the license keys for some paid apps. These apps seem to have reasons to request root privilege for their own functionality. It is also reasonable to believe that users may grant the root privilege to these apps. However, DKFBootKit makes use of the granted root privilege for other malicious purposes, namely compromising the system integrity. Since it has root privilege it can execute any utility commands such as mount, ifconfig, etc., to modify kernel, maliciously. Therefore, it is evident that there is a need for detecting bootkits in android operating system. An attempt has been made in this paper to design and develop one such solution for detecting bootkits.

The rest of the paper has been organized as follows: Section 2 discusses the related works. In Section 3 motivation is given. In Section 4 introduction to init process and its functions, proposed solution, implementation details, android bootkit attack generation and results of execution are discussed in detail. Finally Section 5 concludes the paper.

## 2 Related Work

In Windows 8, Unified Extensible Firmware Interface (UEFI) is used for securing the boot sector of kernel from the bootkit attack [3]. UEFI Framework is composed of a list of handles database and a set of protocol interfaces. The handle database is composed of objects called handles and protocols. Handles are a collection of one or more protocols, and protocols are data structures tagged by a Global Unique Identifier (GUID). The data structure for a protocol may be empty, may contain data field, or may contain services (function pointer), or may contain both. The handle database is the central repository for the objects that are maintained by the UEFI-based firmware. These handles are used to detect the bootkit.

In Mac operating system firmware password and secure boot is used for detection of bootkit [4]. The firmware password mechanism prevents the attackers from executing malicious EFI drivers and applications from devices connected to the USB, FireWire, and network interfaces. But such mechanism does not protect the user from malicious drivers loaded from devices which are connected directly to the PCI bus via ExpressCard or Thunderbolt. The secure boot is a mechanism by which approved vendors sign their drivers, boot loaders, and applications with a cryptographic key. A database of allowed vendor keys is stored in secure, non-volatile storage, and these keys are used to verify the signatures within executables that are to be loaded and executed. Executables that are not signed by approved vendors are refused execution.

The UEFI and firmware password and secure boot security mechanism were broken. Present solution 'secure boot' given by the Microsoft for windows was broken by "*Italian security consultant ITSEC*" in Sept 2012. Extensible Firmware

Interface (EFI) security used by the MAC operating system against bootkit was broken in August 2012 by “*Louas K. BlackHat USA*”. From this literature survey it is evident that there is no detection mechanism for detection of bootkit on any operating system. Even on android operating system only avoidance methods have been suggested.

### 3 Motivation

Android is a modern mobile platform that was designed to be truly open. Android applications make use of advanced hardware and software, as well as local and served data, exposed through the platform. To protect such values, the platform offers an application environment that ensures the security of users, data, applications, the device, and the network. Securing an open platform requires robust security architecture and rigorous security programs. Android was designed with multi-layered security that provides the flexibility required for an open platform, while providing protection for all users of the platform. But still, it is not completely secure. Report given in [5] predicted that in the near future, android could be the target of 92 % of all detected mobile malware threats. This was a significant uptick from 2011 when android made up 47% of all detected threats from 24% in 2010. The increasing percentage of threat in Android operating system [5] was the motivation to take up the research in this paper.

Antivirus, which are generally used to detect any threat in system, gets activated only after the boot process is completed, since they use file system to detect the threat in the system. But Android bootkit cannot be detected by antivirus. As a consequence it needs a technique that gets started during the boot process of the Android to detect any Bootkit. This paper proposes one such solution for the detection of bootkits.

### 4 Proposed Solution

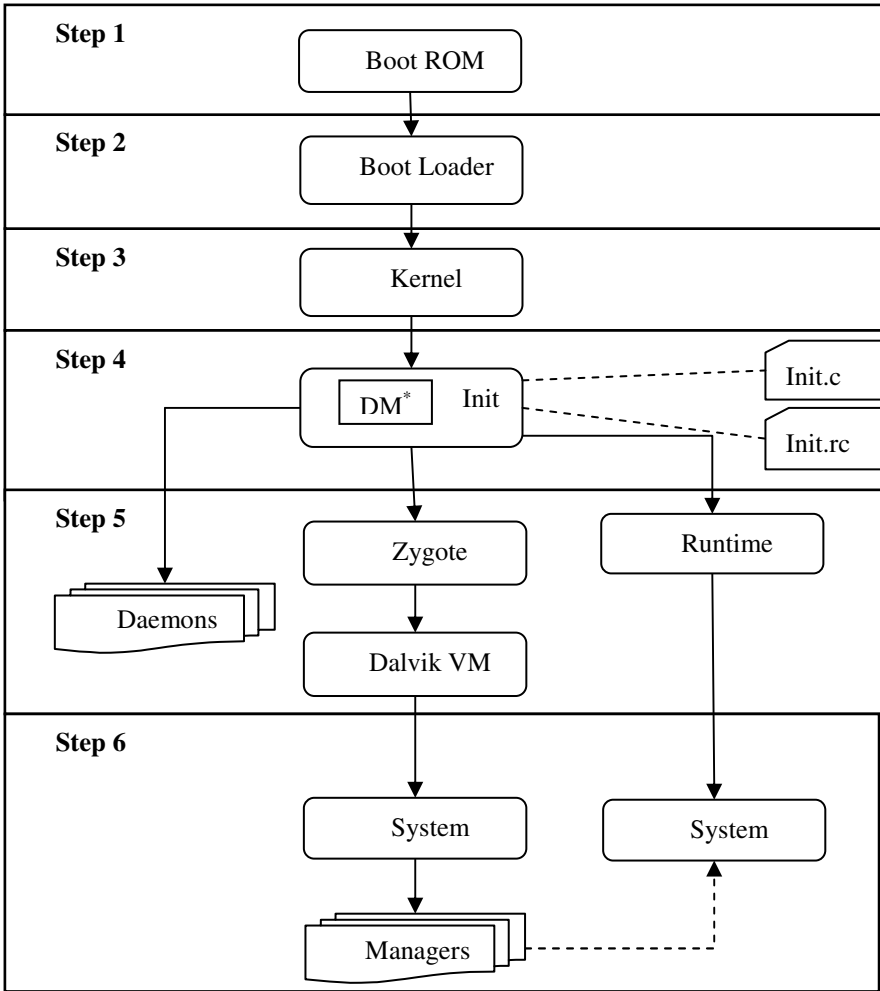
The proposed solution is incorporated in the init process which gets activated during the activation of init process during booting. The positioning of the proposed solution is shown in Figure 1. Figure 1 depicts the step by step process of booting in Android.

#### Step 1: Power Up and System Startup

When power supply of the system is on, the Boot ROM code starts to execute from pre defined location, which is hardwired on ROM. It loads the Boot loader into RAM and starts the execution.

#### Step 2: Bootloader

Bootloader is the first program to be run on the system. It detects external RAM and sets up network and memory. It will load kernel and store it on the RAM.



\*DM- Proposed Solution for detection of android bootkit

**Fig. 1.** Positioning of Proposed Detection module in android boot process

**Step 3: Kernel**

It will set up everything that is needed for the system to run, viz., initialize interrupt controllers, set up memory protections, caches, and scheduling. The “init” process is launched by the kernel.

**Step 4: Init Process**

The init process is the first process to start in system by the kernel. It is called mother of all processes since it creates and executes all processes that run on the Linux

system [6]. The init process is executed in the user space when the Linux kernel is booted. Then, the init process executes all processes that are required for subsequent system operations. Once the system booting is completed, the init process becomes daemon process, running in the background as it monitors other processes. When a process terminates and enters a zombie state, the init process returns the process resources to the system. The following are the four basic functions of init process:

1. Executing init.rc file.
2. Creating device driver.
3. Handling child process termination.
4. Property service.

Init.rc file is used by init process to start Step 5 and Step 6. Android init language is a special script language used by android operating system in its init.rc file [7]. Init.rc file is used by the init process for triggering services, mounting file system, initializing networking framework and properties assigned to the system. It has the following four main classes of statements:

1. Action
2. Commands
3. Services
4. Options

This init.rc file is used in attack generation of android bootkit. The proposed Detection module is made part of init process.

#### 4.1 Proposed Detection Module of Android Bootkit

The Android bootkit gets activated during the boot process, before the android framework for any application has been initialised. During the booting up of the system fixed number of processes is getting initialized. It has been practically verified that the number of processes getting initialized never changes. Their process id also never changes until some new process is included to boot up. This fact has been used in the proposed Android Bootkit detection module, in this paper.

The proposed detection module is designed to differentiate the legal and illegal processes. The algorithm for detection of Android bootkit is as follows:

1. Start
2. Initialize the variables and illegal process record
3. Open the directory
  - a. Read the directory
  - b. Check if the subdirectories in "/proc" directory are numbered
  - c. If it is alphanumeric or alphabetical then goto step 4

- i. Read the subdirectory name and store it as 'id' variable
- ii. Path of file to read is concatenation of "/proc/" + id + "/stat"
- iii. Read the file into local buffer
- iv. Get the process name and process id from the file
- v. Compare process name with database
- vi. If it is found in the record then goto step 4
- vii. Else send the process name to function defined in init.c called `handle_control_message()` .  
*//Illegal process is detected and killed.*
- viii. Check if all subdirectories are read
  1. If yes then goto step 5
  2. Else goto 4

4. End

Legal process record is collected by booting up the system 'n' number of times. After each boot of system, the process names were collected using unix command 'ps'. This record is initialised in the file `init.c` file of android system. `Init.c` file is used by the `init` process of android operating system. Also the detection module is called in `init.c` file when all the booting processes get initialised. When a system boots up some directories are created by `init` process. One such directory is `/proc`. The `"proc"` directory contains details of all the processes running on the system and even kernel related information. Specifically the numeric subdirectories in the `/proc` directory has the processes running on the system. These subdirectories are non-existent once the system is shutdown and get started during next boot up. It contains files with information related to a particular process. Using `"stat"` file, process names and process id are retrieved and compared with the record of legal processes. A process which is not found in the record is the illegal process. This illegal process is suspected as Android bootkit; it is detected and killed by the detection module. Killing of android bootkit is done by calling function named `"handle_control_message ()"`. The parameters to this function are message `"stop"` and illegal process id. This function kills the process, releases the resources, and notifies the `init` process. Thus android bootkit is killed during the booting process itself.

## 4.2 Working Environment

The detection module has to be checked for its working. The working environment was created following the steps given in [8], which is summarized in Table 1. The

detection module was tested by injecting a bootkit attack. The algorithm for bootkit attack is discussed in the Section 4.3.

**Table 1.** Working environment

Name	Version
Android platform version	Android 2.2.3_r2 (froyo)
Android hardware	Goldfish (emulator)
Base operating system	Ubuntu 10.03
Processor	Quad core processor

### 4.3 Attack Generation

The development of the Android bootkit was done using eclipse software [9] and Android software development kit (sdk). Android bootkit gets attached to legal application and it waits till that legal application gets a root system access. Application was given root system access using Superuser software. Generally root file directory has access permission as read only access. Once legal application gets an access, with the help of this access the bootkit changes the root file directory access permission rights as read, write, and execute which is done using the command 'chmod'. Then it searches for the file init.rc file which is used by init process and appends trigger line to end of the init.rc file so that android bootkit gets activated during the boot time. It attaches a script file to user accessible directory; this script file is used to launch the attack. Then it waits till next boot up of the operating system and it gets triggered during boot up process. The purpose of the attack is to damage file system and operating system. The algorithm for bootkit attack generation is as follows:

1. Start
2. Get attached to some legal app
3. Loop till root access is granted
4. If access to root directory is not permitted goto 11.
5. Else, Change the access permission of root directory to read, write, and executable.
6. If access permission is not granted then goto 3, else goto 7.
7. Attach the attack script file to user writable directory.
8. Append attack trigger lines to init.rc file

9. Wait till next boot
10. When system is booting
  - a. Launch the attack
11. End

#### 4.4 Experiments and Results

Android source code of Froyo was downloaded from Google repository, compiled, and built to deploy the detection module on base operating system Ubuntu version 10.03. Android App using Eclipse was installed and through this the bootkit attack was launched. The attack was designed purposefully not to harm the system but to run as a daemon process. During the next boot up of the emulator the detection module detected the bootkit successfully which was running in kernel mode and killed the process. The screenshots at different stages of experimentation were recorded and shown in [10].

## 5 Conclusion

It is evident from the literature of bootkit on different operating system that there is no prevention or detection technique available for android bootkit. Motivated by the non-existence of prevention or detection technique, in this paper, a solution has been proposed. The detection module was added to android boot process using android init process. The proposed detection module was tested by generating a bootkit attack. The bootkit attack was performed through android apps which escalated privileges to get the root privileges. The detection module has been found to detect android bootkit successfully and kill it.

Current research focuses on the prevention method. It should be designed to secure the booting procedure of the android operating system such that any random utility commands are not executed. Essentially it should avoid any android apps to get root access to system.

## References

1. Rootkits, Part 1 of 3: The growing threat: McAfee (2006)
2. Security Alert: New Android Malware — DKFBootKit — Moves Towards the First Android BootKit (March 29, 2012), <http://research.nq.com/>
3. UEFI architecture, <http://www.itsec.it/2012/09/18/uefi-technology-say-hello-to-the-windows-8-bootkit/#!prettyPhoto>
4. Kumar, N., Kumar, V.: Vbootkit: Compromising Windows Vista Security. In: Black Hat Europe 2007, Blackhat 2007 (2007)
5. 600% rise in malicious apps: Study. The Times of India (June 27, 2013)



6. Kim, T.Y., Song, H.J., Park, J.H., Lee, B., Lim, K.Y.: Android Anatomy - Episode1 – The Init Process: Special Edition (2011)
7. The Android Init Language, [https://android.googlesource.com/platform/system/core/+android-2.2.3\\_r2/init/readme.txt](https://android.googlesource.com/platform/system/core/+android-2.2.3_r2/init/readme.txt)
8. Android Open Source Project, <http://android.source.com/>
9. Android App Development, <http://developer.android.com/>
10. Rao, H., Selvakumar, S.: M. Tech. Phase II Thesis Report, Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli (May 2013)

# Optimizing CPU Scheduling for Real Time Applications Using Mean-Difference Round Robin (MDRR) Algorithm

R.N.D.S.S. Kiran, Polinati Vinod Babu, and B.B. Murali Krishna

Swarnandhra College of Engineering and Technology, Narsapur, India 534280  
{scetseta,vinodbabusir,bbmuralikrishna}@gmail.com

**Abstract.** This paper proposes a novel approach called Mean-Difference Round Robin Algorithm, which is meant for optimizing CPU scheduling for real time applications. The proposed algorithm calculates the mean burst time of all the processes in the ready queue. Next, it finds out the difference between a process burst time and calculated mean burst time. This step is repeated for all the processes in the ready queue. Then, the proposed algorithm find out the process having the largest difference value and assigns it to CPU, and execute it for one time slice. Once the time slice of the process expires, the next process with the largest difference value is picked up from the ready queue and executed for one time slice. The process is repeated for all the processes in the ready queue. The experimental results of the proposed algorithm have been compared with other standard scheduling algorithms and the proposed Mean-Difference Round Robin Algorithm is found to have produced optimum scheduling.

**Keywords:** CPU Scheduler, Burst Time, Waiting Time, Turnaround Time, Gantt chart, Pre-emptive Scheduling, non Pre-emptive Scheduling, Round Robin, MDRR.

## 1 Introduction

The need for a scheduling algorithm arises from the requirement for most modern systems to perform multitasking (execute more than one process at a time) and multiplexing (transmit multiple flows simultaneously).

CPU Scheduling is the basis of multi programmed operating systems. By switching the CPU among the processes, the operating system can make the computer system productive [1].

In single processor system, only one process can run at a time; any others must wait until the CPU is free and can be rescheduled. The objective of multiprogramming is to have some processes running at all times, to maximize CPU utilization. The idea is relatively simple. A process is executed until it must wait, typically for the completion of some I/O request. In a simple computer system, the CPU then just sits idle. All this waiting time is wasted; no useful work is accomplished. With multiprogramming, we try to use this time productively. Several processes are kept in memory at one time. When one process has to wait, the

operating system takes the CPU away from that process and gives to another process. This pattern continues. Every time one process has to wait, another process can take over use of the CPU [1].

Scheduling of this kind is a fundamental operating system function. Almost all computer resources are scheduled before use. The CPU is, of course, one of the primary computer resources. Thus, it's central to operating system design.

## 1.1 CPU Scheduler

Whenever the CPU becomes idle, the operating system must select one of the processes in the ready queue to be executed. The selection process is carried out by the short term scheduler or CPU Scheduler. The scheduler selects a process from the processes in memory that are ready to execute and allocate the CPU to that process [1].

## 1.2 Pre-emptive Scheduling and Non Pre-emptive Scheduling

CPU scheduling [1] decisions may take place under the following four circumstances:

1. When a process switches from the running state to the waiting state (for example, as the result of an I/O request of an invocation of wait for the termination of one of the child processes).
2. When a process switches from the running state to the ready state (for example, when an interrupt occurs).
3. When a process switches from the waiting state to the ready state (for example, at completion of I/O).
4. When a process terminates.

When scheduling takes place only under circumstances 1 and 4, we say that the scheduling scheme is non pre-emptive scheduling or cooperative; otherwise it is pre-emptive.

Scheduling Criteria [1]: Different CPU scheduling algorithms have different properties, and the choice of a particular algorithm may favour one class of processes over another. In choosing which algorithm to use in a particular situation, we must consider the properties of the various algorithms.

Many criteria have been suggested for comparing CPU Scheduling algorithms. Which characteristics are used for comparison can make a substantial difference in which algorithm is judged to be best. The criteria include the following:

1. CPU utilization: we want to keep the CPU as busy as possible. Conceptually, CPU utilization can range from 0 to 100 percent. In a real system, it should range from 40 percent (for a highly loaded system) to 90 percent (for a heavily loaded system).
2. Through put: if the CPU is busy executing processes, then the work is being done. One measure of work is the number of processes that are completed per time unit, called throughput.

3. Turnaround time: from the point of view of a particular process, the important criterion is how long it takes to execute that process. The interval time from the submission of a process to the time of completion is the turnaround time. Turnaround time is the sum of the periods spent waiting to get into memory, waiting in the ready queue, executing on the CPU, and doing I/O.
4. Waiting Time: The CPU scheduling algorithm does not affect the amount of time during which executes or does I/O; it affects only the amount of time that a process spends waiting in the ready queue. Waiting time is the sum of the periods spent waiting in the ready queue.
5. Response Time: In a  $n$  interactive system, turnaround time may not be the best criterion. Often, a process can produce some output fairly early and can continue computing new results while previous results are being output to the user. Thus another measure is the time from the submission of a request until the first response is produced. This measure, called response time, is the time it takes to start responding, not the time it takes to output the response.

### 1.3 Scheduling Objective [4]

A system designer must consider a variety of factors when developing a scheduling discipline, such as what type of system and what are user's needs. Depending on the system, the user and designer might expect the scheduler to:

1. Maximize throughput: a scheduling discipline should attempt to service the maximum number process per unit of time.
2. Avoid indefinite postponement and starvation: A process should not experience an unbounded wait time before or while process service.

## 2 Literature Survey

### 2.1 Round Robin Scheduling Algorithm [1]

Round Robin architecture is a preemptive version of first come first served scheduling algorithm. The processes are arranged in the ready queue in first come first served manner and the processor executes the process from the ready queue based on time slice. If the time slice ends and the process are still executing on the processor the scheduler will forcibly pre-empt the executing process and keeps it at the end of the ready queue then the scheduler will allocate the processor to the next process in the ready queue. The pre-empted process will make its way to the beginning of the ready queue and will be executed by the processor from the point of interruption.

A scheduler requires the time management function to implement the round robin architecture and require a tick timer. The time slice is proportional to the period of clock ticks. The time slice length is critical issue in soft real time embedded application as missing of deadlines will have negligible effects in the system

performance. The time slice must not be small which results in frequent context switches and should be slightly greater than average process computation time.

## 2.2 An Optimized Round Robin Scheduling Algorithm for CPU Scheduling [4]

The existing algorithm will be executed in three phase which help to minimize a number of performance parameters such as context switches, waiting time and average turnaround time. The algorithm performs following steps as:

Phase 1: Allocate every process to CPU, a single time by applying RR scheduling with a initial time quantum (say  $k$  units).

Phase 2: After completing first cycle perform the following steps: Double the initial time quantum ( $2k$  units). Select the shortest process from the waiting queue and assign to CPU.

After that we have to select the next shortest process for execution by excluding the already executed one in this phase.

Phase3: For the complete execution of all the processes we have to repeat phase 1 and cycle.

## 3 Methodology

This section discusses the methodology we adapted to develop MDRR Algorithm. The proposed algorithm calculates the mean burst time of all the processes in the ready queue. Next, it finds out the difference between a process burst times and the calculated mean burst time. This step is repeated for all the processes in the ready queue. Then, the proposed algorithm find out the process having the largest difference value and assigns it to CPU, and execute it for one time slice. Once the time slice of the process expires, the next process with the largest difference value is picked up from the ready queue and executed for one time slice. The process is repeated for all the processes in the ready queue. The experimental results of the proposed algorithm have been compared with other standard scheduling algorithms and the proposed Mean-Difference Round Robin Algorithm is found to have produced optimum scheduling. The pseudo code of the proposed algorithm is given below:

**Algorithm: Mean-Difference Round Robin**

**Input:**

*Bt [] is the array of burst times of all process in the ready queue.*

*Pid [] is the array of processes identifiers in the ready queue.*

*N is the total number of process in the ready queue.*

*$\delta$  is the time quantum.*

*Tcxsw is the context switch time.*

**Output:**

*Average waiting time, AWT, of all processes.*

*Average turnaround time, TAT, of all processes.*

*Context switch overhead and CPU efficiency or CPU Cycle time.*

**Method:**

**Step 1:** calculate mean burst time for each process in the ready queue.

$Sum =+ Bt [i];$  // sum is the sum of burst times of all the processes in the ready queue

//  $Bt [i]$  is the burst time of the  $i$ th process in the ready queue

$m = sum/N;$  //  $m$  is the mean of burst times of all the processes

**Step 2:** repeat for each process in the ready queue {

$S [] =m - Bt [i];$  //  $s[]$  is the array of mean differences

}

**Step 3:** Find out the process having the largest difference value and assign it to CPU and execute it for one time slice.

Find  $S \max []$  and  $Pid [];$

//  $S \max$  is the process with the largest difference value,  $Pid []$  is the process identifier of the process.

$Pid [], Bt [i] = Bt [i] - \delta;$

**Step 4:** After finishing Step 3, pick up the next process having the largest difference value from the ready queue and execute it for one time slice.

Find  $Smax-1 []$  and  $Pid[];$

//  $Smax-1$  is the process with the next largest difference value,  $Pid []$  is the process identifier of the process.

$Pid [], Bt [i] = Bt [i] - \delta;$

**Step 5:** Repeat Step 4 for all the processes in the ready queue.

**Step 6:** Repeat Step 1 through Step 5 until all the processes in the ready queue are finished.

**Step 7:** Compute average waiting time of all the processes in the ready queue using the formula

Average Waiting Time (WT) =  $\sum Wi / N$ , where  $Wi$  is the waiting time of  $i$ th process in the ready queue and  $N$  is the total number of processes in the queue.

**Step 8:** Compute average turnaround time of all the processes in the ready queue using the formula.

Average Turnaround Time (TAT) =  $\sum Ti / N$ , where  $Ti$  is the waiting time of  $i$ th process in the ready queue and  $N$  is the total number of processes in the queue.

**Step9:** compute the context switch overhead using the formula

Context switch overhead =  $(Tcxsw * 100) / \delta$

**Step 10:** compute the CPU efficiency or CPU Cycle time using the formula

CPU efficiency =  $\delta / (\delta + Tcxsw)$

In the next section, we discuss the results produced by our proposed Mean-Difference Round Robin algorithm and the standard Round Robin algorithm with the help of two examples:

## 4 Results and Discussion

**Example 1:** Consider the following set of processes, assumed to have arrived at time  $t_0$ , in the order  $p_1, p_2, p_3, p_4$  with the length of the CPU burst given in milliseconds and time quantum ( $\delta$ ) is 10, 15, 20, 25 milliseconds.

**Table 1.** Processes with their burst times in milliseconds

Process Name	Burst Time
p1	53
P2	17
p3	68
p4	24

Average waiting time (AWT) can be calculated using the following formula

$$AWT = \sum Wi / N . \tag{1}$$

Where  $W_i$  is the sum of waiting time of all the processes and  $N$  is the total number of processes.

Average Turnaround Time (ATT) can be calculated using the following formula

$$ATT = \sum T_i / N . \tag{2}$$

Where  $T_i$  is the Turnaround Time of all the processes and  $N$  is the total number of processes.

Scheduling overhead what we call the CPU time spent in making a scheduling decision; we denote it by  $\sigma$  (the Greek letter sigma), Where  $\sigma$  is the scheduling overhead per scheduling decision [3].

Context switch overhead can be calculated using the following formula

$$\text{Scheduling overhead } (\sigma) = (T_{csw} * 100) / \delta . \tag{3}$$

CPU efficiency can be calculated using the following formula

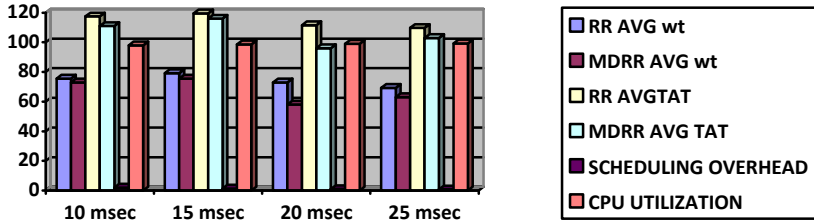
$$\text{CPU efficiency} = \delta / (\delta + T_{csw}) . \tag{4}$$

The following table shows the comparative study of the performance of the standard Round Robin and the proposed Mean-Difference Round Robin Scheduling Algorithm.

**Table 2.** Comparison of the performance of the standard Round Robin and the proposed Mean-Difference Round Robin Scheduling Algorithm

TIME SLICE ( $\delta$ )	10 msec	15 msec	20 msec	25 msec
RR AVG WT	75.5	79.25	73	69.25
MDRR AVG WT	73	75.5	58	63
RR AVGTAT	117.5	119.5	111.5	109.75
MDRR AVG TAT	111	116	96	103
SCHEDULING OVERHEAD	2	1.33	1	0.8
CPU UTILIZATION	98	98.67	99	99.2

The comparative study of the performance of the standard Round Robin and the proposed Mean-Difference Round Robin Scheduling Algorithm is represented in the form of a graph as follows:



**Example 2:** consider the following set of processes, assumed to have arrived at time  $t_0$ , in the order  $p_1, p_2, p_3, p_4, p_5$  with the length of the CPU burst given in milliseconds and time quantum ( $\delta$ ), is 10, 14, 16, 18 milliseconds.

**Table 3.** Processes with their burst times in milliseconds

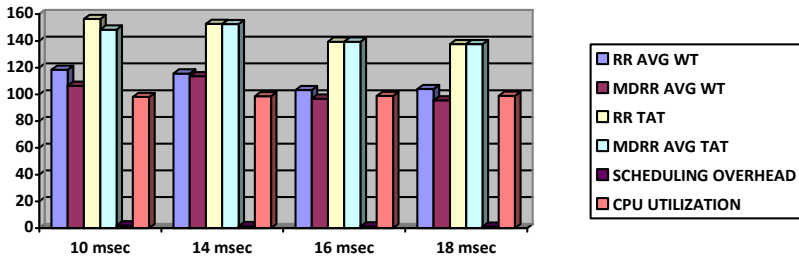
Process Name	Burst Time
P1	48
P2	26
P3	54
P4	16
P5	66

The following table shows the comparative study of the performance of the standard Round Robin and the proposed Mean-Difference Round Robin Scheduling Algorithm.

**Table 4.** Comparison of the performance of the standard Round Robin and the proposed Mean-Difference Round Robin Scheduling Algorithm

TIME SLICE ( $\delta$ )	10 msec	14 msec	16 msec	18 msec
RR AVG WT	118.4	115.6	103.2	104
MDRR AVG WT	106.4	113.6	96.8	95.6
RR TAT	156.4	152.8	139.4	137.6
MDRR AVG TAT	148.4	152.6	139.4	137.6
SCHEDULING OVERHEAD	2	1.4	1.25	1.11
CPU UTILIZATION	98	98.6	98.75	98.89

The comparative study of the performance of the standard Round Robin and the proposed Mean-Difference Round Robin Scheduling Algorithm is represented in the form of a graph as follows:





Effectiveness of the MDRR scheduling algorithm depends on two factors: choice of  $\delta$ , the time slice, and nature of processes in the system, if a system contains  $n$  processes and each request by a process can consumes exactly  $\delta$  seconds, the response time ( $rt$ ) for a request is

$$rt = n * (\bar{C} + \delta) \quad (5)$$

Where  $\bar{C}$  is the scheduling overhead, which can be computed using the equation (3). From equation (3) we can understand that as the value of  $\delta$  increases, the value of  $\bar{C}$  and vice versa. In other words, the scheduling overhead ( $\bar{C}$ ) is inversely proportional to the value of time slice ( $\delta$ ). Therefore we must choose an optimum value for  $\delta$  so that the scheduling overhead can be minimised reasonably. CPU efficiency can be computed using the equation (4) it can be understood that as the value of  $\bar{C}$  decreases the CPU efficiency increases for a given  $\delta$  value.

## 5 Conclusions

In this paper we proposed a novel CPU Scheduling algorithm called Mean-Difference Round Robin Algorithm, which is meant for optimizing CPU scheduling for real time applications. The proposed algorithm calculates the average burst time of all the processes in the ready queue. Next, it finds out the difference between a process burst time and the calculated average burst time. This step is repeated for all the processes in the ready queue. Then, the proposed algorithm find out the process having the largest difference value and assigns it to CPU, and execute it for one time slice. Once the time slice of the process expires, the next process with the largest difference value is picked up from the ready queue and executed for one time slice. The process is repeated for all the processes in the ready queue. The performance of the proposed algorithm with the Standard Round Robin algorithm with the help of a number of examples, of which two are presented in this paper. The experimental results of the proposed algorithm have been compared with other standard scheduling algorithms and the proposed Mean-Difference Round Robin Algorithm is found to have produced more optimum scheduling. The proposed algorithm can also be used for packet scheduling in the network applications and for scheduling jobs in embedded systems.

## References

1. Silberschatz, A., Galvin, P.B., Gagne, G.: Operating system principles, 7th edn.
2. Shibu, K.: Introduction to Embedded Systems. THM (2009)
3. Dhamdhare, D.M.: Operating Systems A concept-based approach. Tata McGraw Hill
4. Sinth, A., Goyal, P., Batra, S.: An optimized Round Robin Scheduling Algorithm for CPU Scheduling. IJCSSE 02(07), 2383–2385 (2010)
5. Yaashuwanth, C., Ramesh, R.: A New Scheduling Algorithm. IJCSIS 6(2) (2009)

6. Noon, A., Kalakecch, A., Kadry, S.: A New Round Robin Based Scheduling Algorithm for Operating Systems: Dynamic Quantum Using the Mean Average. *IJCSI* 8(3(1)) (May 2011)
7. Matarneh, R.J.: Self-Adjustment Time Quantum in Round Robin Algorithm Depending on Burst Time of the Now Running Processes. *American Journal of Applied Sciences* 6(10), 1831–1837 (2009) ISSN 1546-9239
8. Hiranwal, S., Roy, K.C.: Adaptive Round Robin scheduling using shortest burst approach, based on smart time slice. *International Journal of computer Science and Communication* 2(2), 219–326 (2011)
9. Matarneh, R.J.: Self-Adjustment Time Quantum in Round Robin Algorithm Depending on Burst Time of the now Running Processes. *American Journal of Applied Sciences* 6(10), 1831–1837 (2009)

# A Study on Vectorization Methods for Multicore SIMD Architecture Provided by Compilers

Davendar Kumar Ojha and Geeta Sikka

Department of Computer Science and Engineering,  
Dr. B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab  
{ojha.dev, sikkag}@gmail.com

**Abstract.** SIMD vectorization has received important attention within the last few years as a vital technique to accelerate multimedia, scientific applications and embedded applications on SIMD architectures. SIMD has extensive applications; though the majority and focus has been on multimedia. As a result of it is an area of computing that desires the maximum amount of computing power as possible, and in most of the cases, it is necessary to compute plenty of data at one go. This makes it an honest candidate for parallelization. There are many compiler frameworks which allow vectorization such as Intel ICC, GNU GCC and LLVM etc. In this paper, we will discuss about GNU GCC and LLVM compilers, optimization methods, vectorization methods and evaluate the impact of various vectorization methods supported by these compilers and at last note we will discuss about the methods to enhance the vectorization process.

**Keywords:** Intel ICC, GNU GCC, LLVM, SIMD, Vectorization.

## 1 Introduction

There has been a rich body of compiler research and development on exploiting SIMD parallelism for modern CPU and GPU cores with rich and powerful SIMD hardware support [1-5] through compiler auto-vectorization. To efficiently solve the problem low degree of SIMD vectorization caused by the lack of vector registers in the SIMD mathematical function library and the lower speed of memory access, an optimizing method based on vector register data reuse was introduced aiming at relative address access mode [1]. Driven by the increasing prevalence of SIMD architectures in modern CPU and GPU processors, an approach suggested for C and C++ languages, is to provide a thin abstraction layer between the programmer and the hardware with a small set of high-level SIMD vector extensions to, which the programmer can use to cleanly express SIMD parallelism to harness the computational power of SIMD vector units without the low programmer productivity of directly writing in SIMD intrinsic or inline ASM code, whenever the automatic vectorization fails [6].

An application which will make the most of SIMD is one wherever the same value is being added to an oversized number of data points that is a standard operation in several multimedia applications. With a SIMD processor there are two enhancements

to the present process. The primary one is, the data is known to be in blocks, and a number of values may be loaded all at one go, rather than a series of instructions, SIMD processor can have a single instruction. For a number of reasons, this will take abundant less time than traditional hardware design. Another advantage is that SIMD systems usually include only those instructions which will be applied to any or all of the data in one operation. That means, if the SIMD system works by loading up eight data points along at one go, the add operation being applied to the data can happen to all eight values at the exact same time. Current SIMD architectures such as - Intel's MMX, SSE, and SSE2, SSE3, SSE4, AMD's 3dNow, and Motorola's AltiVec provide the x86 architecture with a widened 128-bit data path, which enhances the computation capability of processing unit. Intel SSE allows up to four floating point operations for consecutive memory data while latest Intel's AVX provides a widened 256-bit data path and up to eight floating point operations depending upon data dependencies and consecutive memory. So how to harness the best from such platforms which provides such level of parallelism? The answer may be generating such executables which directly target these platforms. Generation of an executable is mostly done by compilers so we have to tune our compiler architecture so that we can take advantage of these platforms.

The remainder of the paper is organized as follows. In Section 2, we give an overview of optimizations using vectorization techniques supported by the latest compilers. Next, in Section 3, we introduce the applications used as benchmarks for this study and analyze the performance results. In section 4 we will discuss about some vectorization problems which are still to be fixed. In Section 5, we summarize our findings and conclude the paper.

## 2 SIMD Vectorization Techniques

### 2.1 Basic Block and Loop Vectorization

Basically there are two vectorizers in LLVM; first is the Loop Vectorizer, which operates on Loops, and second is the Basic Block Vectorizer, through which straight-line code is optimized. These vectorizers target different optimization opportunities and use different techniques. The Basic Block vectorizer merges multiple scalars that are found in the code into vectors while the Loop Vectorizer widens instructions in the original loop to operate on multiple consecutive loop iterations. It can be enabled through clang using the command line flag "-mllvm vectorize-loops" as it is not enabled by default. Performance of Loops that are not even vectorizable by GCC can be boost using The Loop Vectorizer including many other loops. Following code is a simple example of a loop which can be vectorize by the LLVM Loop Vectorizer.

```
int foo(int *X, int *Y, int n) {
unsigned temp = 0;
    for (int i = 0; i < n; ++i)
        if (X[i] > Y[i])
            temp += Y[i] + 5;
return temp;
}
```

In this example, the Loop Vectorizer uses a number of non-trivial features to vectorize the loop and executes the last few iterations as scalar code thus it increases the code size, as 'n' may not be a multiple of vector width. The variable 'temp' is used by consecutive iterations of the loop. This could prevent vectorization, but the vectorizer can detect that 'sum' is a reduction variable. The variable 'sum' becomes a vector of integers; the elements of the array are added together at the end of the loop to create the correct result. Another challenge that the Loop Vectorizer needs to overcome is the presence of control flow in the loop. The Loop Vectorizer is able to "flatten" the IF statement in the code and generate a single stream of instructions. Vectorization of loops with an unknown trip count is another important feature. The Loop Vectorizer is a target independent IR-level optimization that depends on target-specific information from the different backends. Optimal vector width needs to be selected and it has to decide if vectorization is worthwhile. Certain vector width can be forced using the command line flag "-mllvm -force-vector-width=X". In this command X stand for vector elements, this command uses accurate method for vectors, while other targets use a less accurate method only the X86 backend provides detailed cost information at this moment.

The SLP vectorizer (superword-level parallelism) is a new vectorization pass. SLP vectorizer is different from loop vectorizer, in terms of vectorizing consecutive loop iterations is done by loop vectorizer, where the SLP vectorizer combines similar independent instructions in a straight-line code. The goal of SLP vectorization is to combine similar independent instructions within simple control-flow regions into vector instructions. Using this technique we can vectorize all comparison operations, memory accesses, arithmetic operations and some math functions. The SLP-vectorizer has two phases, bottom-up, and top-down. The top-down vectorization phase is more aggressive, but takes more time to run.

### 3 Benchmarking and Results

We have used Phoronix test suit for benchmarking latest compilers as LLVM Clang 3.2 and GCC 4.7.2. This test suit was originally developed for automated Linux testing, support for many operating systems including Microsoft Windows has been added to test suit. The Phoronix Test Suite consists of pts-core (lightweight processing core) with each benchmark consisting of related resource scripts and XML-based profile. The process is completely repeatable and fully automated from the actual benchmarking, to the parsing of important hardware and software components. We have used LLVM Clang and GCC as our compiler system as both provides broad range of optimizations. The performance measurement is carried out on an Intel® Core™ i5-2450M CPU system, running at 2.50GHz, with 6.0GB RAM, 3M smart cache, 64-bit Windows 7 Home Premium. We have used Mflops and total time of execution as measurement parameters, where greater value of Mflops indicates better use of processor cycle and lesser execution time indicates better target to specific problem.

A compilers usability or modularity cannot be determined only by a single application or the size of generated binary, we have to test it under various applications and various workloads. We have used applications of different areas so that we can target the different optimization options provided by these compilers. These application includes some basic mathematical structures, multimedia extensions, cryptography algorithms and some graphic rendering applications. We have run our simulation over different optimization options and determined the best value provided in all runs with all different optional flags. These results indicate that current compilers are good but not good enough for current SIMD platforms there are some functionalities missing and some needs to be improved. We will discuss about these missing functionalities in next Section.

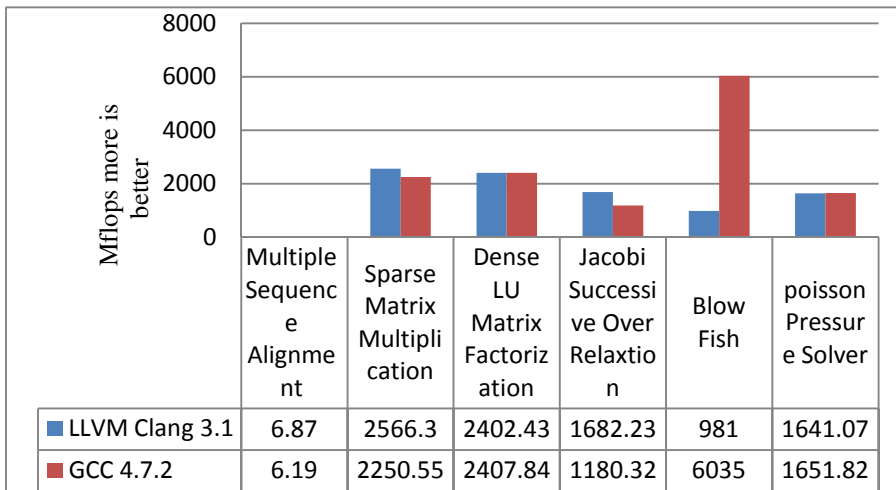


Fig. 1. Mflops Count For LLVM and GCC for various applications

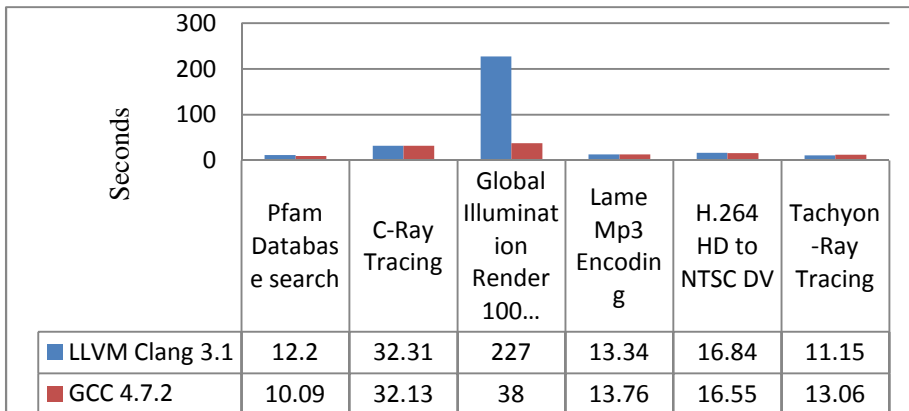


Fig. 2. Total Time of Execution for LLVM and GCC for various applications

## 4 Vectorizer Enhancements

Now a day vectorization is major area of research in compilers. Several problems and major issues has been targeted but still there are some major issues which have to be implemented yet by which the basic vectorizer can be enhanced. An issue that repeatedly comes up during the development of the vectorization is the tension between two conflicting needs. One is the requirement to maintain a high-level, platform-independent program representation. The other is the need to consider platform specific issues and express low-level constructs during the process of vectorization. There are some missing functionalities which has to be addressed are:

1. Vector capabilities to the tree-level for target machine: expose the required target specific information to the tree level. A mapping from scalar operations to the corresponding vector support is also needed for tree level. Introduce new tree-codes (and corresponding optabs); introduce new built-ins that are exposed to the compiler; use target hooks to handle these cases (the hook could return a call to a machine specific built-in function). Also, we need to consider, how much information to expose to the tree-level.
2. Enhance the Built-ins Support: Current compiler systems which are using tree optimizers, because of lack of knowledge the semantics of target specific built-in functions, do not attempt to optimize them. Somehow we should expose the semantics of these built-ins to the compiler.
3. Cost Model: It is required to describe a cost model in order to allow the vectorizer to evaluate whether it is worth to vectorize a given loop. Also can have run time tests to decide which version of the loop to execute (scalar or vectorized).
4. Alignment: One of the areas to improvise is that data accesses need to be properly aligned on a certain boundary. This consists of several stages:
  5. Each memory access has misalignment properties and it needs to be computed.
  6. Build a cost model to decide when to apply loop-peeling and/or loop-versioning to force alignment, and when to generate unaligned vector accesses.
  7. Handle Pointer Aliasing: Create run-time tests for cases where memory anti-aliasing cannot be resolved at compile time. Aliasing and Virtual def-use Chains:- Address the item from the tree-SSA "SSA information for arrays".
8. Specifically for LLVM loop vectorizer need to add additional vectorization features like support for user pragmas, vectorization of function calls, automatic alignment of buffers, and the quality of the generated code to be improved.

## 5 Summary and Conclusion

Throughout our simulation we found that LLVM/Clang 3.2 performed quite well against GCC 4.7. The LLVM Clang leads in a number of the computational benchmarks, but GCC also had several leads. Tests where LLVM Clang wasn't leading were those where applications take benefit of OpenMP, in a majority of the cases it was generally competitive aside from the notable lack of OpenMP support.

Beyond the performance of the generated executable binaries, both compilers have their own set of features and abilities such as debugging and error messages, tuning switches, and other differences that have the potential to impact both developers and end-users of the compiled binaries.

## References

1. Xu, J., Shaozhong, G., Lei, W.: Optimization Technology in SIMD Mathematical Functions Based on Vector Register Reuse. In: 2012 IEEE 14th International Conference on High Performance Computing and Communications, pp. 1102–1107 (2012)
2. Peter, K., Yu, H., Li, Z., Tian, X.: Performance Study of SIMD Programming Models on Intel Multicore Processors. In: 2012 IEEE 26th International Symposium Workshops & PhD Forum on Parallel and Distributed Processing, pp. 2423–2432 (2012)
3. Lee, C.-Y., Chang, J.-C., Chang, R.-G.: Compiler Optimization to Reduce Cache Power with Victim Cache. In: 2012 IEEE 9th International Conference on Ubiquitous Intelligence & Computing and 2012 9th International Conference on Autonomic & Trusted Computing, pp. 841–844 (2012)
4. Stefan, H., Schoeberl, M.: Worst-Case Execution Time Based Optimization of Real-Time Java Programs. In: 2012 IEEE 15th International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing, pp. 64–70 (2012)
5. Desai, N.P.: A Novel Technique for Orchestration of Compiler Optimization Functions Using Branch and Bound Strategy. In: 2009 IEEE International Conference on Advance Computing, pp. 467–472 (2009)
6. Tian, X., Saito, H., Girkar, M., Preis, S.V., Kozhukhov, S.S., Cherkasov, A.G., Nelson, C., Panchenko, N., Geva, R.: Compiling C/C++ SIMD Extensions for Function and Loop Vectorization on Multicore-SIMD Processors. In: 2012 IEEE 26th International Symposium Workshops & PhD Forum on Parallel and Distributed Processing, pp. 2349–2358 (2012)
7. Aho, A.V., Ullman, J.D.: Principles of Compiler Design (Addison-Wesley series in computer science and information processing). Addison-Wesley Longman Publishing Co., Inc. (1977)
8. Lattner, C.A.: LLVM: An infrastructure for multi-stage optimization. PhD dissertation. University of Illinois (2002)
9. GCC optimization options, <http://gcc.gnu.org/onlinedocs/gcc/Optimize-Options.html>
10. LLVM optimization and passes options, <http://llvm.org/docs/Passes.html>
11. Amit, B., Joshi, B.K.: A parallel lexical analyzer for multi-core machines. In: 2012 IEEE Sixth International Conference on Software Engineering, pp. 1–3 (2012)
12. Qawasmeh, A., Chapman, B., Banerjee, A.: A Compiler-Based Tool for Array Analysis in HPC Applications. In: IEEE 41st International Conference Parallel Processing Workshops, pp. 454–463 (2012)



# A Clustering Analysis for Heart Failure Alert System Using RFID and GPS

Gudikandhula Narasimha Rao<sup>1</sup> and P. Jagdeeswar Rao<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering,  
KKR & KSR Institute of Technology & Sciences, A.P., India

<sup>2</sup> Department of Geo-Engineering, Andhra University College of Engineering,  
Visakhapatnam, A.P., India

narasimha66@yahoo.co.in, pjr\_geoin@rediffmail.com

**Abstract.** The “Heart Failure Alert System” has become one of the key emerging businesses in the World. In this paper, a mobile health management system is presented which is the first one integrating a chip in hand, chip is a pulse monitoring sensor with a smart phone. All physiological measurements are transmitted to the smart phone through Bluetooth. The user can monitor his/her own pulse and temperature from the smart phone. Then these data are transmitted to a remote server through the mobile communication of the smart phone, such as HSDPA, Wi-Fi, WiMax, GPRS, etc. The build-in GPS further provides the position information of the monitored person. The remote server not only collects physiological measurements but also tracks the position of the monitored person in real time. Finally this paper estimates classification, which age group of persons may caused by heart attack.

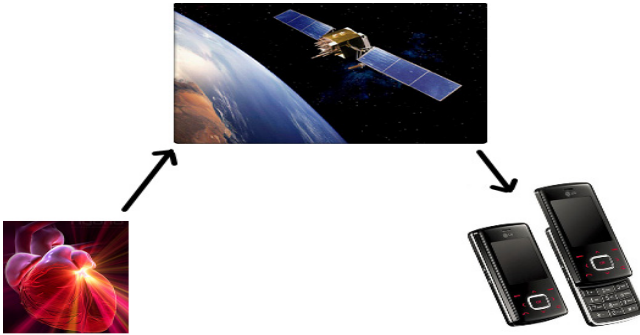
**Keywords:** Classification, Smart phone, Bluetooth, RFID, GPS, HFAS.

## 1 Introduction

It is thought to declare convincingly what is the most important organ of our body. In fact every organ has its own importance contributing and coordinating superbly to keep the wonderful machine the human body functioning smoothly. And one of the primary organs which the body cannot do without is the heart, 72 beats a minute or over a trillion in a lifetime[1]. The pump house of our body pumping the blood to every corner of our body every moment, thus sending oxygen and nutrients to each and every cell[2]. Over a period of time, the heart muscles go weak, the arteries get blocked and sometimes because of a shock a part of the heart stops functioning resulting in what is called a HEART ATTACK. Heart attack is a major cause of death and in today’s tension full world it has become very common. Presently there is no mechanism by which a device monitors a person’s heart 24 hours a day, 7 days a week and gives him instant protection in case of problem[4].

Our primary focus is on people with a history of heart problem as they are more prone to death due to heart failure. In the 1970s, a group of scientists at the Lawrence

Livermore Laboratory (LLL) realized that a handheld receiver stimulated by RF power could send back a coded radio signal[3]. Such a system could be connected to a simple computer and used to control access to a secure facility This system ultimately became one of the first building entry systems based on the first commercial use of RFID. RFID or Radio Frequency identification is a technology that enables the tracking or identification of objects using IC based tags with an RF circuit and antenna, and RF readers that "read" and in some case modify the information stored in the IC memory[5].



**Fig. 1.** Global Structure of Heart Failure Alert System

This paper comprises three major contributions (1) Global structure of Heart Failure Alert System. (2) Working process of HFAS with mobile alert algorithm and performance evolution of RFID and GPS. (3) Clustering analysis of HFAS and to provide patients information from GPS to medical station then medical station to mobile. This paper is organized in the following way: section 2 describes how the HFAS is achieved. Section 3 explains about the analysis of the RFID and GPS. Section 4 discusses the simulation results. Lastly we will conclude and assert the future scope for this work in section 5.

## 2 Related Works

The hardware of the presented consists of a RFID-based ring-type pulse/temperature sensor, a Bluetooth module, and a smart phone. Physiological Sensor, although there is a ring-type pulse monitoring sensor in the market, the measured data are displayed in the LCD and cannot be transmitted out of the ring. In this paper, a RFID wearable ring-type sensor designed by Sino pulsar Technology Inc., Taiwan was adopted, instead. Bluetooth, the data communication between RFID reader and the smart phone is through Bluetooth. HL-MD08A (Bluetooth RS232 Adaptor manufactured by Hot life Technology) is used in the presented system. It supports a wide range of Baud rates from 1.2K to 921.6K bps. Smart Phone, any smart phone which operating system is Windows Mobile 6.1 is suitable for the presented system. The smart phone used in this system is ASUS P552W with built-in GPS. It supports HSDPA 3.6Mbps/EDGE/GPRS/GSM 900/1800/1900.

## 2.1 RFID

RFID is an automated data-capture technology that can be used to electronically identify, track, and store information about groups of products, individual items, or product components. The technology consists of three key pieces:

1. RFID TAGS
2. RFID READERS
3. HOST COMPUTER

RFID tags are small or miniaturized computer chips programmed with information about a product or with a number that corresponds to information that is stored in a databases. RFID readers are querying systems that interrogate or send signals to the tags and receive the responses.

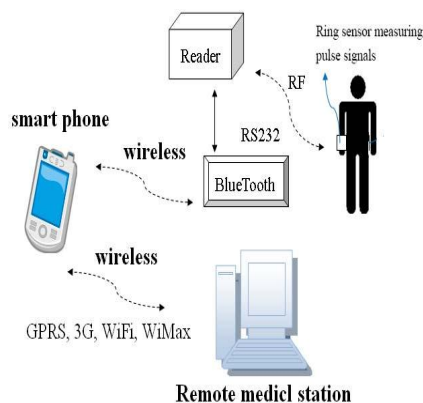
## 2.2 General Model for Heart Failure Alert System

It consists of

- RFID Tag (Implanted into Human body).
- RFID Reader (Placed in a Cellular Phone).
- Global Positioning Satellite System and locating and tracking station
- Mobile Rescue Units.

## 2.3 Working of Heart Failure Alert System

The grain-sized RFID Tag is implanted into the human body, which keeps track of the heart pulse in the form of Voltage levels. A RFID Reader is placed into the Cellular Phone. The RFID Reader sends a Command to the RFID Tag which in turn sends these Voltage pulses in the form of bits using the Embedded Software in the Tag as Response which is a continuous process[6]. These bit sequence is then sent to Software Program in the Cellular Phone as input and checks for the Condition of



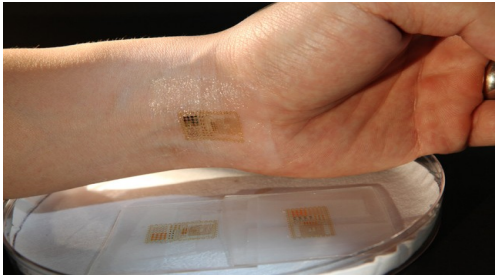
**Fig. 2.** Architecture for Heart Failure Alert System

Heart Failure. If any sign of Failure is sensed then immediately an ALERT Signal will be generated and in turn results in the AUTODIALING to the Locating & Tracking Station. This station with the use of GPS System comes to know the Where abouts of the Victim. The Locating & Tracking Station also simultaneously alerts the Rescue Units[5]

### 3 Analysis of Heart Failure Alert System

#### 3.1 Working of the Implanted TAGS

Passive RFID systems typically couple the transmitter to the receiver with either load modulation or backscatter, depending on whether the tags are operating in the near or far field of the reader, respectively[6].



**Fig. 3.** Grain sized RFID Tag

The reader communicates with the tag by modulating a carrier wave, which it does by varying the amplitude, phase, or frequency of the carrier, depending on the design of the RFID system in question. The tag communicates with the reader by varying how much it loads its antenna[7]. This in turn affects the voltage across the reader's antenna. By switching the load on and off rapidly, the tag can establish its own carrier frequency (really a sub carrier) that the tag can in turn modulate to communicate its reply.

#### **Algorithm:**

Step1: Read the Analog Signals from the Heart.

Step2: Sample the Analog Signal and generate series of pulses based on the results of Sampling based on the Tag Frequency.

Step3: Assign Integral Values to each Sampled Instances generated.

Step4: Consider every Individual Sampled Unit and Compare with the Average Voltage Level of the Heart.

Step5: If the Sampled Instance Value is in between the avg\_pulse Values

Then Assign BIT=0 Otherwise Assign BIT=1

Step6: Generate the bit sequence by considering all the generated Individual Sample Instances.

**Working of RFID inside Cellular Phone**

The RFID reader sends a pulse of radio energy to the tag and listens for the Tag’s response. The tag detects this energy and sends back a response that contains the tag’s serial number and possibly other information as well.

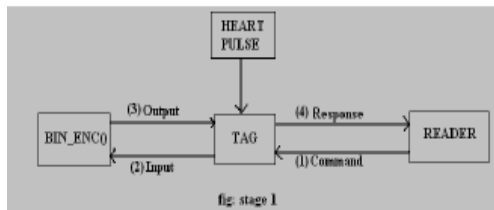
The Reader continuously sends the Command to the tags and in turn receives the Voltage levels in the form of bit sequence as Response from the tags with the help of the BIN\_ENC algorithm.

**Algorithm ALERT**

- Step 1: Read the bit sequence from the Reader.
- Step 2: Count for the no of bit zeros in the data using a counter.
- Step 3: If you encounter a bit one, then set counter to zero.
- Step 4: If the counter is equal to five then go to Step 5 else go to Step 1.
- Step 5: Send alert to the nearest Locating & Tracking Station.

**Stages in Heart Failure Alert System**

Stage 1:



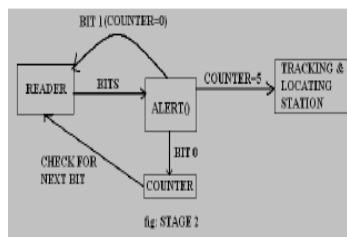
The Tag continuously senses the Heart Pulses, when the Reader sends a Command it sends the output of the BIN\_ENC () as the Response to the Reader.

/\*Module for the Conversion of Analog Signals to Binary digits\*/

```

BIN_ENC()
{
  scanf ("The Value of the generated Sample %f", Value);
  (+Avg_pulse<Value<-Ang_pulse) {Bit=0;}
  else if (Value>+Avg_pulse || Value<-Avg_pulse) {Bit=1 ;}
}
    
```

Stage 2:



ALERT() program to check whether the bit is 'BIT 0' or 'BIT 1'. If a 'BIT 0' is encountered, the counter is incremented and again it checks for the next bit. If a 'BIT 1' is encountered then counter is set to zero and it again checks for the next bit. If counter=5 then it alerts the Locating & Tracking Station[9].

/\*Module for checking the Weak Pulse \*/

```

ALERT ( )
{
    if (bit==0) {counter++; }
    else {counter=0 ;}
    if (counter==5)
    { Printf ("Report 'Weak Pulse Detected' to Locating & Tracking System");
      Counter=0;}}
    
```

The following figure shows the interactions among the smart phone, RFID reader, and ring Tag for those commands used on smart phone. For instance, the SearchTag command instructs the RFID reader (action 1) to search for available ring Tag

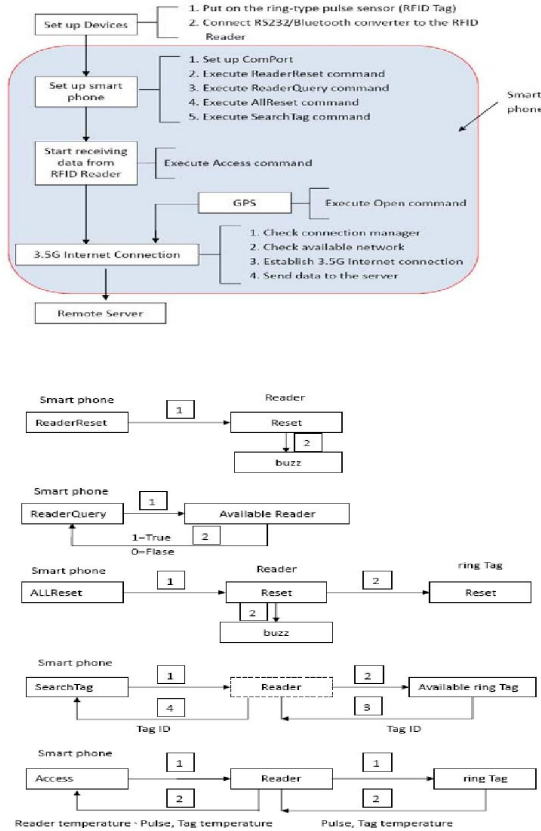


Fig. 4. Interactions among smart phone, Reader, and Tag

## 4 Simulation and Results

The hardware of the presented HFAS is implemented as Fig. 4. The Bluetooth adaptor is connected to the RFID reader and the ring Tag is worn on the user's finger.

- Step 1: On the main page, click the "Menu," go to "Setting," and then choose "RFID setting." RS232 COM port setup page will pop up.
- Step 2: Choose the correct COM port and click "OPEN." Then, click "BACK" to return to the main page.
- Step 3: Click "Reader" and "Reader setup" page will pop up. Click "Reader Reset" to reset the RFID reader, and the RFID reader beeps and "True" is shown on the message box. Then click "Reader Query" to search for the Reader's address. The message box of Reader Query shows the address number.
- Step 4: Click Tag and the "Tag setup" page will pop up. First click "ALL Reset" and turn on the power of the ring Tag. Then click "Search Tag" to search for the ring Tag around the RFID reader.



Fig. 5. Setup steps for communication between the smart phone and RFID reader/Tag

The 3G connection setup page of the smart phone is shown on the left side and the webpage of the server is on the right side where temperature values, pulse, and Google map are displayed. Based on the GPS position information sent from the smart phone[8].



Fig. 6. 3G communication setup and the webpage of remote medical station

### Classification Analysis

GPS always stores patients information regarding their heart pulses as well as temperature. Our classification results shows which type of age persons frequently suffering with heart attacks. This classification based on server updated information. Based this information we can easily find out which type of age person frequently caused by this heart attack. This paper estimate to above 50 years of age persons have more heart attacks when compared to less than 50 years old.

## 5 Conclusion and Future Scope

A Heart Failure Alert System based on a smart phone with build-in GPS and a RFID physiological sensor has been presented. The hardware and software for this has also discussed. All physiological measurements are transmitted to the smart phone through Bluetooth. The user can monitor his/her own pulse and temperature from the smart phone. Then these data are transmitted to a remote server through the mobile communication of the smart phone, such as HSDPA, Wi-Fi, WiMax, GPRS, etc. The sensor placed in the hand is less cost and there is no affect to body. The build-in GPS further provides the position information of the monitored person. The remote server not only collects physiological measurements but also tracks the position of the monitored person in real time. In future we would like to deal with a big data and we will perform best classification results.

## References

1. Stephen, S., Kopp, W.: The top 10 innovative products for 2006: Technology with a human touch. *The Futurist*, pp. 16–20 (July/August 1996)
2. Yu, S.A., Lu, S.S., Lin, C.W., Wang, Y.H.: Personal electronic nurse, *Scientific Development*, 393 (September 2005) (in Chinese)
3. Chang, K.S.: Embedded electrocardiogram measurement system design and its application to personal remote health care. Master thesis, National Cheng Kung University, Taiwan, 32-37 (2004) (in Chinese)
4. Lin, J.L.: Development of wireless sensor network for home health care. Master thesis, National Chiao Tung University, Taiwan (2005) (in Chinese)



5. Wu, J.L.: Implementation of a portable wireless physiological signal measurement system. Master thesis, Southern Taiwan University, Taiwan (2004) (in Chinese)
6. Lin, T.H.: A mechanism integrating electrocardiogram compression and error protection and its application to the Bluetooth transmission in the home care system. Mater thesis, Chung Yuan Christian University, Taiwan (2004) (in Chinese)
7. Shu, Y.L.: Development of intelligent maintenance system for establishing the quality of the elder's life. Engineering Science and Technology Communication 84 (2005) (in Chinese)
8. Ye, C.F.: A PDA-based home care system. Master thesis, National Chiao Tung University, Taiwan (2006) (in Chinese)
9. Lee, R.G., et al.: A mobile care system with alert mechanism. IEEE Transactions on Information Technology in Biomedicine 11(5), 507–517 (2007)

# Globally Asynchronous Locally Synchronous Design Based Heterogeneous Multi-core System

Rashmi A. Jain and Dinesh V. Padole

Electronics Engineering Department, G.H. Rasoni College of Engineering Nagpur (M.S.) India  
{rashmishubhi, dvpadole}@gmail.com

**Abstract.** Multi-core system has wide efficacy in today's applications due to less power consumption and high performance. According to study of different scalable architectures of heterogeneous multi-core system we have been presented two different cores. First one synchronous or clocked core design is still through far the most accepted digital system design methodology. Synchronous core is well understood and supported by the grown-up CAD tools. Now-a-day it is implemented as System-on-Chips (SoCs). Second one Asynchronous Locally Synchronous (GALS) core is a comparatively latest design methodology of VLSI system that promises to merge the advantages of synchronous and asynchronous designs. By different partitioning strategy of the synchronous architecture it is created. To draw comparisons; a general purpose 8-bit synchronous core was designed and then converted into GALS core. These models were implemented in VHDL with Xilinx ISE 13.3 software and simulated using ModelSim tool. The synthesis results show in the same power consumption and a less area, GALS core outperformed the synchronous core of operating frequency which is just about double the operating frequency than the synchronous version. Globally through the proposed and integrate these cores into a single integrated chip. Generate a multi core system.

**Keywords:** Asynchronous, Synchronous or clocked core GALS core, general purpose processor core, low power, microprocessor, SoC.

## 1 Introduction

Now-a-day digital systems are implemented as System-on-Chips (SoCs), where different clock and operating frequencies requirements by functional block. To the entire system distributing a high frequency global clock with low skew, a many design effort, die area and power is a demanding task challenging. As a system integration has occurred to a critical problem. These causes, pure synchronous methodology doesn't appear greatly capable to handle every design requirements. Methodology of Asynchronous [1] is revealed to be a good solution those can solve nearly all the problems linked to clock. Though asynchronous circuits can have better performance, less power consumption and more robustness, due to handshaking they can be larger overheads and also suffer from design difficulty and lack of

programmed CAD tools. GALS methodology [2] is an intermediate result that a both synchronous and asynchronous methodology benefits. A GALS system have synchronous modules of locally clocked i.e. all module can execute operations with its have own clock and these modules are generally developed with synchronous standard CAD tools. Different synchronous modules between communications can be achieved using asynchronous channels i.e. using handshaking techniques. In a GALS design, the clocking is no longer causes the clock skew constraints are reduced. GALS is a promising method for simplifying the design of great performance SoC design. In it the main issue is designing reliable GAL's interfaces to overcome the problem of metastability. And can occur between Asynchronous and synchronous logic domains.

### **1.1 Stretchable Clock**

In this design, use a local clock generator, able to pause or stretch to reduce metastability during data transfer between different clock domains. Stretchable clock is more or less similar to synchronous circuits of clock gating .The approaches of both, on the similar paradigm discarding needless clock cycles.

### **1.2 Asynchronous (Clockless)**

This design technique involves common case without of timing relationship consideration between synchronous clocks.

### **1.3 Loosely Synchronous**

This design technique involves between clocks, also in clock frequency (in the clock phase). Pausible clock design technique is used in this work.

## **2 Related Works**

Many researchers focus their work mostly to pausable clock based GALS system [4], [5] as it has been established a capable approach to SoCs and Network-on-chips (NoCs). Presently, GALS applications mostly target the area of NoCs [6], [7], [8], multiprocessor or core systems [9] and integration of highly complex SoCs [10], [11]. Implementation of GALS is discussed Field programmable gate array (FPGA) [12], [13]. Not several publications are available on the area of design of GALS processors or core except [14] where the authors enclose attempted to manufacture a GALS model of existing synchronous superscalar core architecture with a different partitioning approach. Apart from this, [15] and [16] Design of 8051 microcontroller presents asynchronous core. The synchronous core to design is an easy and simple general purpose core. Its GALS version is formed on the synchronous architecture by applying different partitioning strategy.

## 3 Theory

### 3.1 Design of Synchronous Core

Till date microprocessors and core available in market are all built using synchronous technology. In synchronous systems there is a centralized clock-pulse originator. Each part of the core has to be connected to the same clock signal. This action has some disadvantages.

**1) The clock rate is limited by Lowest component:** E.g. There is an operation which wants four gates and an operation which wants six gates the clock-signal is not allowed to be more rapidly than the required time to pass six gates. In this case, the process with four gates needs 30% longer than required. **2) Worst case:** The processing time for an operation is not stable or constant. It depends on:

**a) Information or data:** Operation time is information or data dependent. For example it is easier to add zero than an 8 bit figure. **b) Voltage or electric energy:** Higher voltage or *electric energy* allows faster switching **c) Temperature:** Gates get slower when temperature or heat is increasing. The slowest possible case (e.g. maximum allowed temperature, lowest voltage) limits the clock rate.

**3) Power or energy consumption:** All parts of a core need energy even if there is nothing to do. At least the clock tree switches each clock. **4) High-frequency emission (radiation):** Higher clock rates also mean more high-frequency radiation. The synchronous design leads to a strengthening of the radiation. **5) Essential chip space:** Much chip space is needed by a clock generator. For higher frequencies the clock generator needs a quartz piezo-electric oscillator. Despite the no. of disadvantages, designers continue to struggle with the synchronous methodology. The reason being this methodology is well understood and tool vendors (e.g. Cadence, Mentor Graphics, Synopsis etc.) provide a plethora of design environments that support in design, verification, synthesis and testing.

### 3.2 Design of Asynchronous Core

Asynchronous or self-timed circuits work under the lack of global clock and the system timing is performed by the elements themselves. This has several possible benefits: **1) Lower (lesser) power:** Synchronous circuits enclose to toggle clock lines, and probably pre-charge and discharge signals. For example, while a unit of floating point unit on a core may not be used in a particular instruction stream. The unit must be operated during the clock. Synchronous circuits require less transition on the computation path than Asynchronous circuits. They usually have transitions just in areas involved in the present computation.**2) No clock skew:** The arrival times are different of the clock signal at the circuit of different parts is known as a clock skew. According to definition asynchronous circuits have no globally distributed clock; there is no need to be bothered about clock skew. But in synchronous systems their circuits often slow down to contain the skew. As feature sizes reduce, clock skew becomes a much larger concern.**3) Average-case in its place of worst-case performance:** Synchronous circuits must wait until every possible computation has

finished earlier than latching the results, yielding worst-case performance. When a computation has finished various asynchronous systems sense allowing them to expose average-case performance. Designed of circuits like ripple-carry adders where the worst-case is extensively worse in comparison to average-case delay, this can end result in a substantial savings. **4) Easing of overall issues of timing:** In a synchronous core, the system have clock, and performance, is dictated by the lowest path. Thus, most portions of a system should be cautiously optimized to get the highest clock rate, together with seldom used portions of the method. Since various asynchronous systems start the circuit path speed of presently in process. **5) Small Electro-magnetic Interference (EMI):** In asynchronous cores, it is doubtful that the majority of transistors switch at the same point in time, which leads to much lower electromagnetic radiation than in synchronous core because the pulsed electromagnetic radiation of millions of concurrently switching transistors of synchronous cores adds up to a large electromagnetic pulse. The potential advantages of asynchronous design, one might question why synchronous systems predominate. The cause of it asynchronous circuits have many problems as well:

**1)** Asynchronous circuits are very complex to design in an adhoc fashion than synchronous circuits. **2)** By setting the long enough period of clock rate, all worries on hazards (undesired signal transitions) and the dynamic state of the circuit are removed. But designers of asynchronous systems of the dynamic state of the circuit must pay a big deal of attention. Hazards must be removed from the circuit. It is not introduced in the first place, to avoid wrong results. **3)** Asynchronous circuits in general cannot leverage off of existing CAD tools and implementation alternatives for synchronous systems. **4)** Timing driven test equipment is mostly not appropriate for testing event-driven asynchronous circuits. In addition, these circuits tend to require logically redundant gates to eliminate hazards, and these are un-testable by normal methods. Unfortunately, asynchronous or clockless designs would not possible be the solution for today's or tomorrow's monster SoCs.

### 3.3 Overview of GALS Core Design

Methodologies of the GALS design have been developed to address several key problems of the generally used synchronous design methodology. GALS basically combines both synchronous design methodologies by the asynchronous design technique. The reason is to combine the advantages of the individual design styles while avoiding their shortcomings. By eliminating the global clock, a major source of power consumption and a design bottleneck is eliminated. The greatest advantages of GALS systems are:

**1)** The possibility to reuse the existing synchronous IP cores and simple to design systems with many clock domains **2)** No global clock distribution problem and the employment of the standard synchronous EDA tools to design and verify new IP cores. **3)** The ability to run SoC components at different frequencies, which contributes to power savings and also solves system addition difficulty. **4)** System only operates when data is available, hence energy efficient. **5)** Testing IC is not a

problem since the no. of asynchronous gates in GALS system is comparatively small and the usual test strategy of scan-based methods can be applied [17]. And GALS inherently generates low EMI. However, GAL's systems have their own drawbacks, e.g. Metastability problem; when through a clock an asynchronous signal is sampled. In order to avoid Metastability several methods are used already discussed in section I

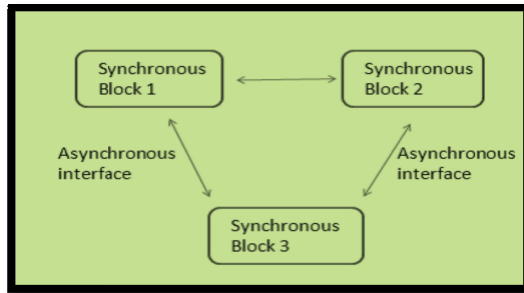


Fig. 1. GAL's architecture is composed of large synchronous blocks (SBs) which communicate asynchronously

### 4 Proposed 8-Bit General Purpose Synchronous Core Design

The idea of designing this core is taken from [18] with several modifications in the coding as well as in the architecture to get the required, errorless simulation and synthesis results. The design of a core can be divided into two main parts: control unit and datapath. In designing a CPU, we should first describe its instruction set and how it is encoded and execute. One time we have decided the instruction set, we can carry on to a datapath designing, that can execute and processing every the instructions in the instruction set. Finally, we can plan to design the control unit (CU). The control unit acts as a finite-state machine which cycles through three main categories: 1) fetch an instruction; 2) decode the instruction; and 3) execute the instruction.

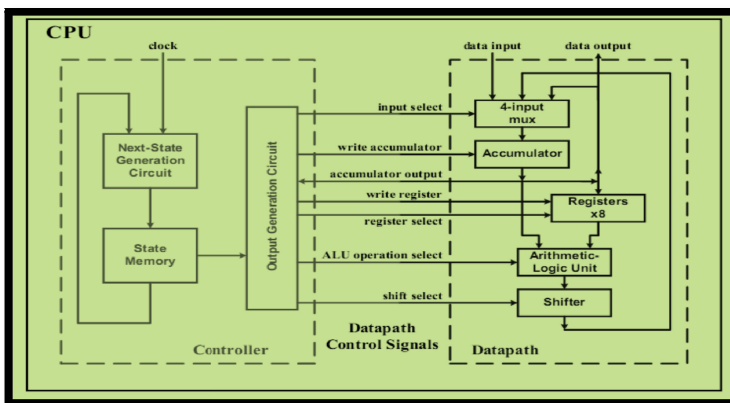


Fig. 2. The architectural block diagram of the 8-bit synchronous core implemented in VHDL

These steps perform by control unit sending the suitable control signals to the datapath blocks. Instructions in a program are generally stored in external memory. There is external memory that is linked to the CPU through a data bus and an address bus. Therefore, fetch an instruction that is the step 1 generally involves situation on the address bus. According to the control unit and the external memory to output the instruction from that memory location on the data bus. The CU at that time by data bus reads the instruction. To remain the design easy, instead of having external memory, the memory is place directly inside the CPU and implemented basically as a 16-byte array. For step 2 (decode of the instruction) the opcode bits extracts by control unit as of the instruction and determines what the present instruction is via jumping to the state. For, executing that instruction it has been assigned .In exacting state, the FSM (finite-state machine) done step 3 by mostly asserting the suitable control signals that is designed for the datapath controlling to execute that instruction.

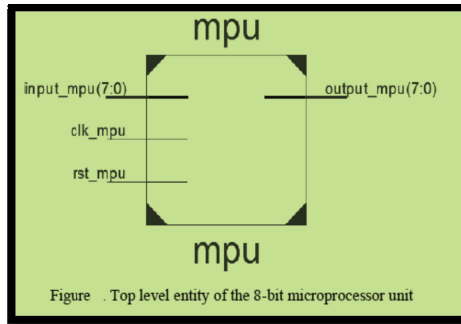


Figure . Top level entity of the 8-bit microprocessor unit

**Fig. 3.** Top level entity of the 8-bit core unit for General purpose synchronous core

**Table 1.** Instruction Set and I/O pin description for General purpose synchronous core

INSTRUCTION	ENCODING	COMMENT	Signals	Width(bit)	I/O	Name/Function
LDA A, rrr	0000 0rrr	Load accumulator from register	Input_mpu	8	I	Input port of the core module
AND A, rrr	1000 0rrr	Acc. AND register	Clk_mpu	1	I	Clock for the core module
OR A, rrr	1001 rrr	Acc. OR register	Rst_mpu	1	I	Active high reset for the core module
ADD A, rrr	0001 0rrr	Acc. + register	Output_mpu	8	O	Output for the core module
SUB A, rrr	0010 0rrr	Acc. - register				
NOT A	1010 0000	Invert acc.				
INC A	0011 0000	Increment acc.				
DEC A	0100 0000	Decrement acc.				
SHFR A	1011 0000	Shift acc.Right and fill with 0				
ROTR A	1100 0000	Rotate acc. Right				
HALT	1111 0000	Halt execution				

To design the core, first the behavior of the various datapath blocks (i.e. 4:1 mux, 8-bit accumulator, 8\*8 register file, 8-bit ALU, 8-bit shifter) was described and then using structural modeling the various datapath components were interconnected. After that control unit was coded as finite-state machine in behavioral modeling. Finally the top level consists of structural modeling description to interconnect control and datapath units. The behavioral simulation results were verified for every the instructions.

### 5 GALS Core Design

Synchronous design is the foundation of the GALS design. Partitioning the synchronous design into locally clocked synchronous modules remains the most critical aspect of GALS. The partitioning has more influence on the performance of the system than all other factors combined. Till date no well-defined methodology has been developed for partitioning. It requires manual intervention to partition the design depending upon the architecture and considering design complexity. Our GALS design consists of 4 synchronous modules (SMs), operating in 4 different frequencies

- SM 1: control unit module
- SM2: mux + acc. module
- SM3: register memory module
- SM 4: functional units (ALU + shifter) module

Handshaking between different clocked modules is handled by the simple request and acknowledge signals which are nothing but the control and data signals that mark the validity of the data. Also, the block which is not required in the current execution cycle can be stalled using pausable clock approach. From the above partitioning we can theoretically say that in GALS, the global clock distribution is replaced by local clock distribution due to which clock related problems will now be limited within a particular synchronous module or within a short area and will not be as cruel as in the case of fully synchronous design.

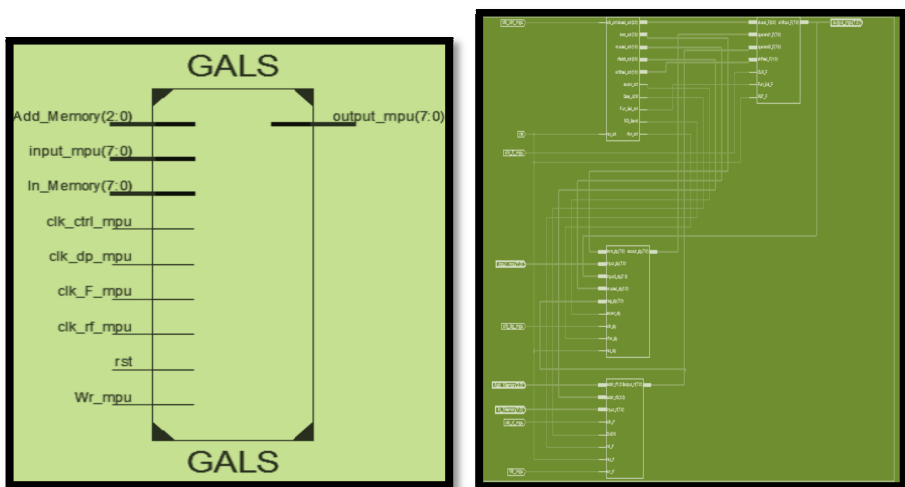


Fig. 4. Top level entity of the 8-bit GALS core and Top level integrated GALS module

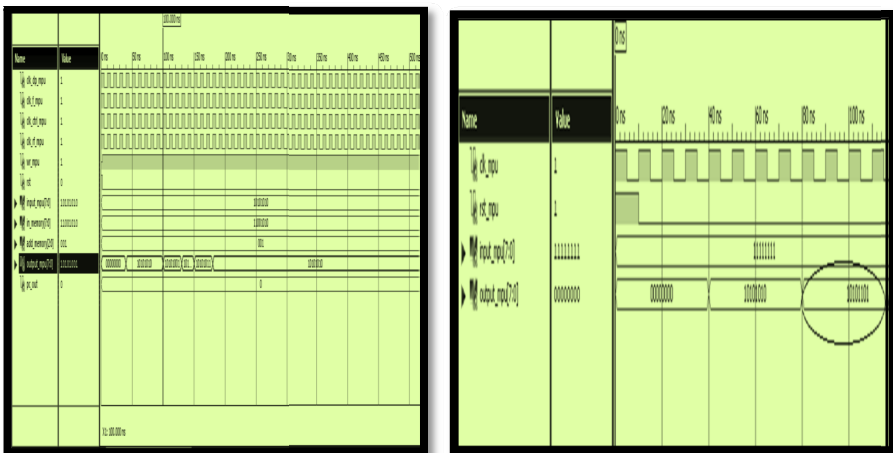


**Table 2.** Shows the instruction set that the proposed core can execute

Signals	Width (Bits)	I/O	Name/Function
Input_mpu	8	I	Input port of the GALS core module
Add memory	3	I	Address of the location where the 8-bit data is to be written into the register memory
In memory	8	I	Input port for providing data to the register memory
Clk_ctrl_mpu	1	I	Clock for the control unit module
Clk_dp_mpu	1	I	Clock for mux + acc. Module
Clk_rf_mpu	1	I	Clock for functional unit module
Clk_rf_mpu	1	I	Clock for register memory module
Rst	1	I	Active high reset for the GALS core module
Wr_mpu	1	I	Active high write signal for the register memory module
Output_mpu	8	O	Output port of the GALS core module

## 6 Simulation Results

The VHDL code for both synchronous and GALS core is synthesized using Xilinx Synthesis Tool (XST) and the functionality is verified using ISE simulator (ISim). Power consumption for both designs is estimated using XPower Estimator (XPE) [19] for accurate power analysis.



**Fig. 5.** Shows the output result for ADD A, R [1] instruction execution in case of synchronous and GALS core respectively

**Table 3.** Comparisons of synchronous and GALS core

Parameters	Synchronous core	GALS core
Operating Frequency	143.698 MHz	269.233 MHz
Power Consumption	Quiescent power=56mW Dynamic power=3mW Total power=59mW	Quiescent power=56mW Dynamic power=3mW Total power=59mW
Area overhead	59 LUTs 31 slices 27 slice flip-flops	67 LUTs 52 slices 69 slice flip-flops

## 7 Conclusion

Technology has seen the development of processor industry right from micro to the latest nano-technology with speed and performance being important criteria, not much attention has been specified to the power requirement for these integrated circuits. The presented GALS core was a simple model of the proposed synchronous core, designed using pausable GALS clocking style. The extra advantage which we found in our GALS core was the reduced handshaking latency which is one of the drawbacks of asynchronous design as well as GALS design. Present fully synchronous cores have evolved with a global clock which is supplied throughout the die; this has resulted in unwanted power consumption and dissipation. The lack of global synchronization makes power savings in the clock net possible. Moreover, the independently clocked synchronous blocks opens up great possibilities for substantial power reduction using Dynamic Voltage and Frequency Scaling (DVFS) techniques which can be applied to this work for further reduction in power and also this architecture can be implemented in FPGA.

## References

- [1] Sparso, J., Furber, S.: Principles of Asynchronous Circuit Design- A System Perspective. Kluwer Academic Publishers (2001)
- [2] Krstic, M., Grass, E., Gurkaynak, F.K., Vivet, P.: Globally Asynchronous, Locally Synchronous Circuits: Overview and Outlook. IEEE Design and Test of Computers 24, 430–441 (2007)
- [3] Teehan, P., Greenstreet, M., Lemieux, G.: A Survey and Taxonomy of GALS Design Styles. IEEE Design and Test of Computers, 418–428 (2007)
- [4] Fan, X., Krstic, M., Grass, E.: Analysis and Optimization of Pausible Clocking based GALS Design. In: International Conference on Computer Design ICCD, pp. 358–365 (2009)
- [5] Rahimian, M.A., Mohammadi, S., Fattah, M.: A High- Throughput, Metastability-Free GALS Channel Based on Pausible Clock Method. In: 2nd Asia Symposium on Quality Electronic Design, pp. 294–300. IEEE (2010)

- [6] Ning, W., Fen, G., Fei, W.: Design of a GALS Wrapper for Network on Chip. In: 2009 World Congress on Computer Science and Information Engineering, pp. 592–595 (2009)
- [7] Lin, S., Li, S., Depeg, J., Lieguang, Z.: Universal GALS platform and evaluation Methodology for networks on chip. *Tsinghua Science and Technology* 14(2), 176–182 (2009) ISSN 1007-0214
- [8] Thonnart, Y., Vivet, P., Clermidy, F.: A Fully-Asynchronous Low-Power Framework for GALS NoC Integration. In: Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 33–38 (2010)
- [9] Yu, Z., Baas, B.: High Performance, Energy Efficiency & Scalability with GALS chip Multiprocessors. *IEEE Transactions on VLSI Systems*, 66–79 (2009)
- [10] Watn, R., NjBlstad, T., Berntsen, F., Lonnum, J.F.: Independent Clocks for Peripheral Modules in System on- Chip Design. In: Proceedings of the IEEE International [Systems-on- Chip] SOC Conference, pp. 25–28 (2003)
- [11] Zhuang, S., Carlsson, J., Wanhammar, L.: A Design Approach for GALS based SoC. In: Proceedings of the 7th International Conference on Solid-State and Integrated Circuits Technology, vol. 2, pp. 1368–1371 (2004)
- [12] Gagné, R., Belzile, J., Thibeault, C.: Asynchronous Component Implementation Methodology for GALS Design in FPGAs. In: Joint IEEE North-East Workshop on Circuits and Systems and TAISA Conference, NEWCAS-TAISA 2009, pp. 1–4 (2009)
- [13] Amini, E., Najibi, M., Jeddi, Z., Pedram, H.: FPGA Implementation of Gated Clock Based GALS Wrapper Circuits. In: International Symposium on Signals, Circuits and Systems, ISSCS 2007, vol. 1, pp. 1–4 (2007)
- [14] Iyer, A., Marculescu, D.: Power and Performance Evaluation of GALS processors. In: Proceedings of the 29th Annual International Symposium on Computer Architecture, pp. 158–168 (2002)
- [15] Kovac, M.: Asynchronous microcontroller simulation model in vhdl (2008)
- [16] Mabry, R.: Asynchronous implementation of 8051 microcontroller. Honour's Thesis
- [17] Gurkaynak, F., Villiger, T., Oetiker, S., Felber, N., Kaeslin, H., Fichtner, W.: A Functional Test Methodology for GALS Systems. In: Proceedings of the Eighth International Symposium on Asynchronous Circuits and Systems, pp. 181–189 (2002)
- [18] Hwang, E.: *Microprocessor design Principles and Practices with VHDL*. La Sierra University (2004) ISBN: 0-534-46593-5
- [19] <http://www.xilinx.com>
- [20] Jain, R.A., Padole, D.V.: Scalable and Flexible heterogeneous multi-core system. (IJACSA) *International Journal of Advanced Computer Science and Applications* 3(12) (2012)

# Recognition of Smart Transportation through Randomization and Prioritization

Bhavya Boggavarapu<sup>1</sup>, Pabita Allada<sup>1</sup>, SaiManasa Mamillapalli<sup>1</sup>,  
and V.P. Krishna Anne<sup>2</sup>

<sup>1</sup>Dept. of Computer Science & Engineering,  
Koneru Lakshmaiah Education Foundation (KL University),  
Vaddeswaram Post, Guntur DT, A.P., India – 522502

<sup>2</sup>CSI – Koneru Chapter  
Koneru Lakshmaiah Education Foundation (KL University),  
Vaddeswaram Post, Guntur DT, A.P., India – 522502  
{bhavyaboggavarapu, allada\_pabita,  
praveenkrishna}@kluniversity.in,  
saimanasa.klu@gmail.com

**Abstract.** Complexity involved in maintenance of proper traffic rules and controlling them certainly is raising as an accruing problematic situation in our daily life. There are more collisions that are been taking place now and then due to the inefficient traffic signaling formats that are laid presently. The proposed Smart Transportation can find a fashion which provides co-ordination with neighboring traffic control points. Even sometimes, due to the poor performance of the signals, they are illogical and inefficient. So, in order to solve this problem we are assigning a technique of organizing a rule of allotting priority to the traffic lines by considering the count of vehicles through randomization technique on that particular line. The lane that has more traffic gets less ‘green’ signal while the ones with less cars are kept open for more time. This current paper is designed to address these issues. This transportation methodology provides an efficient way to manage the traffic for the city traffic management authority by providing the alert signals and can easily manage the traffic flow for the vehicles and pedestrians automatically through prior update for every interval of time kept constant. Even the registration of the police is kept for surveillance in order to provide security and also to fulfill the prior requirements. It is been beneficial by providing the VIP data when authorized police provides.

**Keywords:** Randomization, Prioritization, Automation, security and authorization.

## 1 Introduction

The main view is to match traffic demand to supply with optimal usage of available public resources and concomitant optimization of citizens’ private resources for travel needs. Informally stated for the road traffic, the problem can be seeking to manage traffic optimally on a road network using public resources while allowing citizens to complete their daily travel needs optimally.

An Example of this problem type: For a given day, minimize over-time payments to traffic personnel while minimizing average commute time per km. Service objectives can also be stated like average commute time per km be below 10 min/km.

From this we can conclude that the aim is to maximize the flow of vehicles and reduce the waiting time while maintaining fairness among the other traffic lights. Each traffic light controlled intersection has an intersection control agent that collects information from the sensor nodes generated by sensing the number of vehicles. An intersection control agent manages its intersection by controlling its lights. Multiple intersection agents can exchange information among themselves to control a wider area. Tremendous amount of time and power is wasted due to green traffic light with no cars passing on its lane. We envision a smart road system where the total trip time is minimum due to minimizing the average waiting time on traffic lights.

On making this problem statement proposed into existence can make down many of the complexities to face the end gate. This technique of implementation is striving for the maintenance of the traffic favor to the congestion less approach, which means it provides the risk free zone of involvement. On following this implementation basic problem raised can be solved even like the communication defects since all the actors involved in it will be clearly informed about what should be followed to achieve the proper destination.

## 2 Background

### 2.1 Literature Survey

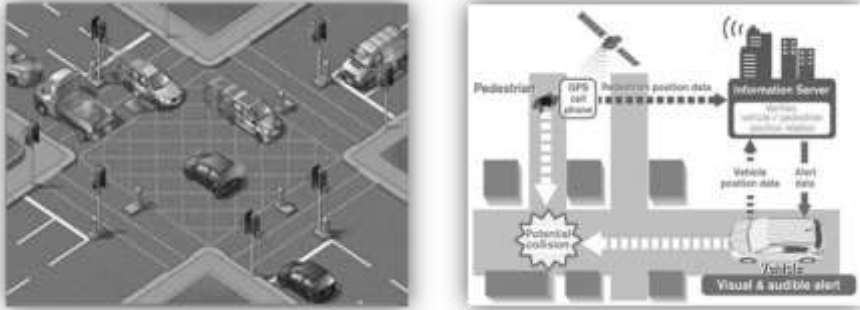
Traditionally, traffic management has been funded only by city governments and they had no framework to access citizen's information related to travel. Both are changing. Given the importance to traffic to citizen's daily lives, traffic information is being provided as a value-added service by business (e.g., radio stations, mobile phone operators) and citizens (and businesses) are willing to pay directly or indirectly. Furthermore, technology is enabling citizen's traffic demands to be available on a more regular basis than previously possible using demographic surveys. So, a more general problem statement is needed, and we believe can be solved. On the side of the city administrators, another way of managing traffic has emerged – Intelligent Transportation Systems. To reiterate, some dimensions along which a city can manage traffic are:

- ❖ **Method – 1.** Creating physical infrastructure, e.g., new roads, expanding capacity of existing roads.
- ❖ **Method – 2.** Making policy changes, e.g., banning traffic movement during major games, changing traffic direction, making road usage chargeable.
- ❖ **Method – 3.** Intelligent Transportation System (ITS), which seeks to use IT infrastructure to measure and manage traffic based on cities policies. It is not one specific technology or usage scenario (e.g., using GPS for Region-based Charging) but a broad term for any IT-based traffic measurement cum management solution.
- ❖ **Method – 4.** Hybrid representing any combination of above methods. For e.g., making a road chargeable may mean a policy change, adoption of ITS.

## 2.2 Solution

- I. Focus area of developed cities
- II. Focus area for cities of emerging geographies.

Traffic management is a key responsibility of any city management as part of its mandate to provide quality infrastructure to its citizens. At the core of traffic management is the problem of knowing the scale of traffic on the roads.



**Fig. 1.** The rule shows the traffic optimization policy (*left*) and the screen of GPS usage in the method – 3 transportation (*right*). *Src:* (www.googleimages.com).

These should be arranged by converting all the complexities to face down by implementing it with the sensor based randomization technique application invoked with prioritization principle.

## 2.3 Existing System

In urban areas where traffic signals are nearby, the co-ordination of adjacent signals is important and gives great benefits to road users by increasing the utilization per unit time in the peak hours. Coordinating signals over a network of conflicting routes is much more difficult than coordinating along a single route.

Early work developed off-line software to calculate optimum signals settings for a signal network. It can be used to compile a series of fixed time signal plans for different times of day or for special recurring traffic conditions.

Traffic control systems like traffic signal controllers, traffic blinkers, solar traffic blinkers, etc., are used as an independent system at isolated intersections or as part of a synchronized chain of controllers for coordinated control traffic.

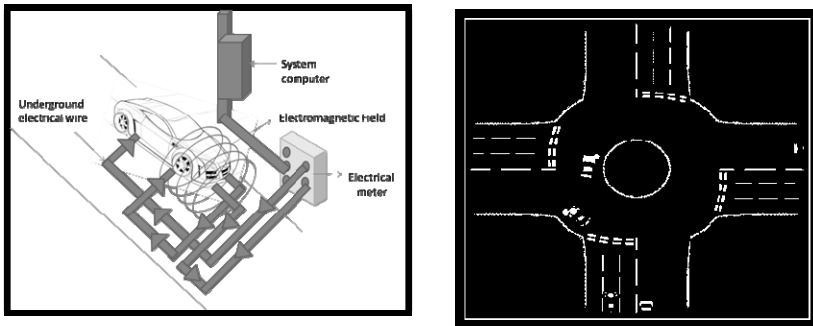
Preparing such signal plans is time consuming, expensive and also requires repetitive surveys to understand the dynamics in traffic flow. Unless plans are updated frequently as traffic increases and patterns change rapidly the signals become less efficient.

## 2.4 Proposed System

The existing systems are quite complex and become obsolete in short time durations. In order to adapt the changing demands of speed and efficiency, a reliable software system which may handle the situation dynamically is needed. For this purpose a model of smart traffic management system is designed in UML, which can be converted into a Real Time System.

Software architecture based on UML models will help in handling complexities and drawbacks of existing systems and also help to better understand the domain.

UML is the de-facto standard visual modeling language which is a general purpose, broadly applicable tool supported, industry standardized modeling language which offers an extensive set of diagrams for modeling. The complexity of the problem domain requires extensive efforts for the clarification of initial problem statement. Moreover, due to the extremely long lifespan of systems, stable and robust analysis models enabling the integration of new operational scenarios are needed which can be efficiently obtained using UML models.



**Fig. 2.** Proposed System of smart transportation through sensor sensing capability (*left*) and by maintaining traffic flow clear (*right*)

## 3 The Smart Traffic Product Perspective Recognition

A descriptive study is a high-level capsule version of the entire system analysis and design process. The study begins by classifying the problem definition. Feasibility is to determine if it's worth doing. Once an acceptance problem definition has been generated, the analyst develops a logical model of the system. A search for alternatives is analyzed carefully.

- ❖ Administrator on Internet will be using HTTP/HTTPS protocol.
- ❖ Traffic Police Will Get the information from the Web Form created to him/her on logging in.

## Product Perspective

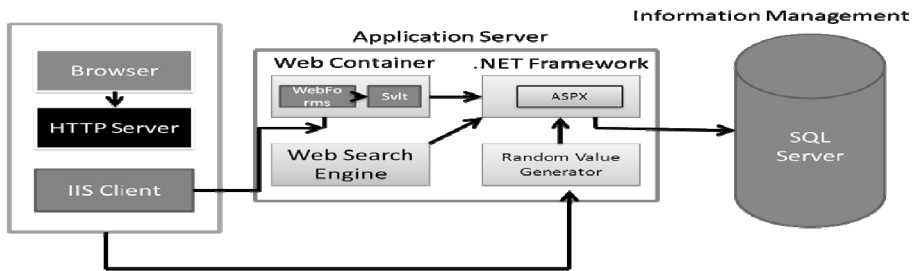


Fig. 3. Proposed System of smart transportation working product perspective

## Work Flow

Flow starts in the way as described above, At first admins locks into his account by entering authorized user-id and password. If the data entered by him is correct then he can access the data from it if not account cannot be opened it will direct to the login page. Later on entering he can process on with another module linked up with it i.e., clicking on sensor update button through this action he can get connected to the server and get information from the line and send it to the admin for allotment of priority for the lines and then to the signals. Then that route is allotted for the motion and followed by the followed routes. In case if any VIP access route then priority is given first to him, then to the other. Pedestrians are also taken into considerations at respective periods according to their presence at particular lane. There our project covers smart traffic management System.

## 4 Experiment

### 4.1 Experiment Method – Roles of All the Users Involved in the Transportation System

**Admin:** Administrator has the authority to move, make wait the car owners and pedestrians in case of any VIP arrival and also can generate view reports and set the line zones which need to be allotted according to the priority schedule. Administrator can get complete details about the traffic present and can make the online traffic movement path allotment. Administrator can also have co-ordination with the neighboring traffic control points.

**Traffic Police:** Traffic Police is one who checks the mode of the vehicles and Traffic Police regulates the traffic rules to be followed perfectly and provides the traffic lines to be accident free.



**Vehicle (Car Owner):** Car Owner follows the road links by searching for the signals and vehicle number can be sensed by the sensors fixed at the regular intervals of distances on the road lines.

**Pedestrians:** Pedestrians follows the road links by searching for the signals and there by choosing Zebra Crossing or foot paths for their travel purposes.

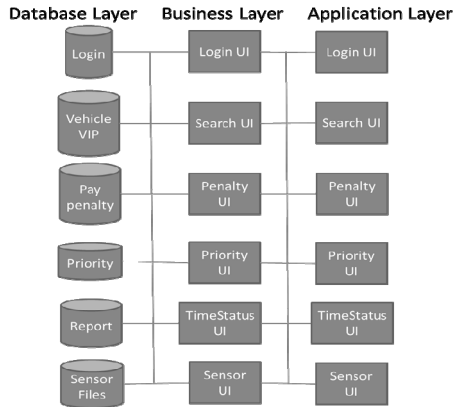


Fig. 4. Architectural Design of the Transportation System implemented

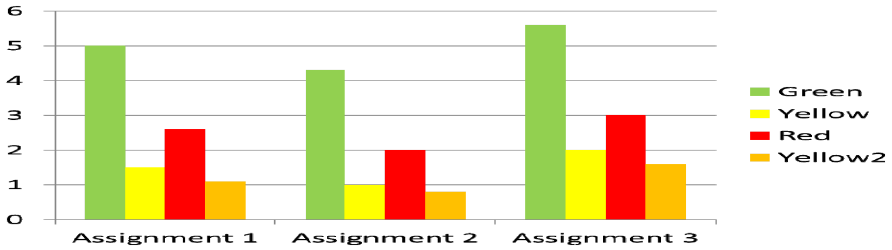
### 4.2 Experimental Process Involved

All the respective goals and properties of the users involved in the system are shown as follows.



Fig. 5. The results of the priority generated allotting R, G, Y signals in various criteria – More the traffic higher the green signal allotment, optimization of waiting time

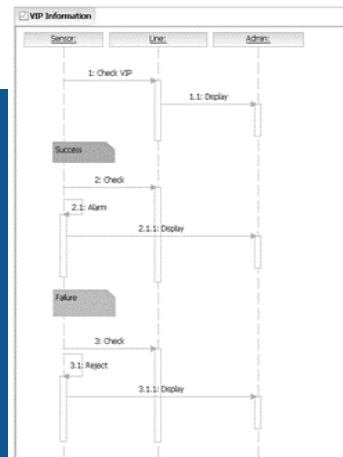
This graph provides the necessary information that can be derived based upon the signal allotment done to the respective lines with more vehicles are allotted with more green signal and the waiting time for the vehicles on the lines are also optimized to a great extent.



Suppose, Consider there are four lines such as shown above: line 1, line 2, line 3, line 4 in which the sensors are placed for the considered area in a equal interval distance say 3 per line and all the vehicles passing through that particular line are sensed by the sensor and updated systematically to the admin or traffic police operating on the system. Since the system is automated for 24X7 availability the regular update of sensor takes place and get stored in the database of the transportation system for every regular intervals of time say 5 min. Accordingly if the traffic on respective lines are say 34, 58, 78, 62. Then the allotment of the signals are as followed: line 3 is allotted with green light, line 4 with yellow, line 2 with yellow time variation existence and finally line 1 with red signal.

### 4.3 VIP Vehicle Information Maintenance

VIP vehicle maintenance is the most specific goal of the transportation system which involves the timely information update provided by sensor detected through the alarm and sound effects of vehicles and even the traffic police or admin can update the VIP information manually on the system and details related to the arrival on which line and considering those factors the priority is given for the ambulance, fire brigades than to the normal vehicles.



**Fig. 6.** The VIP information received through the sensor data and passed onto the automated system of the Smart Transportation for the surveillance of admin and traffic police

## 5 Conclusion

In this paper, we showed that using a sensor and automating the traffic flow for the allotment of the signals in prioritized and randomized fashion. This technique involves less manual power involved and continuous control of the traffic with automatic update of the traffic flow in each and every line. If there is any VIP information available, the traffic police has only to check the priority allotted by the automated system and make the vehicle moved in that respective priority and which can help a better in optimizing the traffic complexity and thereby efficiency of maintenance of the traffic flow also increases gradually. This also calms down the probability of occurrence of accidents. By following this illustrated procedure can improve the traffic errors to be rectified by 80.0 percentage of accuracy.

**Acknowledgements.** This research has been supported in part by TGMC '12 – Imagine Spark and IBM.

## References

1. IBM Research Report - A New Look at the Traffic Management Problem and Where to Start Biplav Srivastava, IBM Research - India
2. Design of Adaptive Road Traffic Control System through Unified Modeling Language. International Journal of Computer Applications (0975 – 8887) 14(7) (February 2011)
3. Ravi, N.D., Mysore, P., Littman, M.: Activity recognition from Prioritization. Traffic rules and adaptivity
4. Min, W., Wynter, L.: Real-time road traffic prediction with spatiotemporal correlations. In: Sixth Triennial Symposium on Transportation, Analysis, Phuket Island, Thailand (2007)
5. Shankar, R., Datta, D.: Jam Session, India Today article (September 6, 2010)
6. Xing-Guang, C., Jing, Z., Zhen-Tao, Z., Hong-Li, X.: Measure of Urban Traffic Supply and Demand Coupling Balance. In: Sixth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2009, Tianjin (2009)
7. Sen, R., Raman, B., Sharma, P.: Horn-Ok-Please. In: The Proc. of ACM MobiSys 2010, San Francisco, USA (2010)
8. Zheng, Y., Li, Q., Chen, Y., Xie, X.: Understanding Mobility Based on GPS Data. In: Proceedings of ACM Conference on Ubiquitous Computing (UbiComp 2008), pp. 312–321. ACM Press, Seoul (2008)
9. Google traffic, <http://maps.google.com>,  
[http://en.wikipedia.org/wiki/Google\\_Maps](http://en.wikipedia.org/wiki/Google_Maps)

# Prioritized Traffic Management and Transport Security Using RFID

V. Kishore<sup>1</sup>, L. Britto Anthony<sup>1</sup>, and P. Jesu Jayarin<sup>2</sup>

<sup>1</sup>Computer Science Department,  
Jeppiaar Engineering College, Jeppiaar Nagar, Chennai, India  
{kishorekalpakkam,brittoanthony.1}@gmail.com

<sup>2</sup>CSE Dept.,  
Jeppiaar Engineering College, Jeppiaar Nagar, Chennai, India  
jjayarin@gmail.com

**Abstract.** The Dynamic Traffic Priority System avoids chaos that usually arise with commonly available traffic control systems, mainly those related to image processing, inductive loop, passive infrared sensors techniques. This RFID technique deals with a infinite vehicle, multilane, multi road junction area. It provides an efficient and intelligent time management scheme, in which a signal is dynamically scheduled out in real time for the passage of each traffic column. The number of vehicles in each column and the priority which is assigned well in advance to the vehicles are the proprieties, upon which the calculations and the decision is based on.

**Keywords:** RFID, traffic sequence, dynamic time schedule, Gate Control, Vehicle Detection.

## 1 Introduction

The present system works with sensors and algorithms[1-2] and [10]. This method has the algorithm that considers that the traffic system is same always. But in cities the traffic density varies according to the working hours of industries and colleges. Hence the present system has to be modified.

A more elaborate approach has been introduced to overcome these problems. It employs real-time traffic flow monitoring with image processing systems [3-4]. But the above described one can only give a quantitative description of traffic flow [5]. Some general issues involved in image processing systems are False Acceptance Rate and False Rejection Rate.

Normally, in case of jam- packed traffic, the computer vision results in false detection [3]. The sensor based approach will be poor as it is based on the line of sight and hence the vehicles can go in blind spot.

## 2 Experimental Arrangement

A prototype of the system has been created to simulate the actual design of a RFID based traffic control using Networking devices and RFID passive tag. The intelligent traffic control system consists of five main parts: the RFID tags, RFID Reader, the access point, the location server and Wide Area Network. The first simulates the moving vehicles, the second detects the RFID tags, the third acts as an RFID software, and fourth simulate the ubiquitous environment and WAN is used for Networking.

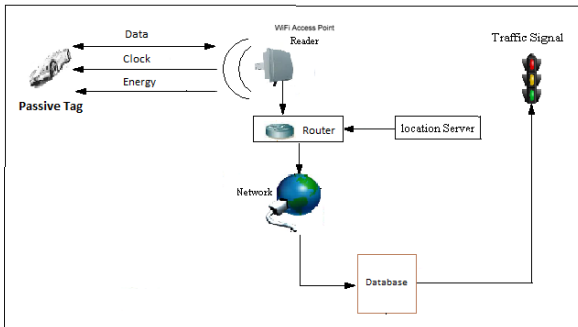


Fig. 1. Dynamic Traffic Signal Simulation

The location server is used as the microcontroller of the traffic signal. It is used to collect the location and time stamp of the data from reader. This data is sent to the database via internet. The database using Dynamic Traffic Priority Algorithm will then send suitable instructions to control the traffic. The layout is shown schematically in figure 1.

## 3 Dynamic Traffic Priority Algorithm

A dynamic algorithm for the traffic signal control can change the priority dynamically and perform traffic control at certain junction efficiently. It is based on an automatic selection of traffic sequence in a multilane traffic flow.

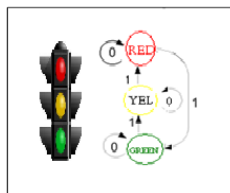
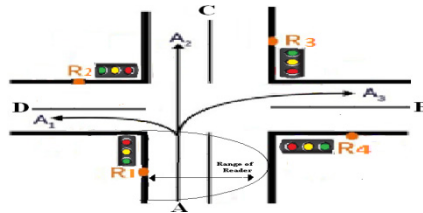


Fig. 2. Signal Transition

The Figure 2 shows how the transition takes between different signals. Assuming A, B, C and D are traffic column in which a vehicle from A can go to B, C and D with time slot that is dynamically determined. The same sequence is then shifted to B, C then D. The decision process for the Dynamic traffic control depends on the information provided by the RFID Reader. The data is also captured and saved in the master database.

A set of readers are implemented to detect and count the vehicle at each junction. The reader captures the time-in and Id for each vehicle passing within its range. Figure 3 shows the practical arrangement of the RFID readers in the four sides of the signals [7].



**Fig. 3.** Range of a RFID reader while implementation

The information such as location, time and unique id are saved as tag reference in the tag reference table. The waiting time of the various junctions can be calculated from the table 1.

**Table 1.** Waiting Time at Each State

state	Waiting time at each state
$A_R$	$A_Y + B_G + B_Y + C_G + C_Y + D_G + D_Y$
$B_R$	$B_Y + C_G + C_Y + D_G + D_Y + A_G + A_Y$
$C_R$	$C_Y + D_G + D_Y + A_G + A_Y + B_G + B_Y$
$D_R$	$D_Y + A_G + A_Y + B_G + B_Y + C_G + C_Y$

However the time for the State Yellow can be kept as 3 seconds which will be enough for a driver to stop the vehicle. Hence the equation can be written as:

$$A_R = B_G + C_G + D_G + 4(3S) \tag{1}$$

$$A_R = B_G + C_G + D_G + 12S \tag{2}$$

But the waiting time for all the junction's are not the same because the it depends on the environment of the junction and the length of the queue based on priority.

$$A_R \neq B_R \neq C_R \neq D_R \tag{3}$$

The decision is not a static one and it based on the priority given to each of the vehicles. The vehicles are classified as Ambulance & Fire Engine, VIP Vehicle College and Office Busses, Public Transport, Car, Bike. All these vehicles are given priority in the descending order in such a way that if a Ambulance or a Fire Engine is found to be nearing the signal then the signal will be immediately made to Green State. The priorities assigned to various vehicles are given in the table 2.

**Table 2.** Priority Table Assigned to Vehicles

Priority	Vehicle
5	School and College Bus
4	Public Transport
3	Car
2	Bike
1	Shared Auto's
	Special Priority tag is given to the Ambulance, Fire Engine and VIP vehicles

By using such RFID Mechanism the important vehicles will be given the priority to go first. The Tag number along with its Time stamp will be sent to the Database. Each Side of a Junction will have a table in the database and the details of ID and Time stamp will be stored in it. The Priority of the vehicles can taken from the time when it is becoming Yellow and calculate the total priority of the side and based on Priority comparison the states will be made Green on one side. For example the table in database will be:

**Table 3.** in Database

Tag ID	Time Stamp
12589	10:30:00
13456	10:30:03

If the ID is already present in the Table then the old entry is deleted and the new entry is retained. The raw data are taken from the reader and passed on the database where it performs the desired calculation. The Calculation are made by taking the Priority of the Vehicle which has passed the reader only after the signal has come to the Yellow state. The Algorithm to calculate the priority to each side is as follows:

```

Priority=0;
While(time stamp of an ID > ((time stamp of yellow)&&
(signal != green)))
    If(ID is already not present in the table
of Other three columns)Then
        Take priorities of those ID's;
    
```

```

        Priority=Priority+Priority[ID];
        Take next ID;
    Else
        Discard the ID;
    End
End
End

```

If a special priority tag is detected then the side is made without any priority calculation. Based on the priority calculation of all the four signals the side with highest priority will be made green based on the following Algorithm.

```

time=0;
1:If((A_priority>B_priority)&&(A_priority>C_priority)&&
(A_priority >D_priority))
Then
    If(A==AG)
        Continue;
    Else
    Then
        B=BR; C=CR; D=DR; A=AG;
    End
    time++;
    If(time<3)
        Goto 1;
    Else
    Then
        Make the other signal Green again based on
their priority for 20s;
        Time=0;      Goto 1;
    End
End
Else
Then
    Select the highest priority and make it Green and
the others Red;
    Time=0;
End

```

In the above algorithm the signal is made green for a maximum of 3 times. After which the other signals are compared and the signal with highest priority among them is made Green for a shorter time order.

The comparison of the existing fixed time technology and the RFID based system for a single signal with higher traffic density at a given time is shown in the figure 4.



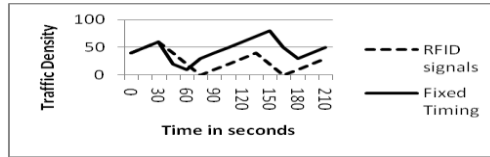


Fig. 4. Comparison Graph

As we store the ID's of the vehicles that cross each signal in the database, we can find the last signal crossed by a stolen vehicle which will be useful for finding the stolen vehicles.

## 4 System Implementation and Requirements

This system can be used to regulate much complicated traffic problems. Implementation of this RFID priority algorithm will give a simpler solution for the traffic congestion problem. For this each vehicle needs to have a RFID tag fixed on it. This requirement may involve decisions by higher authorities and implementation of policies by executive authorities [8]. The tag is generally embedded or placed on the object which needs to be traced; the tag carries necessary information about the vehicle like weight, type, length and other information. The RFID system is enabled by ubiquitous sensors made part of input devices in the traffic management [9].

This system can randomize the reply time so that it can take multiple tags simultaneously and process them. The RFID reader is located in such a way that, it is possible to collect the data from all the vehicles that pass through its reading range. The information is then sent to the master database of the management system. The collected information is processed by the centralized systems; it is also responsible for generating feedback signals that manages the intelligent traffic light sequence. A secure structured network can be used for sharing of data between local databases and central database. [8].

## 5 Railway Gate Control

The number of accidents due to crossing of Railway Gate has lead to a loss of 300+ lives in India. The basic idea is to place a RFID Tag on each and every engine of the train and to place a reader 500 m in front of the crossing gate, if the reader reads the tag then it immediately sends a signal to the nearest crossing gate so that it triggers the gate to be closed by first giving a warning signal. As soon as the train leaves it is again open by passing the signal from the reader kept at the other end.

The implementation of this system is shown in the Figure 5. In the figure 5, where A, B, C, D are the RFID Readers. If the train crosses the Reader A and the Reader B then the Gate will be closed by first raising a warning signal for 30 seconds and if the

train crosses Reader D and then Reader C it gets opened. Similarly if the train comes in the other direction the same principle happens. The algorithm for the Railway Gate system assuming the speed to be 90 Km/hr is as follows:

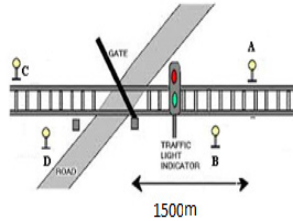


Fig. 5. Railway Track with RFID Reader's

```

If(((A==1) followed by (B==1)) || ((C == 1) followed by
(D == 1)))
Then
    Give a warning signal for 30 seconds and then
    Close the gate by triggering the System
End
If(((B==1) followed by (A==1)) || (D==1) followed by (C ==
1)))
Then
    Open the gate by triggering the System
End
    
```

## 6 Conclusion

Our traffic systems will move onto intelligent systems from being an ordinary system. The system will be efficient enough to solve a lot of common problems that exists in our traffic system. It has the system built in it to handle vehicles tagged with high priority which is very crucial in traffic management; also it will help us to save a lot of man hours which will otherwise be required.

This frequency range can enhance the reading range for RFID tag reading capability. One of the issues that need to be highlighted to the Frequency allocator is to allocate the license control of bandwidth ranging from 430 MHz to 440 MHz for RFID application. By improving this capability, the development of RFID usage in India will grow rapidly.

Once this technology becomes widely accepted, it will be used in development of other new application as a result of which its implementation cost will gradually come down due to increased volumes of production.

## References

1. Albagul, A., Hrairi, M., Wahyudi, Hidayathullah, M.F.: Design and Development of Sensor Based Traffic Light System. *American Journal of Applied Sciences* 3(3), 1745–1749 (2006)
2. Han, T., Lin, C.: Design of an Intelligent Traffic Light Controller (ITLC) with VHDL. In: *Proceeding of IEEE TENCON 2002*, pp. 1749–1752 (2002)
3. Tseng, S.T., Song, K.T.: Real-Time Image Tracking for Traffic Monitoring. In: *IEEE 5th International Conference on Intelligent Transportation Systems*, Singapore (2002)
4. Rabie, T., Shalaby, A., Adbulhai, B., Rabbany, A.E.: Mobile Vision-Based Vehicle Tracking and Traffic Control. In: *IEEE 5th International Conference on Intelligent Transportation Systems*, Singapore (2002)
5. Ferier, N.J., Rowe, S.M., Blake, A.: Real-time Traffic Monitoring. In: *Proceedings of the Second IEEE Workshop on Application of Computer Vision*, pp. 194–199 (1994)
6. Al-Khateeb, K., Johari, J.A.Y.: Intelligent Dynamic Traffic light Sequencing Using RFID. In: *Computer & Communication Engineering (ICCCE)*, Kuala Lumpur (2008)
7. Bandyopadhyay, S.: *Traffic Congestion Management Using RFID & Wireless Technologies*, IIM Kolkata (May 2010)
8. Johari, J., Khateeb, K.: Ubiquitous RFID Network for Highway Monitoring and Management. In: *IEEE International Conference on Computer & Communication Engineering (ICCCE)*, Kuala Lumpur (2006)
9. Want, R.: Enabling Ubiquitous Sensing with RFID. *Computer* (April 2004)
10. Hasan Ibrahim, A., Ismail, M., Keong, T.S., Kader Mastan, Z.B.: Development of Software Planning Tools for an Intelligent Traffic Light Wireless Communication Link Using 5.8GHz WLAN. In: *IEEE Asia Pacific Conference on Applied Electromagnetics Proceedings*, Johor (2005)
11. Bamford, R., Ahad, R., Pruscino, A.: A Scalable and High Available Networked Database Architecture. In: *Proceeding of the 25th VLDM Conference*, Edinburgh, Scotland (1999)
12. Pradip, D., Kalyan, B., Sajal, K.D.: An Ubiquitous Architectural Framework and Protocol for Object Tracking using RFID Tags. In: *IEEE MobiQuitous* (2004)

# A Novel Multi Secret Sharing Scheme Based on Bitplane Flips and Boolean Operations

Gyan Singh Yadav and Aparajita Ojha

Computer Science and Engineering,  
PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur,  
Dumna Airport Road, Khamaria, 482005, Jabalpur, India  
{gyan.yadav,aojha}@iiitdmj.ac.in

**Abstract.** Visual secret sharing is a process of distributing the information into multiple encrypted shares, overlapping those shares according to access structure reveals the secret. In the last two decades several schemes have been proposed for single secret as well multi secret sharing. The present paper proposes a multi-secret sharing scheme based on bitplane flipping and Boolean operations. Image retrieval is lossless and the scheme uses two levels of encryption which makes it more secure. Robustness of the method is demonstrated using numerical results.

**Keywords:** Secret sharing, dependency, X-or operation, security, access structure.

## 1 Introduction

With the fast development of communication technologies and computation devices, information security has become more challenging and complex. In spite of a number of computationally complex cryptographic schemes, possibility of cyber-attacks cannot be ruled out in the modern world of explosive information communication through open channels. Visual Cryptographic (VC) schemes have emerged in the last two decades and have attracted the attention of many researchers due to their potential for applications in image and video encryption, data hiding, digital watermarking etc. While security is guaranteed, VC schemes are less computationally complex and hence the decryption or information retrieval is easier for authorized participants. VC techniques mainly deal with the images or text transformations which makes them unreadable to the attacker. Visual Secret Sharing (VSS) scheme was first introduced in 1994 by Moni Naor and Adi Shamir [1]. The Idea was to share an image secret into  $n$  meaningless shares in such a way that stacking any  $k$  of them together could reveal the secret but any set of less than  $k$  shares could not help in getting any information about the secret. Using this idea, numerous VSS techniques have been proposed in the literature. Most of these schemes propose solutions with emphasis on one of the three concerns apart from information security. These three concerns are (i) Pixel expansion – One pixel is represented by more than one pixel shares (ii) Image contrast – Contrast quality of the retrieved image is compromised (iii) Memory

requirement – Due to multiple shares of an image memory requirement is increased. For a detailed account of VSS schemes addressing one or more of these issues, reference may be made to [12].

In 2005 a multi secret sharing scheme based on Lagrange's interpolation was proposed by Feng *et al* [3]. In this scheme a sharing circle having some properties has been used to generate the shares. Further Wang *et al* proposed a Boolean operation based  $(2, n)$  pair scheme for binary images and  $(n, n)$  pair scheme for gray scale/color images [4]. The  $(2, n)$  scheme is probabilistic and the contrast of the retrieved image is shown to be better than other existing schemes. The  $(n, n)$  pair scheme in [4] uses XOR operations only and is implemented without pixel expansion. Although no unauthorized person could reveal the secret, a little alteration in the image could help detect that the image was tampered for some secret reasons. In 2009 Chang *et al* [5] proposed a verifiable secret sharing scheme using a verification function of the reconstructed secret. Lin *et al* [6] proposed a scheme by breaking the secret randomly into two parts and then applying OR operation between the parts of different secrets. To increase the randomness they also introduced certain camouflaging. The scheme completes the objective without pixel expansion. The secret could be revealed by stacking shares according to the access structure, but the secret contained some noise due to camouflaging. The problem of noise has been successfully removed in [7].

Boolean operation based schemes solve the problem of pixel expansion. However generation of multiple shares leads to significant increase in memory requirement. To overcome this problem, Chen and Wu introduced a multi-secret sharing scheme based on Boolean operations [8] employing  $n+1$  shares for  $n$  secrets. With a similar objective Chen and Tsao proposed a  $(k, n)$  threshold scheme based on random grid [9]. To increase the security and to reduce the number of generated shares, different type of random grids have been proposed in the literature. Chen and Li [10] have used circular grids with different orientations to share the multi secret. It is worthwhile to mention that Chen *et al* [11] have recently introduced a scheme that employs only two shares for four secret images. This leads to remarkable improvement in memory requirement, but the contrast quality of decrypted secret images is very low as compared to the original secrets. Hou *et al* [13] have proposed a block based progressive visual secret sharing scheme. Although the scheme is quite secure contrast of the revealed image is very poor as compared to the original secret image. Further, Li *et al* [14] have proposed a  $(k, n)$  scheme based on the privilege of the participants. Shares are divided in two categories, essential and non-essential based on certain criteria. The secret is revealed only if  $k$ -subset contains a minimum specified number of essential shares.

Most of the schemes discussed above roam around the objectives of balancing the memory requirement, enhancing the security, reducing computational complexity and increasing the contrast of retrieved secret. While one of the objectives is met, other features are deteriorated in general, in all such schemes. The proposed scheme keeps a nice balance between these objectives and features. The proposed scheme uses two levels of encryption which increases the security of the information. In first level we flip some of bit planes which make the shares circularly dependent and random. Further to enhance the security an access structure is employed, that makes the secret

completely random and secure. In Section 2, a brief description of a XOR based VSS scheme used in [4, 8] is given that is used in the subsequent sections. Section 3 is devoted to the proposed scheme of bitplane flipping and Boolean operations. Results are presented in Section 4 followed by conclusion in Section 5.

## 2 Related Work

We begin with the discussion of Boolean operation based single secret sharing scheme introduced in [4].

### 2.1 Single Secret Sharing

Let  $A = \{A(i, j) \in [0, 255], i = 1, \dots, h \text{ and } j = 1, \dots, w\}$  be a secret image of size  $h \times w$ . In order to create  $n$  shares, following steps are performed.

- Step 1. Generate  $n - 1$  random matrices  $R_k, \{k = 1, \dots, n - 1\}$  of size  $h \times w, R_k(i, j) \in [0, 255]$ .
- Step 2. Let the first share be defined as  $S_1 = R_1$ .
- Step 3. Calculate the last share as  $S_n = A \oplus R_{n-1}$ .
- Step 4. Calculate remaining shares as  $S_k = R_k \oplus R_{k-1}, k = 2, \dots, n - 1$ .

Here  $\oplus$  represents the Exclusive - OR (XOR) operation. Fig. 1 shows the original secret image Lena, three generated shares  $S_1, S_2, S_3$  based on the above structure and the retrieved image. In the decryption process the secret can be revealed by applying XOR operation as follows.

$$A = S_1 \oplus S_2 \dots \oplus S_n .$$

Although the scheme is secure enough and provides lossless retrieval, memory requirement is at least twice of the secret. To deal with this issue, Chen and Wu [8] introduced a multi secret sharing scheme based on XOR operation. In the following section, we briefly discuss the scheme presented [8].

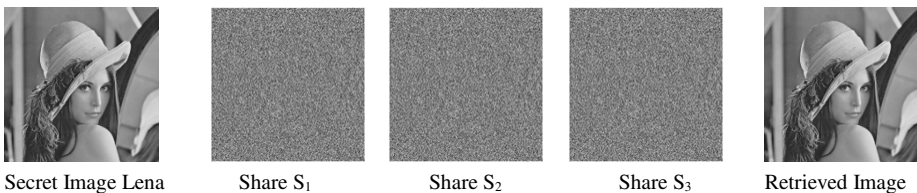


Fig. 1. Secret image Lena, corresponding shares  $S_1, S_2, S_3$  and retrieved image

### 2.2 Multi Secret Sharing

Let the  $n$  secret images be given by  $A_0, A_2 \dots A_{n-1} \{A_k(i, j) \in [0, 255], i = 1, \dots, h \text{ and } j = 1, \dots, w\}$ . Encryption process is performed in the following steps

Step 1. Generate a random matrix  $R_0 \{ R_0(i, j) \in [0, 255], i=1, \dots, h \text{ and } j=1, \dots, w \}$  and assign it to first share  $S_0$ , i.e.,  $S_0 = R_0$ .

Step 2.  $R_k = A_k \oplus S_0 \{ k=1, \dots, n-1 \}$ .

Step 3.  $S_1 = R_1$  and  $S_n = A_0 \oplus R_{n-1}$ .

Step 4. Calculate the remaining shares as  $S_k = R_k \oplus R_{k-1} \{ k=2, \dots, n-1 \}$ .

Decryption is performed as follows.

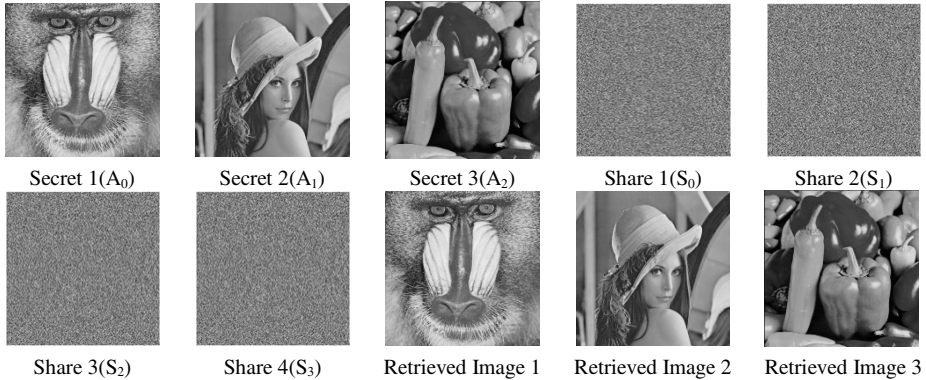
Step 1. First secret can be reconstructed by XORing all the shares  $A_0 = S_1 \oplus S_2 \dots \oplus S_n$ .

Step 2. Generate random matrix  $Rr_1 = S_1$ .

Step 3.  $Rr_k = S_k \oplus Rr_{k-1} \{ k=2, \dots, n-1 \}$ .

Step 4. Reveal  $A_k = Rr_k \oplus S_0 \{ k=1, 2, \dots, n-1 \}$ .

Fig. 2 shows the three secret images  $A_0, A_1, A_2$ , four generated shares  $S_0, S_1, S_2, S_3$  (based on above mentioned steps) and retrieved images (based on above described decryption process).



**Fig. 2.** Secret images  $A_0, A_1, A_2$  corresponding shares  $S_0, S_1, S_2, S_3$  and retrieved images

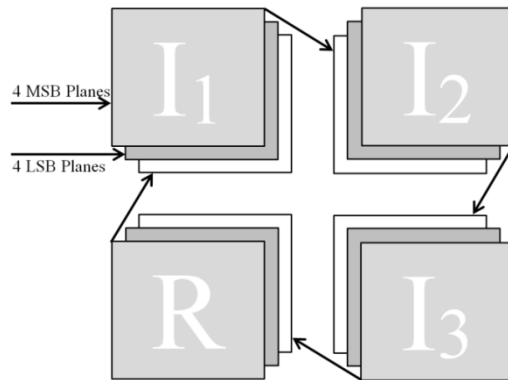
It is a remarkable feature in both the schemes that the image reconstruction is lossless. In case of multi-secret sharing, number of shares are just one more than the total number of secret images. This effectively utilizes the bandwidth without much overhead.

### 3 Proposed Method – Bitplane Flips and XOR-Based Encryption

The proposed scheme provides two levels of encryption. In the first level bit planes of secret images are flipped in a circular order with the help of a random matrix of the same size. To explain the concept of bit plane flipping, consider three secret images  $I_1, I_2$ , and  $I_3$  and a random matrix  $R$ , all of the same size. For the sake of convenience  $R$  is labeled as  $I_4$ . Replace the bit planes in the following order.

- Take four most significant bit (MSB) planes of  $I_4$  and insert them in  $I_1$  in place of four least significant bit (LSB) planes. Push the existing four LSB Planes of  $I_1$  to the place of four MSB planes of  $I_1$ ,
- Take the four overflown MSB planes of  $I_j$  and insert them in  $I_{j+1}$  in place of its four LSB planes. Push the existing four LSB planes of  $I_{j+1}$  to the four MSB planes of  $I_{j+1}, j = 1, 2, 3$ .

The process is explained with the help of a diagram in Fig. 3 below.



**Fig. 3.** Process of generation of immediate shares

After the above structured bit plane flipping process is performed, all the shares appear to be random, call them immediate shares. Now after first level of randomization let the four shares are given by  $A_1, A_2, A_3$  and  $A_4$ . The same process can be adopted for  $n$  secrets also. In general  $i^{th}$  immediate share contains four MSB planes of previous secret (in a circular order) and four LSB planes of itself.

To illustrate the method, we perform encryption on the following three secret images of size  $2 \times 2$  and one random image of same size.

248	51
153	89

45	156
201	2

58	89
129	57

78	9
45	211

$I_1$ 
 $I_2$ 
 $I_3$ 
R

Let us arrange these pixel values in their binary form as follows in arrays as shown below.

```

I1{ 11111000, 00110011, 10011001, 01011001}
I2{ 00101101, 10011100, 11001001, 00000010}
I3{ 00111010, 01011001, 10000001, 00111001}
R{ 01001110, 00001001, 00101101, 11010011}
    
```



As per the proposed algorithm, four MSB planes (underlined) of an image will replace the four LSB planes of next image (shaded) and existing LSB planes will be shifted to the MSB planes of the present image.

$I_1$	<u>1111</u> 1000	0011 <u>0011</u>	<u>1001</u> 1001	<u>0101</u> 1001
$I_2$	<u>0010</u> 1101	<u>1001</u> 1100	<u>1100</u> 1001	<u>0000</u> 0010
$I_3$	<u>0011</u> 1010	<u>0101</u> 1001	<u>1000</u> 0001	<u>0011</u> 1001
$R$	<u>0100</u> 1110	<u>0000</u> 1001	<u>0010</u> 1101	<u>1101</u> 0011

On applying the circular bit plane shifting, the intermediate image is formed as follows.

$$A_1 = \{\underline{1000} \underline{0100}, \underline{0011} \underline{0000}, \underline{1001} \underline{0010}, \underline{1001} \underline{1101}\} \text{ OR } A_1 = \{132, 48, 146, 157\}$$

$$A_2 = \{\underline{1101} \underline{1111}, \underline{1100} \underline{0011}, \underline{1001} \underline{1001}, \underline{0010} \underline{0101}\} \text{ OR } A_2 = \{223, 195, 153, 37\}$$

$$A_3 = \{\underline{1010} \underline{0010}, \underline{1001} \underline{1001}, \underline{0001} \underline{1100}, \underline{1001} \underline{0000}\} \text{ OR } A_3 = \{162, 153, 28, 144\}$$

$$A_4 = \{\underline{1110} \underline{0011}, \underline{1001} \underline{0101}, \underline{1101} \underline{1000}, \underline{0011} \underline{0011}\} \text{ OR } A_4 = \{227, 149, 216, 51\}$$

We next proceed to describe the second level of encryption based on Boolean operations. This is done to enhance the security level of the secret images. Define

$$S_1=A_1, S_2=A_2 \oplus A_1, S_3=A_3 \oplus A_2, S_4=A_4 \oplus A_3;$$

or

$$S_k = A_k \text{ for } k=1 \text{ and} \\ = A_k \oplus A_{k-1} \text{ for } k = 2, \dots, n.$$

It may be noted that the shares now become even more random and security is enhanced.

Decryption process is performed as follows. The immediate shares are generated by the following steps.

$$A_4=S_1 \oplus S_2 \oplus S_3 \oplus S_4 \quad \text{Or } A_3 \oplus S_4 \\ A_3=S_1 \oplus S_2 \oplus S_3 \quad \text{Or } A_2 \oplus S_3 \\ A_2=S_1 \oplus S_2 \quad \text{Or } A_1 \oplus S_2 \\ A_1=S_1$$

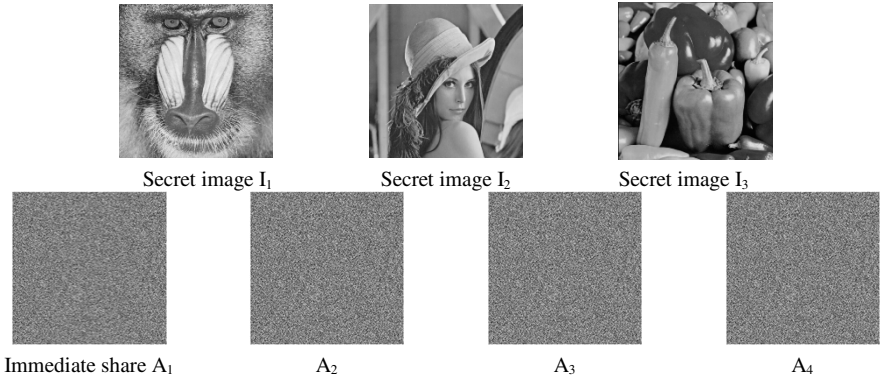
or

$$A_i= S_1 \oplus S_2 \dots \oplus S_i$$

After getting the immediate shares we apply the reverse process for bit plane flipping and the secret images  $I_1, I_2, I_3$  and  $I_4(R)$  are revealed without any loss.

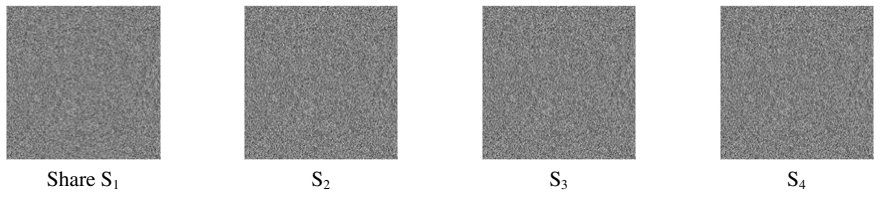
### 4 Results and Comparison

The VSS scheme proposed in section 3 is an efficient multi-level bit plane flipping based multi-secret sharing scheme with no pixel expansion and lossless image retrieval. Fig. 4 shows the original images and the corresponding immediate shares after changing the bit planes for three secrets.



**Fig. 4.** Secret images  $I_1, I_2, I_3$  and immediate shares  $A_1, A_2, A_3, A_4$

Fig. 5 shows the shares generated after applying XOR operation on the immediate shares.



**Fig. 5.** The final generated shares of the secret images

A comparison of features of the proposed scheme is compared with the existing techniques in Table 1 below.

**Table 1.** Performance comparison of the proposed scheme with existing methods

	Wang et al.[4]	Chen et al.[8]	Proposed
Sharing Capacity	1/n	1/(n+1)	1/(n+1)
Level of Security	One Level	One Level	Two levels
Codebook required	No	No	No
Reconstruction	Lossless	Lossless	Lossless
Pixel Expansion	No	No	No

## 5 Conclusion

An  $(n, n)$  multi secret sharing scheme based on bit plane flips and Boolean operations is proposed. The scheme retains the quality measures such as no pixel expansion, double level of security, computational ease, and lossless secret image retrieval and effective use of multiple shares for multi-secret sharing. Initially a random matrix is generated; secret images along with this random matrix interchange some biplanes in circular order to make the intermediate shares which appear to be random. Since the shares are circularly dependent on each other secret cannot be revealed unless all the shares are available. Further, to increase the security a second level of encryption based on XOR operation is applied. Decryption is applied by reversing the process.

## References

1. Naor, M., Shamir, A.: Visual Cryptography. In: Proc. of 13th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Perugia, Italy, pp. 1–12 (1994)
2. Shamir, A.: How to Share a Secret: Comm. ACM 22(11), 612–613 (1979)
3. Feng, J.B., Wu, H.C., Tsai, C.S., Chu, Y.P.: A New Multi-Secret Images Sharing Scheme using Lagranges Interpolation. J. System and Software 76, 327–339 (2005)
4. Wang, D., Zhang, L., Ma, N., Li, X.: Two Secret Sharing Schemes based on Boolean Operations. J. Pattern Recognition 40, 2776–2785 (2007)
5. Chang, C.C., Lin, C.C., Le, T.H.N., Le, H.B.: Sharing a Verifiable Secret Image using Two Shadows. J. Pattern Recognition 42, 3097–3114 (2009)
6. Lin, T.L., Horng, S.J., Lee, K.H., Chiu, P.L., Kao, T.W., Chen, Y.H., Run, R.S., Lai, J.L., Chen, R.J.: A Novel Visual Secret Sharing Scheme for Multiple Secrets Without Pixel Expansion. J. Expert Systems with Applications (2010) (Article in Press)
7. Lin, P.Y., Chan, C.S.: Invertible Secret Image Sharing with Steganography. J. Pattern Recognition Letters 31, 1887–1893 (2010)
8. Chen, T.H., Wu, C.S.: Efficient Multi-Secret Image Sharing Based on Boolean Operations. J. Signal Processing. 91, 90–97 (2011)
9. Chen, T.H., Tsao, K.H.: Threshold Visual Secret Sharing by Random Grids. J. Systems and Software. 84, 1197–1208 (2011)
10. Chen, T.H., Li, K.C.: Multi-Image Encryption by Circular Random Grids. J. Information Sciences. 189, 255–265 (2012)
11. Chen, T.H., Tsao, K.H., Lee, Y.S.: Yet Another Multiple-Image Encryption by Rotating Random Grids. J. Signal Processing 92, 2229–2237 (2012)
12. Yin, Z.X., Lin, C.C., Chang, C.C.: Image Sharing with Steganography and Authentication. In: Cimato, S., Yang, C.N. (eds.) Visual Cryptography and Secret Image Sharing. CRC Press, New York (2012)
13. Hou, Y.C., Quan, Z.Y., Tsai, C.F., Tseng, A.Y.: Block Based Progressive Visual Secret Sharing. J. Information Sciences. 233, 290–304 (2013)
14. Li, P., Yang, C.N., Wu, C.C., Kong, Q., Ma, Y.: Essential Secret Image Sharing Scheme with Different Importance of Shadows. J. Visual Communication and Image Representation 24, 1106–1114 (2013)

# Large-Scale Propagation Analysis and Coverage Area Evaluation of 3G WCDMA in Urban Environments Based on BS Antenna Heights

Ramarakula Madhu<sup>1</sup> and Gottapu Sasi Bhushana Rao<sup>2</sup>

<sup>1</sup> Department of Electronics and Communication Engineering, JNTU College of Engineering, JNTUK, Kakinada – 533003, A.P., India

<sup>2</sup> Department of Electronics and Communication Engineering, College of Engineering, Andhra University, Visakhapatnam – 530003, A.P., India  
madhu.ramarkula@gmail.com

**Abstract.** The design of an efficient cellular system requires a detailed understanding of propagation characteristics of a mobile radio channel. A signal propagating through a channel undergoes two kinds of variations; large-scale and small-scale. Path loss is generally the most important parameter predicted by large-scale variations. In this paper a Large-scale propagation model is described for urban environments, to evaluate the coverage area of 3G WCDMA cellular system based on BS antenna heights. COST 231 Hata model is presented for large-scale analysis of WCDMA systems. The results are evaluated with different parameters; coverage radius, antenna heights and  $E_b/N_0$ . It is observed that, path loss decreases from 167.55 dB to 155.67 dB as the BS antenna height increases from 30 m to 120 m for 8 km coverage radius in a WCDMA system for urban environments.

**Keywords:** Path loss, WCDMA, COST 231 Hata model, coverage radius.

## 1 Introduction

The radio channel places fundamental limitations on the performance of wireless communication systems. The transmission path between the transmitter and the receiver can vary from a simple line-of-sight to one that is severely obstructed by buildings, mountains, and foliage. Modeling the radio channel has historically been one of the most difficult parts of mobile radio system design, and is typically done in a statistical fashion, based on measurements made specifically for an intended communication system or spectrum allocation.

There are two types of variations in a radio signal, *Shadowing* and *Multipath*, when it undergoes through a channel. Shadowing is the gradual decrease of mean signal strength as the signal travels over long distances. Since this variation is slow and are evaluated over a long distance, the variations due to shadowing are termed as large-scale variations. This depends on the presence of obstacles in the signal path, the distance from the base station to the mobile station, frequency of operation, BS and

MS antenna heights. Multipath is a single wave undergoing reflections at various channel objects and reaching the receiver at different time lags and thereby resulting in interference. Since these fluctuations occur over small distances or short time intervals, therefore multipath variations are termed as small-scale variations.

Wideband Code Division Multiple Access(WCDMA) plays a crucial role in the third-generation (3G) cellular systems [1]. WCDMA is a step further in the CDMA technology which is a wide band spread-spectrum channel access method, which utilizes the Direct-Sequence Spread Spectrum (DSSS). It operates in two modes; Frequency Division Duplex (FDD) and Time Division Duplex (TDD). In FDD, two different frequency bands, separated by a guard band are used; one for the uplink and other for the downlink transmission. The spectrum allocation in FDD mode is for uplink 1920-1980MHz and downlink 2110-2170MHz. In TDD, on the other hand, the same frequency band is used for transmission in both directions. The spectrum allocation in TDD mode is 1900-1920MHz and 2010-2025MHz. It uses a 5 MHz wide radio signal and a chip rate of 3.84 Mc/s and supports higher data rates up to 2 Mb/s [4]. Face-to-face video calling, impressive range and versatility of the mobile internet, visual greetings and multimedia on demand are some applications of 3G cellular system.

In order to investigate the propagation characteristics of a mobile channel, several models have been proposed and most of them generate statistics which are very near to the practical values obtained. The empirical propagation channel models are used to model the Large-scale variations and the short term variations by a statistical channel model. In this paper COST 231 Hata model, an empirical propagation model used for large-scale variations is presented for the analysis of coverage area evaluation of 3G WCDMA cellular systems in urban area.

## 2 Large-Scale Propagation Models

The mean received signal strength of a WCDMA signal at any point depends on its distance from the transmitter, the carrier frequency, the type of antennas used, antenna heights, atmospheric conditions, etc. It may also vary because of terrain and clutter such as hills, buildings and other obstacles. This type of signal variation is slow, which is observable over relatively long distances, i.e., a few tens or hundreds of wavelengths of the *radio frequency* (RF) carrier, and known as large-scale variations or shadowing [6].

The mobile radio channel is usually evaluated from 'statistical' propagation models: no specific terrain data is considered and channel parameters are modeled as stochastic variables. The mean signal strength for an arbitrary transmitter-receiver (T-R) separation is useful in estimating the radio coverage of a given transmitter whereas measures of signal variability are key determinants in system design issues such as antenna diversity and signal coding [2]. Electromagnetic waves propagate through environments where they are reflected, scattered and diffracted by walls, terrain, buildings and other objects. The ultimate details of this propagation can be obtained by solving Maxwell's equations with boundary conditions that express the

physical characteristics of these obstructing objects. This requires the calculation of the Radar Cross-Section (RCS) of large and complex structures. Since these calculations are difficult and necessary parameters are often not available, approximations have been developed to characterize signal propagation without resorting to Maxwell’s equations.

The free space propagation model is the basic propagation model which implies equal radiation in all directions from the radiating source and propagation to an infinite distance with no degradation. This model does not consider the effects of propagation over ground. When a radio wave propagates over ground, some of the power will be reflected due to the presence of ground and then received by the receiver. Therefore the path loss at any point depends on a number of factors which include the distance from the transmitter, frequency, atmospheric conditions etc. This dependence is complex and is very difficult to describe with exact mathematical expressions. A number of propagation models based on empirical formulae are available that can be used to estimate the path loss and signal distribution. These models are based on experimental data obtained in numerous measurements of loss characteristics in conditions of open and rural environments for rough and hilly terrain. Hata-Okumura and COST 231 Hata propagation model adapted to the description of large-scale wave propagation in obstructive (clutter) conditions.

The Hata-Okumura model predicts the path loss between two stations for large distances. Hata obtained mathematical expressions by fitting the empirical curves provided by Okumura. Hata-Okumura model can be used for carrier frequencies ranging from 150 MHz to 1500 MHz. It is efficiently used to estimate the Large-scale characteristics of 2G GSM systems since the operating frequency band for uplink is 890-915 MHz and for downlink is 935-960 MHz.

### 3 COST 231 Hata Model

The limitation of Hata-Okumura model is the range of carrier frequencies which vary only from 150 MHz to 1500 MHz. It is not valid for 2G CDMA and 3G WCDMA systems, since the carrier frequencies used in these systems are in the range of 2 GHz. COST 231 Hata model, which is the extension of Hata-Okumura model is used to model the Large-scale variations in CDMA based systems. This model can be used with the carrier frequencies from 1500 MHz to 2000 MHz. The path loss according to COST 231 in urban scenarios is given by [7]

$$L_p = [46.3 + 33.9 \log(f) - 13.82 \log(h_t) - a(h_m)] + [44.9 - 6.55 \log(h_t)] \log(d) + C_m \text{ dB} \tag{1}$$

where,  $f$  is the frequency in MHz,  $d$  is the distance between BS and MS in km,  $h_t$  is the BS antenna height above the ground level in meters and  $C_m$  is the correction factor. The factor  $C_m$  is given as 3 dB for urban and 0 dB for rural and suburban environments.

The factor  $a(h_m)$  for urban environments is defined as

$$a(h_m) [\text{dB}] = 3.2 [\log(11.75 h_m)]^2 - 4.97 \tag{2}$$

The model is valid for the following range of input parameters of Table 1.

**Table 1.** Model parameter range for COST 231 model

Parameter	Range
Frequency	1500 MHz – 2000 MHz
$h_m$	1-10 m
$h_t$	30-200 m
Distance (d)	1 – 20 km

### 4 Coverage Area Determination

Link budget is one of the networks planning process, which helps to determine the required coverage and quality of service requirement in the network [3]. It is used in calculating the maximum allowable propagation path loss between two radio links based on the output power from the transmitter. The link budget considers all the gains and losses that a radio wave experiences along the path from transmitter to receiver. The maximum allowable path loss from link budget analysis is given as

$$L_p = EIRP - \text{total noise power} + G_p - E_b/N_0 - \text{receiver antenna gain} - \text{fast fading margin} \tag{3}$$

where,  $G_p$  is the processing gain and  $E_b/N_0$  is the energy per bit to noise spectral density ratio.

From Eq. (1), (2) and (3), a relationship can be expressed for coverage and data rates in urban case, for  $f=2000$  MHz,  $h_t=20$  m,  $C_m=3$  and  $h_m=5$  m,

$$142 \cdot 17 + 36.37 \log(d) = EIRP - \text{total noise power} + G_p - E_b/N_0 - \text{receiver antenna gain} - \text{fast fading margin} \tag{4}$$

The processing gain  $G_p$ , of a cellular system is defined as

$$G_p = \frac{\text{Chip Rate}}{R_b} \tag{5}$$

where,  $R_b$  is the data rate.

The coverage area can be calculated by using the cell range, d. The coverage area for one cell in hexagonal configuration can be estimated with

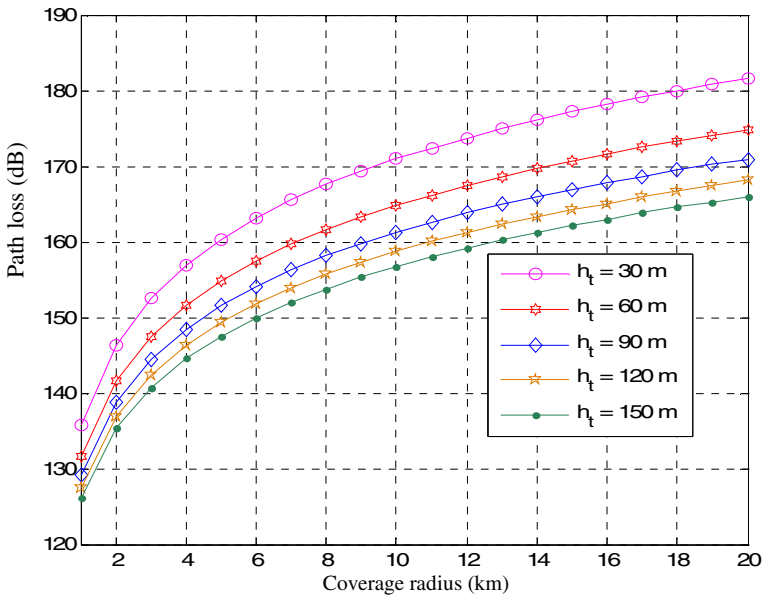
$$A_{\text{coverage}} = \frac{3\sqrt{3}}{2} \times d^2 \text{ m}^2 \tag{6}$$

where,  $A_{\text{coverage}}$  is the coverage area and  $d$  is the maximum cell range.

### 5 Results and Discussion

The propagation characteristics of a 3G WCDMA system are analyzed under Large-scale variations, which are mainly used to predict the coverage area of a cellular network in urban environments. Path loss is the main characteristic parameter for analyzing the Large-scale variations. The path loss analysis has been done using COST 231 Hata model for urban environments. Link budget analysis is used for coverage area determination by estimating the maximum allowable propagation path loss between two radio links.

The path loss in a WCDMA system for  $f = 2000$  MHz, and  $h_m = 5$  m at different base station antenna heights in urban area is shown in Fig. 1. It is observed that for a given cell radius, the path loss decreases with an increase in the height of the base station antenna.



**Fig. 1.** Coverage radius vs. Path loss at different BS antenna heights for urban area in a WCDMA system

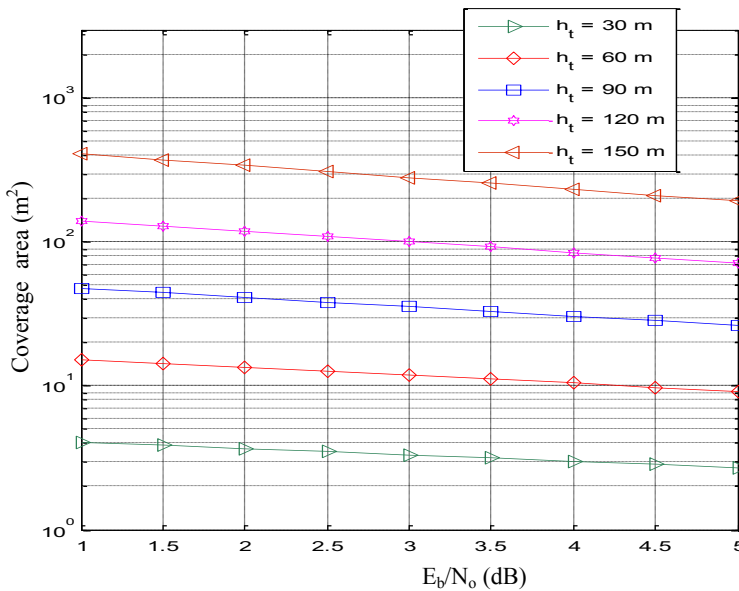
The values of path loss in urban environment for different base station antenna heights are listed in Table 2. The path loss decreases from 167.55 dB to 155.67 dB as the height of the base station increases from 30 m to 120 m for 8 km coverage radius in a WCDMA system under urban environments.



**Table 2.** Path loss in urban area for different BS antenna heights in a WCDMA system

Coverage radius (Km)	Path Loss (dB) in Urban area				
	$h_t = 30$ m	$h_t = 60$ m	$h_t = 90$ m	$h_t = 120$ m	$h_t = 150$ m
2	146.35	141.59	138.81	136.84	135.31
4	156.95	151.60	148.47	146.25	144.53
6	163.15	157.46	154.13	151.76	149.93
8	167.55	161.61	158.14	155.67	153.76
10	170.97	164.83	161.25	158.70	156.73
12	173.76	167.47	163.79	161.18	159.16
14	176.11	169.69	165.94	163.27	161.21
16	178.16	171.62	167.80	165.09	162.98
18	179.96	173.32	169.44	166.69	164.55
20	181.57	174.85	170.91	168.12	165.95

The coverage area in sq. m for different base station antenna heights for a WCDMA system at 2000 MHz in urban areas is shown in Fig. 2. The area covered by a base station decreases with an increase in  $E_b/N_0$ .



**Fig. 2.**  $E_b/N_0$  vs. coverage area for different BS antenna heights for urban area in a WCDMA system

The Fig. 2 represents that, for a given  $E_b/N_0$  value the coverage area of a cell increases with an increase in base station antenna height. The values are tabulated in Table 3.

**Table 3.** Coverage area with different BS antenna heights in urban area in a WCDMA system

$E_b/N_0$ (dB)	Coverage area (m <sup>2</sup> ) for urban environment				
	$h_t= 30$ m	$h_t= 60$ m	$h_t= 90$ m	$h_t= 120$ m	$h_t= 150$ m
1.0	4.03	15.27	47.81	140.95	411.30
1.5	3.83	14.33	44.37	129.46	373.91
2.0	3.64	13.44	41.18	118.91	339.92
2.5	3.46	12.61	38.22	109.22	309.02
3.0	3.29	11.83	35.47	100.32	280.93
3.5	3.13	11.10	32.92	92.14	255.39
4.0	2.97	10.42	30.55	84.63	232.18
4.5	2.82	9.78	28.35	77.73	211.07
5.0	2.68	9.17	26.31	71.40	191.88

For a 2.5 dB  $E_b/N_0$ , as the height of the base station increases from 30 m to 120 m, the coverage area increases from 3.46 m<sup>2</sup> to 109.22 m<sup>2</sup> for a WCDMA system in urban environments.

## 6 Conclusion

In this paper, the modeling of a WCDMA signal under Large-scale variations and coverage area determination in urban environments is presented. COST 231 Hata model is described for the analysis of path loss at the receiver in 3G WCDMA systems in terms of coverage radius, and base station antenna heights. It is observed that the propagation path loss increases with an increase in coverage radius and a low path loss results in a high base station antenna height. As the coverage radius increases from 4 km to 16 km, the path loss increases from 134.76 dB to 155.07 dB in urban environments.

Link budget analysis is described for the determination of coverage area in a WCDMA system. The link budget considers all of the gains and losses that a radio wave experiences along the path from transmitter to receiver. The simulation results are obtained to observe the dependence of coverage area with  $E_b/N_0$  and base station antenna heights. It is observed that for a given  $E_b/N_0$  values the coverage area of a cell increases with an increase in base station antenna height. For a 2.5 dB  $E_b/N_0$ , as the height of the base station increases from 30 m to 120 m, the coverage area increases from 3.46 m<sup>2</sup> to 109.22 m<sup>2</sup> for a WCDMA system in urban environments.

## References

1. Mishra, A.R.: Fundamentals of Cellular Network and optimization 2G / 2.5G / 3 G evolution to 4G. John wiley & Sons, Ltd (2004)
2. Smith, C., Collins, D.: 3G Wireless Networks. McGraw-Hill Book Co. (2002)
3. Lee, C.Y.: Wireless and Cellular Communications, 3rd edn. copyright©2006 (1995). 1887 McGraw-Hill companies, Inc. all rights reserved
4. Rao, G.S.: Mobile Cellular Communication. In: Pearson International (2012)
5. Hata, M.: Empirical formula for Propagation loss in Land Mobile Radio Services. IEEE Trans. Veh. Technol. VT-29(3), 317–325 (2008)
6. Karim, M.R., Sarraf, M.: W-CDMA and cdma2000 for 3G Mobile Networks. McGraw-Hill Publications (2004)
7. Parsons, J.D.: Mobile radio propagation channel. Wiley & Sons (2000)

# Inclination and Pressure Based Authentication for Touch Devices

K. Rajasekhara Rao, V.P. Krishna Anne, U. Sai Chand,  
V. Alakananda, and K. Navya Rachana

Department of Computer Science & Engineering, KL University, Guntur, India

**Abstract.** This paper explores the methodologies that have been implemented in behavioral biometrics when used for the purpose of authentication. With focus on keystroke dynamics and pressure as major factors while inputting password we detail a comprehensive study on the research so far. We also propose a new method which makes use of the accelerometer sensor present in most of the mobile devices along with touch pressure applied by a user to establish an authentication mechanism. Two models that effect the key parameters in establishing a behavior biometric are proposed and compared with respect to the results they produced when tested on a prototype we developed.

**Keywords:** Keystroke Dynamics, Touch pressure, Security, Behavioral Biometrics, Accelerometer.

## 1 Introduction

Keystroke dynamics and Behavioral biometrics have been intensively researched since 1997 [1] with the inception of the idea to use behavior as a biometric that aids in authentication in order to provide robust security. The introduction of keystroke dynamics provided a new paradigm to explore the possibilities of providing a better biometric mechanism that is inexpensive and effective. Revolutionizing technologies and increasing presence of confidential data per individual is a driving factor for the development of this highly accessible technology. The initial idea had been developed through intensive research evolving and catching up with new technologies like touch screens [2] and now provides a lucrative mechanism if implemented on current devices to enhance the users sense of security. The flexibility of this idea lies with its ability to be implemented at various levels and on various scales to provide security in wide range of applications. We briefly re-introduce the behavior biometric via keystroke dynamics by exploring its history and bring you to its current form which is taking leverage of present technological trends. The enhancement that we propose to the existing use is detailed through the experiment and statistical conclusions we derived with the help of a working prototype of the proposed system.

## 1.1 Inception

Keystroke dynamics first introduced by Gianes in 1980 and later published as a comprehensive study by Rick Joyce and Gopal Gupta [1]. Latency timings were a crucial factor that lacked sample data and the conclusions were based on rather small amount of data at that time. Observing data from 7 people in Gianes experiment Leggett, Umpress and Williams recorded 5.5 % False alarm rate and 5% Importer Pass rate by repeating the experiment on 17 programmers. The experiment was to test record the typing patterns of the subjects and later authenticate them by matching their previous pattern with the help of a mean reference signature. Joyce represented the mean reference signature as follows:

$$M=M_{\text{username}}+M_{\text{password}}+M_{\text{first name}}+M_{\text{last name}} \quad (1)$$

In the initial publication [1] a toolkit developed using C++ and X-view Library routines for GUI was presented. This toolkit by Fabian Monrose and Aviel Rubin aided their keystroke experiment by acting as a recognition engine that generated graphs for Matlab and Gunplot for determining the classifiers to analyse keystroke dynamics of the user. In their research the classification algorithms used for recognition were Euclidean distance measure, weighted probability and weighted probability measure elaborated in [1]. This initial research concluded that the recognition score for left hand features was lower than the right handed features and Keystroke dynamics exploited certain weaknesses in protocols like PGP (Pretty Good Privacy).

## 1.2 Re-Visiting Not What You Type But How You Type

The key terms that need to be noted while exploring keystroke dynamics fall under two categories parameters that measure the behavior and factors that determine the accuracy of the methodology used.

The keystroke behavior is studied based on the 3 parameters key press time, key dwell time and flight time. Key press time is the time taken by a user to press and release a key while dwell time measures the time a user lingers on a particular key without releasing it. Flight time on the other hand is the actual time a user takes to move from one key to another. Different methods proposes making use of these parameters to analyze user's typing patterns and use them for authentication purpose.

Factors that determine the accuracy of the keystroke dynamics algorithm used are False Acceptance Rate (FAR), False Rejection rate (FRR) and Equal Error Rate (ERR).

False Acceptance Rate (FAR) which indicates the probability that the system will erroneously grant access to an intruder. False Rejection Rate (FRR) which is the probability that the system will wrongly deny access to a legitimate user. The point at which both FAR and FFR are equal is denoted the Equal Error Rate (ERR). The EER makes it easier to compare the performance of various biometric systems or classifiers, and the lower its value the better the classifier.

The determining factors and basic parameters remain the same throughout the evolution of Keystroke based behavioral biometrics for authentication no matter what new technology is made use for employing this innovation.

## 2 New Implementations of the g’olden’ Idea

Taking leverage of the new and fast changing technologies is done most rapidly by this idea which had its foundations way back in 1980’s. Researchers tried to capture user’s behavior in several ways apart from keystrokes alone. The idea was to keep it simple and far from utilizing expensive hardware so that it can be available to common users who have the right to protect their data from ever increasing threats and data thefts. On the other hand this innovation has the flexibility and potent to be used for high end protection by utilizing just the same resources. Mobiles provided just the required platform for collecting and analyzing behavior to use it as a biometric without any additional resource requirement. The inbuilt sensors provided the data required [3]. Shaking for the purpose of authentication provided a new area for obtaining behavioral biometric for the purpose of authenticating devices. The method provided high entropy from an attacker’s point of view. This also provided an unobtrusive mechanism making it easier to use as well.

### 2.1 Deploying Sensors

The introduction of tri-axis accelerometer enhanced the behavior collection method by providing most encouraging results [5]. The accelerometer readings combined with the uWave study helped achieve 3% FAR, 3% FRR and 3% EER as published in [5]. However leak in gesture pushing the ERR to 10% was a slight drawback in this methodology. Needless to say the study provided a whole new solution space by proposing two most important types of authentication mechanisms Non critical authentication with 98% accuracy and Critical authentication with 3% EER.

In non-critical authentication uWave identifies best matching template from multiple templates created by different users. In critical authentication it matches the distance between gesture and the template gesture representing the claimed user identity (classic binary classifier). The prototype of uWave takes input from Wii Remote with application developed using visual C#.

**Table 1.** Methods with their determining factors

Method	Factors	Remarks
Latency timings	FRR 5.5% , FAR 5%	Small data under analysis.
Tri-Axis accelerometer	FRR 3%,FAR 3%	Gesture leak pushes EER to 10%.
Touch Technology	EER as low as 1%	99% accuracy achieved

## 2.2 The Touch Effect

Re-inventing the basic keystroke idea touch technology provided a whole new discriminating information on user's behavior based on finger pressure [4] [6]. Similar to the dwell time parameter the touch pressure helped determine user's identity with a 1% EER making it almost 99% accurate compared to previous models. The technology being utilized is also widely used and is available to every common user. Multi touch devices further strengthens the behavior model by increasing the space for user input. This feature was intensively explored in [6] by experimenting on a multi touch based android device.

Table 1 provides a summary of results observed in various methods implemented till date. A recent publication [9] introduced cover pad and other techniques to design leak resistant passwords which strengthen the security altogether.

We experimented by implementing a blend of all these ideas used for collecting behavioral biometrics by proposing a new system that makes use of both accelerometer readings and touch pressure for the purpose of authentication. We detail the application used and results obtained in the following sections. Two models are being introduced namely intended behavior and natural behavior model.

## 3 Introducing: Not What You Type But How You Hold and Tap

Inspired by the above variations of behavior biometrics we came up with an idea to blend the sensor and touch technologies to create a simple but robust system for authentication. We synthesized a prototype for the purpose and present the basic results of this collaborative approach. The basic idea is to use both the accelerometer readings and touch pressure for the purpose of capturing user's behavior.

### 3.1 The Method

The behavior of a user while entering passwords on touch based devices can be perceived in two ways the way they hold the device and the way they touch the device. Our idea makes use of this fact and captures user's behavior by collecting data in the form of accelerometer readings and touch pressure readings.

Accelerometers are used in tablet computers and digital cameras so that images on screens are always displayed upright. Single and multi-axis models of accelerometer are available to detect magnitude and direction of the proper acceleration (org-force), as a vector quantity, and can be used to sense orientation (because direction of weight changes), coordinate acceleration (so long as it produces g-force or a change in g-force), vibration, shock, and falling in a resistive medium (a case where the proper acceleration changes, since it starts at zero, then increases). The accelerometer readings can be obtained using several API's and we collected the data to determine how the user is holding the device and its tilt. Almost all the mobile devices now are equipped with accelerometer or at least a tilt sensor. This makes it easier to implement this method without any external hardware requirement.

As for the touch pressure unlike the multi touch implementation we utilized a variation which measures the tap pressure on buttons present on the touch screen. This tap pressure is similar to the dwell time which is the most sensitive parameter in determining the distinction between users.

### 3.2 Proposed Models

We present two models that determine the results of the experiment. Intended behavior model and natural behavior model.

Intended behavior model aims at capturing the voluntary behavior of the user where in the user is not only conscious of the pressure and inclination he is using but intends to apply the deviation which is not natural to him. The user tries to remember the pressure and inclination of the device while registering and re-creates it intentionally to authenticate himself.

Natural behavior is the congenital feature of the user that is recorded through repeated registration or calibration and the system defines a pattern of the user's behavior. The user is aware of the process but need not re-create something from memory but use his natural way which the system recognizes.

The two models effect the results of authentication and the second model can be considered to be more dynamic. A comparison of the models is depicted in Table 2.

**Table 2.** A comparison of Intended and Natural behavior models for authentication

<b>Intended Behavior Model</b>	<b>Natural Behavior Model</b>
User inputs password by recreating the behavior with which he registered from memory.	System recognizes pattern of user behavior by constantly recording and calibrating itself by user's input that is sub-conscious.
Requires little training of the end user.	No particular training is necessary.
Highly suitable for high-end security like military applications where user intends to mislead an attacker.	Applicable for regular use and is easy to use.

## 4 Prototype for Proposed Behavior Capture

A working prototype that collects user's touch pressure and device inclination readings for the purpose of determining the effect of the parameters while authenticating is presented. This prototype developed using jQuery and apple web API can be used to study the variation different users can create while authenticating themselves based on their touch pressure and the way they hold the device.

### 4.1 Prototype Walkthrough

The prototype has two parts one that registers user by requesting input twice to calibrate and set a range for the distinctive parameters and the other that is used for



comparing the authentication result based on the users current input. The basic modules are illustrated in fig 1.

Registration module asks the user to tap in the password twice while holding the device to save the values and set a range for the touch pressure and device inclination parameters. The structure of the password is a typical one that consists of 5 special characters that can be permuted with infinite repetitions. The password space may seem to be reduced in contrast to the normal one but the hidden touch pressure and device inclination can provide a much greater password space. The use of special characters is to make the password difficult to guess except for brute force and impossible to crack using brute force due to the hidden parameters.

Login module requires the user to either recreate the password input behavior from memory (Intended Behavior Model) or tap in the password just in a regular way reflecting behavior (Natural Behavior Model). The login prompts the registered ranges and the current achieved values which help in the study of the suggested parameters and their effects on user authentication.

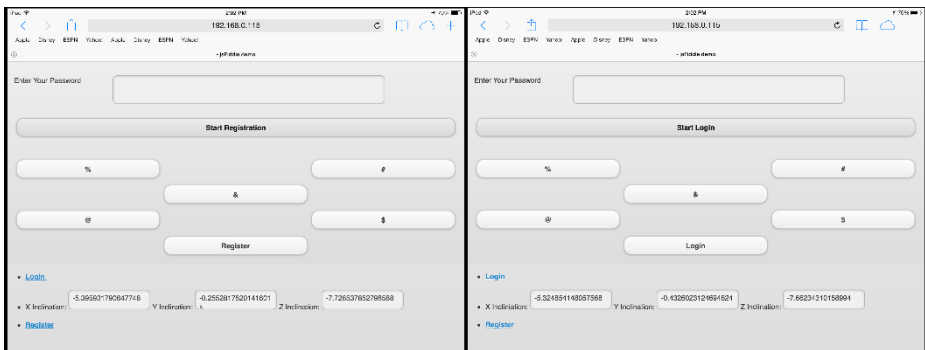


Fig. 1. The basic modules of the prototype

## 4.2 Intended Behavior and Prototype Result

The prototype results when utilized for studying intended behavior model produced results as depicted in fig 2. The experiment was conducted by explaining the model and procedure to 4 team members and recording the results. The user consciously taps on desired password by holding the device at an intended angle and by applying a manually measured or stopwatch measured touch pressure. Manually measured touch pressure is achieved by counting the time in seconds for how long a button is touched and the same is also observed with a variation by using a stopwatch to measure the time. The intended angle is achieved by providing a visual aid of measure being displayed on the prototype itself.

## 4.3 Natural Behavior and Prototype Result

The natural behavior model was tested with the help of subjects who are new to the system by explaining to them how to use it. The usage details are only disclosed

without mentioning how their behavior is captured. The users were made to practice several times by registering without their conscious on the device inclination but only letting them reflect their natural behavior with the system. The results are depicted in fig 3. The readings presented are collected from 5 users completely new to the system and each varied in their own way most distinctively with respect to the inclination reading while touch pressure only slightly varied.



Fig. 2. Results of Intended Behavior model on the prototype



Fig 3. Results of Natural Behavior on the prototype

## 5 Final Views on the Idea

The gathered results suggest that results obtained from both models are greatly influenced by the inclination of the device. The Intended behavior model with proposed method gave long touch pressure readings and more accurate acceptance range when measured using a stop watch. The inclination of the device when revealed while inputting password provided higher chances for legitimate user being authenticated in intended authentication whereas in natural behavior model it can be susceptible to attack by an outsider.

The prototype could simulate the two models proposed and can be used for further study on large sets of data to come up with more accurate results on how the proposed parameters help in acquiring distinctive information about the user in order to authenticate him.

Some average values recorded by the testing on prototype have been summarized in Table 3.

**Table 3.** Summary of the experimental results

Parameters/Factors	Intended Behavior	Natural Behavior
Device Inclination	100% recreation accuracy.	99% falls within registered range.
FRR	0 with consideration to inclination.	1% chance of error
FAR	1% even on gesture leak.	2-3% even on gesture leak.

## 6 Extending the Method

It is to be noted that the results recorded are based on the minimal data set considered on repeated testing on the primitive prototype. Before full-fledged implementation of the idea a study on much larger user base and input data on a much sophisticated working model would be more helpful. The method is sure to provide robust security over different levels and across several applications based on the type of the model utilized. The potent of this method can be fully realized by strengthening the prototype into a full-fledged application that can help gather larger amounts of data and one that can establish a pattern by calculating different classifiers to capture user's behavior to an extent at which it can be used as a biometric.

## References

1. Monrose, F., Rubin, A.: Authentication via Keystroke Dynamics (1997)
2. Davrondzhon, G., Kiris, H., Torkjel, S.: Biometric Gait Authentication Using Accelerometer Sensor. *Journal of Computers* 1(7) (2006)
3. Rene, M., Hans, G.: Shake Well Before Use: Authentication based on accelerometer data (2007)
4. Hataichanok, S., Patrasinne, B.: User Authentication Using Combination of Behavioral Biometrics over the Touch Pad acting like Touch Screen of Mobile Devices. In: *International Conference on Computer and Electrical Engineering* (2008)
5. Liu, J., Zhong, L., Wickramasuriya, J., Vasudevan, V.: User Evaluation of Light Weight User Authentication with a Single Tri-Axis Accelerometer (2010)
6. De Luca, A., Hang, A., Brudy, F., Lindner, C., Hussmann, H.: Touch me once and I know it's you! Implicit Authentication based on touch screen patterns (2012)
7. Angulo, J., Wästlund, E.: Exploring Touch Screen Biometrics for User Identification on Smart Phones. In: Camenisch, J., Crispo, B., Fischer-Hübner, S., Leenes, R., Russello, G. (eds.) *Privacy and Identity 2011. IFIP AICT*, vol. 375, pp. 130–143. Springer, Heidelberg (2012)
8. Li, L., Zhao, X., Xue, G.: Unobservable Re-Authentication for Smartphones (2013)
9. Yan, Q., Han, J., Li, Y., Zhou, J., Deng, R.H.: Designing Leakage-Resilient Password Entry on Touch Screen Mobile Devices. *ACM 978-1-4503-1767* (2013)

# Anticipated Velocity Based Guidance Strategy for Wheeled Mobile Evaders Amidst Moving Obstacles in Bounded Environment

Amit Kumar and Aparajita Ojha

Computer Science and Engineering,  
PDDM Indian Institute of Information Technology, Design and Manufacturing Jabalpur  
Dumna Airport Road, 482005, Jabalpur, India  
{amitku,aojha}@iiitdmj.ac.in

**Abstract.** This paper is concerned with a class of pursuit-evasion game problems (PEGs) amidst moving obstacles in a bounded environment. We concentrate on the evader's strategy, taking into account the following challenges (i) Pursuer and evader are non-holonomic wheeled mobile robots and the evader is slower than the pursuer (ii) Pursuer follows proportional navigation (PN) law (iii) Geometry of the environment is not known to the players, *a priori*. We propose an efficient evader-centric anticipated velocity based guidance strategy for a single evader and a single pursuer. Pursuer's trajectory is anticipated at each step by the evader using quadratic polynomial interpolation. Aim of the evader is to escape interception with the pursuer for maximum possible time. A recently introduced reciprocal orientation method is employed to avoid collision with other moving vehicles in the environment. Efficiency of the proposed strategy is analyzed with respect to the interception time and the distance travelled by the two players.

**Keywords:** Pursuit-evasion game, Proportional navigation, Collision avoidance.

## 1 Introduction

Pursuit-evasion is an interesting class of problems in the field of object tracking, motion planning and control. PEGs have emerged from real life problems and find interesting applications in military surveillance, search and rescue operations, missile combat and path planning in adversarial environment. A Pursuit-evasion game problem deals with conflicting interest of the evader and its pursuer(s) and may have kinematic / dynamic constraints. In the present study, both the pursuer and the evader are chosen to be non-holonomic wheeled mobile robots. Pursuit-evasion games in various geometric settings have been studied extensively to capture an evader under any circumstance ([1], [2], [3], [4]). Most of the pursuer-centric guidance laws proposed by researchers typically fall into one of the following categories – line-of-sight (LOS) pursuit, proportional navigation guidance (PNG), optimal linear control

and differential methods ([5], [6], [7], [8]). PNG is the most common and efficient class of methods used to intercept maneuvering target. True PNG, Ideal PNG, Augmented Ideal PNG (AIPNG) and Optimal PNG are some of the variants used in intercepting the target ([7], [9], [10], [11]). In [12], it has been reported that AIPNG method exhibits the best performance among all the proportional navigation based guidance methods. Recently, a novel approach called angular acceleration guidance (AAG) law is proposed in [13] that is capable of catching an evader more efficiently than the previously proposed PNG laws. This motivated us to employ the AAG method for the pursuer in the present study.

Whereas a rich number of pursuer-centric navigation strategies are available in the literature for efficient tracking and targeting an evader, very little work is done in favor of the evader ([1], [14], [15], [16], [17]). Recently, studies on evader-centric problems have gained momentum and numerous approaches have been suggested for the evader to avoid intelligent pursuers and to survive in an environment with certain constraints ([18], [19], [20]). However most of the suggested strategies have focused on evasion from multiple pursuers. Challenges with multiple pursuers are coordination and communication among pursuers while navigating through the environment and seeking the evader. A large number of resources and a central controller are also needed to operate with multiple pursuers ([16], [18]). Pursuit-evasion problems with a single pursuer and a single evader are also of equal importance. Range of applications of such problems is also quite vast starting from computer games to serious military and surveillance operations. Task of evader becomes more difficult when the pursuer is faster than the evader and both players are navigating in a bounded environment. The present paper proposes an approach for interception-free navigation of evader for maximum possible time against a fast pursuer. It is assumed that both the players can sense the precise locations of each other at all the times during the operation. It is further assumed that the maximum velocity of the pursuer is greater than that of the evader. This assumption ensures that the interception would occur within a finite time frame. The proposed method is based on prediction of the pursuer's trajectory using polynomial extrapolation. The environment also contains other non-holonomic wheeled mobile robots with capabilities similar to the pursuer and the evader. These are treated as dynamic obstacles. A recently introduced reciprocal orientation method [21] is adopted to avoid collision. The algorithm assigns equal responsibility to each of the mobile robots to ensure collision-free navigation. Performance of the proposed method is analyzed in terms of the interception time and the distance travelled against fast pursuer. Simulation results demonstrate that the proposed method for pursuit-evasion is effective.

Rest of the paper is organized as follows. In Section 2, the pursuit-evasion game problem is formulated in a bounded environment containing dynamic obstacles, a single pursuer and a single evader. An overview of AAG method, which is used for catching the evader, is discussed in Section 3. In Section 4, the proposed method is presented that enables an evader to escape from a fast pursuer for maximum possible time. In Section 5, collision avoidance is discussed using reciprocal orientation method to deal with dynamic obstacles. Experimental results and the performance of the evader are detailed in Section 6. Concluding remarks and future scope of work are highlighted in Section 7.

## 2 Problem Formulation

Let  $P(t)$ ,  $E(t)$ ,  $V_p(t)$  and  $V_E(t)$  denote positions and velocities of the pursuer and the evader robots at the time  $t$  respectively. It is assumed that both the players know the position of each other at any time instance  $t$ . It is further assumed that the maximum velocity of the pursuer robot is greater than that of evader, i.e.,  $V_{\max P} > V_{\max E}$ . The pursuer and the evader are also supposed to avoid dynamic obstacles during the operation. Mathematically, it is expressed as

$$d(Q(t), O_i(t)) > (2 \times \text{rad} + \Delta) \text{ for } i \in [1, 2, \dots, n], \tag{1}$$

where  $n$  is the number of dynamic obstacles,  $O_i(t)$  is the position of the  $i^{\text{th}}$  moving obstacle at time  $t$ ,  $Q(t) = P(t), E(t)$  and  $d(Q(t), O_i(t))$  denotes the distance between the center of the player robot  $Q$  and the center of the dynamic obstacle robot  $O_i$  at the time  $t$ . Parameter  $\text{rad}$  is the radius of the mobile robot and  $\Delta$  is a small number to indicate that the distance between the two robots is safe to avoid collision. The problem is stated as follows.

While the pursuer constantly tries to catch the evader in minimum possible time, the evader tries to maximize the interception time. Given the initial positions of the pursuer and evader, find  $t$  such that  $d(P(t), E(t)) \leq (2 \times \text{rad} + \Delta)$ . This condition leads to potential collision and the game ends when such a time  $t$  is achieved.

As mentioned in Section 1, an efficient interception method known as angular acceleration guidance (AAG) has been used to guide the pursuer robot in order to capture the evader in minimum possible time [13]. The method is briefly discussed in the following section.

## 3 Angular Acceleration Guidance Law for Pursuer

The geometry engaged in the AAG method is shown in Fig. 1. For the sake of convenience, the relative position vector is expressed as  $R = \sqrt{x_s^2 + y_s^2}$ .

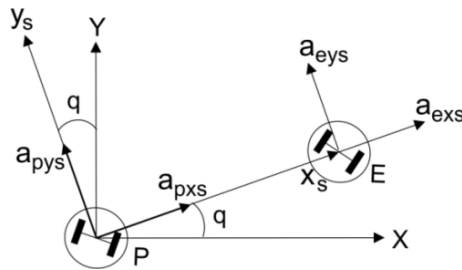


Fig. 1. Geometry engaged in AAG method

Moreover, the relative velocity and the relative acceleration can also be computed as  $\dot{R} = (x_s \dot{x}_s + y_s \dot{y}_s) / \sqrt{x_s^2 + y_s^2}$  and  $\ddot{R} = R \dot{q}^2 + a_{exs} - a_{xs}$ , respectively. Here  $q, \dot{q}, \ddot{q}$  denote the angle, angular rate of change and angular acceleration of the line-of-sight (LOS) (Fig. 1). Further  $a_{pxs}, a_{pys}$  are the acceleration components of the pursuer in the frame of LOS, treating this as the X-axis.  $a_{Exs}, a_{Eys}$  are defined similarly.

The pursuer is continuously being guided by the AAG law using an acceleration command which determines the acceleration and the direction of the pursuer robot to move further in order to catch the evader robot. The acceleration command generated by the AAG law, is formally defined as follows [13] –

$$a_{AAG} = \frac{\lambda}{2} k (R \ddot{q} + 2 \dot{R} \dot{q} + a_{pys}) - \lambda \dot{R} \dot{q}, \tag{2}$$

where  $\lambda$  is the navigation gain and  $k$  is the equivalence factor to compensate for the loss due to the basic assumption that the evader is navigating with constant velocity. The AAG law takes into consideration the angular acceleration of LOS which is easier to estimate accurately as compared to estimation of evader’s acceleration employed in many augmented proportional navigation based guidance laws. This law has been effectively used to intercept even fast maneuvering evaders. In order to have a competitive strategy for the present evader-centric problem, AAG has been used to design the pursuer’s strategy. This throws a greater challenge for the relatively slower evader, to maximize the evasion time in the given bounded environment.

## 4 Anticipated-Velocity Based Guidance Strategy for Evader

The pursuer robot tracks the evader by continuously updating its acceleration and direction based on AAG law. Suppose the current positions of the pursuer and the evader at time  $t$  are represented by the coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$  respectively. Also let  $P_{prev} = (x_0, y_0)$  be the previous position of the pursuer (one unit of time before the current position) as shown in Fig. 2. The proposed approach predicts the next position of the pursuer by assuming that the pursuer would move towards the evader on a curve joining the points  $(x_0, y_0), (x_1, y_1)$  and  $(x_2, y_2)$ . Polynomial interpolation is used to create the probable trajectory of the pursuer as detailed below.

### 4.1 Polynomial Interpolation to Predict the Future Behavior of the Pursuer

Consider a polynomial  $q(t) = a_0 L_0(t) + a_1 L_1(t) + a_2 L_2(t)$  of degree  $\leq 2$  which satisfies  $q(t_i) = (x_i, y_i)$  for time instance  $t_i, i = 0, 1, 2$ . Here  $L_i(t), i = 0, 1, 2$  is the  $i$ -th Lagrange basis polynomial of degree 2 given by

$$L_0(t) = \frac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)}, L_1(t) = \frac{(t-t_0)(t-t_2)}{(t_1-t_0)(t_1-t_2)}, L_2(t) = \frac{(t-t_0)(t-t_1)}{(t_2-t_0)(t_2-t_1)} .$$

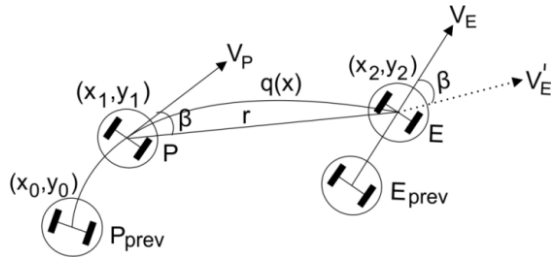


Fig. 2. The evader estimates moving direction of the pursuer at time  $t$

As per Lagrange interpolation formula, coefficients of  $q$  are given by  $a_i = (x_i, y_i)$ ,  $i = 0, 1, 2$ . In order to define the polynomial curve one needs to have a parameterization, or in other words, length of the time intervals  $[t_{i-1}, t_i]$ ,  $i = 1, 2$ . One can choose a uniform parameterization or chord length parameterization. In the present problem we have chosen the well-known chord-length parameterization ([22], p. 180). The generated curve  $q(t)$  is shown in Fig. 2. The evader predicts the moving direction of the pursuer by computing the tangent of the curve  $q(t)$  at the point  $(x_1, y_1)$  (at the current time  $t$ ). The anticipated-velocity direction of the pursuer is denoted by  $V_p$  as shown in Fig. 2. Using this predicted tangent direction, the evader makes its own strategy to turn and run away from the pursuer. We are set to describe the evader's strategy in detail as follows.

### 4.2 Evader's Strategy

The evader computes an angle, which is denoted by  $\beta$  between  $V_p$  and the line segment which connects  $(x_1, y_1)$  to  $(x_2, y_2)$  (see Fig. 2). If  $|\beta| \geq \pi/2$  then it is augmented as  $\beta = \text{sgn}(\beta) \times \pi/2$  where  $\text{sgn}(\beta)$  denotes the sign of the angle  $\beta$ . The proposed evader's strategy takes the angle  $\beta$  into account and generates two separate acceleration commands applied to the evader. First acceleration command  $a_{nE}$  is in the direction of the evader's body axis. The vector  $a_{nE}$  is responsible for rotating the evader robot in a safe direction away from the pursuer and is defined as

$$a_{nE} = \begin{cases} w \cos \beta n_E & \text{if } \beta \leq 0 \\ -w \cos \beta n_E & \text{if } \beta > 0 \end{cases} , \tag{3}$$

where  $n_E$  is the unit vector to the body axis of the evader and  $w$  is a real valued constant used to calibrate the computed value of  $a_{nE}$  with the angle by which the evader robot will turn in appropriate direction. Sign of the angle  $\beta$  is taken positive in



anti-clockwise direction as usual. New moving direction of the evader after applying the generated first acceleration command is denoted by  $V'_E$  and is shown in Fig. 2.

Second acceleration command is the tangential acceleration command, denoted by  $a_{tE}$ , and is in the direction of the evader's orientation. The evader robot is accelerated linearly by applying the tangential acceleration command  $a_{tE}$ , which is expressed as

$$a_{tE} = |\max(V_E)|\eta n_T, \tag{4}$$

where  $n_T$  is the unit vector in the direction of the evader's orientation and  $\eta$  is a constant which is inversely proportional to the value of the first acceleration command i.e.,  $\eta \propto 1/|a_{nE}|$ . When the evader robot will turn with a large angle, linear velocity of the evader will be less. Range of  $\eta$  is specified as  $0 < \eta \leq 1$ . The generated acceleration command  $a_{tE}$  manages the evader's linear velocity from zero to the specified maximum value. The proposed evader's strategy based on turn and run away (as discussed above) makes the evader survive in a bounded environment as long as possible.

Evader and pursuer are also required to avoid collision with the moving vehicles in the environment. These moving vehicles are treated as dynamic obstacles. In order to avoid dynamic obstacles, both players use reciprocal orientation algorithm which is discussed in brief in the following section. A detailed description of the reciprocal orientation algorithm can be found in [21].

### 5 Avoiding Dynamic Obstacles

The reciprocal orientation algorithm is inspired by the concept of reciprocal velocity but does not rely on the complete information of the robots velocities, rather it takes the orientation of the robots (one robot is a player and the other is dynamic obstacle) into account and imposes equal responsibility to each of the mobile robots for safe navigation. The reciprocal orientation algorithm consists of following steps.

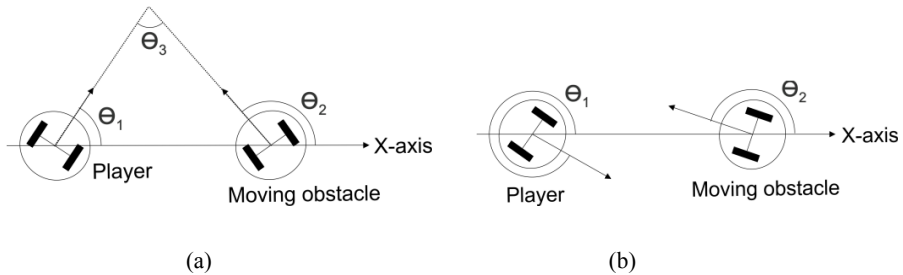


Fig. 3. Possible scenarios of collision of two robots

**Step 1.** In the first step, the algorithm works on finding the possibility of collision between two mobile robots using their orientations. Then the algorithm guides the two potentially colliding robots to change their orientations. Consider the potential collision cases as shown in Fig. 3.

**Case 1** (Fig. 3(a)). If the angle  $\theta_3$  between the robots directions is less than  $90^\circ$  and  $(\theta_1 + \theta_2) < 180^\circ$  then the player robot is rotated in a counter-clockwise direction whereas the moving obstacle is rotated in a clockwise direction and vice versa if  $(\theta_1 + \theta_2) > 180^\circ$  and  $\theta_3 < 90^\circ$ .

**Case 2** (Fig. 3(b)). If the angle  $\theta_3$  between the robots directions is greater than  $90^\circ$  then both robots rotate in the same direction. If  $(\theta_1 + \theta_2) < 180^\circ$  then both robots are rotated in a clockwise direction, otherwise they are rotated in a counter clockwise direction.

**Step 2.** Next, the trajectories of robots are calculated. If any pair robots is satisfying condition stated in equation (1), no change is made in the future trajectory of either of the robot. Else, moving direction of two colliding robots (a player and a dynamic obstacle) is changed as suggested in Step 1.

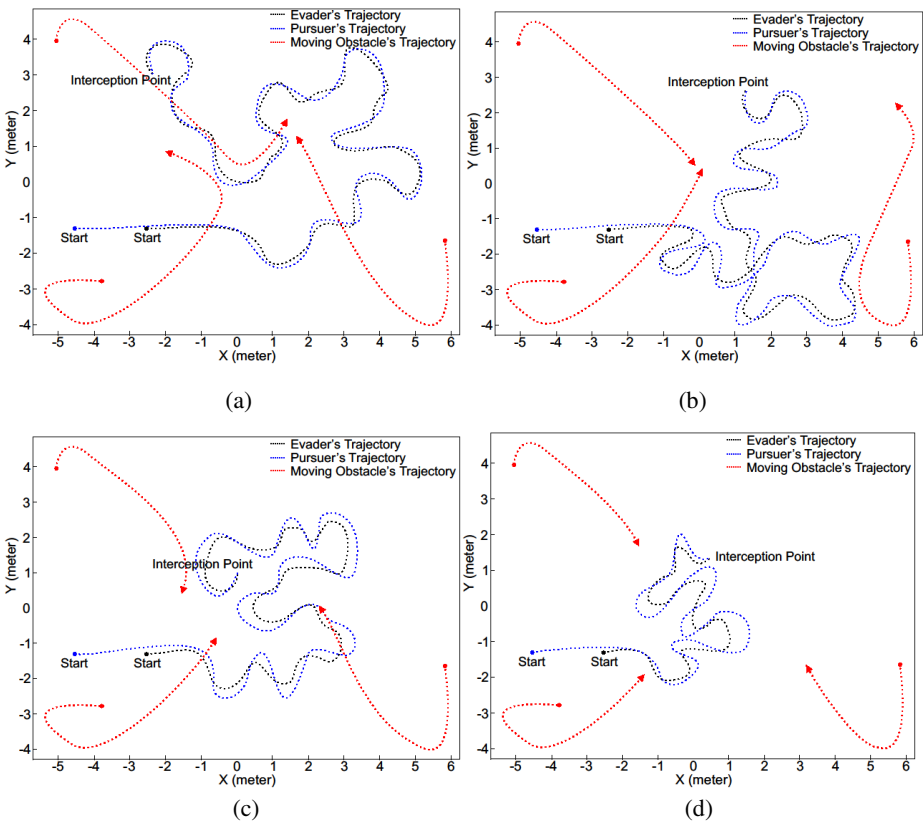
The reciprocal orientation algorithm described above is applied on two robots aligned according to the horizontal axis only (Please refer Fig. 3). A two-dimensional transformation is further used to make the algorithm work for all alignments. According to this, virtual positions of robots are calculated by rotating the actual positions of the robots to the horizontal direction. Center of the rotation is one of the robots and the radius is equal to the distance between them. Now the future trajectories of the robots are computed as discussed above on the virtual positions of the robots to find the new virtual positions with no collision between them. Finally, the new virtual positions of the robots are transformed to their original positions by applying the inverse rotational transformation.

While avoiding moving obstacles, the evader's main objective of escaping from the pursuer is analyzed in terms of the interception time and the distance travelled against the fast pursuer.

## 6 Experimental Results

In this section, simulation results are presented on implementation of anticipated-velocity based guidance scheme presented in Section 4. Maximum velocity of the pursuer is assumed to be  $\alpha$  - times greater than that of the evader. For the present simulation,  $\alpha \in \{1.2, 1.3, 1.4, 1.5\}$ . Three mobile robots are considered as dynamic obstacles in a bounded environment. The algorithm was tested on a Pentium® Dual Core CPU @ 1.86 GHz computer with 2GB RAM. The code was written in C++ using Player/Stage simulator. Choice of  $\alpha$  is an important factor, a small value of  $\alpha$

(say  $\alpha = 1.1$ ) limits the maximum velocity of the pursuer almost equal to the evader's maximum velocity. In this case, the pursuer takes more time to intercept the evader and so the pursuer will have to travel large distance by consuming more energy. On the other hand, a large value of  $\alpha$  ( $\alpha > 1.5$ ) accelerates the pursuer, resulting in missing the target at the estimated interception point especially when the evader is fast maneuvering as in our case. Velocity of other moving obstacles is assumed to be equivalent to that of the evader. A few simulation results are displayed in Fig. 4. Although the proposed method is regardless of the initial positions of the robots but we keep the start positions and orientations of the players and the moving obstacles the same for different values of  $\alpha$ , to analyze the performance of players which is shown in Table 1.



**Fig. 4.** Evader robot escaping from the pursuer robot for the value of (a)  $\alpha = 1.2$ , (b)  $\alpha = 1.3$ , (c)  $\alpha = 1.4$ , and (d)  $\alpha = 1.5$ . Initial distance between players is two meters and both players are avoiding dynamic obstacles using reciprocal orientation algorithm.

**Table 1.** Performance analysis of anticipated-velocity based scheme against the fast pursuer directed by AAG interception law

Value of $\alpha$	Simulation results shown in	Distance travelled by the		Interception time (seconds)
		Evader robot (meters)	Pursuer robot (meters)	
1.2	Fig. 4(a)	28.67	32.78	37.4
1.3	Fig. 4(b)	23.02	28.37	29.1
1.4	Fig. 4(c)	19.67	25.49	25.3
1.5	Fig. 4(d)	13.11	21.87	18.8

## 7 Conclusion and Future Work

A novel technique based on anticipated-velocity of the pursuer is proposed for a non-holonomic wheeled mobile evader to navigate and escape from the pursuer in a bounded environment with dynamic obstacles. The pursuer robot is guided by AAG method and the pursuer's maximum velocity is assumed to be  $\alpha$  times greater than the evader's maximum velocity, where  $\alpha > 1$ . Both players use reciprocal orientation algorithm to avoid collision with other mobile robots in the environment. Simulation results demonstrate effectiveness and efficiency of the evader. These methods have the potential for application to holonomic mobile robots and can also be extended to the three dimensional problems. In future, we plan to extend this method to higher dimensions and would also work on implementing the algorithms on hardware.

## References

1. Isaacs, I.: *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization*. John Wiley and Sons, New York (1965)
2. Chung, T.H., Hollinger, G.A., Isler, V.: Search and Pursuit-Evasion in Mobile Robotics - A Survey. *Auton. Robot* 31(4), 299–316 (2011)
3. Bhadauria, D., Klein, K., Isler, V., Suri, S.: Capturing an Evader in Polygonal Environments with Obstacles: The Full Visibility Case. *Int. J. Robot. Res.* 31(10), 1176–1189 (2012)
4. Bakolas, E., Tsiotras, P.: Relay Pursuit of a Maneuvering Target using Dynamic Voronoi Diagrams. *Automatica* 48(9), 2213–2220 (2012)
5. Belkhouche, F., Belkhouche, B., Rastgoufard, P.: Line of Sight Robot Navigation Toward a Moving Goal. *IEEE Trans. on Syst.* 36(2), 255–267 (2006)
6. Kung, C.C., Chiang, F.L., Chen, K.Y.: Design A Three-dimensional Pursuit Guidance Law with Feedback Linearization Method. *World Acad. Sci. Engg. and Tech.* 55, 136–141 (2011)
7. Ghose, D.: True Proportional Navigation with Maneuvering Target. *IEEE Trans. on Aero. and Elec. Syst.* 30(1), 229–237 (1994)
8. Croft, E.A., Benhabib, B., Fenton, R.G.: Near-time Optimal Robot Motion Planning for On-line Applications. *J. Robot. Syst.* 12(8), 553–567 (1995)

9. Mehrandezh, M., Sela, M.N., Fenton, R.G., Benhabib, B.: Robotic Interception of Moving Objects using Ideal Proportional Navigation Guidance Technique. *J. Robot. Auton. Syst.* 28, 295–310 (1999)
10. Mehrandezh, M., Sela, M.N., Fenton, R.G., Benhabib, B.: Robotic Interception of Moving Objects using an Augmented Ideal Proportional Navigation Guidance Technique. *IEEE Trans. on Syst.* 30(3), 238–250 (2000)
11. Jiali, G., Wanchun, C.: Optimal Proportional Navigation Guidance based on Generalized Predictive Control. In: *Proc. 16th International Conference on System Theory, Control and Computing, Sinaia*, pp. 1–6 (2012)
12. Keshmiri, M., Keshmiri, M.: Performance Comparison of Various Navigation Guidance Methods in Interception of a Moving Object by a Serial Manipulator Considering its Kinematic and Dynamic Limits. In: *15th International Conference on Methods and Models in Automation and Robotics, Miedzyzdroje*, pp. 212–217 (2010)
13. Song, Y., Chen, W., Yin, X.: A New Angular Acceleration Guidance Law with Estimation Approach based on Sliding Mode Observer Against High Maneuvering Target. *App. Mech. and Material*, 110–116, 5249–5256 (2012)
14. Meschler, P.A.: On Constructing Efficient Evasion Strategies for a Game with Imperfect Information. *IEEE Trans. on Auto. Cont.* 15(5), 576–580 (1970)
15. Rzymowski, W.: Avoidance of One Pursuer. *I. of Math. Analy. and App.* 120(1), 89–94 (1986)
16. Chodun, W.: Differential Games of Evasion with Many Pursuers. *I. of Math. Analy. and App.* 142(2), 370–389 (1989)
17. Murrieta-Cid, M., Monroy, R., Hutchinson, S., Laumond, J.P.: A Complexity Result for the Pursuit-Evasion Game of Maintaining Visibility of a Moving Evader. In: *IEEE International Conference on Robotics and Automation, California*, pp. 2657–2664 (2008)
18. Jin, S., Qu, Z.: Pursuit-Evasion Games with Multi-pursuer vs. One Fast Evader. In: *Proc. 8th World Congress on Intelligent Control and Automation, Jinan*, pp. 3184–3189 (2010)
19. Ibragimov, G.I., Salimi, M., Amini, M.: Evasion from Many Pursuers in Simple Motion Differential Game with Integral Constraints. *Euro. J. Oper. Res.* 218(2), 505–511 (2012)
20. Liu, S.Y., Zhou, Z., Tomlin, C., Hedrick, K.: Evasion as a Team Against a Faster Pursuer. In: *American Control Conference, Washington DC*, pp. 5368–5373 (2013)
21. Rashid, A.T., Ali, A.A., Frasca, M., Fortuna, L.: Multi-robot Collision-free Navigation Based on Reciprocal Orientation. *Robot. and Auton. Syst.* 60(10), 1221–1230 (2012)
22. Farin, G.: *Curves and Surfaces for CAGD: A Practical Guide*. Morgan Kaufmann, California (2002)

# Reconfigurable Multichannel Down Converter for on Chip Network in MRI

Vivek Jain and Navneet Kumar Agrawal

Department of Electronics and Communication,  
College of Technology and Engineering,  
Maharana Pratap University Agriculture and Technology,  
Udaipur, India  
{vivekjain297,navneetctae}@gmail.com

**Abstract.** Reconfigurable multi channel down converter for on- chip network in MRI has been designed and developed in this paper. The converter is extendable up to 128 channels using direct Re-use FIR filter, requiring less routing area, less static and dynamic power and finally, less delay. The system developed is superior to existing model in terms of high power consumption, large delay & limited channels up to 8 channels. We have reduced the power consumption in this model using placement & routing algorithm. It is done by creating the separate P blocks in the existing model. The existing 8 channel model consumes 3092.1mw (static power 3092.1mw & dynamic 42.6 mw). After applying three different routing & placement algorithms on 8 channels model namely global routing, channel routing and river routing the power consumed results 2608.2mw (static power 2608.2mw & dynamic 34.6mw), 04.5mw (static power 68.8mw & dynamic 35.7mw) and 100.5mw (static power 64.9mw & dynamic 35.6mw) respectively. In an extended work the authors have tried and successfully executed the model and system for 128 channels for MRI applications. The proposed model is first designed on simulink platform using Xilinx blackest and then it is transferred on FPGA platform using system generator. The complete circuit is synthesized, implemented, simulated using Xilinx design suite.

**Keywords:** DDC, DUC, FIR filter, LTI system. Routing Algorithm, P block.

## 1 Introduction

Reconfigurable down converter is used for base band echo signals to regulate the sample rate in medical imaging applications, such as ultrasound and MRI. Extremely oversampled unique echo signals are down sampled by an integer factor, which can transform at run time. Usually as the decimation rate increases one challenge in designing reconfigurable down converter for medical imaging is the support to process most of the channels simultaneously. While current profitable MRI scanners available with 8 channels of data, Next generation scanners may support up to 128 channels of data simultaneously [2].

### 1.1 Resampling Filter Basics

Sample rate conversion has a wide range of applications including wireless-communications, medical imaging, and military applications. Down convertor is generally computationally exhaustive and needed equivalent processing of a huge number of self-governing data channels, making a high-performance down convertor implement on suitable FPGA family [2].

### 1.2 Existing Model of Sample Rate Converters

In Fig 1 Sample rate conversion refers to decimation, interpolation, or a combination of down sampling and up sampling for fractional rate change factors.

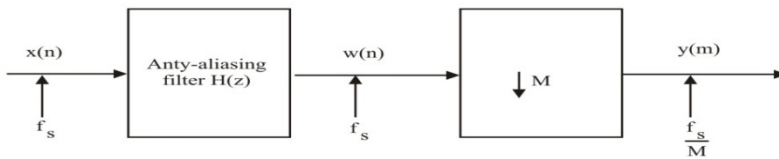


Fig. 1. Sample Rate Converters

Decimation requires low-pass filtering on high rate data first, followed by periodic down sampling to remove unused data. Low-pass filtering shapes the signal spectrum to prevent is at a lower sample rate [2].

### 1.3 Existing Model of Single Channel FIR Filter

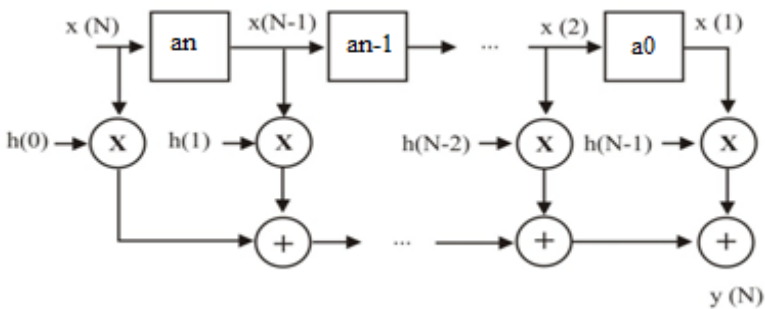


Fig. 2. Single channel FIR Filter [1]

In Fig 2 the delay line coefficient  $a_0$ ,  $a_{n-1}$ ,  $a_n$  of the FIR filter is calculated using sampling frequency, order of the filter, cut off frequency and window techniques. [1], [3].

## 1.4 Existing Model for Multichannel FIR Filter

### 1.4.1 Typical Structure of Multi-channel FIR Filter

In Fig 3 Multi channel FIR filter the single channel active at a time with help of multiplexing mechanism, the advantage of that structure is data rate is increased by  $M$  times of single channel filter. Here the single channel sample rate and data rate is  $f_1$ . That process is completely automatic. In multichannel structure each channel independently works to each other. That structure required  $M*N$  registers,  $M*N$  Multipliers and  $M*(N-1)$  adders [1].

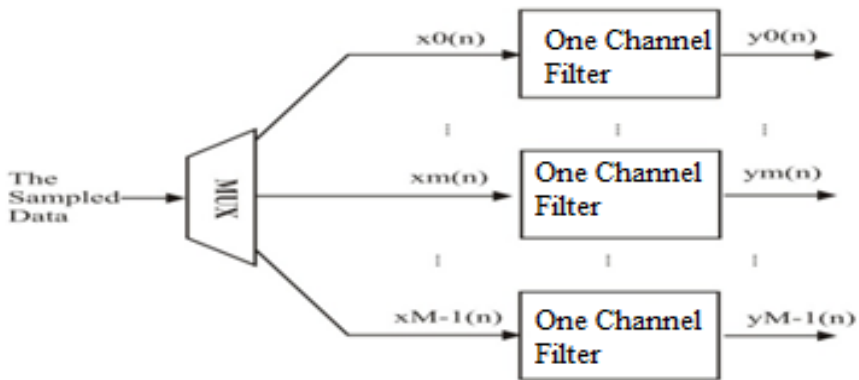


Fig. 3. Multi-channel FIR filter [1]

### 1.4.2 Structure of Direct Re-use FIR Filter

In digital signal processing required sample rate is very low generally in terms of KHz. But the clock frequency of FIR filters at MHz To fulfill that need in direct Re-use FIR filter operate at  $f_2$  clock frequency and complete the process in  $p$  clocks where  $p$  is equal to  $f_2/f_1$  [1].

Fig 4 shows the structure of direct re-use FIR filter. The direct re-use filter has  $N$  registers,  $[N/p]$  multipliers and  $[N/p-1]$  adders,  $N$  registers,  $[N/p]$  multipliers and  $[N/p-1]$  adders. The data rate of input data to the single channel FIR filter can be optimized to  $f_1$  [1].



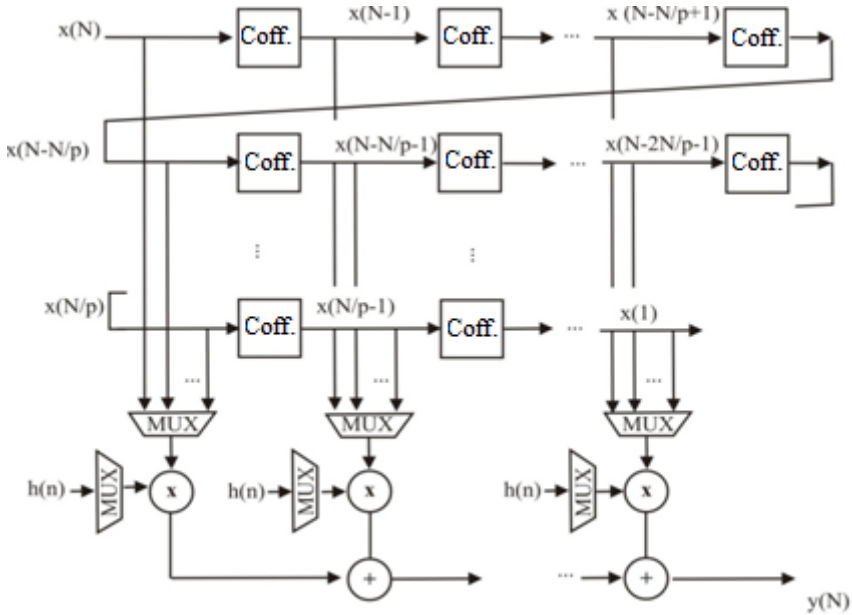


Fig. 4. Direct Re-use FIR filter [1]

## 2 Limitation of Existing Model of MRI

- Number of Input Channels is limited.
- Power consumption is very high

## 3 Proposed Model

The proposed structure is modifying at different stages using new algorithms extensively. The following discussion includes the suggestive alterations.

### 3.1 Multiplexed Structure of Reconfigurable FIR Filter

The number of channels is increased up to 128 channels using time division multiplexing applying at 8 channel direct reuse FIR filter. In the propose model first we design 8 channel direct reuse FIR filter. Then 8 channel direct reuse FIR filter extended up to 128 channels with the help time division multiplexing mechanism. Fig 5 shows 8 channels direct reuse FIR filter .In that all 8 channels share the same recourses so that requirement of the adders ,multipliers reduced on FPGA board. So it also reduced the power consumption.

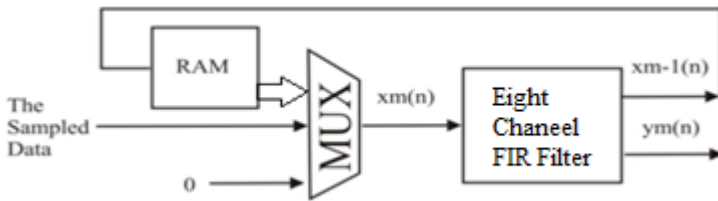


Fig. 5. 8 channel Direct Reuse Multi-channel FIR filter [1]

To obtain the 16, 32, 64, 128 channels reconfigurable down convertor we use 2, 4, 8, 16 blocks of 8 channel direct-reuse filter blocks respectively. Each block transmits their data using time division multiplexing. In Fig 6 shows the 16 channels direct Reuse FIR filter.

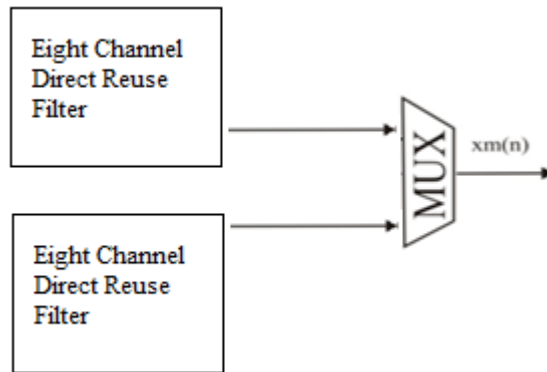


Fig. 6. 16 channel Direct Reuse Multi-channel FIR filter

## 3.2 Routing Algorithm

To decrease the power consumption we apply routing algorithms namely global routing, channel routing, river routing on the different blocks of reconfigurable down converter.

### 3.2.1 Global Routing

The algorithm used here by PI is essentially shortest path algorithms. Our paths will hop among the points that are midpoints of the channels edges, beginning and ending at ports belonging to the same net. The length of any edge in a path is the ordinary Euclidean distance between points. Edges available to take us between channels is deemed to both channels, a fact that allows path to pass from one channel to next.

### 3.2.2 Channel Routing

Having selected positions for all crossing points, PI can route the wires within each channel, one at a time, knowing that what it does in one channel does not affect any other. Moreover, it is no longer necessary to distinguish between points on the border

of the channel that are due to ports of a cell, and points representing wall crossings. Of course since the widths of the channels were selected arbitrarily, there is no guarantee that a given channel can be routed by even the cleverest algorithm. PI uses a set of the channel routing algorithm, starting with a trivial one that in essence runs wires for each net independently.

### 3.2.3 River Routing

River routing is channel routing restricted in the following way.

- Wires run in one layer only; thus do not cross,
- Each net consists of two points; one on top and one on the bottom .there are no ports on the sides.
- The order of the nets is the same on top and bottom, as it must be if we are to run wires in one layer.

## 4 Reconfigurable Convertor Design

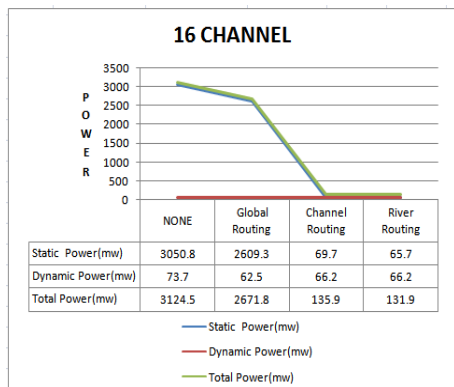
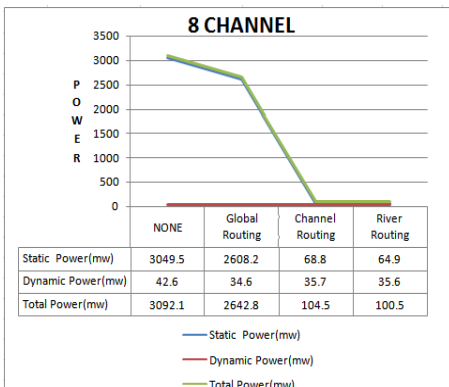
In order to carry out the work, following steps were involved.

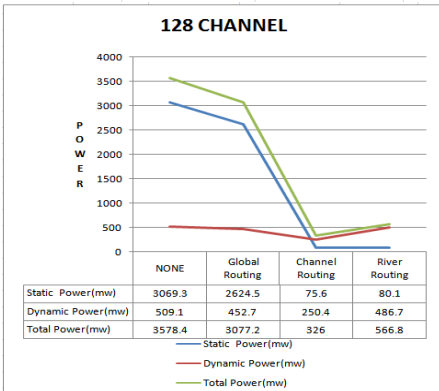
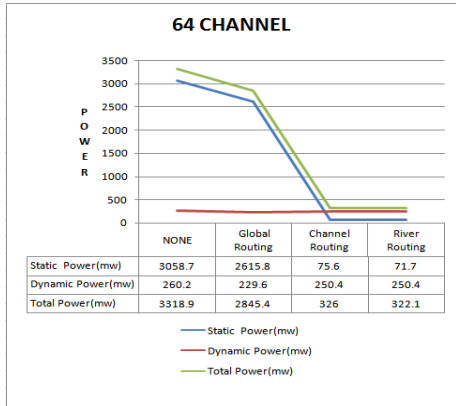
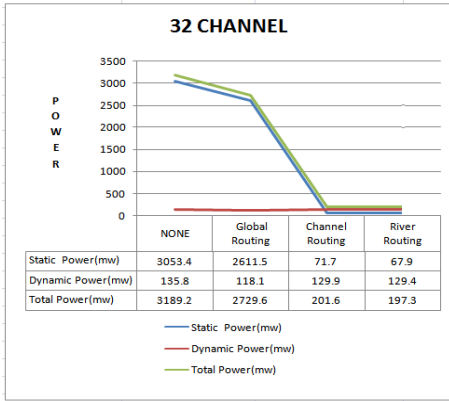
- 1 Developed the MATLAB subscript for the down convertor in MRI.
- 2 Developed the simulink model for the down convertor in MRI.
- 3 Generate the HDL code using system generator.
- 4 Implement the model with the help of HDL language using Xilinx Design suite.
- 5 Applying routing & Placement algorithm on the model by creating P blocks with the help of the Xilinx Plan Ahead.
- 6 Estimate the power using Xilinx Plan Ahead.

## 5 Result and Analysis

### Power Consumption

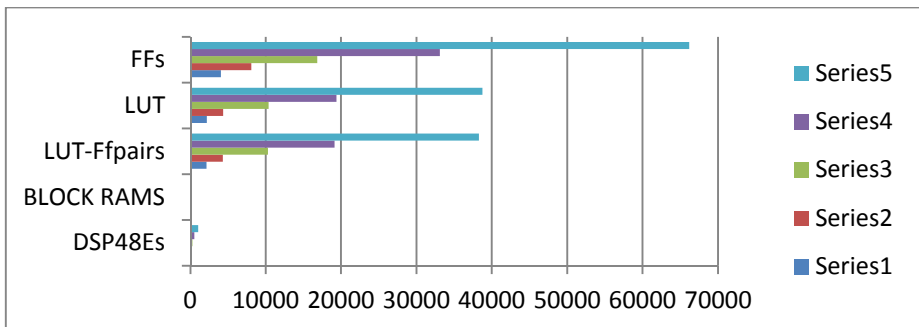
Resultant graph for power consumption in 8,16,32,64,128 channel MRI without algorithm &with algorithm





From the resultant graph authors observe that when applying River routing algorithm, down converter in MRI consume minimum power for 8, 16,32,64,128 channels.

### Resource Utilization Graph



## 6 Conclusion and Future Work

Number of channels can be increased by using Multiplexed direct Reuse FIR Filter up to 128 channels. River routing algorithm is the best algorithm in terms of power consumption. Resource utilization increases linearly with number of channels. Research work is still going on to minimize the delay of the Reconfigurable down convertor and develop up convertor with optimum power consumption & delay. The future work will focus on fractional rate convertor which will get its applications in wireless communication.

**Acknowledgement.** The authors would like to thank all the faculty members of Department of Electronics & Communication, College of Technology & Engineering, Udaipur, for on time support to carry out the study and providing requisite laboratory facilities to perform our experiment.

## References

1. Benkrid, A., Benkrid, K.: Novel Area-Efficient FPGA Architectures for FIR Filtering With Symmetric Signal Extension. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 17(5), 709–722 (2009)
2. AN 623: Using the DSP Builder Advanced Blockset to Implement Resampling Filters, ALTERA (2010)
3. Francis, M.: Infinite Impulse Response Filter Structures in Xilinx FPGAs White Paper, Spartan®-3A DSP, Xilinx (2009)
4. Creaney, S., Kostarnov, I.: Designing Efficient Digital Up and Down Converters for Narrowband Systems, XAPP1113 (v1.0) Xilinx (November 21, 2008)
5. Patronis, S.G., DeBrunner, L.S.: Sparse FIR Filters and the Impact on FPGA Area Usage. In: 42nd Conference on Signals, Systems and Computers, pp. 1862–1866 (2008)
6. Video and Image Processing Design Using FPGAs, WP-VIDEO0306-1.1 by Altera, ver. 1.1 (March 2007)
7. Tarn, H., et al.: Designing Efficient Wireless Digital Up and Down Converters Leveraging CORE Generator and System Generator, Xilinx® Application Note, XAPP1018
8. Cheney, E.W.: *Introduction to Approximation Theory*. American Mathematical Society, Rhode Island (1998)
9. Biased Discriminate Euclidean Embedding for Content-Based Image Retrieval IEEE (2010)

# Author Index

- Acharya, Sasmita 351  
Adarsh, V.S. 613  
Agrawal, Navneet Kumar 459, 799  
Akshatha, G.B. 613  
Alakananda, V. 781  
Allada, Pabita 749  
Ambaw, Lubak M. 11  
Anne, V.P. Krishna 749, 781  
Anthony, L. Britto 757  
Anupama, K.S.S. 331  
Archana, M. 621  
Arora, Priyanka 653  
Arora, Sparsh 301  
Arsekar, Shubharaj 671  
Arya, Rakshit 695  
Atique, Mohammad 109  
Atta, Soumen 195  
Avadhani, P.S. 11, 473
- Babu, Inampudi Ramesh 393  
Babu, Polinati Vinod 713  
Babu, Samitha R. 523  
Bala, Mohammad Irfan 505  
Balamurugan 663  
Balamurugan, S. Appavu Alias 179  
Balkish, Naziya 575  
BalvinderSingh, B. 69  
Banupriya, C.V. 187  
Basha, Zareena Noor 31  
Basu, Subhadip 133  
Benala, Tirimula Rao 85  
Bennet, Suja 149  
Bhagat, Amol 109  
Bhaskari, D. Lalitha 473
- Bisoy, Sukant Kishoro 371  
Biswas, Papun 205  
Boggavarapu, Bhavya 749  
Boopathy, D. 555
- Chandrashekar, Sanjana 613  
Choudhary, Amit 159  
Choudhary, Naveen 417
- Das, Nibarani 133  
Desai, Madhav 671  
Devi, Vijeyta 481, 497  
Dharmalingam, Jeya Mala 663  
Donala, Swetha Reddy 621
- Francis, F. Sagayaraj 513
- Gebeyehu, Worku B. 11  
Ghodke, Ajit 21  
Gouda, Bhabani Sankar 311  
Govardhan, A. 167  
Gowri, S. Sri 331
- Hanah Ayisha, V. 49  
HariPriya, M. Vani 85  
HimaDeepthi, V. 69
- Jagadeesh Kannan, R. 321  
Jagetiya, Anurag 259  
Jain, Lokesh 409  
Jain, Rashmi A. 739  
Jain, Vivek 799  
Janvale, Ganesh B. 21  
Jayaraju, Ch. 31  
Jayarin, P. Jesu 757

- Jeeru, Syama Sundar 59  
 Judy, M.V. 241  
  
 Kadam, Megha 267  
 Kale, Vijay 21  
 Kalyani, M. 31  
 Kannan, E. 179  
 Karpagavalli, S. 187  
 Khaikum, Samitha 679  
 Khamrui, Amrita 251  
 Khasnabish, Jyotiska Nath 543  
 Kiran, R.N.D.S.S. 713  
 Kishore, V. 757  
 Kotteeswaran, R. 117  
 Koushik, S. 467  
 Krishna, B.B. Murali 713  
 Krishna, Ch.V. Phani 125  
 Krishnakumar, U. 241  
 Kudchadkar, Gaurav 671  
 Kulkarni, Krutika G. 523  
 Kumar, A. Dinesh 49  
 Kumar, Amit 789  
 Kumar, Harish 401, 409  
 Kumar, Kailash 259  
 Kumar, Mousumi 219, 639  
 Kumar, N.R. Shashi 687  
 Kumaraswamy, Y.S. 679  
 Kumari, G. Rosline Nesa 41, 59  
 Kundu, Mahantapas 133  
 Kushwaha, Satpal Singh 259  
  
 Limkar, Suresh 267  
  
 Madan 441  
 Madhavarao, E. 31  
 Madhu, Ramarakula 773  
 Mahapatra, Priya Ranjan Sinha 195  
 Mahitha, O. 583  
 Majumder, Koushik 233  
 Mall, Rajib 85  
 Mamillapalli, SaiManasa 749  
 Manasa, N.L. 167  
 Mandal, J.K. 251  
 Manokar, Nandhini Vigneshwar 49  
 Maruthuperumal, S. 59  
 Mehta, Neha 417  
 Mishra, Sarojananda 449  
 Mohanty, Prases Kumar 361  
 More, Amruta 489  
 Mouli, P.V.S.S.R. Chandra 93  
  
 Mukhopadhyay, Debajyoti 489, 505  
 Murali, T. Satya 331  
 Murthy, J.V.R. 1  
 Murthy, M.S. Narasimha 605  
 Murty, M. Ramakrishna 1  
  
 Nagalakshmi, Vadlamani 481, 497  
 Naganathan, E.R. 77  
 Naik, Anima 1  
 Nair, T.R. Gopalakrishnan 563, 593, 687  
 Nandini, C. 433  
 Narayanan, A.G. Hari 241  
 Narayanan, S. 77  
 Naresh, K. 321  
 Nasipuri, Mita 133  
 Nathan, Sabari 663  
 Navamani, T.M. 287  
 Navya Rachana, K. 781  
 Nedunuri, S.S.S.N. Usha Devi 393  
 Nidhya, R. 49  
  
 Ojha, Aparajita 765, 789  
 Ojha, Davendar Kumar 723  
  
 Padole, Dinesh V. 739  
 Pai, Anusha 671  
 Pal, Bijay Baran 205, 219, 639  
 Pandey, Krishna K. 361  
 Pandey, Shilpa 417  
 Pani, Niroj Kumar 449  
 Parhi, Dayal R. 361  
 Parwekar, Pritee 277, 301  
 Patil, Annapurna P. 441, 467  
 Patnaik, Prasant Kumar 371  
 Prabakaran, N. 321  
 Prasad, A.M. 575  
 Premamayudu, B. 425  
  
 Qureshi, Salim Raza 535  
  
 Raja, N. Sri Madhava 141  
 Rajagopalan, S. 77  
 Rajan, R. Arokia Paul 513  
 Rajanikant, K. 441  
 Rajinikanth, V. 141  
 Rajkumar, S. 93  
 Rangan, P. Shashikanth 613  
 Rani, G. Vakula 149  
 Ranjani, R. Siva 473  
 Rao, B. Prabakara 331

- Rao, Gottapu Sasi Bhushana 773  
 Rao, G. Sivanageswara 125  
 Rao, Gudikandhula Narasimha 729  
 Rao, Harsha 703  
 Rao, K. Rajasekhara 125  
 Rao, K. Rajasekhara 781  
 Rao, K. Venkata 425  
 Rao, P. Jagdeeswar 729  
 Rao, Shrishya 543  
 Rebello, Ross 671  
 Reddy, K. Hemant Kumar 311  
 Reddy, M.A. Eswar 11  
 Reddy, P.V.G.D. Prasad 1  
 Rishi, Rahul 159  
 Roy, Abhinaba 133  
  
 Sabarish 441  
 Sahoo, Anirudha 341  
 Sahoo, Rashi Ranjan 233  
 Sai Chand, U. 781  
 Saini, Hemant Kumar 259  
 Sardar, Abdur Rahaman 233  
 Sarkar, Ram 133  
 Sarkar, Subir Kumar 233  
 Satapathy, Suresh C. 1  
 Satyanarayana, Ch. 167  
 Sekaran, K. Chandra 523  
 Selvakumar, S. 703  
 Sen, Shyamal 219  
 Senthilkumar, P. 101  
 Shanti, Chilukuri 341  
 Sharma, Ankita 381  
 Shial, Rabindra Kumar 311  
 Shivasankaran, N. 101  
 Sikka, Geeta 695, 723  
 Sing, Jamuna Kanta 233  
  
 Singh, Dharm 417  
 Singh, Moutushi 233  
 Singh, Nirmal 629  
 Singh, Raman 401  
 Singh, Sarbjeet 629, 653  
 Singla, R.K. 401, 409  
 Sivakumar, L. 117  
 Soni, Yashwant Kumar 459  
 Sree, Pokkuluri Kiran 393  
 Srikavya, P. 85  
 Srimani, P.K. 149  
 Srinivas, P.V.S. 621  
 SrinivasaRao, V. 69  
 Sudheer, M. 41  
 Suma, V. 563, 575, 583, 593, 605, 687  
 Sundaresan, M. 555  
 Surabi 441  
 Swain, Tanmaya Kumar 371  
  
 Tajuddin, Mohammed 433  
 Tamilkodi, R. 41  
 Tanna, Keval 671  
 Tripathy, C.R. 351  
  
 Vaidehi, M. 563, 593  
 Varma, P. Suresh 425  
 Vashistha, Sumit 381  
 Veeraswamy, A. 179  
 Venkatesh Raja, K. 101  
 Verma, Ananda Prakash 543  
 Vernekar, Vishal 671  
 Vij, Sheetal 489, 505  
  
 Waghmare, Vishal 21  
  
 Yadav, Gyan Singh 765  
 Yogesh, P. 287